U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Hyperspectral Imaging for Remote Marine Litter Detection and Classification using Learning based Approaches

Sara Costa Freitas

Supervisor: Doctor Hugo Miguel Gomes da Silva
Co-Supervisor: Doctor Eduardo Alexandre Pereira da Silva

Submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Electrical and Computer Engineering
Faculdade de Engenharia do Porto
2023

Funding Institutions: Fundação para a Ciência e Tecnologia
Grants SFRH/BD/139103/2018 and COVID/BD/152190/2021

Abstract

Nowadays, Remote Sensing (RS) is one of the most prominent research topics in Earth Observation. The combined use of satellite and airborne platforms for data collection extends evermore the range of RS applications to solve daily life problems on the planet. One of the most utilised sensors is the hyperspectral imaging sensor due to its extended range capabilities (wavelength) coupled with the capability to acquire massive amounts of data. The computer vision community has witnessed a shift from heuristic and physical model development to more data-driven research based on machine learning and deep learning models over the last decade. This phenomenon was pushed by the increase in the availability of data together with the development of more powerful CPU and GPU units. The same happened to hyperspectral imaging systems, where novel methods for the detection and classification of dissimilar materials started to appear at the same time these methods were being used and developed for RGB (visible spectral images) cameras. However, due to a series of internal and external factors related to the availability of the data, most of the works on hyperspectral imaging are focused on a series of available datasets and are not explicitly targeted to solve immediate societal problems.

The work conducted in the framework of this thesis precisely addresses this problem by developing novel learning-based methods using remote hyperspectral imaging data to detect and classify marine litter in ocean environments. We developed and assembled a hyperspectral imaging system using two pushbroom hyperspectral image sensors capable of acquiring data from the 400-2500 nm wavelength. This system, together with the computational and navigation unit, was the basis for the data acquisition during the thesis and in the scope of the ESAPLASTICS project.

To evaluate the hyperspectral imaging system's capability to detect and classify marine litter samples, a set of samples was collected from a marine litter hotspot (Porto Pim Bay Azores). These samples were analysed in-situ in an outdoor laboratory environment in dry and wet sampling experiments. These experiments allowed us to collect the spectral response of the marine litter material for later comparison with the remote sensing hyperspectral data that was collected during the airborne dataset campaigns. An extensive airborne dataset campaign with aeroplanes and

drones using the hyperspectral imaging system was performed in the Azores and allowed to acquire over 2 TB of hyperspectral imaging data of marine litter materials placed on the artificial targets at the test site.

Having collected the hyperspectral imaging data, learning-based approaches were developed to find ways of automatically detecting and classifying marine litter samples at different flight altitudes. We started by using a standard software library to implement Random Forest (RF) and Support Vector Machine (SVM) methods to evaluate Supervised methods for detecting and classifying marine litter samples using airborne data. This implementation was important to establish a benchmark of the state-of-the-art performance in this dataset. We then passed on to more advanced methods using deep learning methodologies, mainly focusing on a novel CNN-3D implementation that combines spectral and spatial information.

To finalise the work conducted in the thesis, we developed a novel (first) Zero-Shot Learning method for detecting and classifying marine litter samples using hyperspectral imaging data. The method is capable of distinguishing different marine litter samples, even the ones that are unknown to him during the training phase. The results are in line with the ones obtained by Zero-Shot Learning state-of-the-art methods using visible spectrum images.

Keywords: Convolutional Neural Networks, Deep Learning, Hyperspectral Imaging, Marine Litter, Random Forest, Remote Sensing, Supervised Approaches, Support Vector Machine, Unmanned Aerial Vehicles, Unsupervised Approaches, Zero-Shot Learning

Resumo

A análise e estudo dos fenómenos que ocorrem no planeta Terra é uma temática complexa, sendo que atualmente este problema tem vindo a ser colmatado com o recurso a técnicas de deteção remota. A utilização de diversas plataformas para a aquisição de dados remotos, tais como satélites e aviões, tem permitido a expansão das aplicações e do tipo de sensores usados. Um exemplo disto é a utilização de câmaras hiperespectrais, que devido à sua capacidade de adquirir dados numa vasta gama do espectro eletromagnético permitem a sua aplicação na resolução de problemas quotidianos.

O aumento da capacidade computacional dos sistemas de processamento permitiu uma mudança no paradigma das técnicas de processamento de dados utilizadas, tendo evoluído de modelos heurísticos e físicos para técnicas de aprendizagem computacional. Esta evolução verificou-se em vários domínios da visão computacional, incluindo no processamento de dados hiperespectrais. Considerando a quantidade de dados que estes sistemas são capazes de retornar, o aumento da capacidade computacional permitiu a adaptação de novas abordagens de deteção e classificação de materiais distintos, que inicialmente apenas eram usadas para imagens do espectro visível. Contudo, devido à indisponibilidade e dificuldade em obter dados hiperespectrais, a maior parte dos algoritmos são testados sempre nos mesmos conjuntos de dados, que não foram recolhidos para a resolução de nenhum problema em específico. Assim sendo, a aplicação destes algoritmos não é capaz de demonstrar o potencial dos dados hiperespectrais para a resolução de problemas específicos da sociedade atual.

O trabalho desenvolvido no âmbito desta dissertação visa o desenvolvimento de novos métodos de aprendizagem computacional para dados hiperespectrais, que permitam a deteção e classificação de lixo marinho em ambientes marítimos. Para isso, foi desenvolvido um sistema hiperespectral utilizando duas câmaras distintas em combinação com um sistema de aquisição de dados e de navegação. Este sistema é capaz de adquirir dados na gama espectral dos 400 aos 2500 nm, tendo sido montado em diferentes estruturas de forma a servir de base à aquisição de dados necessária para o desenvolvimento desta dissertação e ainda no âmbito do projeto ESAPlastics.

De forma a avaliar a capacidade deste sistema detetar e classificar as amostras de lixo marinho, foi recolhido um conjunto de amostras num local bastante afetado pelo problema do lixo marinho (Baía de Porto Pim, Açores). Estas amostras foram analisadas em laboratório, recorrendo ao sistema desenvolvido instalado num suporte no exterior, para que as amostras secas e submersas em água fossem testadas com luz natural. Estas experiências permitiram a recolha da resposta espectral das amostras de forma a comparar com os dados obtidos remotamente, em que o sistema se encontrou instalado em duas aeronaves distintas: uma tripulada e outra não tripulada. Esta campanha de recolha de dados resultou na aquisição de mais de 2TB de dados hiperespectrais referentes a alvos artificiais construídos com materiais recolhidos do local de testes, idênticos às amostras analisadas em laboratório.

Considerando a diferença nas aeronaves usadas, foi possível recolher dados a diferentes altitudes de voo. Isto permitiu desenvolver novas metodologias de aprendizagem computacional para detetar e classificar o lixo marinho automaticamente, sendo que a análise iniciou-se pela aplicação de dois métodos de estado-da-arte, *Random Forest* (RF) e *Support Vector Machine* (SVM). Isto permitiu a avaliação da utilização de métodos supervisionados para a deteção e classificação de lixo marinho, de forma a estabelecer uma linha de comparação para a avaliação dos resultados. De forma a continuar a análise de técnicas supervisionadas, foi a implementada uma nova CNN-3D, que acrescenta a informação espacial à espectral.

Por fim, desenvolveu-se o primeiro método de *Zero-Shot Learning* para deteção e classificação de lixo marinho com dados hiperespectrais. Este método é capaz de distinguir diferentes tipos de lixo marinho, mesmo aqueles que pertencem a classes às quais o modelo não teve acesso durante a fase de treino. Os resultados obtidos demonstram ser semelhantes aos dos métodos de estado-da-arte para esta técnica, utilizando imagens de espectro visível.

Palavras-Chave: Abordagens Não-Supervisionadas, Abordagens Supervisionadas, Aprendizagem Computacional, Deteção Remota, Imagem Hiperespectral, Lixo Marinho, *Random Forest*, Redes Neuronais Convolucionais, *Support Vector Machine*, Veiculos Aéreos Não Tripulados, *Zero-Shot Learning*

Acknowledgements

A realização desta tese de doutoramento não teria sido possível sem a minha família, amigos e colegas, aos quais agradeço por todo o apoio ao longo desta jornada.

Primeiro, gostaria de agradecer ao meu orientador, Doutor Hugo Silva, por todo o apoio incessante ao longo do meu percurso em investigação científica. Sem o seu incentivo e acompanhamento não teria fechado este capítulo. Estarei para sempre grata por tudo o que fez e me ensinou durante estes últimos sete anos. Ao meu Co-Orientador, Professor Eduardo Silva, pelo seu conhecimento, preocupação e disponibilidade ao longo desta jornada. Agradeço todas as suas intervenções ao longo deste percurso, com as quais tanto aprendi.

Aos Professores José Miguel Almeida e Alfredo Martins, pelos seus esforços para criar melhores condições de trabalho e motivação.

O ambiente de união e entre-ajuda que sempre se viveu no Laboratório de Sistemas Autónomos (LSA) foi determinante para o desenvolvimento desta tese. Gostaria de começar por destacar os meus colegas de ano curricular, por todo o companheirismo: André Faria, Bruno Matias, Carlos Almeida, João Ribeiro e Tiago Santos. Aos meus colegas e amigos Luís Lima e Pedro André Silva, por todo o apoio e motivação ao longo desta jornada.

Não teria sido possível desenvolver este trabalho sem a aquisição de dados realizada no âmbito do projeto ESAPlastics. A todos os que diretamente ou indiretamente contribuíram para o sucesso do projeto, o meu agradecimento. Em especial aos meus colegas, que fizeram todos os esforços possíveis para que tivesse os dados necessários ao desenvolvimento desta tese: Alexandre Amaral, Carlos Almeida, Diana Viegas, Paulo Rodrigues e Tiago Santos. À equipa do Aeroclube de Torres Vedras, que participou com tanta dedicação na recolha dos dados do projeto ESAPlastics: João Francisco, Filipa e Duarte. Agradeço também a toda a equipa do Okeanos – Universidade dos Açores, orientada por Christopher K. Pham, cujo empenho permitiu a colocação dos alvos artificiais no mar, apesar das condições adversas.

À minha família que sempre acompanhou com entusiasmo, amor e carinho todo o meu percurso académico, em especial aos meus avós e irmã.

Aos meus pais, a quem estarei para sempre grata por me terem motivado desde

criança a estudar e lutar por um futuro melhor. Por todos os esforços, todas as vezes que apoiaram as minhas decisões, irei estar para sempre em dívida por tudo o que fizeram por mim.

Ao meu marido, Carlos Ferrás, que me apoiou ao longo de todos os dias. Obrigada pela paciência, apoio, carinho, compreensão e motivação ao longo de toda a minha jornada académica.

Contents

Abstract **i**

Resumo **iii**

Acknowledgements **v**

Acronyms **xix**

1 Introduction **1**
1.1 Motivation 1
1.2 Objectives 4
1.3 Contributions 5
1.4 Publications 6
1.5 Thesis Overview and Organisation 9

2 Related Work **11**
2.1 Introduction 11
2.2 Taxonomy 11
2.3 Remote Sensing Platforms using Hyperspectral Imaging Sensors (HSI) 13
2.3.1 Satellite 14
2.3.2 Unmanned Aerial Vehicles (UAV) 16
2.4 Hyperspectral Imaging Classification 20
2.4.1 Supervised Hyperspectral Imaging Classification 20
2.4.1.1 Probabilistic Approaches 21
2.4.1.1.1 Multinomial Logistic Regression (MLR) . . 21
2.4.1.1.2 Sparse Multinomial Logistic Regression (SMLR) 22
2.4.1.1.3 LOgistic Regression via Splitting and Augmented Lagrangian (LORSAL) 22
2.4.1.1.4 Multinomial Logistic Regression (MLRsub) 22
2.4.1.2 Kernel Methods 23
2.4.1.2.1 Support Vector Machines (SVM) 23

2.4.1.2.2 Kernel Fisher Discriminant (KFD) 23
2.4.1.3 Decision Trees 24
2.4.1.4 Ensemble Methods (Multiple Classifiers) 24
2.4.1.4.1 Random Forests 24
2.4.1.5 Sparse Representation Classifiers (SRCs) 25
2.4.1.6 Neural Networks (NN) 25
2.4.1.7 Deep Learning 25
2.4.1.7.1 Autoencoder (AE) and Stacked Autoencoder (SAE) 26
2.4.1.7.2 Deep Belief Network (DBN) 26
2.4.1.7.3 Convolutional Neural Network (CNN) . . . 26
2.4.2 Unsupervised Hyperspectral Imaging Classification 28
2.4.2.1 Centroid-based Clustering 29
2.4.2.1.1 Hard Partition-based Clustering 29
2.4.2.1.2 Soft Partition-based Clustering 29
2.4.2.2 Density-based Clustering 30
2.4.2.3 Probability-based Clustering 30
2.4.2.4 Bionics-based Clustering 31
2.4.2.5 Intelligent Computing-based Clustering 31
2.4.2.5.1 Single-Objective-based Clustering 31
2.4.2.5.2 Multiobjective-based Clustering 31
2.4.2.6 Graph-based Clustering 32
2.4.2.6.1 Complete Graph-based Clustering 32
2.4.2.6.2 Bipartite Graph-based Clustering 32
2.4.2.6.3 Abbreviated Graph-based Clustering . . . 33
2.4.2.7 Subspace Clustering 33
2.4.2.7.1 Spectral-Spatial Subspace Clustering . . . 33
2.4.2.7.2 Multiview Subspace Clustering 34
2.4.2.7.3 Kernel Subspace Clustering 34
2.4.2.8 Hybrid Mechanism-based 34
2.5 Zero-Shot Learning 35
2.5.1 Inductive Zero-Shot Learning 37
2.5.2 Transductive Zero-Shot Learning 39
2.5.3 Few-Shot Learning 40
2.5.4 Generalised Zero-Shot Learning 41
2.5.4.1 Embedding-based Methods 42
2.5.4.2 Generative-based Methods 43

2.5.5 Semantic Representations in Zero-Shot Learning 44
2.5.6 Hyperspectral Imaging Zero-Shot Learning Applications . . . 46
2.6 Classification Methods Discussion 47
2.7 Application Oriented Perspective (Marine Litter) 49
2.8 Summary 56

3 Fundamentals **57**
3.1 Introduction 57
3.2 Hyperspectral Imaging 57
3.2.1 Working Principle 58
3.2.1.1 Whiskbroom cameras 58
3.2.1.2 Tunable filter cameras 59
3.2.1.3 Pushbroom cameras 59
3.2.2 Data Acquisition 62
3.3 Hyperspectral imaging classification techniques 64
3.3.1 Dimensionality Reduction 64
3.3.1.1 Principal Component Analysis (PCA) 64
3.3.2 Machine Learning-based Methods for Hyperspectral Imaging Classification 65
3.3.2.1 Support Vector Machines 65
3.3.2.2 Random Forests 68
3.3.3 Deep Learning-based Methods for Hyperspectral Imaging Classification 68
3.3.3.1 Autoencoder (AE) and Stacked Autoencoder (SAE) 69
3.3.3.2 Convolutional Neural Network (CNN) 70
3.3.3.3 Zero-Shot Learning 71
3.4 Performance Metrics 71
3.5 Summary 73

4 Data Acquisition Setup **75**
4.1 Introduction 75
4.2 ESAPlastics Project 75
4.3 Hardware Configuration Acquisition Setup 77
4.4 In-situ Characterisation 78
4.4.1 In-situ Marine Litter Samples Characterisation 79
4.5 Flight Data Acquisition 80
4.5.1 Dataset Campaign 80

4.5.2 Hyperspectral Imaging System Setup - Manned Aircraft (F-BUMA) . . . 82
4.5.2.1 F-BUMA Flight Data . . . 83
4.5.3 Hyperspectral Imaging System Setup - Unmanned Aircraft (Grifo-X) . . . 84
4.5.3.1 Grifo-X Flight Data . . . 84
4.6 Summary . . . 88

5 Supervised Approaches for Hyperspectral Imaging Classification 89
5.1 Introduction . . . 89
5.2 Support Vector Machines and Random Forests - Implementation . . . 89
5.2.1 Results . . . 92
5.2.2 Discussion . . . 95
5.3 CNN-3D Spectral/Spatial Marine Litter Detection and Classification - Implementation . . . 96
5.3.1 Results . . . 98
5.3.2 Discussion . . . 98
5.4 Summary . . . 99

6 Zero-Shot Learning Marine Litter Detection 101
6.1 Introduction . . . 101
6.2 Implementation . . . 101
6.2.1 Feature Extractor . . . 102
6.2.2 Classifier . . . 103
6.2.3 Decoder . . . 104
6.2.4 Discriminator . . . 104
6.3 Marine Litter Dataset Acquisition and Experimental Setup . . . 105
6.4 Results . . . 109
6.5 Discussion . . . 111
6.6 Summary . . . 116

7 Conclusions and Future Work 117
7.1 Conclusions . . . 117
7.2 Future Work . . . 119

A Hyperspectral Imaging Datasets 121

B Dataset Campaign Logbook 123

C Hyperspectral Cameras Characteristics **127**

List of Figures

1.1 INESCTEC Centre for Robotics and Autonomous Systems robotics platforms for water and air application scenarios. The platforms are equipped with multiple sensors, such as GPS/INS, cameras, sonars, and other sensors. 2

1.3 Main absorption bands for PolyEthylene Terephthalate (PET), Low-Density Polyethylene (LDPE), PolyPropylene (PP), PolyVinyl Chloride (PVC), and PolyStyrene (PS) plastic types 4

2.1 Related Work Taxonomy 12

2.2 Timeline representing satellites with multispectral or hyperspectral imaging systems, with the following information: satellite launch date, spectral range and the number of bands, ground sampling distance and swath width, respectively. 17

2.3 Supervised approaches taxonomy. 21

2.4 Unsupervised approaches taxonomy. 28

2.5 Marine Litter found in the test site, Porto Pim Bay (Faial Island, Azores). 50

3.1 Hyperspectral data cube 58

3.2 Hyperspectral data cube 59

3.3 Whiskbroom hyperspectral camera acquisition scheme 59

3.4 Tunable filter hyperspectral camera acquisition scheme 60

3.5 Pushbroom hyperspectral camera acquisition scheme 60

3.6 UAV pushbroom scan methodology. 61

3.7 Specim FX10 (400-1000 nm), HySpex Mjolnir S-620 (970-2500 nm) and Cubert S 185 (450-950 nm), respectively (not to scale). 61

3.8 Hyperspectral data processing pipeline. 63

3.9 Support Vector Machine linear example 66

3.10 Support Vector Machine not linearly separable example 66

3.11 Random Forest algorithm scheme. 68

3.12 Autoencoder layer example. 69
3.13 Convolutional Neural Network (CNN) architecture example. 70
3.14 Zero-Shot Learning (ZSL) scheme for hyperspectral imaging data . . 72

4.1 ESAPlastics project scheme: the artificial targets are placed in the water, while the manned and unmanned aircraft and the satellite are acquiring data. 76
4.2 Connections between the hyperspectral cameras, data acquisition computers and GPS/INS. 77
4.3 Marine litter samples collected in the test site - Porto Pim bay, Faial Island, Azores 78
4.4 Test setup used during the in-situ tests. 79
4.7 Hyperspectral imaging system mounted in the manned aircraft (F-BUMA). 83
4.8 Waterfall image of the flyby analysed from the 18/9 afternoon F-BUMA flight. 83
4.9 Results for the 18/9 F-BUMA afternoon flight, analysing three different pixels (bright pixel-afternoon 1, a medium pixel-afternoon 2 and a dark pixel-afternoon 3) for each material, and comparing them to the laboratory results. 84
4.10 Hyperspectral imaging system mounted in the unmanned aircraft (Grifo-X). 84
4.11 Waterfall image of the flyby analysed from the 25/9 Grifo-X flight 1. 85
4.12 Results for the 25/9 Grifo-X first flight, analysing three different pixels (bright pixel-1, a medium pixel-2 and a dark pixel-3) for each material, and comparing them to the laboratory results. 86
4.13 Comparison between the bright pixels for the 25/09 flight 1 with Grifo-X, 20/09 F-BUMA flight and the laboratory morning results for all the target materials. 87
4.14 Comparison between the rope target results for F-BUMA flight (600 m altitude), Grifo-X flight (20 m altitude) and laboratory setup, with a focus on HySpex Mjolnir S-620 bands (970 - 2500 nm). 87

5.1 Grifo-X flying over the targets during the dataset campaign. 96
5.2 Flowchart of the proposed CNN-3D architecture. 97
5.3 Left image flyby 4 waterfall image, right image flyby 5 waterfall image. 98
5.4 Left image flyby 4 ground-truth image data, right image flyby 4 CNN-3D result. 98

5.5 Left image flyby 5 ground-truth image data, right image flyby 5 CNN-3D result. 98

6.1 Schematic of the algorithm used: first, the PCA method is used to reduce the data dimensionality for both the training and prediction phase. First, a feature extractor with channel attention modules after each convolutional layer is used to train the algorithm to create a semantic space Z. Then, this semantic space is used for the classifier, a fully connected neural network, and by the decoder to create a reconstructed spectrum. The classifier and the discriminator are used to perform the prediction, giving a predicted label to each pixel. . . 103
6.2 On the left-hand image, the F-BUMA aircraft, and on the right side of the image, the hyperspectral imaging payload installation on the aircraft landing gear. 106
6.3 Image acquired from the aircraft during a flight with the artificial targets placed in the water. 106
6.4 Reconstructed image for each hyperspectral camera used. (Left image shows the Specim FX-10e data, while the right image shows the HySpex Mjolnir S-620 data. 110
6.5 Left image holds the ground truth while the right image shows the ZSL results. 111
6.6 ZSL results zooming on one of the artificial target's detections (including the two unknown classes (class 1, identified with the colour red, and class 3, identified with the colour pink). 112

A.1 Hyperspectral imaging datasets: the images on the first line are obtained using only one wavelength, while the images on the second line represent the ground truth of each dataset. 121

B.1 The first line of images shows the Grifo-X and F-BUMA platforms during the dataset campaign. The second line shows the targets being prepared and placed in the marine litter hotspot for the dataset campaign. 123

C.1 Specim FX10 (400-1000 nm) and HySpex Mjolnir S-620 (970-2500 nm), respectively (not to scale). 127

List of Tables

2.1 Division of ZSL methods 37
2.2 Inductive ZSL methods summary 38
2.3 Transductive ZSL methods summary 39
2.4 Marine litter detection state-of-the-art. 54

3.1 Confusion matrix example. 72

4.1 Targets used during the dataset collection and their division between the classes. 80
4.2 Other artefacts captured during the dataset campaign. 82

5.1 Targets used during the dataset collection and their division between the classes. 90
5.2 Number of pixels (NP) present for each class in 18/9 F-BUMA flight: class 1 corresponds to the orange target data, class 2 to the white target and class 3 to the rope target data. Class 0 was created with all the remaining pixels containing land and water data. 90
5.3 Number of pixels (NP) present for each class in 20/9 F-BUMA flight: class 1 corresponds to the orange target data, class 2 to the white target and class 3 to the rope target data. Class 0 was created with all the remaining pixels containing land and water data. 91
5.4 Results obtained for the test data (20/09/2020 flight) for Random Forests (RF) and Support Vector Machine (SVM) algorithms. 93
5.5 Figures show the results obtained for the test data (20 September 2020 flight) for RF and SVM algorithms. The first column illustrates the ground truth for each flyby, the second column shows the prediction made by the random forest algorithm, and the last column represents the predicted values by the support vector machine algorithm. . . . 94
5.6 Number of pixels (NP) present in the Grifo-X dataset used for each class 97
5.7 Flyby 4 CNN-3D Overall Results 99

5.8 Flyby 5 CNN-3D Overall Results 99

6.1 Classes used for the ZSL algorithm, and main characteristics e.g., description, known and unknown information, image and number of pixels. 107
6.2 ZSL Results. 111
6.3 Results comparison using RF, SVM, CNN and ZSL methods. 113

A.1 Brief description of the most used hyperspectral and multispectral imaging datasets, adapted from [Shafaey et al., 2021]. 122

B.1 F-BUMA aeroplane dataset campaign details 124
B.2 Grifo-X aircraft dataset campaign details 125

C.1 Specim FX10e Characteristics 128
C.2 HySpex Mjolnir S-620 Specifications 128

Acronyms

AE	AutoEncoder
AMD	Anthropogenic Marine Debris
CNN	Convolutional Neural Network
EO	Earth Observation
ESA	European Space Agency
FN	False Negative
FP	False Positive
FTIR	Fourier-Transform Infra-Red Spectroscopy
GAN	Generative Adversarial Network
GPS	Global Positioning System
GPU	Graphics Processing Unit
GZSL	Generalised Zero-Shot Learning
HDPE	High-Density PolyEthylene
INS	Inertial Navigation System
ISODATA	Iterative Self-Organising Data Analysis Technique Algorithm
LDA	Linear Discriminant Analysis
LDPE	Low-Density Polyethylene
LIBS	Laser-Induced Breakdown Spectroscopy
LIDAR	Light Detection And Ranging
ML	Machine Learning
MS	MultiSpectral
NN	Neural Network
OA	Overall Accuracy
PCA	Principal Component Analysis
PE	PolyEthylene

PET PolyEthylene Terephthalate

PP PolyPropylene

PS PolyStyrene

PVC PolyVinyl Chloride

ReLU Rectified Linear Unit

RF Random Forests

RGB Red Green Blue

SVM Support Vector Machine

SWIR Short-Wave InfraRed

TOA Top-Of-Atmosphere

TN True Negative

TP True Positive

UAV Unmanned Aerial Vehicle

VNIR Visible and Near-InRrared

XRF X-Ray Fluorescence

ZSL Zero-Shot Learning

1 Introduction

1.1 Motivation

The Centre for Robotics and Autonomous Systems[1] has been involved in cutting-edge field robotics applications. The application domains extend from the deep sea to space using underwater, surface and aerial vehicles, see Figure 1.1. Applications include environmental monitoring, inspection, rescue and security, close underwater caves exploration and mapping, deep sea landers for geological and seismic monitoring, bathymetric surveys in open water environments and underwater mining, to remote hyperspectral imaging systems for marine litter detection.

The improvement of computers and sensors has allowed the use of different technologies applied to multiple autonomous systems. One of these technologies is the multispectral and hyperspectral cameras, which, until recently, were only used in static indoor controlled environments. The development of novel hyperspectral imaging systems, combined with their reduction in weight and dimensions, has promoted their application in different domains, such as remote sensing, industry, inspection, and surveillance. Hyperspectral imaging systems can collect information in the spatial and spectral domain, acquire information along the electromagnetic spectrum for each pixel, and create a 3D hyperspectral cube that combines spatial and spectral information. New hardware systems, lighter and smaller, have enabled the implementation of new parallel and complex algorithms capable of processing the huge

[1] `https://www.inesctec.pt/en/centres/cras`, accessed on 7 September 2022



Figure 1.1: *INESCTEC Centre for Robotics and Autonomous Systems robotics platforms for water and air application scenarios. The platforms are equipped with multiple sensors, such as GPS/INS, cameras, sonars, and other sensors.*

amount of acquired hyperspectral imaging data.

Multispectral and hyperspectral imaging systems have been widely used in satellites for multiple purposes, such as environmental monitoring, forestry, land use/land cover (LU/LC) and water analysis. However, existing multispectral systems installed in the available satellites do not meet the spatial and spectral resolution requirements for some applications, such as marine litter detection and classification, which is the application scenario of this thesis. Furthermore, the data collected from these systems at higher altitude is highly influenced by atmospheric factors, such as clouds between the satellite and the scene of interest.

Multispectral imaging systems are similar, however, they collect information in fewer wavelengths, influencing the amount of data that can acquire. These improvements in hyperspectral imaging systems combined with the ones of the robotic vehicles allowed the use of hyperspectral imaging technology in different platforms, such as their use in Unmanned Aerial Vehicle (UAV) and satellites for remote sensing.

Large marine litter accumulation zones, namely marine litter windrows [Ruiz et al., 2020], are appearing in the oceans, such as Atlantic [Ryan, 2015a] and Pacific [Lebreton et al., 2018, McIlgorm et al., 2011]. Marine litter is one of the biggest threats to marine ecosystems, as shown in Figures 1.2(a) and 1.2(b). The characteristics of the marine litter found in the oceans has been changing over the years and are primarily made up of human-based plastic waste that threatens marine life. Therefore, maritime stakeholders and research communities are now working on active/passive systems for detecting and/or capturing marine litter debris. A recent research topic is in detecting and monitoring marine litter accumulation zones, since

the marine litter accumulation zones are spread over large areas. Earth Observation (EO) satellite-based technological resources are currently employed to detect and monitor marine litter. However, these efforts are still in the early stage of their development and still require further experimental validation using complementary detection solutions, using remote and/or in-situ observations. To address this problem, the research community is always looking for novel tools, methods, and systems to detect, classify and ultimately remove marine litter accumulation zones from the sea. To accomplish this goal, it is essential to characterise marine litter, in order to discover its origin, and help create mitigation strategies to decrease its presence in the oceans.



(a) (b)

Figure 1.2: *The image on the left shows a bird surrounded by plastic (photo by Matthew Chauvin), while the one on the right shows a turtle entangled in a fishing net (photo by Francis Perez).*

Marine litter hotspots appear all around the globe. Therefore, retrieving information from remote sensing airborne and satellites can be an interesting solution. These systems can have different payload sensors, such as multi/hyper-spectral imaging systems, visible cameras or RADAR. Despite being an expensive and complex technology, hyperspectral imaging systems can cope better with exogenous situations such as solar reflections, which are a significant concern for RGB cameras due to their higher probability of returning a completely saturated image. It is also insensitive to physical changes in marine litter due to sun exposure and erosion and can collect information in hundreds of wavelengths, from visible to infrared spectrum.

Although hyperspectral imaging systems are now smaller, their price varies with the electromagnetic spectrum range. Infrared cameras tend to be much more expensive, which limits the number of research groups dedicated to developing novel methods for Short-Wave InfraRed (SWIR) classification. However, as depicted in Figure 1.3, most of the principal plastic types present in the marine litter (PET, PP, PVC, PS. PolyEthylene (PE) and LDPE) have their features in the SWIR range. Hyperspectral imaging marine litter datasets are scarce, and even the sys-

tems with the spectral and spatial requirements needed for marine litter detection and classification are practically non-existent.

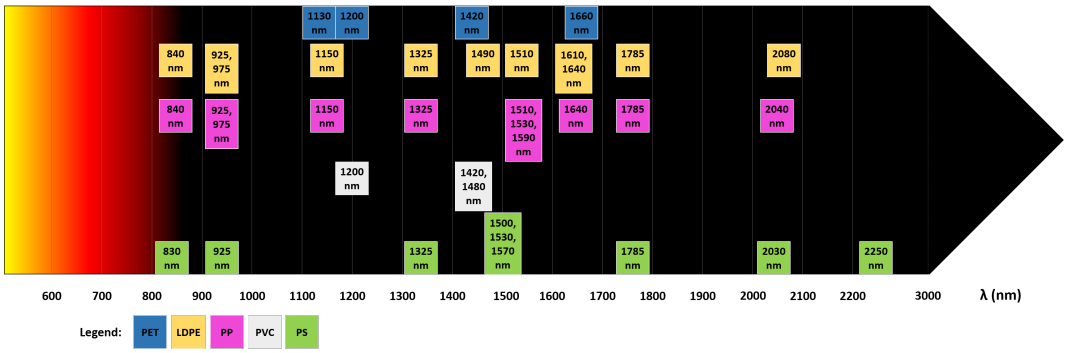


Figure 1.3: *Main absorption bands for PET, LDPE, PP, PVC, and PS plastic types*

Hyperspectral imaging data processing has evolved from anomaly detection methods to machine learning and, nowadays, to deep learning techniques. Deep learning techniques have shown excellent potential to classify this type of data. However, most approaches are supervised and require extensive data to train the model. This makes it difficult to organise a marine litter hyperspectral imaging dataset campaign due to the resources needed (systems, aerial vehicles, permissions to fly over the hot spots, among others) and the number of different types of plastics present in the marine litter. Therefore, it is necessary to create new methodologies to classify marine litter plastic samples for which they never received information during the training process.

In this thesis, we used a customised hyperspectral imaging system to perform extensive data acquisition using remote sensing platforms, to develop novel automated methods for detecting and classifying marine litter in an oceanic environment.

1.2 Objectives

The thesis main objective is to contribute to the hyperspectral image classification remote sensing topic by using robotic platforms for civil space applications. We are particularly focused on studying and developing novel learned-based methods that use hyperspectral imaging information to detect, classify and identify different types of marine litter. To accomplish this task, several objectives need to be fulfilled, namely:

1. Perform the study and implementation of a novel hyperspectral imaging system, including the study of the related work concerning learning-based ap-

proaches for hyperspectral imaging that can be used for marine litter detection and classification;

2. Create a digital hyperspectral imaging database (build a dataset) using manned and unmanned aerial vehicles, in a marine litter hotspot, for marine litter detection. This objective will allow not only to perform extensive data collection for the thesis but also to serve as a benchmark for future developments in this area of research;

3. Study, develop and implement supervised and unsupervised methods to perform hyperspectral image classification. We start it by studying state-of-the-art techniques, using machine learning and deep learning techniques. Followed by novel optimization techniques, image processing and machine learning techniques, deep learning, and optical sensing information. This will provide the knowledge fundamentals for the thesis and will allow the development and exploration of different types of methods and approaches for extracting a variety of spectral, textural and geometrical attributes of the images;

4. Development of a novel semi-supervised technique to perform hyperspectral imaging classification of unknown data. These will allow to a decrease the data needed to train the algorithm and also to classify new samples without previous information about their classes;

5. Apply the developed methods in a real demonstration scenario and carefully evaluate the obtained results.

1.3 Contributions

This work contributes to developing a novel semi-supervised algorithm for marine litter classification using hyperspectral imaging systems.

- The thesis's first contribution is the acquisition of a novel hyperspectral marine litter dataset using manned and unmanned aerial vehicles. This dataset was acquired using a hyperspectral imaging system with a 400-2500 nm spectral range in a marine litter hotspot (Porto Pim bay, Faial Island, Azores). The hyperspectral imaging system consisted of two cameras: Specim FX10e (400-1000 nm) and HySpex Mjolnir S-620 (970-2500 nm). To be captured by Sentinel-2, three artificial targets with marine litter found in the test site were used and, before the dataset campaign, were analysed under laboratory conditions. This dataset campaign resulted in approximately 2TB of data from

both hyperspectral cameras and the Global Positioning System (GPS)/Inertial Navigation System (INS) system.

- Evaluate state-of-the-art hyperspectral detection and classification methods (Random Forests/Support Vector Machines) [Freitas et al., 2021b] to create a benchmark for the marine litter dataset acquired. These methods are machine learning supervised approaches, which implies providing the algorithm with labelled data during the training stage. Both methods were tested using acquired data from the artificial targets, resulting in an overall accuracy of over 97% for all passages with both methods;

- Development of a novel CNN-3D [Freitas et al., 2021a] based on our previous work [Freitas et al., 2019], using spatial and spectral information for marine litter detection. This is a supervised approach, meaning that all the classes must be labelled and given to the model for training proposes. This Neural Network was tested with the data from the artificial targets, and the results show over 84% overall accuracy in the detection and classification of the different types of marine litter present in the artificial targets;

- The last contribution of this thesis is the first semi-unsupervised hyperspectral imaging classification method used for marine litter detection [Freitas et al., 2022], based on a generalised Zero-Shot Learning approach. The main difference of this method is that some of the marine litter samples were hidden from the model during the training phase, denoted as unknown classes. However, these samples were classified into new classes during the classification phase. The results show a marine litter automated detection for all the classes, including in the worst case of an unknown class of over 56% precision rate and overall accuracy of 98.71%.

1.4 Publications

This section briefly states the publications made within the thesis scope and their abstracts.

Convolutional neural network target detection in hyperspectral imaging for maritime surveillance

Sara Freitas, Hugo Silva, José Miguel Almeida and Eduardo Silva
International Journal of Advanced Robotic Systems May-June 2019
(10.1177/172988141984299)

This work addresses a hyperspectral imaging system for maritime surveillance using unmanned aerial vehicles. The objective was to detect the presence of vessels using purely spatial and spectral hyperspectral information. To accomplish this objective, we implemented a novel 3-D convolutional neural network approach and compared against two implementations of other state-of-the-art methods: spectral angle mapper and hyperspectral derivative anomaly detection. The hyperspectral imaging system was developed during the SUNNY project, and the methods were tested using data collected during the project final demonstration, in São Jacinto Air Force Base, Aveiro (Portugal). The obtained results show that a 3-D CNN is able to improve the recall value, depending on the class, by an interval between 27% minimum, to a maximum of over 40%, when compared to spectral angle mapper and hyperspectral derivative anomaly detection approaches. Proving that 3-D CNN deep learning techniques that combine spectral and spatial information can be used to improve the detection of targets classification accuracy in hyperspectral imaging unmanned aerial vehicles maritime surveillance applications.

Hyperspectral Imaging System for Marine Litter Detection

Sara Freitas, Hugo Silva, Carlos Almeida, Diana Viegas, Alexandre Amaral, Tiago Santos, André Dias, Pedro A. S. Jorge, Christopher K. Pham, José Moutinho and Eduardo Silva
OCEANS 2021 San Diego - Porto
(10.23919/OCEANS44145.2021.9705953)

This work addresses the use of hyperspectral imaging systems for remote detection of marine litter concentrations in oceanic environments. The work consisted on mounting an off-the-shelf hyperspectral imaging system (400-2500 nm) in two aerial platforms: manned and unmanned, and performing data acquisition to develop AI methods capable of detecting marine litter concentrations at the water surface. We performed the campaigns at Porto Pim Bay, Fail Island, Azores, resorting to artificial targets built using marine litter samples.During this work, we also developed a Convolutional Neural Network (CNN-3D), using spatial and spectral information to evaluate deep learning methods to detect marine litter in an automated manner. Results show over 84% overall accuracy (OA) in the detection and classification of the different types of marine litter samples present in the artificial targets.

Remote Hyperspectral Imaging Acquisition and Characterization for Marine Litter Detection

Sara Freitas, Hugo Silva and Eduardo Silva
MDPI Remote Sensing 2021
(10.3390/rs13132536)

This work addresses the development of a remote hyperspectral imaging system for detection and characterization of marine litter concentrations in an oceanic environment. The work performed in this paper is the following: (i) an in-situ characterization was conducted in an outdoor laboratory environment with the hyperspectral imaging system to obtain the spatial and spectral response of a batch of marine litter samples; (ii) a real dataset hyperspectral image acquisition was performed using manned and unmanned aerial platforms, of artificial targets composed of the material analyzed in the laboratory; (iii) comparison of the results (spatial and spectral response) obtained in laboratory conditions with the remote observation data acquired during the dataset flights; (iv) implementation of two different supervised machine learning methods, namely Random Forest (RF) and Support Vector Machines (SVM), for marine litter artificial target detection based on previous training. Obtained results show a marine litter automated detection capability with a 70–80% precision rate of detection in all three targets, compared to ground-truth pixels, as well as recall rates over 50%.

Hyperspectral Imaging Zero-Shot Learning for Remote Marine Litter Detection and Classification

Sara Freitas, Hugo Silva and Eduardo Silva
MDPI Remote Sensing 2022
(10.3390/rs14215516)

This work addresses the development of a novel zero-shot learning method for remote marine litter hyperspectral imaging data classification. The work consisted of using an airborne acquired marine litter hyperspectral imaging dataset that contains data about different plastic targets and other materials and assessing the viability of detecting and classifying plastic materials without knowing their exact spectral response in an unsupervised manner. The classification of the marine litter samples was divided into known and unknown classes, i.e., classes that were hidden from the dataset during the training phase. The obtained results show a marine litter automated detection for all the classes, including (in the worst case of an unknown

class) a precision rate over 56% and an overall accuracy of 98.71%.

1.5 Thesis Overview and Organisation

The first chapter starts with an introduction of the thesis context, motivation and how the work developed in the scope of this thesis relates to marine litter detection and classification. It also describes clearly the objectives and contributions of the thesis.

Chapter 2 described the related work, starting with a historical overview of the hyperspectral imaging systems in Remote Sensing, from Satellite applications to Unmanned Aerial Vehicles (UAV). Then it briefly analyses the hyperspectral imaging classification algorithms by dividing this topic into Supervised and Unsupervised approaches. This chapter also describes the Zero-Shot Learning (ZSL) approaches in all configurations: inductive, transductive, Few-Shot Learning and Generalised Zero-Shot Learning (GZSL). Finally, it covers the application-oriented perspective of marine litter, showing the works developed in this scope.

Chapter 3 describes the key fundamental concepts of hyperspectral imaging, such as their working principle and data acquisition, followed by the description of the hyperspectral imaging classification techniques used in this thesis.

Chapter 4 describes in detail the data acquisition setup, starting with the integration of this setup in the ESAPlastics project, followed by a description in detail of the hardware configuration of the data acquisition setup. It also covers the in situ characterisation of the marine litter samples and the flight data acquisition, starting with the dataset campaign using a manned and an unmanned aircraft.

Chapter 5 presents the Supervised approaches for hyperspectral imaging classification used in this thesis. It starts by describing the implementation and results of the two state-of-the-art methods: Random Forests and Support Vector Machines. It also shows the implementation of a novel CNN-3D using spectral and spatial information for marine litter classification and the obtained results.

In chapter 6 is shown a novel Zero-Shot Learning method for marine litter detection based on a generalised approach. This is a semi-supervised approach, where the model is trained with only samples from a few classes. During the classification stage, the algorithm can classify samples from classes that were never seen by the model, denoted as unknown classes. The results obtained with the data collected during a marine litter dataset campaign are presented and analysed.

The last chapter presents the conclusions and outcomes of the thesis and the relevant future work.

Appendix A briefly compares the main hyperspectral imaging datasets most used in the scientific community.

Appendix B presents a detailed description of the flights realised in the scope of the marine litter dataset campaign using both manned and unmanned aircraft.

Appendix C details the hyperspectral cameras used in the data acquisition setup by showing the main characteristics of both cameras.

2
Related Work

2.1 Introduction

This chapter introduces the major topics of the related work concerning the thesis subject. We will perform a hierarchical approach to the related work, defining a taxonomy and mentioning significant contributions from using hyperspectral imaging systems in remote sensing applications. Starting with the distinct types of sensors utilised, namely multispectral and hyperspectral imaging systems, followed by the remote sensing platforms where the sensors are utilised as payload and how this influences the obtained results. Afterwards, we go into a more technical description of methods with their pros and cons related to the spectral/spatial analysis of the hyperspectral data for classification tasks, which is the core topic of the work conducted during this thesis. Finalising with an in-depth discussion of the different methods and approaches for marine litter detection and classification, which is the end motive of our targeted remote hyperspectral imaging learning-based systems.

2.2 Taxonomy

This thesis addresses research topics whose related work is spread over different research domains and is usually tailored to tackle specific applications. To be able to aggregate all the relevant state-of-the-art, we defined a taxonomy to classify the different elements of the state-of-the-art more coherently. For that, we divided the

related work into three major groups as depicted in Figure 2.1, namely:

- Platforms
- Sensors
- Methods

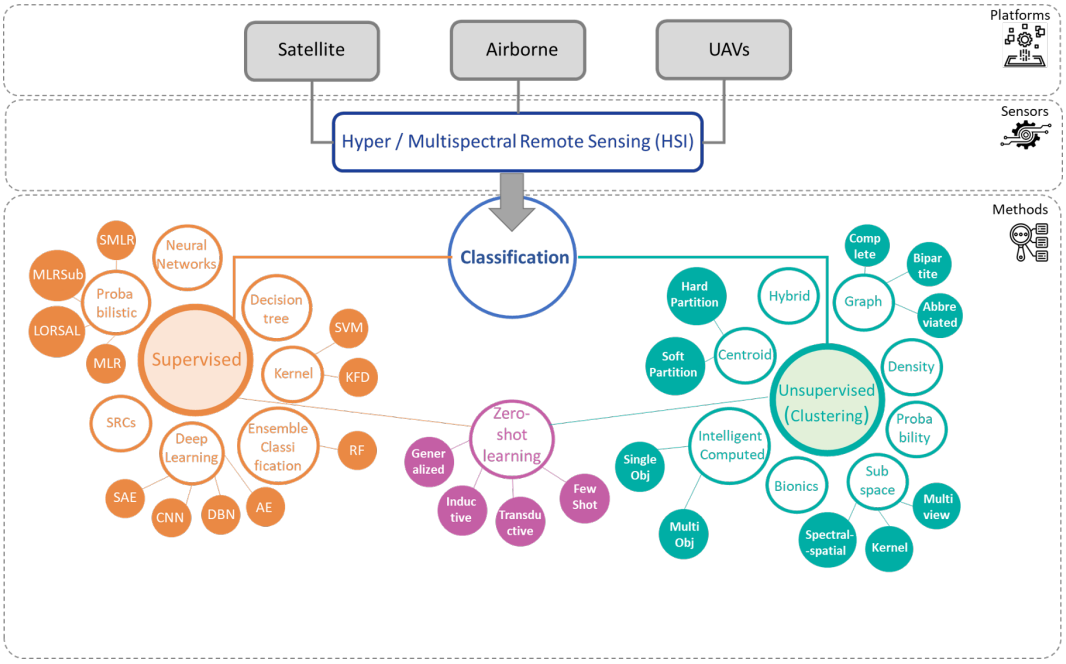


Figure 2.1: *Related Work Taxonomy*

In the Platforms and Sensors groups, there is an introduction to the type of sensors and platforms used for performing remote sensing using multispectral and hyperspectral imaging sensors. It starts with a historical perspective from the first satellite applications to the more recent drone systems and how the miniaturisation of the hyperspectral sensors allowed to foster the range of applications for remote hyperspectral imaging systems. Concerning the developed methods and their characterisation, the state-of-the-art is related to the use of spatial and spectral information to classify hyperspectral imaging data. The mentioned work is divided into three clusters to accomplish this objective: Supervised, Semi-Supervised, and Unsupervised techniques. For each cluster applications, methods that resort to hyperspectral spatial and spectral data are mentioned, including the advantages and disadvantages of each technique. Since one of the major novelties of this thesis is

the use of remote sensing semi-supervised methods for marine litter detection, a more detailed discussion introducing the concepts of ZSL and their taxonomy is performed, including a detailed description of the available works for marine litter detection that includes hyperspectral imaging and other sensors information.

2.3 Remote Sensing Platforms using Hyperspectral Imaging Sensors (HSI)

Hyperspectral and Multispectral Image Sensors share the same working principles while varying the number of available wavelengths (bands). Multispectral image sensors have few wavelengths, generally between 4 to 13 wavelengths. In contrast, hyperspectral image sensors can collect information from hundreds of bands. Despite that, both can provide valuable information according to the desired target application. Results are also influenced by other sensor parameters and operational parameters, such as atmospheric conditions, altitude, and sensor resolution, to name a few. Remote Sensing applications resorting to Multispectral/Hyperspectral Imaging Sensors started in the early seventies with the NASA Landsat-1 satellite. The satellite was launched and went into orbit in July 1972 and was designed and developed to perform Earth Observation using two instruments: a camera system and multispectral sensor with four spectral bands (two on the visible range, red and green bands, and two other infrared bands). Despite being a breakthrough at the time with applications from forest canopies to water analysis [Nelson, 1983, Malila, 1980, Carpenter and Carpenter, 1983], researchers soon realised that the system was unable to capture the information of all relevant wavelengths, for example, when applied to minerals identification [Goetz, 1995].

The discovery motivated the development of novel systems for airborne and satellite applications, such as the first airborne hyperspectral imaging system built in the 1980s, named the Airborne Imaging Spectrometer (AIS) [Gross and Klemas, 1986]. It successfully performed flights for mineral identification, encouraging the development of new solutions, such as the well-known AVIRIS (Airborne Visible InfraRed Imaging Spectrometer) [Vane et al., 1993]. This system has flown on four different aircraft: NASA's ER-2 jet, Twin Otter International's turboprop, Scaled Composites Proteus and NASA's WB-57. Using a hyperspectral whiskbroom acquisition scheme, the AVIRIS system collects information on 224 wavelengths between 380 and 2500 nm, with 677 spatial pixels. Considering the instrument installed in NASA's ER-2 jet, with an altitude of 20 km, each pixel covers approximately an area of 20 meters in diameter on the ground. The data is grouped in scenes containing

512 lines, corresponding to approximately 10 km on the ground. The AVIRIS instrument has been used for several applications, such as vegetation analysis [Cheng et al., 2006, Smith et al., 2003], mineral mapping [Kumar et al., 2020, Kruse, 2002], mapping coastal waters [Neuenschwander et al., 1998], micro and macro plastics analysis [Garaba and Dierssen, 2018], among others. The sensor was used to collect the data for the most utilised hyperspectral imaging datasets in the scientific community, namely, the Indian pines [Baumgardner et al., 2015], Salinas and Kennedy space centre[1].

Although airborne platforms present a good solution regarding payload capacity, each campaign requires a specialised pilot and staff, permits to take off, airport and aircraft availability, and also each campaign usually only covers a specific area. To overcome some of these challenges, other platforms are also utilised, namely satellites and UAV. These platforms will be described in more detail in the following sections.

2.3.1 Satellite

The first hyperspectral imaging system onboard a satellite was launched in the late 1990s. The SPIMs 1-5 hyperspectral system was installed onboard the military use MSX satellite, with a spectral range from 110 to 900 nm, and records data in 1360 spectral bands simultaneously, with a ground sampling distance of 770 meters [Carbary et al., 1994]. Another hyperspectral imaging sensor is the Hyperion sensor, installed in the EO-1 satellite, launched in the year 2000 and finished its operation in 2017. It was the first remote sensing hyperspectral imaging system installed on a satellite to give more visibility and information to the scientific community through data availability. The sensor collected information of 220 wavelengths from 357 to 2576 nm, with a thirty meters ground sampling swadth width [Middleton et al., 2013]. The data from this hyperspectral sensor was used for several applications, such as vegetation classification [Miura et al., 2003, Asner and Heidebrecht, 2003, Thenkabail et al., 2004] detection and mapping [Carter et al., 2009, Pengra et al., 2007], biophysical parameters estimation [Pu et al., 2005, Pu et al., 2008], geology applications [Hubbard and Crowley, 2005, Cudahy et al., 2001], soil applications [WENG et al., 2010], land cover applications [Petropoulos et al., 2012a, Ben-Arfa et al., 2015], urban applications [Weng et al., 2008, Falcone and Gomez, 2005], water resource applications [Dekker et al., 2001a, Giardino et al., 2007] and disaster prevention and post-monitoring applications [Yoon and Kim, 2007, Roberts et al., 2003].

[1] `https://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes`, accessed on March 16th 2022

The European Space Agency (ESA) Project for On-Board Autonomy (PROBA) satellite was launched in 2001, and it is still operational and acquiring data. The initial goal was to prove technologies in space. However, it exceeded all expectations by surviving more than the expected two years. It carries two Earth Observation imaging instruments: the Compact High-Resolution Imaging Spectrometer (CHRIS) and the High-Resolution panchromatic Camera (HRC). CHRIS has up to 62 spectral bands between the 415 to 1050 nm spectral range, and the ground sampling distance varies between 25 to 50 meters [Barnsley et al., 2004]. Depending on the operation mode, the number of bands, wavelength range, spectral bandwidth and ground sampling distance can change. This sensor has been used for several applications, such as forest mapping [Glowienka and Zembol, 2022], canopy biophysical and biochemical parameters monitoring [Stagakis et al., 2010], water monitoring [Mannheim et al., 2004, Lavigne and Ruddick, 2021], macroalgae communities analysis [Casal et al., 2011], land cover (LU/LC) classification [Jin et al., 2012].

ESA launched the ENVISAT satellite on March 1st 2002, carrying a hyperspectral imaging system, MERIS [Bézy et al., 2015]. The payload consists of five identical imaging spectrometers with a spectral range from 390 to 1040 nm, capturing 520 wavelengths. However, it only transmits fifteen bands (each representing a mean value over eight to ten original wavelengths) to decrease the amount of transmitted data. The ground sampling distance was three hundred meters, and it ended its mission in 2012. It was developed for oceans colour analysis, including coastal areas, to assess water quality and to study the oceanic biology [Woerd and Wernand, 2015, Schroeder et al., 2007, Moore et al., 1999, Schiller and Doerffer, 2005, Attila et al., 2013, Campbell et al., 2011].

Another interesting mission from ESA is the Sentinel, which already counts with seven missions. However, only two carried multispectral or hyperspectral imaging systems for remote sensing applications. Sentinel-2 (S2) mission is composed of a constellation of two polar-orbiting satellites, phased at 180° to each other, launched on June 23rd 2015, and March 7th 2017 [Spoto et al., 2012]. The main goal is to monitor the variability of land surface conditions, such as agriculture, land ecosystems, forest management, inland and coastal water quality, disaster mapping and civil security. Both satellites (S2A and S2B) were equipped with a multispectral instrument with thirteen bands between 492 and 1376 nm. The spatial resolution varies between 10 and 60 meters depending on the wavelength. The data collected from this system is used for vegetation classification [Vaglio Laurin et al., 2016, Karlson et al., 2020], soil applications [Castaldi et al., 2016], land cover [Jakovels et al., 2016, Torma et al., 2012b, Torma et al., 2012a, Hunger et al., 2016], water resource applications [Van-

hellmont and Ruddick, 2016, Toming et al., 2016, Du et al., 2016, Beck et al., 2016] and disaster analysis [Fernández-Manso et al., 2016].

Sentinel-3 (S3) is the second satellite of this mission equipped with a pushbroom hyperspectral imaging system, denoted as Ocean and Land Color Imager (OLCI) [Donlon et al., 2012], and was launched on February 2016. The OLCI can capture information from 390 to 1040 nm, with 520 bands and 300 meters of ground sampling distance. However, it only transmits twenty-one bands, using a similar scheme to the MERIS system. The main goal of this system was to provide the following measurements: ocean, inland sea, coastal zone colour, sea surface temperature, and topography [Peschoud et al., 2017, Smith et al., 2018, Kravitz et al., 2020].

The Italian satellite, PRISMA, was launched on March 2019, [Loizzo et al., 2018]. The hyperspectral imaging system carried by this satellite is a pushbroom scanner with two spectrometers: one in the Visible and Near-InRrared (VNIR) from the 400 to 1010 nm range and another in the SWIR from the 920 to 2505 nm wavelength. It has 30 meters ground sampling distance (GSD) and can retrieve information from 237 bands. It has been used in different fields, such as agriculture and land monitoring [Pignatti et al., 2013], geological applications [Tripathi and Garg, 2021], marine litter detection and classification [Taggio et al., 2022] and water analysis [Bresciani et al., 2022].

EnMAP is a German spaceborne imaging spectrometer mission to monitor and characterise the Earth environment [Guanter et al., 2015]. It was launched on April 2022 and carries a hyperspectral imaging system capable of collecting 244 bands of information from 420 to 2500 nm. It was developed to extract geochemical, biochemical and biophysical measurements to model dynamic Earth processes and better understand climate and land cover changes, water availability, natural resources, and hazard and risk assessment.

The aforementioned mission considerably impacts the scientific community due to multispectral/hyperspectral data availability. Still, other missions also have an impact and are depicted in a timeline in Figure 2.2. The Figure also holds additional information, including the satellite launch date, spectral range, number of bands, ground sampling distance and swath width.

2.3.2 Unmanned Aerial Vehicles (UAV)

New remote sensing applications have used UAV platforms. The technology has become more affordable, and their respective payloads include multispectral/ hyperspectral imaging systems. Using these platforms also allowed to bypass of problems with the satellite data, such as the low spatial and spectral resolution that limits

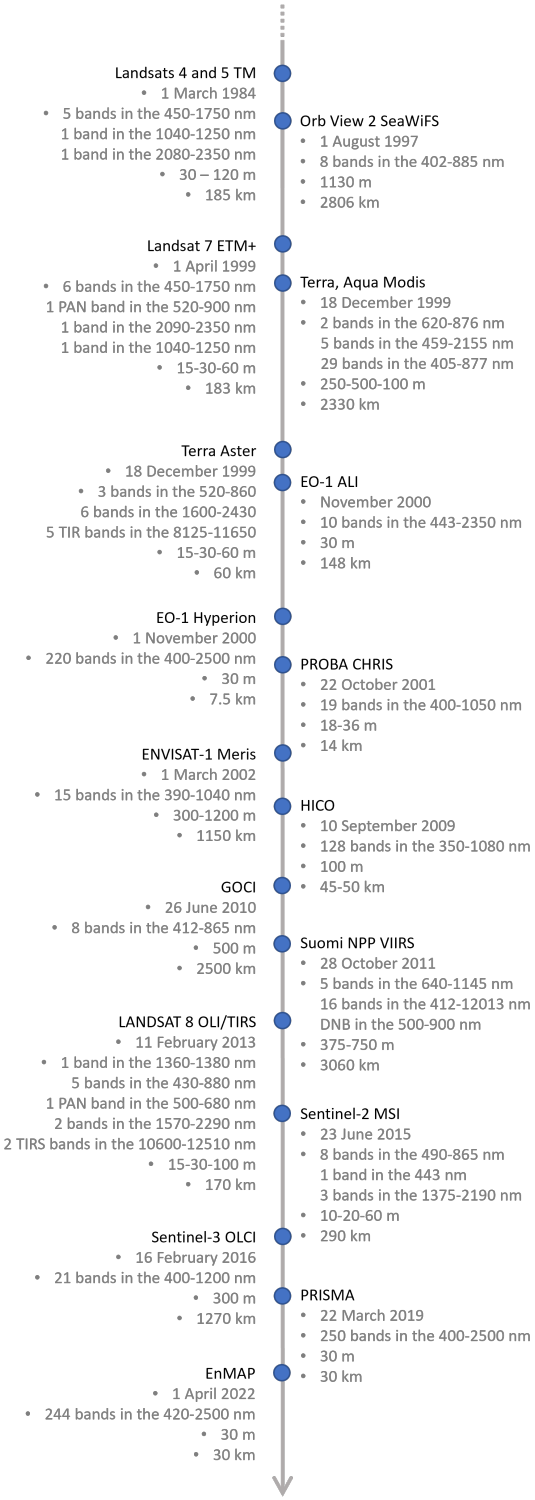


Figure 2.2: *Timeline representing satellites with multispectral or hyperspectral imaging systems, with the following information: satellite launch date, spectral range and the number of bands, ground sampling distance and swath width, respectively.*

results for more complex environments applications [Banerjee et al., 2020].

To accomplish this objective, the hyperspectral imaging manufacturers made an effort to decrease the size of the system to allow their usage in UAV, some examples of these systems are:

- Cubert S 185 [2]: Launched in 2012, this 490 g light-weight system has a spectral range from 450 to 950 nm, with 125 bands. The field of view varies according to the lens selected: 30°, 20°, 13°, 7°. It was the first hyperspectral tunable filter camera installed on a UAV for aerial mapping purposes.

- Headwall Photonics Micro-Hyperspec [3]: This pushbroom camera was launched in 2014, and offers multiple configurations for the wavelength range: 400 to 1000 nm, 900 to 1700 nm, 600 to 1700 nm and 900 to 2500 nm. The weight varies accordingly to the wavelength range from 0.7 kg to 2 kg and the number of bands from 100 to 369 nm wavelength.

- Resonon Pika II [4]: This camera covers the spectral range from 400 to 900 nm, and it was also launched in 2014. It collects information on 240 wavelengths and weighs 1.3 kg.

- Rikola [5]: The prototype developed by VTT (Technical Research Center of Finland) and Rikola Ltd, in 2012, was introduced into the market in 2013. This tunable filter hyperspectral camera weighs less than 600 g and has a 37° field of view, with a spectral range from 500 to 900 nm.

These systems were the first to be developed and used as UAV payload. As can be observed, most of them are in the visible range (approximately from 400 to 1000 nm) and can capture information of 125 wavelengths. The main difference between these systems when compared to older models is in the size and weight, which made it possible for their integration into different platforms, including UAV [Deng et al., 2020].

UAV based hyperspectral imaging systems have been used in several applications, such as leaf area index estimation in vineyards [Kalisperakis et al., 2015], where a VNIR pushbroom hyperspectral camera was installed in a UAV to estimate canopy

[2] `https://www.cubert-hyperspectral.com/products/firefleye-185`, accessed on May 11th 2022

[3] `http://www.analytik.co.uk/wp-content/uploads/2016/03/micro-hyperspec-datasheet.pdf`, accessed on May 11th 2022

[4] `https://cdn.exiteme.com/exitetogo/www.sk-advanced.com/userfiles/files/ResononHyperspectralCameras_Datasheet.pdf`, accessed on May 11th 2022

[5] `http://www.ursus-airborne.com/Home/Index/56`, accessed on May 11th 2022

levels. Land vegetation monitoring [Sankey et al., 2018] where the authors proved that it is possible to use a pushbroom hyperspectral camera with 272 spectral bands ranging from 400 to 1000 nm in order to perform the fusion with Light Detection And Ranging (LIDAR) data and generate a LIDAR point cloud with the ground and vegetation canopies classification. There are several studies about plant diseases analysis and surveillance using hyperspectral imaging systems installed in UAV, such as [Vanegas et al., 2018] where the authors perform plant pest surveillance in vineyards and crops, [Abdulridha et al., 2019] in which the system is used to detect citrus canker disease and [Abdulridha et al., 2020] that shows the possibility of detecting powdery mildew disease in squashes by using hyperspectral data. All these works resort to hyperspectral imaging systems in the visible and near-infrared spectral range (400-1000 nm). [Zhang et al., 2019b] resort to an UAV to install a snapshot hyperspectral camera for yellow rust fungal disease detection in winter wheat. The hyperspectral camera that was used collected a 50 $\times$ 50 pixel image with 125 bands between 450 and 950 nm spectral range.

UAV equipped with hyperspectral imaging systems have also been used for ecological monitoring [Arroyo-Mora et al., 2019] using a novel sensor with 288 bands within the 401 to 996 nm spectral range. It has also been extended to water quality assessment, such as suspended solids concentration analysis to assess the degree of water pollution [Wei et al., 2019], river transparency analysis [Wei et al., 2020] 400-1000, and mapping of intertidal macroalgae [Rossiter et al., 2020], by using hyperspectral cameras in the 400 to 1000 spectral range.

However, these applications using hyperspectral imaging systems installed in UAV did not contemplate the SWIR (1000 to 2500 nm) range due to the sensor characteristics and needs in terms of size and price. Given that UAV also evolved into versions with higher payload capacities, it enabled installing bigger or multiple systems in the same aircraft. These improved capacity allowed novel applications such as marine litter detection [Balsi et al., 2021, Cocking et al., 2022], that needs wider spectral ranges to detect the right features [Garaba and Dierssen, 2018, Balsi et al., 2018].

Despite the evolution of hyperspectral systems, there is still a lack of low-height and size SWIR systems. One example of these systems is the HySpex Mjolnir S-620 [6]: Launched in 2018 by HySpex, this is a hyperspectral imaging system composed by the hyperspectral camera, GPS/INS and data acquisition unit, with a 6.5 kg weight including batteries. The hyperspectral camera can collect information on 300 bands

[6] `https://www.hyspex.com/hyspex-products/hyspex-mjolnir/hyspex-mjolnir-s-620/`, accessed May 11th 2022

from 970 to 2500 nm wavelength.

2.4 Hyperspectral Imaging Classification

Hyperspectral image classification methods aim to classify each pixel as belonging to a single class. Most hyperspectral imaging data classification methods use spectral information or a combination of spectral and spatial information.

Hyperspectral imaging systems can acquire huge amounts of data and are, therefore, suitable candidates for developing machine learning and deep learning models to perform classification tasks. Usually, these techniques are divided in the literature into two major categories, namely: (1) Supervised approaches, which require training a model before the classification stage. Making it necessary to plan the datasets to have enough unique data for training and fine-tuning the models; (2) Unsupervised approaches, where there is no need to have a model. Usually, based on clustering techniques group the data according to their similarity as they are assigned to different clusters. Supervised techniques tend to have better results than Unsupervised ones, however, their dependency on the data before classification can be a drawback in some applications. Hybrid approaches try to take advantage of the strengths of both supervised and unsupervised approaches, denoted as semi-supervised approaches. Typically, these approaches resort to Supervised techniques to train an initial model and then to unlabelled samples to improve the model, or by using Unsupervised techniques to classify samples from other classes during the classification phase. An example of a Semi-Supervised approach is the Zero-Shot Learning method, which is one of the main subjects of this thesis and will be described in detail later in this chapter. In the following sections, we will detail both Supervised and Unsupervised algorithms for classification using hyperspectral imaging.

2.4.1 Supervised Hyperspectral Imaging Classification

Supervised methods require information about classes during the training phase so they can learn their characteristics. However, despite this requirement, it is still possible to divide these methods into parametric and non-parametric methods.

Parametric methods have a finite number of parameters (fixed structure). These methods assume that sample data follows a given distribution based on fixed parameters, such as linear regression, logistic regression, and supervised maximum likelihood classifier.

The non-parametric methods do not make strong assumptions about the form,

and the number of parameters is (potentially) infinite. The complexity of the model grows with the amount of training data. Some examples of these methods are K-Nearest Neighbour, Decision Trees, Support Vector Machines, Neural Networks, and Random Forests, among others. Figure 2.3 shows the taxonomy of the Supervised approaches presented in this work.

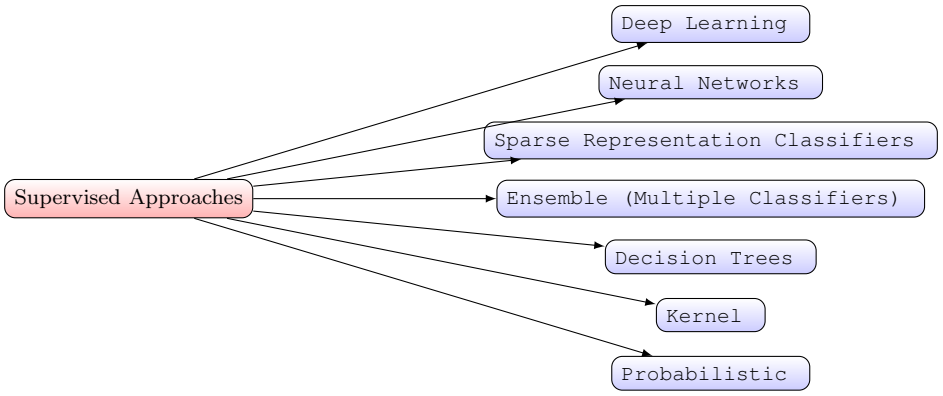


Figure 2.3: *Supervised approaches taxonomy.*

2.4.1.1 Probabilistic Approaches

A common method to classify hyperspectral data is using probabilistic approaches. The main difference of these methods is that instead of returning a class for each data sample, the output is a probability of the pixel being a member of each possible class. The disadvantage is that these approaches require prior knowledge about the statistical distribution of the classes. This section presents some probabilistic approaches-based methods, including Multinomial Logistic Regression, Sparse Multinomial Logistic Regression, Logistic Regression via Splitting and Augmented Lagrangian and MLRsub. The study of these methods has two main advantages: a suitable generalisation of the algorithm with fast computational execution and using the open and flexible structure of the Multinomial Logistic Regression Classifiers.

2.4.1.1.1 Multinomial Logistic Regression (MLR)

This algorithm can be considered a semi-supervised method used for high-dimensional data. There are two main steps associated with the implementation of the Multinomial Logistic Regression (MLR) [Li et al., 2010]:

1. Learning of the posterior class distributions, using a semi-supervised approach - unlabelled training samples are used to improve the class distributions;

2. Segmentation, using an image of class labels from a later distribution built on the learned class distributions and a Markov Random field. The posterior class distributions are modelled with Multinomial Logistic regression, and the regressor resorts to labelled and unlabelled data with a graph-based technique. Classification results obtained from the regression model are improved by performing a maximum posterior segmentation as it considers the spatial information of the hyperspectral image. The prior image of labels uses a multilevel logistic model, enforcing segmentation results in which neighbouring labels belong to the same class.

This method is suitable for situations where training data is reduced and is not labelled. This classifier can directly learn the posterior class distributions and deal with the high dimensionality of the hyperspectral data.

2.4.1.1.2 Sparse Multinomial Logistic Regression (SMLR)

This method starts with the Multinomial Logistic Regression, used to calculate the posterior class probabilities modelled by this algorithm [Zhong et al., 2008, Borges et al., 2011]. To improve this method for hyperspectral classification, it is possible to adopt a Laplacian prior to enforcing sparsity, which leads to good machine generalisation capabilities in hyperspectral classification, despite some drawbacks related to computational limitations.

2.4.1.1.3 LOgistic Regression via Splitting and Augmented Lagrangian (LORSAL)

This method allows the processing of hyperspectral images that are diversified in the number of available classes and can also deal with large datasets in an efficient manner [Bioucas-Dias and Figueiredo, 2009, Li et al., 2011]. It is possible to learn a sparse regressor with a Laplacian prior distribution of the SMLR. Combining these two methods (SMLR e LORSAL) is one of the most effective ways to deal with the high dimensionality of hyperspectral imaging data.

2.4.1.1.4 Multinomial Logistic Regression (MLRsub)

This Supervised method integrates spectral and spatial information in a Bayesian framework. First, using a Multinomial Logistic Regression (MLR) to learn the posterior probability distributions from the spectral information, with a subspace projection method to characterise better noise and highly mixed pixels [Li et al., 2012]. Then the segmentation step infers an image of class labels from a posterior

distribution built on the learned subspace classifier and a multilevel logistic prior to the label image. Finally, the output is based on a maximum a posteriori (MAP) segmentation process.

2.4.1.2 Kernel Methods

The Kernel methods are based on mapping data from the original input feature space to a Kernel feature space of higher dimensionality and solving a linear problem. These methods allow interpreting (and designing) learning algorithms geometrically in the kernel space (which is nonlinearly related to the input space), thus effectively combining statistics and geometry. Their functional performance also matches this theoretical elegance.

2.4.1.2.1 Support Vector Machines (SVM)

SVM is a Supervised learning method commonly used in classification tasks. SVM has been used in different fields, and in the case of hyperspectral data, this method quickly became a reference because SVM was able to obtain accurate results similar to much more complex computational algorithms, with a lower computational time [Petropoulos et al., 2012b]. The goal is to find the optimal hyperplane in n-dimensional classification space with the highest margin/separation between classes. The margin corresponds to the distance between the hyperplane and the nearest data point from either class. In the next chapter, this issue is addressed detail. This algorithm was one of the methods selected to establish a state-of-the-art benchmark in the detection and classification of marine litter samples.

2.4.1.2.2 Kernel Fisher Discriminant (KFD)

Fisher Discriminant analysis is a standard in statistical data analysis that allows defining the separation between two distributions (each one for one class), and the separation between these two distributions is defined as the ratio of the variance between the classes to the variance within the classes [Yang et al., 2007].

However, this is a linear method, which means it has the same problem as linear SVM: it does not work if the data is nonlinear. Based on this limitation, the kernel method was introduced, which allows to map of the data to another space and performs the Fisher Discriminant Analysis in this new space.

2.4.1.3 Decision Trees

Decision Trees are another non-parametric approach that can be used for classification. During the training phase, the training set is split into an increasing number of smaller and more homogeneous groups. The advantage of this method is that the decision tree is structurally simple, and training only requires a short time [Goel et al., 2003].

With this method is possible to directly interpret class membership decisions concerning the impact of individual features. The classifier can be improved using classifier ensembles. [Chan and Paelinckx, 2008].

2.4.1.4 Ensemble Methods (Multiple Classifiers)

The Ensemble methods represent the effective combination of classifiers such that each can benefit the others while avoiding the weakness of each. An ensemble of classifiers is trained, and their individual results are combined in a voting process [Joelsson et al., 2007, Waske et al., 2012]. This is used to increase classification accuracies.

2.4.1.4.1 Random Forests

RF can be seen as a particular case of Decision Trees. However, RF is composed of multiple classifiers, which implies special features that make this method different from traditional classification trees.

To train RF usually is used the general technique of bootstrap aggregating, or bagging, to tree learnings [Breiman, 1996]. This technique is used for training data creation by randomly resampling the original dataset, with replacement [Rodriguez-Galiano et al., 2012]. Strongly correlated trees are produced if many trees are trained on a single training set. The bootstrap sampling can decorrelate the trees by showing them different training sets. RF uses the trees as base classifiers $h(x, \theta_k), k = 1, ...,$, where x and θ_k are the set of input vectors and the independent and identically distributed random vectors. To make the classifier more robust is possible to improve the classifier stability by using more than once the same data for training [Waske et al., 2009].

To generate the RF model, there are only two parameters: the number of trees, a free parameter that can be selected because of the size and nature of the training set, and the number of predictor variables. This method is flexible, can deal with different scenarios (large number of features, small or large dataset, among others), and is easy to evaluate. This method was used as one of the state-of-the-art methods

to benchmark the performance of the marine litter classification methods developed in the scope of this thesis.

2.4.1.5 Sparse Representation Classifiers (SRCs)

Suppose the input signal is represented by a sparse linear combination of samples (atoms) from a dictionary, where the training data is usually used as a dictionary. In that case, it is called Sparse Representation Classifiers (SCR) with dictionary-based generative models [Huang et al., 2017]. This method's advantage is that it avoids the heavy training procedure that a Supervised classifier usually requires since the classification is performed directly on the dictionary.

2.4.1.6 Neural Networks (NN)

Neural Network (NN) approaches began to attract many researchers in the 1990s. The main advantage of these algorithms over the parametric methods is that they do not require a fixed distribution - no prior knowledge about the statistical distribution of classes is needed. Multiple weights and non-linearities describe the NN, and the weights are computed using an iterative training procedure. There are multiple approaches in the Supervised and Unsupervised domains. However, all are non-parametric [Benediktsson et al., 1990,Merényi et al., 2014,Del Frate et al., 2007,Ratle et al., 2010, Zhong and Zhang, 2012].

An example of a widely used NN is the feed-forward NN, which has been well studied since the introduction of the backpropagation algorithm [Rumelhart et al., 1986], a first-order gradient algorithm to parameter optimisation.

However, the backpropagation algorithm has two main problems: slow convergence and the possibility of falling in local minima. To solve this problem, multiple second-order optimisation-based algorithms have been created [Moller, 1993].

2.4.1.7 Deep Learning

In recent years, multiple deep models have been used for hyperspectral data feature extraction and classification. The architecture design is the most crucial part of a successful deep-learning model, one of the problems with this type of method is that it may lead to overfitting, which means that the results can be perfect in the training data but have poor accuracy when benchmarked using the test data. Examples of deep learning techniques are:

2.4.1.7.1 Autoencoder (AE) and Stacked Autoencoder (SAE)

This is an Unsupervised method, which means that it uses unlabelled training data. This algorithm uses the restricted Boltzmann machine (RBM) technique to pre-train the network, and then the obtained weights are finely tuned via backpropagation. An Autoencoder aims to learn a representation (encoding) for a set of data that minimises information loss. The AutoEncoder (AE) is composed of one visible layer of d inputs, one hidden layer of L units, and one reconstruction layer of d units [Özdemir et al., 2014].

When multiple layers of AE are stacked together layer by layer is called a stacked Autoencoder. [Windrim et al., 2019].

2.4.1.7.2 Deep Belief Network (DBN)

The restrict Boltzmann Machine (RBM) is used as a layer-wise training model in constructing a DBN [Chen et al., 2015, Zhong et al., 2016]. The weights between the visible and hidden units are learned using the method of constructive divergence. However, because of the complexity of the hyperspectral data, RBM is not the best way to capture the features. After the training phase of RBM, the learned features can be used as input for the next RBM, building a DBN. This method is very useful when the number of training samples is limited. This type of NN solves the backpropagation problem, which can often lead to local minima. This is solved using an extra step during the training phase, called pre-training. This step is done before backpropagation, and to get close to the final solution, backpropagation is used to reduce the error rate slowly.

2.4.1.7.3 Convolutional Neural Network (CNN)

This algorithm was inspired by neuroscience, and a complete CNN stage is constituted by a convolutional layer with the nonlinear operation and a pooling layer [Romero et al., 2016, Acquarelli et al., 2018]. In the case of spectral classifiers, one-dimensional CNN is only needed. When multiple convolution layers with a nonlinear operation and several pooling layers, it is possible to formulate a CNN. This can be used to extract the features of inputs hierarchically, which is usually invariant and robust.

CNN hyperspectral imaging classification approaches can be divided according to their data inputs, such as:

- If the method only uses spectral information as presented by [Li et al., 2014, Hu et al., 2015b], it is possible to use a 1-D CNN. In this case, the spectral feature

of the original image data is used as an input vector, and the 1-D CNN receives N 1 input vectors (N is the number of spectral bands). [Chen et al., 2016, Zhang et al., 2016b, Ghamisi et al., 2017, Hu et al., 2015a];

- Another approach is based solely on spatial information. This model considers the neighbouring pixels of a certain pixel in the original input image to extract the spatial feature representation. This results in a 2-D CNN, where the input data is a patch of $P \times P$ neighbouring pixels only in the XY space (spatial domain) [Chen et al., 2016];

- Another approach is a hybrid approach that considers both types of data (spectral and spatial), where it is possible to combine more information and improve the classification accuracy. This architecture is known as 3-D CNN, and the input data is a patch with all spectral data associated ($P \times P \times N$, where P are the patch size, which defines the width and height in the spatial domain, and N is the number of spectral bands) as presented by [Chen et al., 2016].

Hyperspectral imaging classification has been focused on developing novel methods mainly using 3-D configurations, taking advantage of both spatial and spectral information. Some examples are the work developed by [Chen et al., 2016] where the authors use multiple convolutional layers to extract the nonlinear and invariant deep features in the spectral-spatial domain, using well-known hyperspectral datasets, see Appendix A. They also investigated some strategies to avoid overfitting during the training model phase. Another example of spectral-spatial feature classification using deep learning is given by [Zhao and Du, 2016], where the authors automatically use a CNN to find spatial-related features at higher hierarchical levels. The authors also employ a dimension reduction method before the deep CNN application. [Yue et al., 2015] performs similar work by developing a deep CNN to learn spectral and spatial features. As for [Zhang et al., 2017], it shows an application in which a CNN based on spectral information is used to extract the hierarchical spectral features automatically. Then a 2-D CNN is used to extract the hierarchical space-related features. The authors use a softmax regression classifier to combine the information, returning the predicted classification results. A data augmentation scheme is used to overcome the problem of limited hyperspectral training samples. [Gao et al., 2018] presents a novel work where a CNN architecture benefits from the multiple inputs corresponding to various image features and the concurrent exploitation of both spectral and spatial contextual information. This allows the creation of a robust and efficient network even with a small number of training samples available. [Paoletti

et al., 2018] developed a deep 3-D CNN architecture for spatial-spectral classification of hyperspectral data and also compared all the results obtained with other 1-D, 2-D, and 3-D CNNs implemented. More recent work is presented by [Firat et al., 2022], where a novel 3-D residual spatial-spectral convolution network, 3D-RSSCN, for extract deep spatio-spectral features by combining a 3-D CNN with ResNet18 architecture. Another approach is presented by [Zhang et al., 2022], where the authors resort to an Unsupervised spatial-spectral CNN-based feature learning approach to classify hyperspectral imaging datasets. The novel algorithm is divided into two branches: decoder and clustering. The training is performed based on the error obtained in the image reconstruction and pseudo-label classification, allowing learning features from unlabeled samples and consequently improving the feature discrimination in the different categories.

2.4.2 Unsupervised Hyperspectral Imaging Classification

Unsupervised methods do not require any information about the classes before the classification and do not need a training phase. These methods are usually called clustering techniques due to their characteristics. There are nine main categories when analysing clustering techniques for hyperspectral imaging classification [Zhai et al., 2021]: centroid-based, density-based, probability-based, bionics-based, intelligent computing-based, graph-based, subspace clustering, deep learning-based, and hybrid mechanism-based. Figure 2.4 shows the taxonomy of the Unsupervised approaches presented in this work.

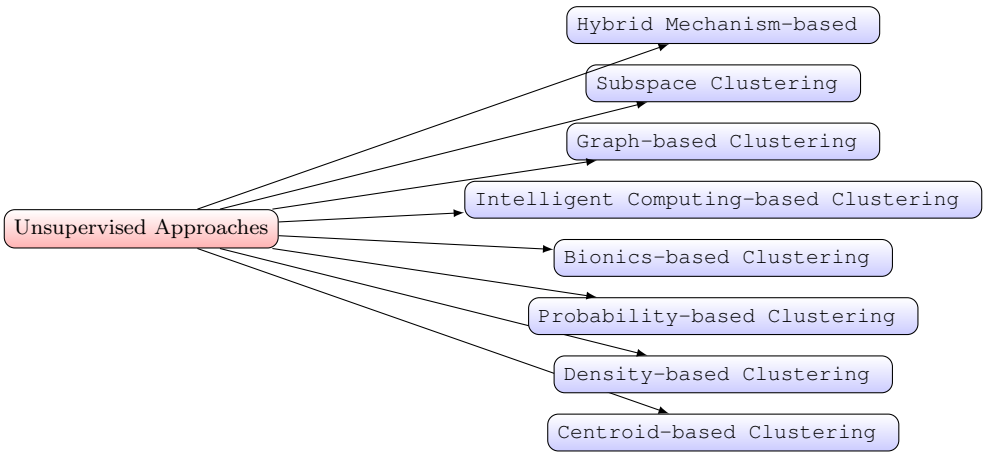


Figure 2.4: *Unsupervised approaches taxonomy.*

2.4.2.1 Centroid-based Clustering

This technique organises the data into non-hierarchical clusters by minimising the within-cluster sample distance. However, these are sensitive to initialisation and outliers and usually are unable to provide robust performance. These techniques can be divided into hard partition clustering and soft partition clustering.

2.4.2.1.1 Hard Partition-based Clustering

In this method, each pixel will belong to a single class and pixels are assigned to the nearest cluster. One algorithm that uses this rule is K-Means [Lloyd, 1982], which divides the pixels into clusters by minimising the sum of squared distances between the pixel and the centroids of all the classes. This algorithm starts with random centroids that will be updated iteratively, which means that it is sensitive to initialisation and outliers. Based on K-Means, other algorithms were developed, such as the Iterative Self-Organising Data Analysis Technique Algorithm (ISODATA) [Rahman, 2015]. The goal was to integrate a dynamic adjustment mechanism of clusters in the clustering process to improve the clustering methodology. Other improvements to the K-Means algorithm for hyperspectral imaging classification include distributed K-Means [Haut et al., 2017], Kernel K-Means [Zhao et al., 2017a], Neighbourhood-Constrained K-Means [Zhang et al., 2013], Two-Stage K-Means combined with a neighbouring union histogram [Yang et al., 2017], and Improved K-Means [Ren et al., 2019].

2.4.2.1.2 Soft Partition-based Clustering

Unlike hard-partition-based clustering, in the case of soft partition-based clustering, by considering the uncertainty of the sum of squared distances between the pixel and the centroids of all the classes, each pixel can be assigned to multiple classes. These techniques resort to a fuzzy membership assigned to each pixel in the range of [0, 1]. The sum of the memberships of all the classes is equal to one. One algorithm that uses this method is the FCM model [Ahn et al., 1999]. Several algorithms show improvements to this model, such as the weighted FCM model [Kuo and Landgrebe, 2004, Hung et al., 2011] and uncertainty analysis-based FCM model [Wang and Shi, 2013]. However, these algorithms were sensitive to initialisation.

Spatial model for fuzzy clustering [Pham, 2001], conditional FCM [Pedrycz, 1996], neighbourhood constraint clustering [Li et al., 2013], FCM with spatial information [Wang and Bu, 2010], fuzzy local information C-Means clustering model [Krinidis and Chatzis, 2010] and the fuzzy local double neighbourhood information C-

Means clustering [Zhang et al., 2018] was developed to use the combination between spectral and spatial information.

2.4.2.2 Density-based Clustering

This method resorts to a contiguous region of a high density of objects to create each cluster, which means that each cluster is considered as dense point sets separated by sparse areas in the feature space. Although these methods are less prone to noise effects and cluster shapes, finding the density peak in a sparse feature space such as the one returned in the hyperspectral imaging data is difficult. The clustering by the fast searching and finding of density peaks [Rodriguez and Laio, 2014] is one of the most used density-based algorithms. It considers that each cluster centroid is surrounded by lower-density pixels and far from higher-density pixels. Each pixel will have two quantities that must be calculated: local density and relative distance, which are used to search the optimal centroids. Enhanced versions of this algorithm were presented in [Chen et al., 2017], density analysis ensemble clustering, and [Bäcklund et al., 2011], density-based spatial clustering of applications with noise. Another example of this methodology is the enhanced spectral-spatial diffusion geometry-based clustering method [Murphy and Maggioni, 2020].

2.4.2.3 Probability-based Clustering

In this approach, the pixel division uses some likelihood criteria. However, hyperspectral imaging data usually do not follow any probabilistic distribution, making these methods perform poorly in hyperspectral imaging classification tasks. An example is using a multivariate conditional distribution to model each cluster, and the joint probability distribution can model the hyperspectral imaging data. By maximising the likelihood function with expectation maximisation, maximum posterior probability or the Bayesian rule can obtain the final clustering. One example of this algorithm is the Gaussian mixture model [Acito et al., 2003], which considers that hyperspectral imaging data pixels follow a Gaussian distribution, modelling each cluster with a multivariate Gaussian conditional distribution. Given that hyperspectral imaging data usually do not follow a Gaussian distribution, some algorithms try to model each cluster as non-Gaussian distributions, such as n independent component analysis mixture model [Shah et al., 2004, Shah et al., 2007], weighted principle component analysis ICA [Li et al., 2017] and non-parametric stochastic expectation maximum [Celeux and Diebolt, 1985].

2.4.2.4 Bionics-based Clustering

These algorithms resort to biological models to replicate the intrinsic structure of hyperspectral imaging data. However, this information suffers from spectral variability, which can reduce the model accuracy and decrease the clustering performance. The self-organising map is one of the most utilised bionics-based algorithms for hyperspectral imaging classification [Zhong et al., 2006, Baraldi and Parmiggiani, 1995]. This algorithm learns to group pixels based on their similarity and usually is composed of two layers, an input layer and a competitive layer. This last layer comprises learning and clustering stages: the first uses the Euclidean distance to select the winning neurons, updating their and neighbouring weights. During the second stage, similar pixels are mapped to the neighbouring neurons. Some approaches developed in the scope of this method are the Unsupervised Artificial Immune Classifier [Zhong et al., 2006], the Unsupervised Artificial Immune Network for remote sensing classification [Zhong et al., 2011], and Context-Aware Unsupervised discriminative extreme learning machine [Xu et al., 2018].

2.4.2.5 Intelligent Computing-based Clustering

These algorithms usually combine with other clustering methods, such as the centroid-based clustering model. The optimal global solution for the clustering model uses advanced intelligent computing algorithms, such as genetic evolution, differential evolution, and particle swarm optimisation. Given the objective functions present in the optimisation problem, it can be divided into single-objective-based clustering and multiobjective-based clustering.

2.4.2.5.1 Single-Objective-based Clustering

This method only uses one objective function for optimisation, such as fuzzy clustering using an improved differential evolution [Das et al., 2008]. This method uses a validation index to search for the optimal solution based on the differential evolution algorithm. This algorithm has some improvements, such as the modified differential evolution fuzzy clustering [Maulik and Saha, 2010, Zhong et al., 2013].

2.4.2.5.2 Multiobjective-based Clustering

Multiobjective-based clustering resorts to more than one optimisation technique, such as the multiobjective PSO [Paoli et al., 2009]. This algorithm solves three problems: clustering the statistical parameter estimation, searching for the best discriminative bands, and estimating the number of clusters using three different

optimisation criteria. Although this algorithm may have better results than traditional clustering methods, their usage is complex, narrowing their application.

2.4.2.6 Graph-based Clustering

These approaches model the relationships between the hyperspectral imaging pixels and an adjacency matrix denoted as a similarity graph. Each element of this matrix represents the similarity between a pair of pixels. This adjacency matrix's accuracy defines the results' quality, usually computed by ε-ball [Carreira-Perpiñán and Zemel, 2004], KNN, and full connection [Liu et al., 2017] strategies. Resorting to one graph cut algorithm can minimise the adjacency matrix's total cutting cost, obtaining the final clustering result. Although these are one of the most interesting hyperspectral imaging clustering methods, it is restricted by the computational power required to use this.

2.4.2.6.1 Complete Graph-based Clustering

These algorithms consider that the adjacency matrix $\tilde{\mathrm{W}}$ resorts to all pixels, containing the similarity between any pair of pixels. Spectral clustering [Shi and Malik, 2000, von Luxburg, 2007], which uses a normalised cut algorithm, is one example of this method. Another example of these algorithms is the graph-based nonlocal total variation [Zhu et al., 2017], which explores the spatial information of the hyperspectral imaging data.

2.4.2.6.2 Bipartite Graph-based Clustering

This method represents an extended version of the previous one. The main difference is that this algorithm models the relationships between two sets (denoted as the anchor and pixel set) to determine a structured similarity matrix $\hat{\mathrm{W}}$ at a larger size. [Liu et al., 2016] presents the spectral-spatial co-clustering based on a bipartite graph by extracting anchors from the cluster centroids obtained with the K-Means algorithm, which can construct a bipartite graph with centroids and pixels. Another example of these algorithms is presented in [Huang et al., 2019b], bipartite graph partition-based co-clustering with joint sparsity, which resorts to a joint-sparsity-constrained optimisation problem to determine the bipartite graph, improving the clustering accuracy.

2.4.2.6.3 Abbreviated Graph-based Clustering

This algorithm aims to decrease the computational power needed for complete graph-based clustering by selecting only the critical and representative points, denoted as key points, to construct a smaller similarity graph. These algorithms are composed of two main steps: anchor selection, corresponding to the representative pixels or cluster centroids and relation matrix construction, which models the relationships between anchors and pixels. Although these methods are more efficient, the adjacency pixel's accuracy is lower due to the smaller number of key points used to represent the hyperspectral imaging structure information. One example of this algorithm is the fast SC with an anchor graph algorithm [Wang et al., 2017a], which randomly selects the anchors from the hyperspectral imaging data.

2.4.2.7 Subspace Clustering

Subspace Clustering (SC) algorithms model the pixels from the same class with multiple spectra into a subspace and determine an approximation of the hyperspectral imaging internal structure by using a sub-spaces union, which can help to decrease the large spectral variability and consequently improve the modelling accuracy. One of the first subspace clustering methods is the SSC [Elhamifar and Vidal, 2013, Elhamifar and Vidal, 2009]. However, it has some issues, such as not considering the spatial information and the nonlinearity of the hyperspectral imaging data. The development of new algorithms to solve these issues can be divided into three main types: spectral-spatial subspace clustering, multiview subspace clustering and kernel subspace clustering. Although these methods have shown promising results for hyperspectral imaging data clustering, they require high computational complexity, time and memory.

2.4.2.7.1 Spectral-Spatial Subspace Clustering

These methods resort to spatial information to improve the subspace clustering. An example is the spectral-spatial SSC based on 3D edge-preserving filtering algorithm [Li et al., 2019a], which resorts to 3D edge-preserving filtering for the sparse coefficient matrix obtained in the previous algorithm, spectral-spatial SSC algorithm [Zhang et al., 2016a], improving the clustering results. Due to the high computational requirements needed for this method, a total variation-regularised collaborative representation clustering algorithm was developed with a locally adaptive dictionary for hyperspectral imaging data [Zhai et al., 2019].

2.4.2.7.2 Multiview Subspace Clustering

To improve the clustering performance, this method resorts to information extracted from different domains of hyperspectral imaging data: usually, each one corresponds to a specific feature domain (spectral, shape, among other feature domains). An example of this method is the spectral-spatial based multiview low-rank SSC algorithm [Tian et al., 2018b], which creates the spectral view by obtaining the correlated bands with spectral partitioning, and morphological processing to obtain the spatial view. To overcome the nonlinearity of the hyperspectral imaging data, this model was enhanced with a kernel version, named kernel-spectral-spatial based multiview low-rank SSC algorithm [Tian et al., 2018a].

2.4.2.7.3 Kernel Subspace Clustering

Due to the hyperspectral imaging nonlinear characteristics, the classes are not linearly separable. The kernel subspace clustering resorts to the junction of multiple linear subspaces to determine hyperspectral imaging data complex nonlinear internal structure. However, this will result in a significant systematic error and poor separability between the different classes, decreasing the clustering performance. This method has been improved to a kernel self-representation model to improve this issue. One example of this approach is the kernel SSC algorithm with a spatial maximum pooling operation [Zhai et al., 2017], which improves the spectral-spatial subspace clustering to include the spatial neighbourhood information by maximum spatial pooling, generating more discriminative features.

2.4.2.8 Hybrid Mechanism-based

This technique combines two or more models, combining the strengths of multiple clustering techniques to improve the clustering performance. One example of this method is the K-GMM [Neagoe and Chirila-Berbentea, 2016], which combines centroid (K-Means algorithm [Lloyd, 1982]) and probability (GMM algorithm [Acito et al., 2003]) based clustering. Another example is the fast density peaks-based clustering [Xie et al., 2018], which combined K-Means and CFSFDP [Rodriguez and Laio, 2014] algorithms. By combining the strengths of different approaches, these techniques can obtain better clustering performances.

2.5 Zero-Shot Learning

Humans can distinguish at least 30000 basic object categories [Biederman, 1987] and a considerable amount of subordinate categories, such as dog breeds. They can also dynamically create categories based on only a few examples of high-level descriptions. However, some of the most used visual computational applications, such as Convolutional Neural Networks, require using a hundred or more labelled samples for each object class to create a recognition model to identify objects.

ZSL is used in different research areas, such as face verification [Kumar et al., 2009], object recognition [Lampert et al., 2014], video understanding [Xu et al., 2016, Fu et al., 2012, Liu et al., 2011, Fu et al., 2014b] and natural language processing [Blitzer et al., 2009]. The main goal of this approach is to replicate the human capacity to recognise objects with just their description, without seeing any visual aspect of the object. The objective is to be able to classify data from unknown classes - unseen data. This technique resorts to attributes, usually given by the user or databases with class information, and the relation between the *seen*, source domain, and *unseen* classes, in the target domain. The intermediate semantic layer designates this, and ZSL approaches are designed to learn this intermediate semantic layer and apply this at inference time to predict a new class of data. This is usually done by projecting the test data to a new space (using regression or classification), and a similarity measure with each unseen class is used for prediction. Some approaches use a database with descriptive properties for each class, such as the WordNet hierarchy [Mensink et al., 2012, Rohrbach et al., 2011, Palatucci et al., 2009]. Some approaches learn this semantic representation directly from unannotated data, as in [Frome et al., 2013]. In the training stage, knowledge about the attributes is captured, and at the inference stage, this knowledge is used to categorise instances among a new set of classes.

The attributes describe the visual appearance of the concept or instance by assigning labelled visual properties. It is easy to transfer this information from seen to unseen classes, encoding linguistic context. That is why it is usually defined as a semantic attribute and given a predefined semantic representation, and an attribute vector can represent each class name. This type of representation is called a class prototype.

In the beginning, ZSL methods resorted to hand-crafted feature representations for objects. However, this type of representation has been replaced by features extracted from deep CNN, sometimes using pre-trained models to extract the features. It is usually used as input to their embedding semantic space or an intermediate

space as the embedding space. ZSL methods rely on the existence of a labelled training set of seen classes and the knowledge about how an unseen class is semantically related to the seen classes. The labelled images are only available for the source categories, while the target domain does not have access to this information.

It is possible to find different opinions in the literature. Some authors claim that extracting features from a dataset that contains zero-shot classes, that is, unseen classes [Xian et al., 2017] violates the concept of ZSL. On the other hand, some authors [Song et al., 2018, Verma and Rai, 2017, Zhang and Saligrama, 2016a, Zhao et al., 2017c] created a new division for the existing ZSL methods:

- **ZSL**: This is considered the standard approach for ZSL, which uses only the unseen domain for testing, and can be divided into two major groups:
  - **Inductive ZSL** : Only the data from the seen classes (source domain) is used during the training phase to adjust the parameters. This is the standard application of the ZSL approach (DAP/IAP [Lampert et al., 2014], VSAR [Yang Long and Shao, 2016], SJE [Zhang and Saligrama, 2016b], LatEm [Xian et al., 2016], JLSE [Zhang and Saligrama, 2016b], ConSE [Norouzi et al., 2013], SSE [Zhang and Saligrama, 2015], ALE [Akata et al., 2016], DeVISE [Frome et al., 2013], ESZSL [Romera-Paredes and Torr, 2015], SynC [Changpinyo et al., 2016], QFSL$^-$ [Song et al., 2018]);
  - **Transductive ZSL** : In this case, the unlabelled data from the unseen classes is also used during the training phase [Zhang and Saligrama, 2016a]. Although this configuration has access to more information, it is only through unlabelled data (UDA [Kodirov et al., 2015], SP-ZSR [Zhang and Saligrama, 2016c], TMV [Fu et al., 2014a], SMS [Guo et al., 2016], QFSL [Song et al., 2018]).

  In both cases, all the test data is only from the unseen classes;
- **Few-Shot Learning**: This configuration allows the use of some labelled samples from each unseen class [Salakhutdinov et al., 2013, Mensink et al., 2014]. In this case, the test data comes only from the unseen classes;
- **Generalised Zero-Shot Learning (GZSL)**: In the case of the standard ZSL and the Few-Shot learning approaches, the test data can only be from the unseen classes. However, in this case, the test data can also be from the seen classes [Chao et al., 2016, Felix et al., 2018, Verma et al., 2018].

In Table 2.1 is possible to see a scheme of the data needed during training and evaluation by the multiple ZSL methods.

		Seen Classes	**Unseen Classes**
Training	Inductive ZSL	X	
	Transductive ZSL	X	Some Unlabeled Samples
	Few-Shot Learning	X	Some Labeled Samples
	Generalised ZSL	X	
Evaluation	Inductive ZSL		X
	Transductive ZSL		X
	Few-Shot Learning		X
	Generalised ZSL	X	X

Table 2.1: *Division of ZSL methods*

As can be observed in 2.1, it is possible to verify that the inductive setting does not require data from the unseen classes during the training stage. However, in this configuration during the evaluation phase, none of the seen classes are classified, which makes the inductive setting an irrealistic configuration. Given this, the most realistic configuration of ZSL is the Generalised ZSL, in which, during the evaluation, both seen and unseen classes are classified. The following sections will analyse some of the historical inductive and transductive models, as well as the Few- and Generalised-shot learning approaches. Since the work developed belongs to the Generalised setting, the analysis in the framework of this thesis will be more focused on that approach. These methods have recently received much attention from the scientific community, developing several ZSL algorithms, mainly for visible spectrum imaging [Xie et al., 2022b].

2.5.1 Inductive Zero-Shot Learning

The inductive ZSL approach is the standard configuration for ZSL, where only the data from the seen classes is used during the training phase. This method uses data from unseen classes during the test phase.

Table 2.2 shows a brief summary of some historical inductive ZSL methods.

Table 2.2: *Inductive ZSL methods summary*

DAP [Lampert et al., 2009]	Attribute-based algorithm: to classify a new image, the attributes are predicted using the learned classifiers, and the attribute scores are combined into class-level scores. This is a probabilistic model which assumes independence, with two steps: 1) learns to attribute specific classifiers; 2) Combining attribute scores into class level scores
IAP [Lampert et al., 2014]	Compared to DAP, this algorithm indirectly estimates attribute probabilities of an image by first predicting the probabilities of each training class and then multiplying the class attribute matrix.
VSAR [Yang Long and Shao, 2016]	Removes some semantic ambiguities, improving the results. Adopt an embedding approach without dealing with the effectiveness of each attribute concept, and due to this is possible to apply VSAR to existing intermediate semantic representations.
SynC [Changpinyo et al., 2016]	Introduces "phantom" classes that connect the gap between seen and unseen classes, allowing the construction of the classifiers of the seen and unseen classes from the same base classifiers as for the phantom classes, using the same coefficient functions. This allows learning the classifier synthesis mechanism leveraging labelled data of the seen classes and applying it to the unseen classes.
SJE [Zhang and Saligrama, 2016b]	For source and target domain data exists a corresponding latent space with a similarity function independent of class labels.
LaTEM [Xian et al., 2016]	Resorts to a ranking-based loss, which ensures that different latent matrices are updated frequently. This method learns a nonlinear, piece-wise linear compatibility function for the ZSL classification.
DeVISE [Frome et al., 2013]	Trainees a visual recognition model with both labelled and a comparatively large and independent dataset, semantic information from unannotated text data. This algorithm is based on a bi-linear compatibility function that associates visual representation and semantic features. This maps the CNN visual feature vector to a semantic space by a hinge ranking loss or least square loss.

ConSE [Norouzi et al., 2013]	This algorithm uses pre-trained classifiers for seen classes and their probabilistic outputs to infer the semantic embeddings of each test example. The main difference compared to the DeVISE algorithm is the design of the NN. The classifier scores are treated as weights in a convex combination of word vectors, and the NN with a softmax classifier generates the weights for the linear combination.
SSE [Zhang and Saligrama, 2015]	Express source data as a mixture of seen class proportions, based on the principle that if the mixture proportion from the target domain is similar to the one from the source domain, both must come from the same class. With this principle, they create semantic similarity embedding functions from training data and optimise predefined parameters using cross-validation on hold-out seen class data.
ALE [Akata et al., 2016]	Aims to solve the problem of using the two-step method as DAP algorithm by learning a bilinear compatibility function between the image and the attribute space using ranking loss.
ESZSL [Romera-Paredes and Torr, 2015]	This algorithm is equal to ALE but uses the square loss to learn the bilinear compatibility and explicitly regularises the objective, resorting to the Frobenius norm.

2.5.2 Transductive Zero-Shot Learning

The transductive ZSL approach is similar to the method mentioned above. However, it uses the unlabeled data from the unseen classes during the training phase. As in the previous implementation, the test data is only from the unseen classes.

Table 2.3 shows a brief summary of some of the first approaches of ZSL transductive methods.

Table 2.3: *Transductive ZSL methods summary*

UDA [Kodirov et al., 2015]	Creates a new regularised sparse coding framework using the target domain class labels projections in the semantic space to regularise the learned target domain projection and effectively overcome the projection domain shift problem.
TMV-BLP [Fu et al., 2014a]	This method tries to solve the projection domain shift problem by rectifying the biased projection in an embedding space and also independently constructing local neighbourhood graphs for each view.

TMV-HLP [Fu et al., 2014a]	As the previous method, it also tries to solve the projection domain shift problem with a new distributed representation of graph structure, with the name heterogeneous hypergraph. With this, the authors can exploit the complementary of different semantic and low-level feature views, and it also proposes a distributed representation of pairwise similarity using heterogeneous hypergraphs.
SP-ZSR [Zhang and Saligrama, 2016c]	The authors resort to a structured prediction approach to seek a better assignment structure between clusters and unseen classes in test time. With this, they can classify unseen target data by considering the learned similarities and the empirical distribution of unlabeled target data.
SMS [Guo et al., 2016]	Transfer the information directly from source classes into the target classes. This is possible because the attributes are shared between classes, allowing the learning of class models that can directly generate the class labels from images instead of attribute models that recognise attributes. The authors also learned a shared model space for source and target classes that enables the generalisation of the recognition model using the class attributes.
QFSL [Song et al., 2018]	This method takes advantage of the transductive configuration combined with the generalised setting of ZSL to project the labelled source data into the points specified by the source classes in the semantic space, building a relationship between the visual embeddings and the semantic embeddings, and also to force the unlabelled target data to be projected to other points specified by the target classes.

2.5.3 Few-Shot Learning

Few-Shot learning is a related problem to ZSL, where some labelled unseen class samples are considered during the training phase [Lake et al., 2013, Fei-Fei et al., 2006]. This will alleviate the issue of ZSL. However, it does not solve it, as is shown in [Wang et al., 2018]. Most of the ZSL methods can be easily extended to few-shot learning, but these methods were not evaluated in few-shot learning settings.

This configuration of ZSL methods aims to learn specific classes using one or a few examples. In contrast, ZSL aims to learn even when no examples of the classes are presented, relying only on side information provided by other domains, such as textual descriptions. The few-shot learning configuration will allow a more realistic

approach to these algorithms. However, it reduces the benefits of the ZSL approach, where some classes without samples should be correctly identified and classified.

2.5.4 Generalised Zero-Shot Learning

Usually, standard ZSL only allows the use of unseen classes during the test phase. However, with the generalised ZSL configuration, it is possible to generalise the ZSL to the case where seen and unseen classes are used during the test time. During the training, this relies on the semantic features of the seen and unseen classes and the visual representation of only the seen classes.

This configuration is considered hard by some authors [Chao et al., 2016], and these authors tried to apply the standard ZSL models. However, this approach does not perform well in the GZSL approach since the trained models are strongly biased to predict the seen classes. They also propose a simple but effective method to adapt the standard ZSL approaches for GZSL by introducing a calibration factor to calibrate the classifiers for both seen and unseen classes to balance the recognising data from seen classes and those from unseen classes, called calibrated stacking.

In [Felix et al., 2018], Generative Adversarial Network (GAN) are used to generate seen and unseen synthetic examples for training GZSL models, showing clear advantages over previous approaches. However, the authors also state that the unconstrained nature of the sample generation from unseen classes can produce models that may not work robustly for some unseen classes. To tackle this problem, the authors constrain the sample generation from unseen classes, improving the GZSL classification accuracy. GZSL algorithms using have advantages when compared to non-based GZSL approaches.

The authors of [Verma et al., 2018] present a generative framework for GZSL considering that the training and test classes are not necessarily disjoint. In this work, they use a conditional variational autoencoder-based architecture and augment it with a discriminator-driven feedback mechanism to enable their model to generate class-specific exemplars from unseen classes, which can then be used in any classification model. This approach allows the learned classification model to depend on the labelled data from seen classes and synthesised examples from unseen classes.

The GZSL methods can be divided into two groups: embedding-based and generative-based [Pourpanah et al., 2022].

2.5.4.1 Embedding-based Methods

Embedding-based methods learn a projection or embedding function to associate the low-level visual features of seen classes with the semantic vectors. The learned projection function gives the similarity between the prototype and predicted representations of the samples in the embedding space.

These algorithms can resort to:

- Knowledge graphs: preserves the geometric structure of features in the latent space (SRG [Zhao et al., 2017b], AGZSL [Wang et al., 2020], DAGDA [Li et al., 2020], Xie et al. [Xie et al., 2020]) by representing each class as a node and connect it to other nodes by using the relationship edges, preserving the geometric structure of features in the latent space;

- Autoencoders: These algorithms resort to an encoder to learn an embedding space and then a decoder to reconstruct the inputs to a representation the most similar possible to the original one (Biswas and Annadani [Biswas and Annadani, 2018], LSE [Yu et al., 2019], PQZSL [Li et al., 2019d], MCGM-VAE [Shao and Li, 2020], low-rank semantic autoencoder [Liu et al., 2022]);

- Meta-learning: The goal is to collect transferable knowledge from auxiliary tasks, to allow a division of the model and consequently mitigate the overfitting issue. By dividing the training classes in two (support and query set), each one corresponding to the seen and unseen classes (Sung et al. [Sung et al., 2018], [Verma et al., 2021a], [Verma et al., 2020], [Yu et al., 2020], [Liu et al., 2021c], [Verma et al., 2021b]), several train tasks will use classes randomly from both sets, helping to mitigate the projection domain shift issue;

- Attention methods: By focusing on learning the most crucial image regions to obtain the visual features (DAZLE [Huynh and Elhamifar, 2020], AREN [Xie et al., 2019], S^2GA [Yu et al., 2018], Xie et al. [Xie et al., 2020], Yang et al. [Yang et al., 2020a], Zhu et al. [Zhu et al., 2019a], GEM [Liu et al., 2021b], SELAR [Yang et al., 2020b], SGMA [Zhu et al., 2019b], Ji et al. [Ji et al., 2022], IGSC [Yeh and Li, 2020], Paz et al. [Paz-Argaman et al., 2020]), these methods can add parameters to the deep learning model that will maximise the importance of the most discriminant features of the data;

- Compositional learning: the main objective is to learn a model that can recognise combinations of unseen compositions of known objects, such as horse and zebra, and primitive states, such as cute and old, which has been recently

used for ZSL applications (Kato et al. [Kato et al., 2018], [Chen et al., 2013], TMN [Purushwalkam et al., 2019], Sylvain et al. [Sylvain et al., 2019], [Atzmon et al., 2020]);

- Bidirectional learning methods: By using the bidirectional projections, these methods can fully use the information contained in the data and learn more general projections, which allow a better distinction between seen and unseen classes (JIL [Liu et al., 2019], DTNet [Ji et al., 2020], Guo et al. [Guo et al., 2019], LAF [Liu et al., 2020], Zhang et al. [Zhang et al., 2019a]). In this case, the visual and semantic spaces are projected together to a shared subspace. After this, the spaces are reconstructed separately using bidirectional projection learning;

- Out-of-distribution detection methods: These methods work similarly to anomaly detection techniques by identifying outliers from the dataset. In the first stage, these techniques distinguish between seen and unseen classes. To classify the seen classes, samples are used standard classifiers, while for the unseen class, samples are used standard ZSL approaches ([Lee et al., 2018], [Atzmon and Chechik, 2019], Boundary-based OOD classifier [Chen et al., 2020a], Semantic Encoding classifier [Ding et al., 2021b], Domain Detector based on Entropy Gate [Felix et al., 2019] and Integrated OOD techniques [Mandal et al., 2019, Keshari et al., 2020]);

2.5.4.2 Generative-based Methods

The Generative-based methods are used in ZSL to create new visual samples for the unseen classes by taking advantage of their semantic representations. These newly created samples must be semantically related to the real samples and discriminative to allow the classification algorithm to classify the test samples easily. Two of the most promising methods to apply this method and satisfy these conditions are:

- GAN: By computing the joint distribution $p(y, x)$ of the samples and resorting to the class conditional density $p(x|y)$ and class prior probability $p(y)$, can create new samples. This is achieved by combining a random uniform or Gaussian noise with the semantics attributes, creating a new visual feature and a discriminator to distinguish between real and the newly created samples ([Felix et al., 2018], WGAM model [Arjovsky et al., 2017], improved WGAM model [Gulrajani et al., 2017], f-CLSWGAN [Xian et al., 2018], optimal transport-based approach [Wang et al., 2021], TFGNSCS [Lin et al., 2019], LisGAN [Li et al.,

2019b], SR-GAN [Ye et al., 2019], CIZSL [Elhoseiny and Elfeki, 2019], MGA-GAN [Xie et al., 2022a], MKFNet-NFG [Xiang et al., 2021], GMMN [Li et al., 2015], AC-GAN [Odena et al., 2017], DASCN [Ni et al., 2019], CANZSL [Chen et al., 2020b], AFC-GAN [Li et al., 2019c], Boomerang-GAN [Li et al., 2022a], TACO [Chandhok and Balasubramanian, 2021]);

- Variational Autoencoders (VAE): a posterior probability $p(z|x)$ is calculated using the parameterised distribution $q_\phi(z|x)$ (where x correspond to the samples, and z to the distribution of the latent representation). VAE also must have two components: an encoder $q_\phi(z|x)$ with parameter ϕ, and a decoder $p_\theta(x|z)$ with parameter θ. While the encoder maps the sample space x to the latent space z (corresponding to class c), and the decoder maps the latent space to the sample again (CVAE [Sohn et al., 2015], CVAE-ZSL [Mishra et al., 2018], SE-GZSL [Verma et al., 2018], GDAN [Huang et al., 2019a], CADA-VAE [Schönfeld et al., 2019]);

- Combined GAN and VAE: Given that VAEs can create blurry images due to the use of element-wise distance in their structures [Larsen et al., 2016], while the GAN training process is not stable [Salimans et al., 2016], one option is to combine both methods to achieve more robust algorithms, such as VAEGAN [Larsen et al., 2016], Zero-VAE-GAN [Gao et al., 2019].

2.5.5 Semantic Representations in Zero-Shot Learning

There are two main groups for semantic representation in ZSL: semantic attributes and semantic representations beyond attributes. Inside these two main groups, there are different types:

- **Semantic attributes**: This is the standard attribute type, which aims to bridge the semantic gap and address data sparsity through transferring attribute knowledge in image and video understanding tasks. This can refer to an intrinsic characteristic of an instance or a class (as a bird) [Fu et al., 2012], or can also be used to indicate properties of the class or even annotations of an image or object [Lampert et al., 2014]

  - **User-defined attributes**: These attributes are given by human experts or concept ontology. Depending on the task, the attributes are different, such as facial and cloth attributes [Wang and Zhang, 2011, Rudd and Boult, 2016, Wang et al., 2016, Datta et al., 2011], age and gender

attributes [Wang et al., 2017b], size, colour and price attributes, and three-dimensional shape attributes [Fouhey et al., 2016].

- **Relative attributes**: Instead of using only a single value to represent the importance of an attribute possessed by one instance or class, this type of attribute can be a more informative way to express richer semantic meaning, which will result in a better represent visual information. The relative attributes appear for the first time in [Parikh and Grauman, 2011], and it was proposed to learn a ranking function capable of predicting the relative semantic strength of a given attribute. There are also some authors [Fu et al., 2016] that extended the relative attributes to subjective visual properties with a learning-to-rank model of pruning the annotation outliers/errors in crowdsourced pairwise comparisons. Other authors [Singh and Lee, 2016] used only weakly supervised pairwise image comparisons and developed an end-to-end deep Convolutional Neural Network to localise and rank relative visual attributes simultaneously.
- **Data-driven attributes**: These attributes represent more discriminative intermediate representations from visual data and have only been used in a few previous works. Data-driven attributes allow us to extend the attribute concepts in images and augment the attributes when used with user-defined attributes [Farhadi et al., 2009]. However, data-driven attributes may rediscover the patterns in the user-defined attributes because they are mined from data we do not know the corresponding semantic attribute names for the attributes. Because of this, they cannot be used directly in ZSL, which inspired more work in this area [Fu et al., 2012, Fu et al., 2014b].

- **Semantic representations beyond attributes**: Represents the semantic word vector and concept ontology, and these representations are directly learned from textual descriptions of categories, as Wikipedia pages [Elhoseiny et al., 2013, Ba et al., 2015] or sentences descriptions [Reed et al., 2016]

  - **Concept ontology**: WordNet is one of the most concept ontology, built from a large lexical dataset of English, where nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms, synsets, that indicate distinct concepts. One example of this application is given in [Fergus et al., 2010] [Fergus et al., 2010] where the WordNet is used to define the semantic distance between any two categories for sharing labels in classification;

– **Semantic word vector**: The train is done using linguistic knowledge bases, such as Wikipedia, using natural language processing models [Huang et al., 2012], allowing the projection of each class textual name into this space, and these projections can be used as prototypes for ZSL. There are also distributed word vector representations, which facilitate the modelling of syntactic and semantic regularities in language. This representation also enables vector-oriented reasoning and vector arithmetics. In this case, for example, the Vec("Lisbon") is much closer to Vec("Portugal")+Vec("capital") than Vec("Portugal") or Vec("capital") in the semantic space, and these syntactic and semantic regularities can be justified by the distributional hypothesis [Harris, 1981].

2.5.6 Hyperspectral Imaging Zero-Shot Learning Applications

ZSL with hyperspectral imaging data is still in the beginning, mainly due to the lack of datasets. However, some works are showing some advances in this field. The few-shot learning configuration is one of the ZSL approaches that have been used more with hyperspectral imaging data. One example is the algorithm proposed in [Qu et al., 2019] that resorts to Dirichlet-net based on multitasking transfer learning, aiming to preserve classification accuracy in different domains. Using a transfer learning scheme to share intrinsic representations from the same type of objects in different domains, the authors developed a few-shot learning approach for hyperspectral imaging data, with only a few samples used for training purposes. The few-shot learning algorithm was tested in the Pavia University dataset.

[Tong et al., 2020] shows a new algorithm, Attention-weighted Graph Convolution (AwGCN), where the data's internal relationships are explored for semi-supervised label propagation. First, the partial self-attention mechanism is used to learn a thorough relational graph structure of the data, then nodes labels are propagated between train and test sets using semi-supervised Graph Convolutional Networks. The algorithm was tested in three public hyperspectral imaging datasets (Indian Pines, Pavia Center and University of Pavia, see A for more information about these datasets).

[Liu et al., 2021a] presents a new approach for few-shot learning for hyperspectral imaging, MDL4OW, using a multitask deep learning method to classify and reconstruct the data simultaneously. The classification task will return the probability of known classes, while the reconstruction allows the estimation of the unknown score. The algorithm was tested in three public hyperspectral imaging datasets (Indian Pines, Salinas Valley and the University of Pavia, see A for more information

about these datasets).

[Ding et al., 2021a] shows a novel approach to mitigate over-fitting by using an extra set of unlabelled samples with soft pseudo-labels to enhance the feature extractor training on a few labelled samples. This method was tested in three public hyperspectral imaging datasets (Pavia University, Salinas and Indian Pine, see A for more information about these datasets).

The algorithm shown in [Li et al., 2022b] developed a new method called Deep Cross-domain Few-Shot Learning (DCFSL) to mitigate the domain shift issue. This is done by resorting to a conditional adversarial domain adaptation strategy and a few-shot learning application strategy to discover transferable knowledge between the classes. The algorithm was tested in five public hyperspectral imaging datasets (Chikusei, University of Pavia, Pavia centre, Salinas and Indian Pines, see A for more information about these datasets).

Another work [Wang et al., 2022] proposes Heterogeneous Few-Shot Learning (HFSL) for hyperspectral imaging classification. First, using the mini-ImageNet dataset, few-shot learning is used to extract transferable knowledge. Then, by using both spectral and spatial information, resorts to a residual network to extract spectral information and to a convolution network for the case of spatial information. The few-shot learning extracts discriminative features that help improve classification performance. This method was tested in two public hyperspectral imaging datasets (Pavia University and Houston, see A for more information about these datasets).

A more recent work [Pal et al., 2022] proposes a novel Outlier Calibration Network (OCN) combined with a feature extraction module integrated into the meta-training phase. This feature extraction module is composed of a novel Residual 3D Convolutional Block Attention Network (R3CBAM) to improve spectral-spatial feature learning. This novel approach focuses on the few-shot open-set recognition problem and uses a Variational Autoencoder (VAE) to augment the query set.

The cost of having all pixels from a hyperspectral imaging dataset labelled is high, and the number of artefacts present in some scenes can also increase this cost. ZSL approaches can decrease the number of samples needed to train the model, given that not all classes are needed for this step.

2.6 Classification Methods Discussion

The last sections presented some classification algorithms already used for hyperspectral imaging classification. This was performed by dividing the approaches into

two main groups: Supervised and Unsupervised approaches. Supervised approaches require more data, mainly due to the need for labelled data to train a model. Due to the lack of hyperspectral imaging datasets and the complexity of collecting hyperspectral data, this can be difficult to fulfil. In some cases, although Unsupervised techniques, namely clustering algorithms, usually result in worse classification results, they can be a good alternative to eliminate the need for additional data to train a model and also the additional processing time needed in this step.

In the scope of the Supervised approaches, the ones that stand out are the Support Vector Machine, Random Forests and Convolutional Neural Network algorithms, mainly because they have been widely tested in hyperspectral datasets. However, due to the particular characteristics of hyperspectral data, it is also necessary to consider the need to extract the most relevant information from the data. This is typically done with dimensionality reduction techniques, such as Principal Component Analysis (PCA).

Although there are no perfect alternatives for all situations, there is a gap between Supervised and Unsupervised methods. Considering that to obtain a better classification, we have to resort to Supervised methods, which implies that for all the classes we intend to classify, we need to have labelled samples for all classes. If a new class appears during the classification stage of the Supervised model, depending on the model, it will be misclassified as one of the classes that the model knows and has been trained on.

ZSL appears to fill this gap by training a model with data from the seen classes, but that will also recognise unseen classes, i.e. samples from classes that were not used to train the model. These algorithms are useful for classifying unseen samples in applications where new classes, such as marine litter detection, can appear during the classification stage.

However, ZSL has been used mainly for visible image applications by using semantic attributes to build the dictionary that will allow the passing of knowledge to classify samples of unknown classes. However, it can be challenging to replicate this solution in the case of hyperspectral data due to the hyperspectral image resolution, which in most cases is acquired from considerable altitudes, which reduces their spatial resolution. In addition, due to the characteristics of hyperspectral cameras, those that offer the best solution for data acquisition for remote sensing are of the pushbroom type, which means that the data will be acquired line by line. Therefore, the characteristics of the dataset will be different from the datasets used for applying ZSL to visible images (one of the most used visible image datasets for ZSL is composed of several images of animals taken relatively close to them). Therefore,

finding a new way to extract meaningful information from the hyperspectral data becomes necessary, allowing the creation of a novel dictionary that can transfer its knowledge to classify the unseen classes.

The work developed in this thesis starts by using two state-of-the-art methods, namely RF and SVM, to serve as benchmarks. Then, we implemented a novel CNN-3D that combines spatial and spectral information to perform classification. Finally, a ZSL method based on a generalised configuration was implemented, being capable of classifying samples from classes never seen by the model.

2.7 Application Oriented Perspective (Marine Litter)

Marine litter concentrations can be traced back to different origins. The origin can be caused by professional and domestic activities and can be intentional or not. Usually, ocean pollution sources include the general public consumer attitude and other activities ranging from coastal tourism and recreation to waste management policy on beaches. Economic activities usually installed in near-shore locations such as fisheries or aquaculture explorations are also commonly cited as potential sources, harbours, and other ship docking infrastructure [Veiga et al., 2016].

Is it possible to classify the marine litter samples according to their size [Lee et al., 2015]: small microplastic, <1 mm; large microplastic, 1 - 5 mm; mesoplastic, 5.1 - 25 mm; macroplastic, >25 mm.

The effect of ocean currents and harsh meteorological conditions, such as direct exposure to ultraviolet light during extended periods [Ryan et al., 2009], can alter the marine litter's physical characteristics. Marine debris tends to deteriorate and break apart, turning macroplastic samples into microplastics and submerging from the sea surface, entering the water column [Goddijn-Murphy et al., 2018], or resting in the seabed mainly due to biofouling. The taxonomy adopted for the classification of marine litter samples is important for two main reasons:

- To understand the sinking and fouling of marine litter samples. Small plastic items start sinking sooner than larger ones, while larger debris will float for more time. This is due to the relation of buoyancy with item volume. Biofouling can also increase the density of the marine litter to the point where the debris sinks, see [Ryan, 2015b];

- To trace the marine litter point of origin since there are studies [van Sebille et al., 2015] referring to the wind as the main effect of the ocean currents, which will influence the marine debris trajectory and allows identification of

origin patterns [Derraik, 2002, Garrity and Levings, 1993, Willoughby et al., 1997] regarding macroplastics accumulation in the ocean.

This is of fundamental importance since marine litter composition raises significant concerns about plastic contamination in the ocean due to their chemically engineered durability, which decreases their bio-degradation rate [Clark et al., 1989]. The marine litter composition will increase the life span of these synthetic polymers, which will stay on the ocean for longer [Shahidul Islam and Tanaka, 2004], thus producing irreversible damage to the marine ecosystem. There are multiple examples of environmental impacts of marine debris [Zielinski et al., 2009], including entanglement of marine fauna, ingestion by seabirds or fish [Brekke and Solberg, 2005], as well as the possible dispersion of microbial and colonising species to non-native waters [Zielinski et al., 2006, Dekker et al., 2001b]. Unfortunately, there are several marine litter hotspots worldwide. Due to its location and geological configuration, Porto Pim bay (Faial Island, Azores) is one of these marine litter hotspots and was the one selected to be analysed during the work developed during this thesis. Figure 2.5 shows marine litter found at the test site in Porto Pim bay.



Figure 2.5: *Marine Litter found in the test site, Porto Pim Bay (Faial Island, Azores).*

This section will focus on remote marine litter detection, including a summary of the previous work on marine litter detection in Table 2.4, focusing on previous research that performed tests in real environment conditions. The table describes the related work considering the following items: types of sensors used to detect marine litter, test site location, platform, target objects (marine litter samples) and developed algorithms.

Briefly summing up the table content, most of the previous work is being carried out using drones equipped with RGB cameras [Papakonstantinou et al., 2021, Wolf et al., 2020, Gonçalves et al., 2020a, Gonçalves et al., 2020b, Martin et al., 2018, Garcia-Garin et al., 2021]. All these works resort to deep learning algorithms, such as Convolutional Neural Networks (CNN). Although using RGB cameras can be

considered an easy and low-cost solution, it has operational constraints for detection, such as water reflections due to sun glare and difficulty achieving the required image resolution for marine litter detection. Recent literature results seem to be shying away from visual-only systems as they are not the most suitable for extended ocean coverage. Other works result from NASA initiatives, such as the one that uses AVIRIS [Garaba and Dierssen, 2018], an airborne solution provided by NASA. In [Garaba and Dierssen, 2018], the authors first performed an in-situ analysis of dry and wet marine-harvested microplastics to detect the absorption bands of each sample on the spectral wavelength to apply to the collected AVIRIS data.

Other authors based their solutions on satellite data, such as satellite WorldView 3 (WV3) for marine litter detection [Acuña-Ruz et al., 2018]. In this work, in addition to the data collection with the satellite WV3, the authors also collected some samples from the test site that was subject to extensive laboratory analysis. To perform the atmospheric correction in the WV3 data, the *Atmospheric Compensation* algorithm from DigitalGlobe was used. The results were convolved using the RSRCalculator, which allowed to estimate the Anthropogenic Marine Debris (AMD) reflectance values for each spectral band that afterwards were compared to the laboratory results. The convolved spectral signatures were then used to train automated classification algorithms such as SVM, RF and Linear Discriminant Analysis (LDA). The results showed a precision of 80% in general (considering all algorithms applied). However, this precision is only to distinguish between a mixture of AMD and expanded polystyrene AMD.

Five other studies use Sentinel-2 to detect marine litter [Topouzelis et al., 2019, European Space Agency and Plymouth Marine Laboratory (PML), 2020, Biermann et al., 2020, Themistocleous et al., 2020, Basu et al., 2021]. Four studies use artificial targets, while one uses plastic aggregations in the test site. Given the spatial resolution of the multispectral sensor installed in the satellite, 10 m $\times$ 10 m, there was the need to construct artificial targets that would meet the sensor requirements. However, in addition to the unsuitable spatial resolution for marine litter detection, the Sentinel-2 multispectral sensor also has a small number of spectral bands. In [Topouzelis et al., 2019], the Sentinel-2A data atmospheric correction was performed using the ACOLITE and Sen2Cor algorithms. The authors also provide a comparison between the pixel coverage for each target. However, they do not perform any automated classification tasks. In work developed in [European Space Agency and Plymouth Marine Laboratory (PML), 2020], the authors use both Sentinel-2B and airborne data. They compared three different algorithms for the airborne data: ENVI true colour composite, greyscale HI index and greyscale NDHI index. The

test was conducted only with beach targets, and the authors did not perform any automated classification task. In [Biermann et al., 2020], the authors developed the FDI algorithm, a new index created especially for Sentinel-2 data. This index allows easy and fast detection and serves as input to the classification stage using naive Bayes. The classification was performed between the following classes: water, seaweed, timber, plastic, foam and pumice. The results presented show 86% accuracy for "suspected plastics", which do not identify the plastic type. [Themistocleous et al., 2020] resort to multiple indexes, such as the Normalized Difference Water Index (NDWI) [McFEETERS, 1996], Water Ratio Index (WRI) [Shen and Li, 2010], Normalized Difference Vegetation Index (NDVI) [Rouse et al., 1973], Automated Water Extraction Index (AWEI) [Feyisa et al., 2014], Modified Normalized Difference Water Index (MNDWI) [Xu, 2006], Normalisation Difference Moisture Index (NDMI) [Wilson and Sader, 2002], Simple Ratio (SR) [Themistocleous et al., 2020], Plastic Index (PI) [Themistocleous et al., 2020], and Reversed Normalized Difference Vegetation Index (RNDVI) [Themistocleous et al., 2020]. These indexes allowed to understand the capability to distinguish between water and plastics, but they did not identify the plastic type. [Basu et al., 2021] present four different approaches for classification: K-Means hard clustering, fuzzy C-Means, Support Vector Regression (SVR) and semi-supervised fuzzy C-Means (SFCM), achieving an accuracy from 64.3% for the SFCM algorithm, and the best one being 98.4% with the SVR method. However, all these models were trained only to distinguish between two classes: plastic or no plastic, which does not allow for classification of the plastic type.

A recent study [Kremezi et al., 2021] uses the PRISMA satellite data to perform tests with several pansharpening methods to improve the spatial resolution of the PRISMA hyperspectral sensor. The authors used several artificial targets with different sizes to test the possibility of using the PRISMA data for marine litter detection and the minimum spatial resolution when using pansharpening methods. Although some plastic materials spectra are similar to the water spectra, the authors were able to produce marine plastic litter indexes, which successfully detected the plastic targets placed in the test site. [Taggio et al., 2022] also presents a study using PRISMA satellite data by resorting to twelve square floating targets with three sizes. Each size corresponds to a different material: white, yellow and green tarps built-in High-Density PolyEthylene (HDPE), transparent water bottles and green oil bottles made with PET and cyan sheets for building isolation in PS. The authors implemented the K-Means and Light Gradient Boosting methods, and the second one returned the best results with 96% accuracy in identifying plastic. However, the

authors did not perform the plastic-type classification.

Kikaki et al. [Kikaki et al., 2022] created MARIDA, a benchmark for marine debris detection from Sentinel-2 remote sensing data. They designed and followed a specific process over a seven-year period (2015-2021), which allowed them to gather marine litter and plastic pollution reports across coastal areas and river mouths in multiple countries. These reports, combined with the data collected from Sentinel-2 (which was subjected to processing, annotation and statistics), resulted in a new dataset. Afterwards, some algorithms were applied to create a benchmark, such as RF and a U-Net architecture. The first was trained in three different RF models: the first based on spectral signatures of each pixel (RF_{SS}), the second based on spectral signatures and calculated spectral indices (RF_{SS+SI}) and the last one with spectral signatures, spectral index, and extracted grey level co-occurrence matrix textural features ($\mathrm{RF}_{SS+SI+GLCM}$).

In [Tasseron et al., 2021], the authors perform an in-situ analysis by resorting to a double camera setup in the range of 400 to 1700 nm, used to analyse the samples in a controlled environment. By collecting data from the samples floating in the water with the linear discriminant analyses (LDA) algorithm, the authors determined the most influential bands to discriminate between water by mixing floating debris and vegetation and plastics. The results were also compared with Sentinel-2 and WordlView3 satellites, allowing the identification of the most promising bands.

[Balsi et al., 2021] the authors present a novel pushbroom hyperspectral imaging device, composed of a spectrometer ImSpector NIR17 from Specim and a Xenics Bobcat 32 SWIR imaging sensor, for collecting 320 wavelengths from 900-1700 nm. This system was installed in a DJI Matrice 600 UAV and tested with several materials, such as PET bottles, plastics containers, PP containers, floor tile, LDPE bags, and HDPE bottles. The authors resorted to Linear Discriminant Analysis (LDA), a linear classifier, to classify the data. The authors could distinguish between PE and PET objects, but they do not present the accuracy results for the classification.

The work developed and presented in this thesis follows the work described in Table 2.4. It is the first to present airborne and drone hyperspectral imaging data results within visible and SWIR range, and it was developed under the framework of the ESAPlastics project. We developed three stages of learning-based approaches for the detection and classification of marine litter samples: (1) RF and SVM methods for establishing a benchmark of the state-of-the-art approaches using the acquired data; (2) a CNN-3D to display the potential of classification using both spatial and spectral information; (3) the first ZSL hyperspectral imaging method for detection and classification of marine litter.

Table 2.4: *Marine litter detection state-of-the-art.*

Ref.	Sensor	Test Site	Satellite/Vehicle	Target Objects	Algorithm Used
[Garaba and Dierssen, 2018]	(1) FTIR spectroscopy; (2)Airborne visible-infrared imaging spectrometer (AVIRIS) retrieved from the AVIRIS online data portal (224 bands between 360–2500nm)	California Sunshine Canyon Landfill (USA)	AVIRIS	Plastic that was present in the test site	(1) Spectral similarity: Spectral contrast angle; (2) Spectral mixing: used a simplified linear mixing simulation to analyse how spectral properties of the dry and wet marine-harvested microplastics were influenced by varying pixel coverage
[Martin et al., 2018]	12 MP RGB Imaging Sensor	Northern Red Sea, South Arabia	DJI Phantom 3	Plastic items of different shapes (bottles, bags, cups, bottle caps and food containers), colour (transparent to brighter), from 2 cm to 60 cm	Random Forests
[Acuña-Ruz et al., 2018]	(1) Laboratory: HyLogger 3 reflectance spectrometer system and APOGEE PS-300 (2) WorldView 3, using a multispectral with a spectral range from 397 to 2373 nm)	Chiloe Archipelago (northwestern area of Chilean Patagonia)	WorldView3	Plastic that was present in the test site	Atmospheric correction of the WV3 data: Atmospheric Compensation from Digital Globe- The spectral collected using the WV3 were convolved using RSR Calculator to estimate the samples reflectance values for the spectral bands to compare with the spectral measurements from the laboratory
[Topouzelis et al., 2019]	(1) S900 DJI Hexacopter: MultiSpectral (MS) SLANTRANGE 3P (S3P), Parrot Sequoia MS camera (4 bands in the 450–850 nm), FLIR Duo R, and RGB camera (2) Sentinel-2A: MS camera (13 bands: 492–1376 nm) (3) Sentinel-1: SAR	Tsamakia beach of Mytilene on Lesvos Island, Greece	Drone; Sentinel-2A; Sentinel-1	(i) plastic bottles; (ii) plastic bags; (iii) plastic fishing net. Each one of these plastic types was held by a 10 m x 10 m frame	Atmospheric correction for the Sentinel-2A data: ACOLITE and Sen2Cor
[European Space Agency and Plymouth Marine Laboratory (PML), 2020]	(1) Sentinel-2B: MS camera (13 bands: 492–1376 nm) (2) Airborne: Specim Aisa Fenix (300–2500 nm)	Whitsand Bay (UK)	Sentinel-2B; Airborne	(i) 2 plastic targets from NERC-FSF with colours grey and black; (ii) 2 types of agricultural fleece plastic targets	Specim data analysis: ENVI true colour composite, greyscale HI index and grey scale NDHI index
[Biermann et al., 2020]	Sentinel-2A and 2B Multispectral sensors: 13 bands between 492–1376 nm	Accra; Ghana; Barbados; Caribbean Da Nang; Vietnam Late Island, Tonga; Mytilene, Greece Gulf Islands; Canada; Scotland; Durban, South Africa	Sentinel-2A; Sentinel-2B	(i) Plastic aggregations that were present in the test site: (ii) Deployed plastic targets	Acolyte to the correction phase, FDI to detect and Naive Bayes to the classification stage
[Themistocleous et al., 2020]	Sentinel-2A and 2B Multispectral sensors: 13 bands between 492–1376 nm	Limassol Old Port, Cyprus	Sentinel-2	Target with 3 m × 10 m filled with multiple size water bottles, held together by nylon string and framed by PVC pipes	Multiple indexes: NDWI, WRI, NDVI, AWEI, MNDWI, NDMI, SR, PI, RNDVI
[Gonçalves et al., 2020a, Gonçalves et al., 2020b]	20 MP RGB Imaging Sensor	Cabedelo beach, Figueira da Foz, Portugal	DJI Phantom 4 Pro	Plastic that was present in the test site	(1) RF; (2) CNN
[Wolf et al., 2020]	20 MP RGB Imaging Sensor	Phnom Penh, Sihanoukville and Siem Reap in Cambodia	DJI Phantom 4 Pro v2	Plastic that was present in the test site	APLASTIC-Q: (i) plastic litter detection (PLD-CNN) + (ii) plastic litter quantifier (PLQCNN)
[Papakonstantinou et al., 2021]	RGB Camera	Xabelia beach in Lesvos, Greece	DJI Phantom 4 Pro v2	Plastic that was present in the test site	Deep learning algorithms – CNN (VGG16, VGG19, DenseNet121, DenseNet169, DenseNet201)
[Taddia et al., 2021]	RGB Camera; Micasense RedEdge M MS camera with five bands (blue, green, red, RedEdge and NIR wavelengths)	Sacca di Goro - Foce del Po di Volano, Nothern Italy	DJI Phantom 4 RTK; DJI Phantom 3 Professional	Plastic that was present in the test site	(1) Digital Surface Model (DSM); (2) NVDI

[Garcia-Garin et al., 2021]	(1) RGB camera Sony ILCE-6000; (2) RGB camera Sony Alpha 7 R; (3) RGB camera FC300S (4) RGB camera FC220 (5) Canon EOS REBEL SL1	Between Delta de l'Ebre and Cap de Creus, Spain	UAVs: (1) fixed-wing HP1 (2) a multi-rotor Topografia (3) DJI Phantom 3 (4) DJI Mavic Pro and aircraft: (5) Partenavia P-68	Crate, bottles, cans, sack, tetra-brik packages, bags, ball, boards, drum, turtle carapax, six-pack rings, trays, jar, balloon, towels and polystyrene debris	CNN
[Basu et al., 2021]	Sentinel-2A and 2B Multispectral sensors: 13 bands between 492–1376 nm	(1) Limasol, Cyprus; (2) Mytilene, Greece	Sentinel-2	Limasol: 10 m filled with multiple size water bottles, held together by nylon string and framed by PVC pipes; Mytilene: (1) 10 m × 10 m targets with water bottles; (2) four 5 m × 5 m with plastic bottles and plastic bags; (3) four 5 m × 5 m, and one 1 m × 10 m plastic targets, composed by plastic bottles and a combination of plastic bags and plastic mesh; (4) four 5 m × 5 m plastic targets	(1) K-Means hard clustering; (2) fuzzy C-Means; (3) Support vector regression; (4) Semi-supervised fuzzy C-Means
[Balsi et al., 2021]	Spectrometer Specim ImSpector NIR17 OEM with a Xenics Bobcat 32 SWIR imaging sensor, 320 bands between 900-1700 nm	shore of Platanoma and Porto Ferro, Sardinia, Italy	DJI Matrice 600	PET bottles, plastics containers, PP containers, floor tile, LDPE bags, HDPE bottles	LDA
[Kremezi et al., 2021]	PRISMA HS with 234 bands from 400-2500 nm and 30 m spatial resolution	Tsamakia beach, Lesvos Island, Greece	PRISMA satellite	12 floating targets (5.1 m x 5.1 m, 2.4 m x 2.4 m, 0.6 m x 0.6 m), with HDPE, PET and PS	Pansharpening methods
[Andriolo et al., 2021]	20 MP RGB Imaging Sensor	Cabedelo beach, Portugal; San Rossore beach, Italy	DJI Phantom 4 Pro	Plastic that was present in the test site	Manual characterisation made by sixteen qualified researchers, and the comparison between the annotations using the coefficient of concordance
[de Vries et al., 2021]	GoPro Hero6 Black (4000 x 3000 pixels RGB Imaging Sensor)	Great Pacific Garbage Patch	Maersk Transporter Vessel; UAV; Other vessels	Plastic that was present in the test site	(1) Faster R-CNN (with a pre-trained network); (2) Ultralytics YOLOv5
[Kikaki et al., 2022]	Sentinel-2A and 2B MS sensors: 13 bands between 492–1376 nm	Guatemala; Honduras; S. Domingo; Haiti; Indonesia; Vietnam; Philippines; Scotland; South Africa; South Korea; Indonesia; China	Sentinel-2A and 2B	Marine debris present in the test site (sargassum macroalgae, ships, natural organic material, waves, wakes, foam, different water types (clear, turbid, sediment-laden and shallow) and clouds)	(1) Random Forest (based on spectral signatures of each pixel (RFss), based on spectral signatures and calculated spectral indices (RFss+si), and with spectral signatures, spectral indices, and extracted Gray-Level Co-occurrence Matrix (RFss+si+GLCM); (2) U-Net architecture
[Cocking et al., 2022]	BaySpec OCI-F SWIR HS push-broom camera, with 36 bands between 900 and 1700 nm	Oban, Scotland	The camera was attached to a taut line at approximately 5 m above the beach	Plastic items (PET, PE, PP, EPS, ABS, PLA, PVC, PE/PP blend, Nylon 6, Melamine) placed along dry sandy regions of the shoreline	(1) Spectral Angle Mapper (SAM); (2) Spectral Correlation Mapper (SCM)
[Gonçalves and Andriolo, 2022]	RGB camera; Micasense Altum MS sensor with five bands (blue, green, red, RedEdge and NIR wavelengths)	Leirosa beach, Figueira da Foz, Portugal	DJI Matrix 210 RTK V2	42 litter items on the upper beach (representing the most common litter found at the test site	SAM
[Taggio et al., 2022]	PRISMA HS with 234 bands from 400-2500 nm and 30 m spatial resolution	Tsamakia Beach, Mytilene and Geras Gulf, Lesvos island, Greece	PRISMA satellite	Twelve square floating targets, with three different sizes. Each size was built with a different material: HDPE (tarps in white, yellow and green), PET (transparent water bottles, green oil bottles), PS (sheets for building insulation, in cyan)	(1) K-Means; (2) Light Gradient Boosting Model

2.8 Summary

Recently, marine litter has become one of the main threats to marine life. It has become a threat to the human specie, and it is getting more attention from the scientific community which is developing methods to detect, classify and remove marine litter from our oceans. Usually, marine litter hotspots are located in remote areas, so the scientific community started to analyse marine litter using already available satellites. However, it is difficult to perform this task due to the lack of spatial and spectral information in the data collected by the multispectral and hyperspectral imaging systems present in these satellites, which are not tailored to address the marine litter phenomena.

Some authors decided to use Red Green Blue (RGB) cameras installed in UAV to detect marine litter. Although it is a cheaper solution, it will not return the necessary information to distinguish the different types of marine litter. Hyperspectral imaging systems can collect the required information to distinguish between different types of marine litter. However, hyperspectral imaging systems return massive amounts of data that must be labelled pixel by pixel, which is very time-consuming. Since different types of marine litter are continuously appearing, making it is impractical to keep an up-to-date database of labelled datasets and models trained to have the information from all marine litter types.

To overcome this problem, there is a need to develop new methodologies capable of classifying data that were never seen during the model's training. This will help decrease the number of labelled data needed to implement a detection and classification algorithm for marine litter using hyperspectral imaging and recognise a new class of marine litter when it is present.

3
Fundamentals

3.1 Introduction

This chapter introduces some fundamental concepts for a better understanding of the thesis work. The idea is not to be exhaustive but briefly explain the topics to non-experts readers. For an in-depth discussion, some related work will be pointed out for each research topic.

3.2 Hyperspectral Imaging

Hyperspectral imaging systems can collect a considerable amount of information. Due to this feature, these systems can be employed for different applications. Hyperspectral imaging systems collect information over the electromagnetic spectrum, from visible to SWIR wavelengths for marine litter detection and classification tasks. Hyperspectral sensors have a fine spectral resolution at given wavelengths of illumination, allowing the measurement of the radiation reflected from each pixel in many wavelengths (bands) from different ranges in the electromagnetic spectrum. To do this, the hyperspectral imaging system uses a spectrometer. This instrument divides the radiation into several finite bands and measures the energy in each band. Based on the spectrometer design, the captured bands may be contiguous or not and with equal width.

The combination of spectroscopy and image forming results in hyperspectral cameras that, during the data acquisition, create an electromagnetic profile by capturing the intensity of radiation emitted from each pixel at various wavelengths. The collected information is usually combined into a hyperspectral image data cube with

two spatial and one spectral dimension. This cube is composed of a stack of images layered on each other. Each pixel (in spatial dimension XY) is a spectrum over the different wavelengths observed by the camera. Combining all the pixels and lines from one wavelength results in an image, as shown in Figure 3.1.

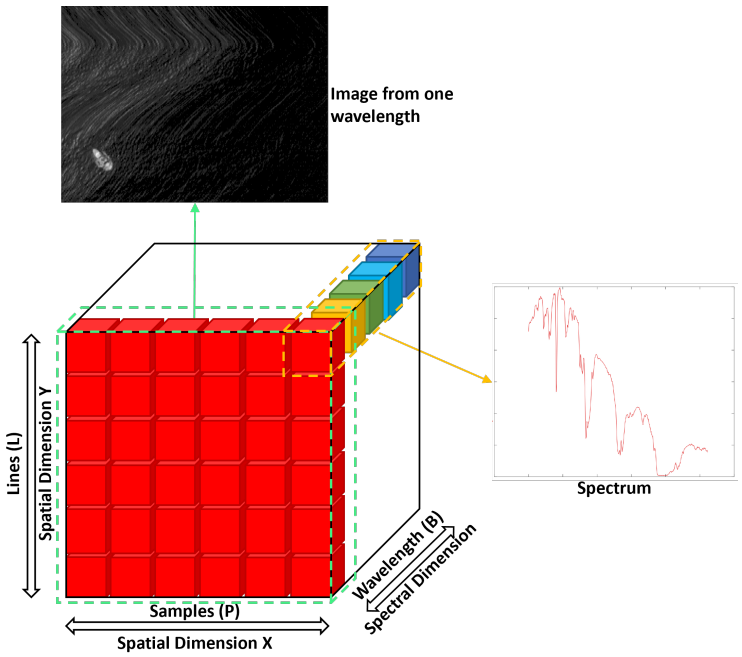


Figure 3.1: *Hyperspectral data cube*

Figure 3.2 shows an example of the combination of all the pixels and lines from one wavelength - resulting in a fake visible image, and the combination of all the wavelengths from one single pixel - resulting in a spectrum.

3.2.1 Working Principle

There are three data acquisition schemes for hyperspectral cameras: whiskbroom, tunable filter and pushbroom.

3.2.1.1 Whiskbroom cameras

This acquisition scheme allows one single pixel in the spectral and spatial domain at a time. Some of the most traditional hyperspectral sensors, such as AVIRIS[1] and HyMap [Cocks et al., 1998] use this type of acquisition. This acquisition scheme results in an instrument with a simple design, wide swath (determined by the scanning mechanism) and easy calibration. However, it is required to have a mechanical scanner containing moving parts, to post-process the data for spatial incongruence and to have some constraints of high spectral and spatial resolution due to the low

[1] `https://aviris.jpl.nasa.gov/`

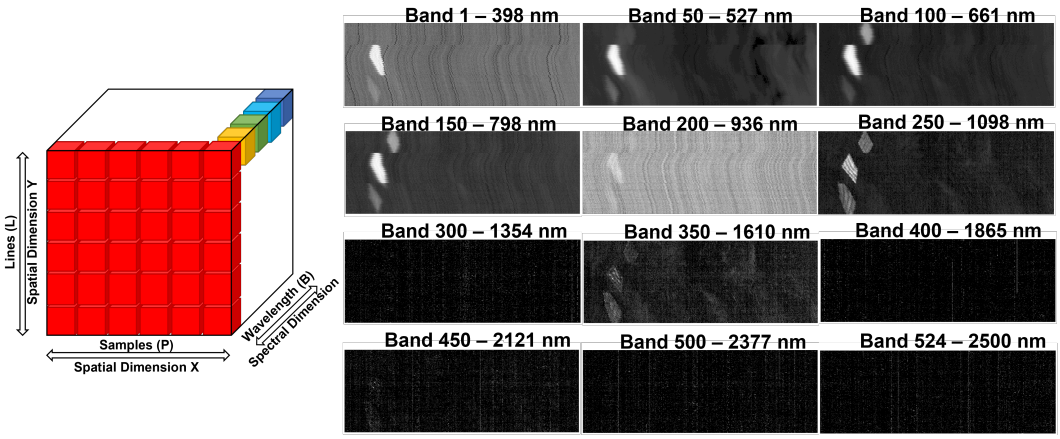


Figure 3.2: *Hyperspectral data cube*

integration time. An acquisition scheme of this type of camera is shown in Figure 3.3.

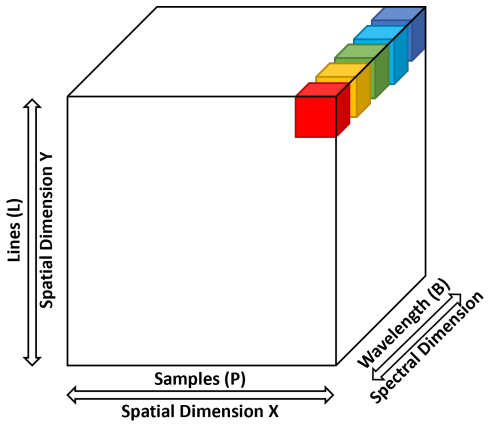


Figure 3.3: *Whiskbroom hyperspectral camera acquisition scheme*

3.2.1.2 Tunable filter cameras

The tunable spectral filter allows a 2D image at a time, wherein each 2D image corresponds to a wavelength. Only when all wavelength (2D image) is acquired is it possible to obtain each pixel's spectrum, as shown in Figure 3.4.

3.2.1.3 Pushbroom cameras

This method simultaneously allows complete spectral data acquisition, with spatial line scanning over time. This allows the simultaneous acquisition of all spectral information, being insensitive to instrument/sample movement. Also, the instrument

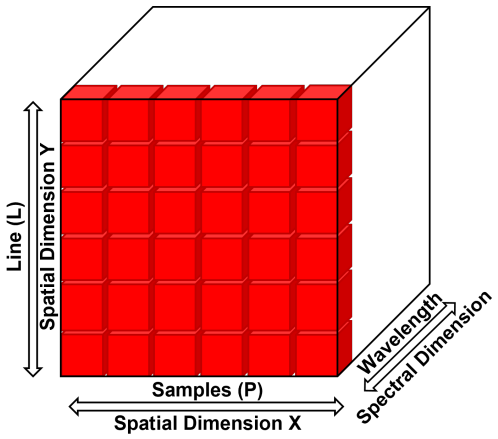


Figure 3.4: *Tunable filter hyperspectral camera acquisition scheme*

has no moving parts and can collect light from samples 5 to 20 times more efficiently than the tunable spectral filter. Furthermore, only a line across the sample needs to be illuminated, which leads to 10 to 30 times more light and, consequently, more speed. The data is received line by line for all wavelengths at the same time. It means that for each line, we obtain the full spectrum for each pixel. However, this acquisition scheme has a complex optical design and focal plane, and the number of spatial pixels limits the swath width. It also needs calibration, which is not an easy procedure, while it has spectral (smile) and spatial (keystone) distortions. This is depicted in Figure 3.5.

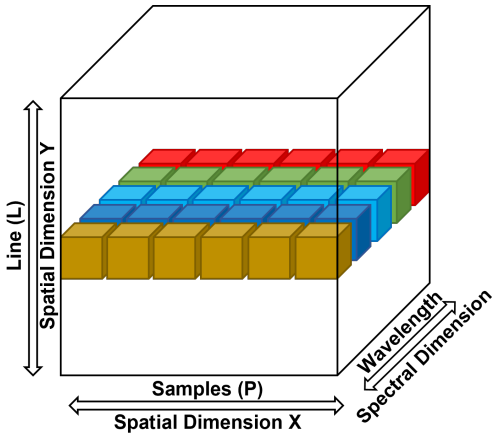


Figure 3.5: *Pushbroom hyperspectral camera acquisition scheme*

Figure 3.6 displays the UAV trajectory and scan line acquisition for a hyperspectral pushbroom sensor.

In Figure 3.7, it is possible to observe examples of pushbroom hyperspectral cameras from three different manufacturers: Specim, HySpex and Cubert.

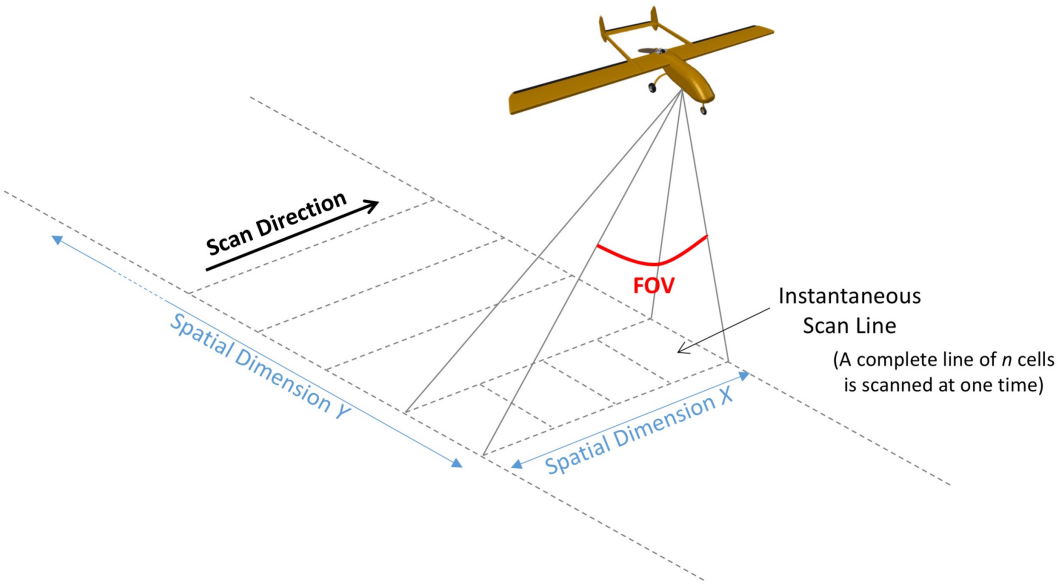


Figure 3.6: *UAV pushbroom scan methodology.*



(a) (b) (c)

Figure 3.7: *Specim FX10 (400-1000 nm), HySpex Mjolnir S-620 (970-2500 nm) and Cubert S 185 (450-950 nm), respectively (not to scale).*

3.2.2 Data Acquisition

Given the type of data collected by this camera, it is necessary to consider the data storage and dispatch for the processing unit. There are three distinct configurations created specifically for this type of data:

- Band Sequential (BSQ): is the most straightforward format, where each data line is followed immediately by the next line in the same spectral band. This is optimal for spatial (x,y) access to any part of a single spectral band;
- Band-interleaved-by-pixel (BIP): stores the first pixel for all bands in sequential order, followed by the second pixel for all bands, followed by the third pixel for all bands, and so forth, interleaved up to the number of pixels. This format provides optimum performance for spectral (z) access to the image data;
- Band-interleaved-by-line (BIL): stores the first line of the first band, followed by the first line of the second band, followed by the first line of the third band, interleaved up to the number of bands. Subsequent lines for each band are interleaved similarly. This format provides a compromise in performance between spatial and spectral processing.

These three data organisation types were defined by ENvironment for Visualising Images (ENVI)[2], which is the most utilised software for hyperspectral data visualisation and processing. The hyperspectral camera manufacturer also follows these data configurations.

Considering a pushbroom hyperspectral camera using the BIL type, the data from all bands related to a line will be received simultaneously, forming all the spectral signatures observed in that line. This offers the best compromise between spatial and spectral processing, being the most used. Each acquired line is added to the hyperspectral cube.

The data returned by these systems are considered to be raw data, i.e., digital numbers. These values are used to describe pixel values that have not yet been calibrated into physically meaningful units, with the influence of different factors such as light source, sensor, atmosphere and surface material. To obtain the spectrum with less external interference is possible to correct the data to have a measurement of radiance or reflectance.

Radiance refers to the amount of radiation coming from an area. A gain and offset must be applied to derive a radiance value from a digital number. These values for hyperspectral imaging are obtained for all pixels and depend on the wavelength. These values are typically retrieved from the image metadata or received from the data provider. Radiance values normally have units of $w(m^2\ SR\ nm)$. Transformation of the digital numbers to radiance is denoted by calibration and is performed following equation 3.1 for each pixel.

$$L_\lambda = Gain \times Pixel\ Value + Offset \tag{3.1}$$

[2] `http://www.harrisgeospatial.com/ProductsandSolutions/GeospatialProducts/ENVI.aspx`

Calibration performs the transformation from raw data to the radiance domain. However, in addition to the surface material, this influences the light source and atmosphere. For some applications, this value is enough for the implementation, but if the spectrum of the material is needed, it is necessary to convert it to reflectance. Reflectance is the proportion of radiation striking a surface to the radiation reflected on it. Some materials can be identified with the reflectance spectra, so it is common to correct an image to reflectance as a first step in identifying the materials in a scene. Top-Of-Atmosphere (TOA) reflectance corresponds to the reflectance measured by a space-based sensor flying higher than the earth's atmosphere. Although, these reflectance values will include influence from clouds and atmospheric aerosols and gases. Apparent reflectance is another type of reflectance that describes the combination of earth/atmosphere behaviour for the reflected solar irradiation. Convert radiance into reflectance is denoted by atmospheric correction. In this work, the data used is in the radiance domain. Our goal is to develop new algorithms that, in the future, can be used in real-time. The atmospheric correction is a complex procedure, limiting its complete and correct application in real time. Therefore, in this thesis, all the algorithms were developed to work in the radiance domain.

Figure 3.8 shows the pipeline used to process the raw hyperspectral data into reflectance data.

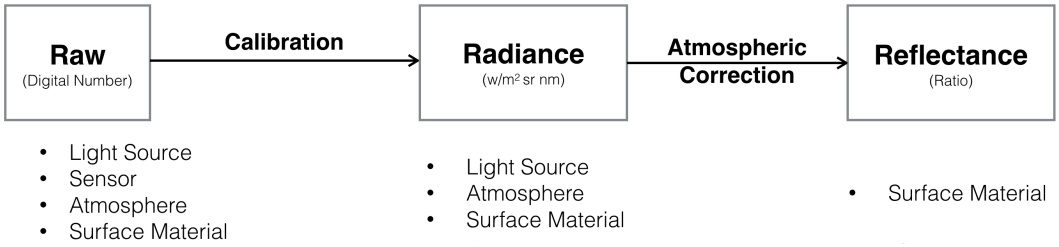


Figure 3.8: *Hyperspectral data processing pipeline.*

However, other issues can influence the data quality [Ghamisi et al., 2017], namely:

- Curse of dimensionality: Due to the high quantity of data collected by hyperspectral systems, there are some issues reported [Scott, 1992, Wegman, 1990, Jimenez and Landgrebe, 1998] showing that with the increase in the dimensionality, the corners will contain most of the hypercube volume; and also with this increase, the volume of the hypersphere will be more noticeable in the outside shell. Although higher bands mean more separability of the classes, it will also imply that a higher-dimensional number of statistics should be estimated. This will result in a decrease in the accuracy of the statistical estimation, resulting in a decrease in the classification accuracies. It is expected that the classification algorithms have methodologies to extract the most meaningful information from the dataset, allowing to overcome the curse of dimensionality issue;

- Uncertainties: The atmospheric conditions, radiometric and spatial resolutions

during the data collection, and others cause data uncertainties that can affect the classification results [Canters, 1997,Dungan, 2002,Friedl et al., 2001,Wang, 2015];

- Influence of spatial resolution: higher spatial resolutions can minimise the mixed-pixel issue (having spectra from multiple materials in the same pixel) [Woodcock and Strahler, 1987].

3.3 Hyperspectral imaging classification techniques

Recently, several hyperspectral imaging classification algorithms have been developed in the past years, mainly using machine learning and deep learning techniques. This section presents a brief overview of the ones that will be used in this work: SVM, RF, PCA, CNN and ZSL.

3.3.1 Dimensionality Reduction

One of the main issues with hyperspectral imaging processing is the high dimensionality of the data [Ghamisi et al., 2016]. This problem can be mitigated using dimensionality reduction techniques, such as feature extraction and band selection. Although it is not a mandatory step, it improves the extraction and selection of meaningful information from the data, increasing the classifiers accuracy. These algorithms can also be used to understand the most meaningful bands that need to be analysed for a specific scenario, such as remote marine litter detection. This information makes it possible to define the requirements for new sensors dedicated to marine litter detection and classification. This work uses a feature extraction method: PCA, which will be described in detail.

3.3.1.1 Principal Component Analysis (PCA)

The PCA is a multivariate statistical technique for hyperspectral imaging feature extraction. This algorithm can analyse data with several inter-correlated dependent variables. In the case of hyperspectral imaging, this data is the spectrum along all wavelengths. PCA will extract the most meaningful information and express it in a set of new orthogonal variables, denoted as principal components, which are linear combinations of the original data. The principal components are organised by importance, which means that the first one has extracted more information from the original data.

To implement the PCA algorithm, it is required to follow the following steps:

- Standardise the data before fitting to make the features have the characteristics of a standard normal distribution with a mean equal to zero, with a standard deviation of one.

  In order to:

$$\mu_{data_norm} = 0 \text{ and } \sigma_{data_norm} = 1$$

We need to apply the following:

$$data_norm = \frac{data - \mu}{\sigma}$$

- Calculate the covariance matrix of the standardised data;
- Determine the covariance matrix eigenvectors and eigenvalues;
- Organise the eigenvalues in descending order;
- Organise the principal component vectors according to the eigenvalues arranged in descending order - the principal component vectors that have the highest value of eigenvalue are those which contain more information about the data;
- Project the points to the new space created by PCA using the first two principal components.

The associated eigenvalue gives the variance that each principal component represents. However, it is still possible to obtain a percentage of variance that each principal component represents using the following equation:

$$\%\ variance = \frac{\lambda_i}{\sum_{n=1}^{d} \lambda_n}$$

Where d is the total number of principal components, and i represents the principal component that should be determined the variance.

PCA will be applied before all algorithms used in this thesis, allowing to decrease in the number of spectral wavelengths needed to process the data.

3.3.2 Machine Learning-based Methods for Hyperspectral Imaging Classification

3.3.2.1 Support Vector Machines

SVM is a supervised learning method commonly used in classification tasks [Melgani and Bruzzone, 2004, Pal and Foody, 2010]. SVM has been used in different fields. In the case of hyperspectral data, this method quickly became a reference because SVM was able to obtain results similar to much more computationally complex algorithms using lower processing time [Petropoulos et al., 2012b]. The goal is to find the optimal hyperplane in n-dimensional classification space with the highest margin/separation between classes. The margin corresponds to the distance between the hyperplane and the nearest data point from either class.

For linearly separable data, when the optimal hyperplane is found, data points that are nearest to the hyperplane (on the margins) are called *Support vectors*. If these points are removed, they will change the position of the hyperplane, so they are critical elements of the dataset. The complexity of the SVM model is only affected by the number of support vectors selected. Therefore, it is not affected by

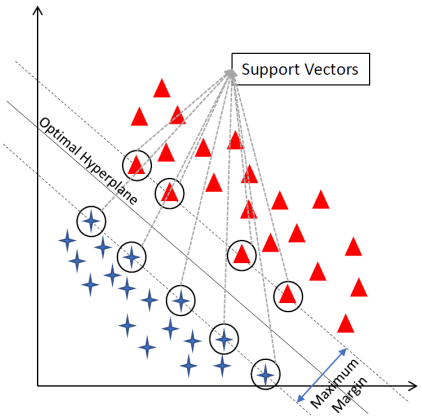


Figure 3.9: *Support Vector Machine linear example*

the number of features in the training data. Figure 3.9 shows an example of a linear SVM.

There are cases where the data is not linearly separable, which causes a linear SVM to have poor results. [Hsu et al., 2003] However, the data may not be linearly separable in $\mathbb{R}^2$, but linearly separable in $\mathbb{R}^3$, which would make them linearly separable by a plane. It would be possible to use a linear SVM classifier that finds a good boundary decision. However, the hyperplane found in $\mathbb{R}^3$ is nonlinear when projected back to $\mathbb{R}^2$, as can be observed in Figure 3.10.

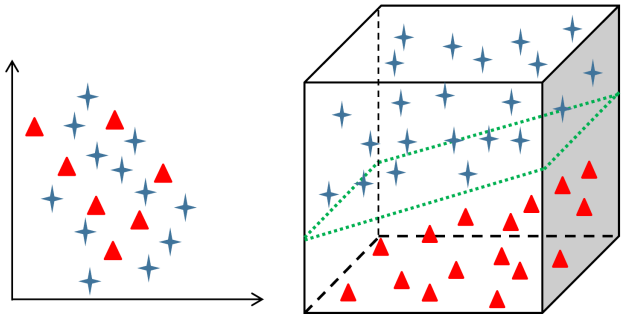


Figure 3.10: *Support Vector Machine not linearly separable example*

With this, is possible to conclude that a dataset D which is not linearly separable in $\mathbb{R}^N$ can be linearly separable in a higher-dimensional space $\mathbb{R}^M$ (where $M > N$).

This simple formulation may be impractical for large-dimension datasets. If the value of M rises quickly relative to the value of N, then learning SVM through dataset transformations will incur serious computational and memory problems. For example, for the polynomial kernel, in the case of a dataset in $\mathbb{R}^2$, a two-degree polynomial kernel (implicitly) performs the transformation $[x_1, x_2] = [x_1^2, x_2^2, \sqrt{2} \cdot x_1 \cdot x_2, \sqrt{2 \cdot c} \cdot x_2, c]$, which will add three additional dimensions.

However, during training or testing, SVM do not need to work specifically on

the higher-dimensional space. During training, [Hofmann, 2006], the optimisation problem only uses the training data to calculate pair-wise dot products $< \vec{x}_i, \vec{x}_j >$, where $\vec{x}_i, \vec{x}_j \in \mathbb{R}^N$.

In this way, it becomes possible to resort to functions that, for two given vectors v and w in $\mathbb{R}^N$ implicitly calculate the dot product between v and w in a higher-dimensional $\mathbb{R}^M$, without explicitly transforming v e w for $\mathbb{R}^M$. This functions are called kernel functions, $K(\vec{x}_i, \vec{x}_j)$, where:

- When using a kernel $K(\vec{x}_i, \vec{x}_j)$, we can implicitly transform datasets into a higher-dimensional $\mathbb{R}^M$ without using more memory, and with a small effect on computing time;
- Due to the previous point, we can efficiently learn nonlinear decision boundaries for SVM, being only necessary to replace all dot products in the SVM computation with $K(\vec{x}_i, \vec{x}_j)$.

In this context, a kernel function is a function $K : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$. As explained previously, mapping the input space to the feature space is called the "Kernel Trick" and is used in most real SVM applications. The kernel function works like a metric of similarity between the input objects. Although it is possible to create new kernels, there are already some *off-the-shelf*: polynomial, radial basis function, and sigmoid are the most used.

The advantages of SVM are:

- Accuracy;
- This algorithm works well with a clear margin of separation and smaller, cleaner datasets;
- Effective in high dimensional spaces and in cases where the number of dimensions is greater than the number of samples;
- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient;
- Produce sparse solutions - the model that defines the decision boundary is finally expressed as a function of a subset of training samples;
- SVM are resistant to getting stuck in local minima since the convexity of the cost function enables the classifier to identify the optimal solution;
- Do not assume any statistical distribution for the data to be classified [Mountrakis et al., 2011].

However, this method also has some disadvantages, such as when the dataset is large, the training time is higher, and it is less effective on noisier datasets with overlapping classes. Another drawback of SVM is the tuning of the kernel parameters, which can affect the classifier performance.

3.3.2.2 Random Forests

RF classifier is an ensemble learning method, meaning that it is composed of multiple classifiers, and their predictions are combined to identify the most promising classification. This algorithm is composed of a combination of tree classifiers. Each one is created using a random vector sampled independently from the input vector, having a unit vote for the most popular class, allowing the classification of an input vector.

To train RF usually is used the general technique of bootstrap aggregating, or bagging, to tree learnings [Breiman, 1996]. This technique is used for training data creation by randomly resampling the original dataset with replacement. [Rodriguez-Galiano et al., 2012] If many trees were trained on a single training set, strongly correlated trees are produced. The bootstrap sampling can decorrelate the trees by showing them different training sets. RF uses the trees as base classifiers $h(x, \theta_k), k = 1, ...,$, where x and θ_k are the set of input vectors and the independent and identically distributed random vectors. To make the classifier more robust, it is possible to improve the classifier stability by using more than once the same data for training [Waske et al., 2009].

To generate the RF model, there are only two parameters: the number of trees, a free parameter that can be selected because of the size and nature of the training set, and the number of predictor variables. This method is flexible and can deal with different scenarios (large number of features, small or large dataset, among others) and is easy to evaluate. Also, this algorithm does not assume any probability distribution for the input data and is insensitive to noise in the training labels.

Figure 3.11 shows an example of the RF algorithm implementation.

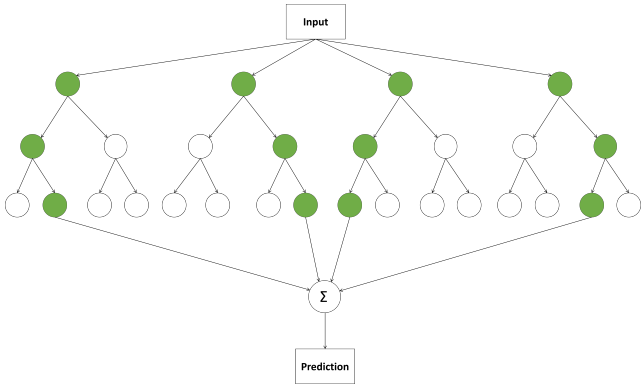


Figure 3.11: *Random Forest algorithm scheme.*

3.3.3 Deep Learning-based Methods for Hyperspectral Imaging Classification

In recent years, multiple deep models have been used in hyperspectral data feature extraction and classification. However, the architecture design is the most critical

part of a successful deep learning model. One problem with this method is that it may lead to overfitting, which means that the results can be perfect in the training data but poor in the test data. It is still possible to combine multiple methods with deep learning algorithms, to create more robust models.

3.3.3.1 Autoencoder (AE) and Stacked Autoencoder (SAE)

AE is an unsupervised method, which means that it uses unlabelled training data. This is constituted by a visible layer of n inputs and one hidden layer of h units, using an activation function f. The training phase will map the input data, $x \in R^d$, to the hidden layer, obtaining the latent representation returned by the hidden layer, $y \in R^h$. This representation is mapped to the output layer (the same size as the input layer), performing the reconstruction phase, $z \in R^d$. The hidden layer h is between the input and the output and describes a code used to represent the input data. The encoder can be represented by the function $h = f(x)$, while the decoder will create a reconstruction $r = g(h)$.

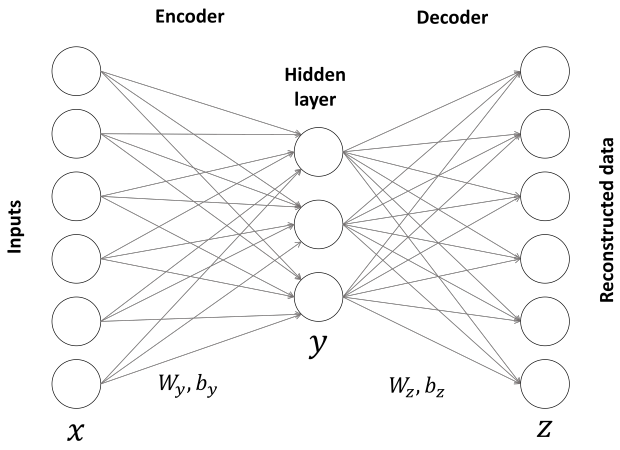


Figure 3.12: *Autoencoder layer example.*

Figure 3.12 shows an example of one layer of an AE. Considering that W_y and W_z represent the input-to-hidden and hidden-to-output weights, b_y and b_z the bias of hidden and output units, and $f(\cdot)$ the activation function (apply element-wise to its arguments), it can be described mathematically as:

$$y = f(W_y x + b_y) \tag{3.2}$$

$$z = f(W_z y + b_z) \tag{3.3}$$

Given that this algorithm tries to reconstruct the input data, the training phase goal is to minimise the error between the input and reconstructed data: $\arg\min_{W,b_y,b_z}[c(x,z)]$, where z is dependent of W, b_y, b_z, and x represents the input data. $c(x,z)$ represents

the error, which can be obtained with different loss functions, such as the mean squared error or L1 loss.

When multiple layers of AE are stacked together layer by layer is called a stacked Autoencoder [Windrim et al., 2019].

3.3.3.2 Convolutional Neural Network (CNN)

CNN were inspired by neuroscience and are constituted by several neurons that self-optimise during the training phase. CNN is composed of three different types of layers stacked:

- Convolutional layers with a nonlinear operation, such as Rectified Linear Unit (ReLU): used to extract features from the data. The first convolutional layer will extract low-level features (edges, colour, gradient information, among others) in visible images. The following convolutional layers will extract high-level features. The selected kernel will be used to perform convolutions along the input data.
- Pooling layers: used to decrease the size of the convolutional layer, mainly for computational resources. It will extract the most dominant features. It can be used as the maximum or the average pooling: the first will return the maximum value of the part covered with the kernel, working as a noise suppressant, while the second will return the average.
- Fully-connected layers: These layers learn the nonlinear combinations of the features extracted by the previous layers.

CNN is composed of blocks of convolutional layers followed by pooling layers and, in the end, the fully connected layers. These last layers are followed by a flatten layer to transform the image into a column vector fed to a feed-forward neural network, applying backpropagation to every training iteration. This is considered an epoch, and after a few, theoretically, the model can recognise the most meaningful features and use them to classify the data.

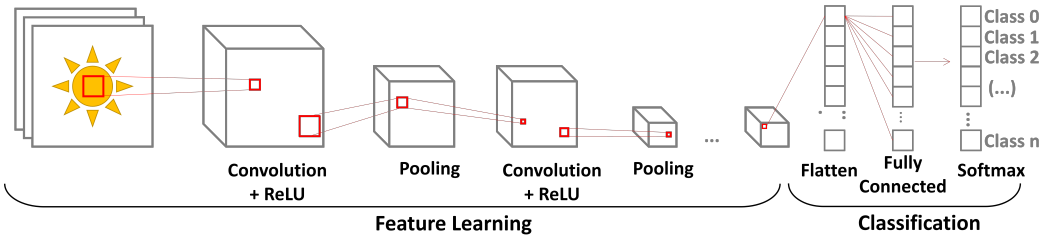


Figure 3.13: *CNN architecture example.*

In the case of spectral classifiers, CNN can be 1-D (using only spectral information), 2-D (using only spatial information) or 3-D (combining both spectral and

spatial information). Compared to visible images, the usage of spectral information only affects the number of channels in the input data. In the first case, there are only three channels, corresponding to red, green and blue. However, in the case of hyperspectral imaging, the number of channels varies according to the sensor used to collect the data. Usually, the hyperspectral data is constantly analysed pixel by pixel. However, to combine spectral and spatial information, there is a need to include spatial information. Due to the hyperspectral data characteristics and acquisition scheme, the spatial information is usually a region around the pixel, called the pixel neighbours. Given the number of neighbours of each pixel, it is possible to construct multiple 3-D blocks where the central pixel is the one that is being used to train and evaluate the model or classification.

3.3.3.3 Zero-Shot Learning

ZSL approaches aim to classify samples that were never seen during the training stage by replicating the Human capability to recognise objects with just their description, without having to see any visual aspect of the objects. This is done by creating an intermediate layer composed of distinguishable information of the seen classes. This layer is then used during the classification to understand the relationship between the seen and unseen features and then perform the prediction, which could include creating new classes.

In the case of RGB images, this information is usually composed of visual descriptions of the objects. For example, a zebra will have stripes, black and white colours, and other visual features that could be extracted from it. This information is then compared to the one present in the unseen classes, allowing the creation and aggregation of the samples into different classes, including new classes composed only of unseen class samples.

In the case of hyperspectral imaging, and given the remote sensing context of the work developed, it is difficult to have the detail level needed to capture relevant spatial information to give visual features to the algorithm. In this case, the distinguishable features can be obtained from the spectral information by using deep learning approaches that can extract the most meaningful features from the data, such as CNN. Figure 3.14 shows a brief example of the primary goal of ZSL approaches for hyperspectral imaging data.

3.4 Performance Metrics

To analyse the performance of the algorithms, some metrics need to be used to evaluate this. In the work developed in this thesis, we will use four different metrics: Overall Accuracy (OA), precision, recall and F1-score. The OA corresponds to the number of correct predictions in the universe of all predictions made. At the same time, the precision indicates the ability of the model to classify correctly (True Positive (TP)) in the universe of all the positive class values predicted (TP plus False Positive (FP)). A model with a low precision rate can be a sign of a high number of false positives. Recall indicates the number of samples correctly classified

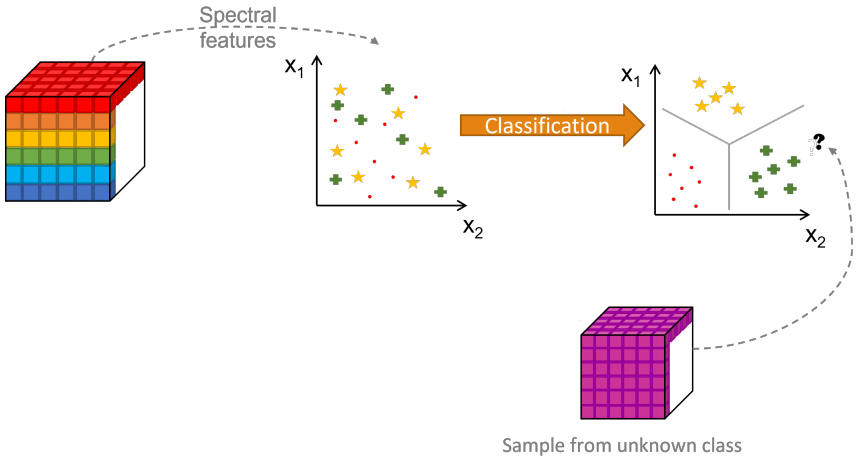


Figure 3.14: *ZSL scheme for hyperspectral imaging data*

(TP) in the number of positive class values existing in the dataset (TP plus False Negative (FN)). F1-Score shows a balance between the precision and recall rates.

All these metrics depend on the number of True Negative (TN), TP, FP and FN, which can be obtained by filling in the confusion matrix. Table 3.1 shows an example of a confusion matrix, where it is possible to see that:

- TP: corresponds to the true positives, the number of samples that were correctly classified as positive;
- TN: true negative, the number of samples that were correctly classified as negative;
- FP: false positive, the number of samples that were classified as belonging to one class but in reality its true label is another class;
- FN: false negative, number of samples that should have been classified as being of that class but were incorrectly classified as belonging to another class.

Table 3.1: *Confusion matrix example.*

		Predicted Label	
		Positive	**Negative**
True Label	**Positive**	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

With this, it is possible to determine the results for each metric by using Equation 3.4 to determine the OA, Equation 3.5 the precision, Equation 3.6 the recall and Equation 3.7 the F1-Score.

$$Overall\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.4}$$

$$Precision = \frac{TP}{TP + FP} \tag{3.5}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.6}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3.7}$$

3.5 Summary

This chapter describes the fundamental concepts that were utilised during the work performed in this thesis. Briefly summarising, it discusses the fundamental working principles of a hyperspectral camera and its data acquisition schemes. Furthermore, it also describes some of the basic approaches of the developed methods during this thesis.

4

Data Acquisition Setup

4.1 Introduction

This chapter describes the work performed to acquire the remote hyperspectral imaging data used to develop and carry out the main scientific work of the thesis. The remote data acquisition with airborne and satellite platforms was performed in the framework of the ESAPLASTICS[1] project, which is a project financed by the ESA General Support Technology Programme. This project aimed to develop and assemble a remote hyperspectral imaging payload and then perform extensive data acquisition using multiple platforms: satellite, airborne and UAV. This data collection will allow to detect and classify marine litter using novel post-processing machine/deep learning algorithms.

4.2 ESAPlastics Project

ESA is currently looking and fostering the development of spaceborne solutions and campaigns for detecting and classifying marine litter in the oceans. Through the Open Space Innovation Platform (OSIP)[2], several projects have been funded to research novel ways to detect, classify, and remove marine litter using heterogeneous sensors, e.g., RGB cameras, radar, multispectral, hyperspectral, and using different remote sensing platforms, e.g., boats, drones, aeroplanes, and satellites.

One of the projects that address this subject is the ESAPLASTICS project, see

[1] https://www.aircentre.org/esaplastics/

[2] https://www.esa.int/Enabling_Support/Preparing_for_the_Future/Discovery_and_Preparation/The_Open_Space_Innovation_Platform_OSIP

concept Figure 4.1. The project main objective was to evaluate which wavelengths (bands) within the visible and SWIR spectral range are more suitable for remotely detecting marine litter from space.

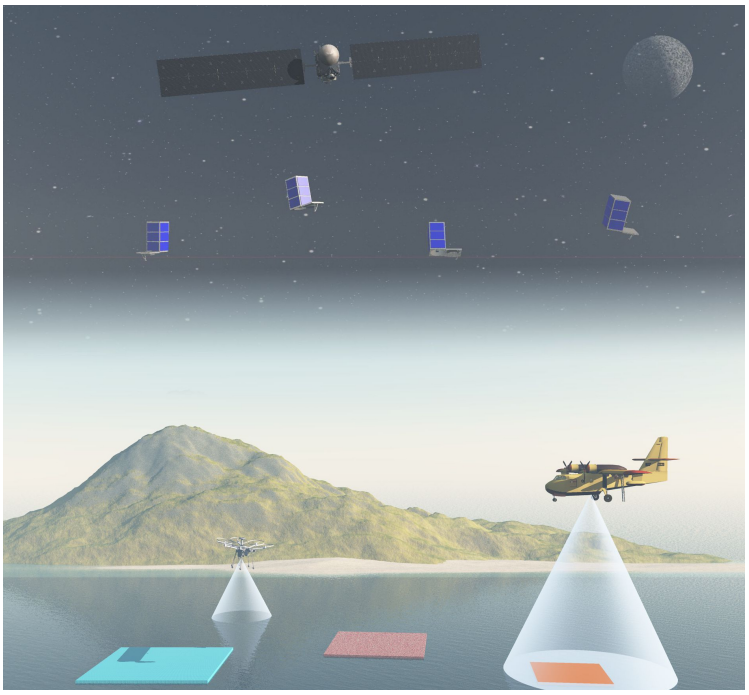


Figure 4.1: *ESAPlastics project scheme: the artificial targets are placed in the water, while the manned and unmanned aircraft and the satellite are acquiring data.*

The ESAPLASTICS project comprises several tasks, namely:

- Evaluate the state-of-the-art technological approaches for marine litter detection using remote sensing applications;
- Develop and assemble a remote hyperspectral imaging payload for synchronised data acquisition;
- Perform extensive data acquisition using satellite, UAV (drones) and manned aircraft in a marine litter hotspot using artificial (marine litter-based) targets;
- Develop and test post-processing machine learning/deep learning algorithms to detect marine litter based on the spectral/spatial response of the acquired data. Compare the data obtained with Sentinel-2 multispectral image information;
- Evaluate marine litter samples in a laboratory environment using Fourier-Transform Infra-Red Spectroscopy (FTIR), Raman Spectroscopy, X-Ray Fluorescence (XRF), Laser-Induced Breakdown Spectroscopy (LIBS) (out of the subject of this thesis);
- Evaluate compressive sensing techniques for developing a single-pixel hyperspectral camera.

Not all of these tasks were useful and performed during the scope of this thesis. Therefore, we will focus on the first four tasks, which constitute the main core of the scientific work related to marine litter applications performed during the thesis.

4.3 Hardware Configuration Acquisition Setup

The hyperspectral imaging system setup holds two pushbroom hyperspectral cameras [Freitas et al., 2018], a CPU unit, GPS, and an inertial system. One of the hyperspectral cameras is a Specim FX10e, a GigE camera with a spectral range from 400 to 1000 nm wavelength and a spatial resolution of 1024 pixels. The other camera is a HySpex Mjolnir S-620, a hyperspectral system composed of the hyperspectral camera, a data acquisition unit, GPS, a dedicated inertial system, a control board, power control and distribution, and a RGB camera, integrated into the same closed box. This hyperspectral camera has 620 pixels of spatial resolution and covers a spectral range from 970 to 2500 nm wavelength. For more details about the characteristics of the camera, see Appendix C.

To perform the dataset acquisition using the two sensors, we need to synchronise the data acquisition. Given that HySpex Mjolnir S-620 also has a controller board connected to the GPS/INS to have a synchronised external trigger, it was possible to use this trigger of the HySpex system to trigger the Specim FX10e. This allows for the acquisition of synchronised data between both cameras, with the GPS/INS as a global time reference. This system also has a data acquisition unit, allowing hyperspectral data acquisition up to 100 Hz. The framerate should always be related to the flight altitude and speed. In the case of Specim FX10e, the data needs to be acquired from an external data acquisition unit due to the high framerate. For lower framerates, it can be acquired using the data acquisition unit from the HySpex system. In Figure 4.2 is possible to see an illustration of the hardware connections of the system.

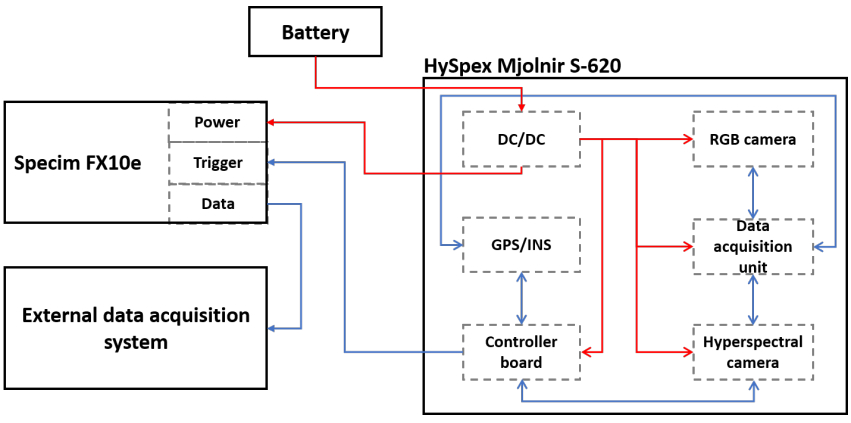


Figure 4.2: *Connections between the hyperspectral cameras, data acquisition computers and GPS/INS.*

Concerning the data acquisition, the HySpex system stores a HySpex file con-

taining all the raw data collected by the camera. After the flight, the data is post-processed using the HySpex Rad software, which calibres and converts the raw data to radiance by removing the effects of the sensor during the raw data acquisition. For the SPECIM FX10e, the raw data is acquired using our own Robotic Operating System (ROS) node customised data acquisition software (driver). Afterwards, the calibrated manufacturer gains are used to convert the raw data into radiance data. The GPS/INS system saves all the flight data into multiple files. Afterwards, using the Applanix POSPac [Hutton et al., 2016, Hutton et al., 2007] software is possible to post-process the data and obtain one position and orientation for each acquired hyperspectral imaging data line from the pushbroom system.

The assembled hyperspectral imaging setup was used to perform a set of tests with three different platforms: in-situ, aircraft and airborne.

4.4 In-situ Characterisation

The first set of tests consisted of an in-situ characterisation of different samples found in the marine litter test site. Figure 4.3 shows a set of 34 catalogued marine litter samples collected from the test site. Although all samples were analysed using different techniques such as FTIR, Raman Spectroscopy, XRF and LIBS in the scope of the ESAPlastics project, only a set of samples were selected by ESA to set up the targets for the in-situ analysis and dataset campaign.



Figure 4.3: *Marine litter samples collected in the test site - Porto Pim bay, Faial Island, Azores*

To perform the in-situ tests, we develop a physical layout to hold both cameras

at the same height. The structure also can change the distance from the cameras to the samples, and all the tests were performed with a one-meter distance between the cameras and the marine litter samples. Figure 4.4 shows the structure mounted with both cameras while one sample is being analysed.



Figure 4.4: *Test setup used during the in-situ tests.*

The entire structure was built in black to minimise the sunlight reflections captured by the camera. In addition, the water placed in the tank is seawater collected in the Atlantic Ocean, the same ocean present in the test zone. The tests were performed at different times of the day, both with the dry samples and submerged in the seawater. The following section details the in-situ tests.

4.4.1 In-situ Marine Litter Samples Characterisation

The main goal of the in-situ tests was to perform laboratory tests before the flight campaigns and create a ground-truth spectral response for comparison of different marine litter samples. The test setup was mounted in an outdoor environment, allowing the characterisation of the spectral response of the marine litter samples under direct sunlight and varying natural light conditions, i.e., morning and afternoon, using dry and wet samples. The study objective was to understand better the water and light incidence effects in the spectral analysis of the marine litter samples. The marine litter samples used in the test are identical to the ones used in the artificial targets on the airborne dataset campaigns, and the goal of using wet samples is to mimic the effects of the water in the artificial targets. Using a water tank to submerge the samples makes it possible to imitate the phenomena that happen with ropes, for example, where they are partially submerged by water.

The results obtained for the artificial target materials during the morning and afternoon tests for different light conditions are displayed in Figure 4.5. The results concerning the dry and wet samples are very similar, noticing an attenuation on the energy of the signal in the wet samples due to water effects, but without significant variation in the absorption of spectral bands wavelength.

4.5 Flight Data Acquisition

The second and third tests were performed using the hyperspectral imaging setup in two different aircraft: Grifo-X and F-BUMA. This section describes the dataset campaign using the airborne platform for data acquisition.

4.5.1 Dataset Campaign

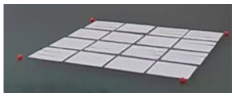
The dataset campaign occurred in Faial Island Azores from the 16th until the 25th of September 2020. The artificial target setups were placed in the Porto Pim Bay area for conducting the hyperspectral image data acquisition. In Figure 4.6 is possible to observe the location of three artificial targets built using the marine litter samples during the campaign.

The targets dimension was chosen to be 10 × 10 m, corresponding to the minimal resolution (1 × 1 pixel) to be detected by Sentinel-2. These targets were composed of different types of plastic, namely:

- Orange polypropylene plastic used for oyster aquaculture. This type of plastic is widely found in the hotspot analysed and can be observed in the first Figure presented in Table 4.1;
- White polyethylene plastic used to cover greenhouses, shown in the second Figure of Table 4.1;
- Blue polypropylene ropes used for the fishery, but with two different diameter sizes. This type of plastic is also widely found in the hotspot tested and can be observed in the third Figure of Table 4.1.

The targets were mounted in different configurations, namely using one type of material for each target, or combining different types of materials in one target. During the F-BUMA flights, the targets were used in the first configuration: each target was composed of only one type of material, and Grifo-X flights used both targets configuration. Table 4.1 displays an example of the targets mounted with the first configuration, where each target is composed of a single plastic type.

Table 4.1: *Targets used during the dataset collection and their division between the classes.*

Orange Plastic Target (Orange)	**White Plastic Target (White)**	**Ropes Target (Ropes)**
		

The dataset campaign also allowed us to capture information about more artefacts on the campaign site, such as water, concrete pier, trees and boats. Table 4.2 shows an example of each one of these artefacts.

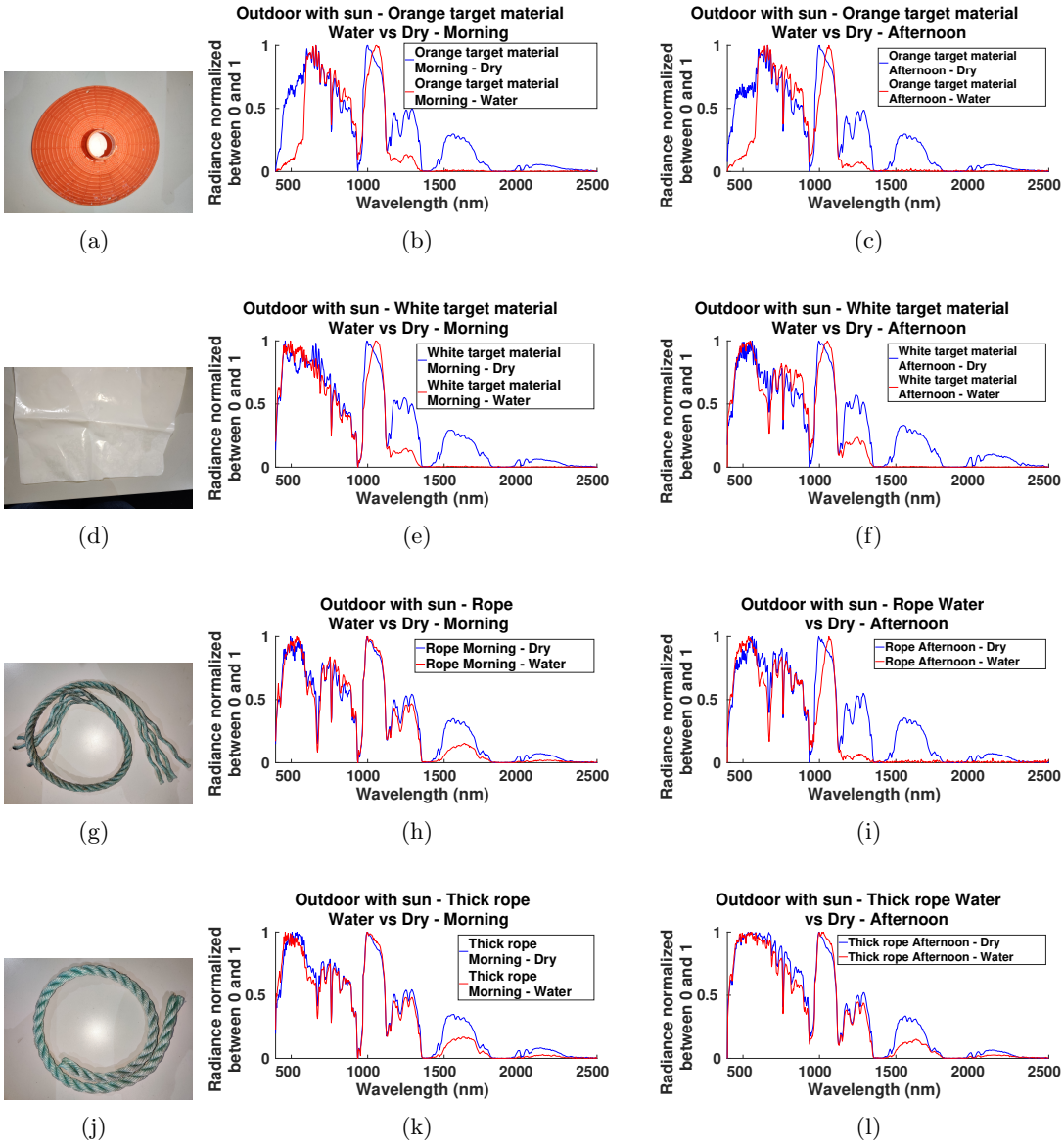


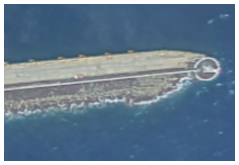
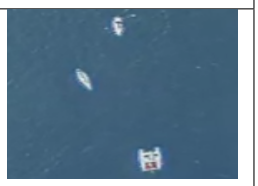
Figure 4.5: *Plastics used in the targets and the laboratory results in normalised radiance for each target material: (**a**) orange target material (with approximately 20 cm diameter), (**b**) and (**c**) laboratory results for the orange target material, respectively; (**d**) white target material (with approximately 50 cm width, with variable length), (**e**) and (**f**) laboratory results for the orange target material, respectively; (**g**) thin rope target material (3 cm diameter, with variable length), (**h**) and (**i**) laboratory results for the orange target material, respectively; (**j**) thick rope target material (5 cm diameter, with variable length), (**k**) and (**l**) laboratory results for the orange target material, respectively;*



(a) (b)

Figure 4.6: *Artificial targets: (**a**) image of the three targets during the trials. From left to right: low-density polyethylene orange target (with used oyster spat collectors), white plastic film target and rope target. (**b**) view of the artificial targets from the F-BUMA aeroplane during the flight.*

Table 4.2: *Other artefacts captured during the dataset campaign.*

Water	**Concrete pier**	**Trees**	**Boats**
			

The dataset campaign was two weeks long, and the weather conditions during the course of the two weeks were not perfect, therefore, not all days were used to perform the dataset flights. The F-BUMA aeroplane flew four times in three days. The flight altitude varied from 500 to 750 meters, while the speed range was between 20 m/s and 30 m/s. The data acquisition campaign using the drone, Grifo-X, was performed over three days, resulting in five flights. The Grifo-X flight altitude varied from 25 to 35 meters, while the speed was one m/s. The detailed log book of both F-BUMA and Grifo-X UAV are presented in appendix B.

4.5.2 Hyperspectral Imaging System Setup - Manned Aircraft (F-BUMA)

The first aerial platform utilised in the data acquisition during the ESAPlastics project was a manned aircraft, a Cessna F150L (F-BUMA). Figure 4.7 shows the system already assembled in the aircraft's landing gear. The GPS antenna was placed in the rear window of the aircraft, inside the aeroplane and facing up. The structure that holds the system was built to ensure that the lenses were at the same height, keeping the cameras physically aligned.



Figure 4.7: *Hyperspectral imaging system mounted in the manned aircraft (F-BUMA).*

4.5.2.1 F-BUMA Flight Data

This section presents the preliminary F-BUMA flight data and a first approximation to characterise the obtained hyperspectral imaging results. In Figure 4.8 it is possible to observe the waterfall image (obtained by resorting to all the lines acquired, considering only a few bands to create a false RGB image) from the flight, with 500-600 m altitude.

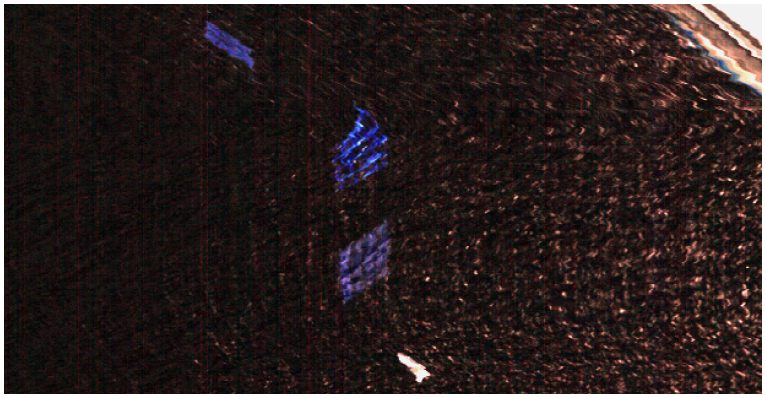


Figure 4.8: *Waterfall image of the flyby analysed from the 18/9 afternoon F-BUMA flight.*

The flight and laboratory data were compared using three different pixels from the three distinct targets: a bright pixel, a medium-intensity pixel and a dark pixel. The pixels were manually selected using the hyperspectral waterfall image. In Figure 4.9 is possible to see the spectral response (in normalised radiance) for all three targets, compared with laboratory results for the wet samples. As can be observed in these results, the absorption peaks are the same for each target material, regardless of where the data was acquired. The attenuation of the signal is due to two different factors:

- The change in altitude/distance from the camera to the samples;
- There were differences between the in-situ and the airborne tests. While in the in-situ tests, the weather was clear, during the dataset campaign, the sky was cloudy. This difference is more clear in the visible part of the spectrum.

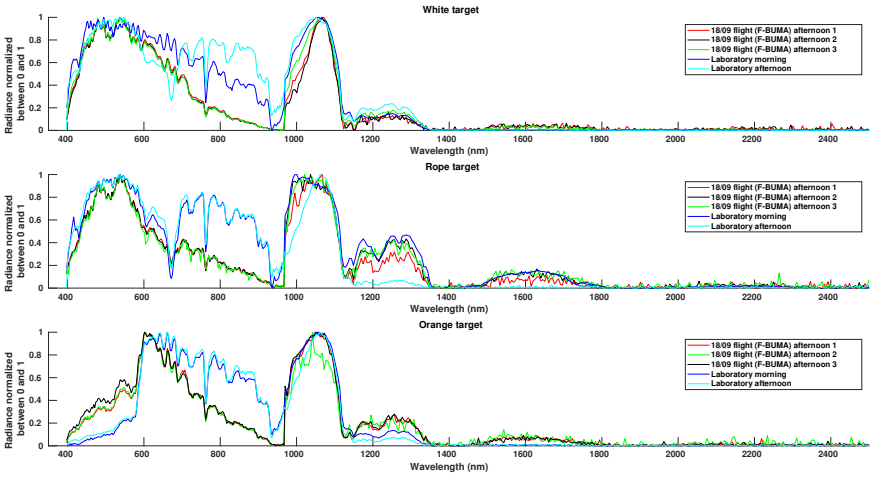


Figure 4.9: *Results for the 18/9 F-BUMA afternoon flight, analysing three different pixels (bright pixel-afternoon 1, a medium pixel-afternoon 2 and a dark pixel-afternoon 3) for each material, and comparing them to the laboratory results.*

4.5.3 Hyperspectral Imaging System Setup - Unmanned Aircraft (Grifo-X)

The second aerial platform utilised in the project was an Unmanned Aerial Vehicle (UAV), Grifo-X, developed at CRAS (Center for Robotic and Autonomous Systems) from INESCTEC. The GPS antenna was installed in a custom support on top of the UAV frame. Figure 4.10 shows the hyperspectral imaging system mounted onto the Grifo-X UAV before a flight in Porto Pim bay.



Figure 4.10: *Hyperspectral imaging system mounted in the unmanned aircraft (Grifo-X).*

4.5.3.1 Grifo-X Flight Data

This section presents the Grifo-X flight data preliminary results and a first approximation to characterise the obtained hyperspectral imaging data results, similar to the performed for F-BUMA. In Figure 4.11 is possible to see the waterfall image for this flight at 25-35 meters altitude.

A similar comparison between the flight and laboratory data was performed using the same type of pixels: a bright pixel, a medium intensity pixel and a dark

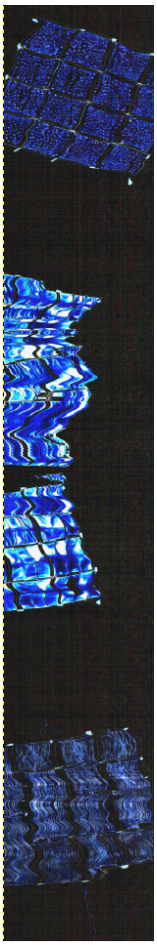


Figure 4.11: *Waterfall image of the flyby analysed from the 25/9 Grifo-X flight 1.*

pixel, selected from visual inspection using the waterfall image. In Figure 4.12 is possible to see the spectral response (in normalised radiance) for all three targets, compared with laboratory results for the wet samples. The effects and conclusions are the same as previously presented for F-BUMA. Although the flight altitude is lower, the attenuation is still verified, showing that it is mainly in the visible part of the spectrum due to the sun exposure factor.

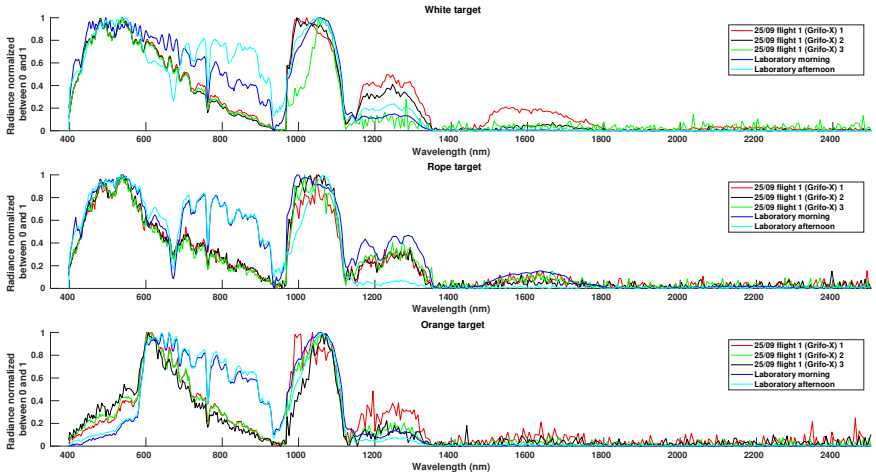


Figure 4.12: *Results for the 25/9 Grifo-X first flight, analysing three different pixels (bright pixel-1, a medium pixel-2 and a dark pixel-3) for each material, and comparing them to the laboratory results.*

Figure 4.13 compares the bright pixels for the 25/09 flight 1 with Grifo-X, 20/09 F-BUMA flight and the laboratory morning results for all the target materials.

Both aircraft flew close to the Sentinel-2 satellite revisiting time. However, the data collected by Sentinel-2 on the flight days were not as good as expected due to cloud coverage. In our experimental setup, the cloud coverage did not significantly affect the hyperspectral imaging system acquisition, as shown in Figure 4.14 where it is possible to see that the aeroplane, the UAV, and the laboratory setup produce similar results concerning the spectral profile of one of the targets. However, it is known that cloud coverage affects hyperspectral data acquisition results and can reduce the amount of direct sunlight onto the targets, influencing the reflections' energy and shifting absorption bands as described in more detail in [Arroyo-Mora et al., 2021].

In Figure 4.15a is possible to see the trajectory of the F-BUMA aircraft during the 20 September 2020 flight, and in Figure 4.15b is possible to observe the smaller trajectory of the GRIFO-X UAV during the 25 September 2020 flight.

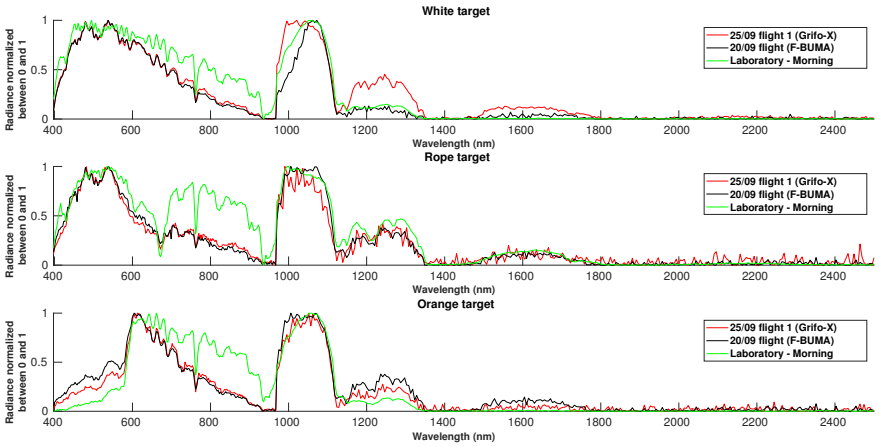


Figure 4.13: *Comparison between the bright pixels for the 25/09 flight 1 with Grifo-X, 20/09 F-BUMA flight and the laboratory morning results for all the target materials.*

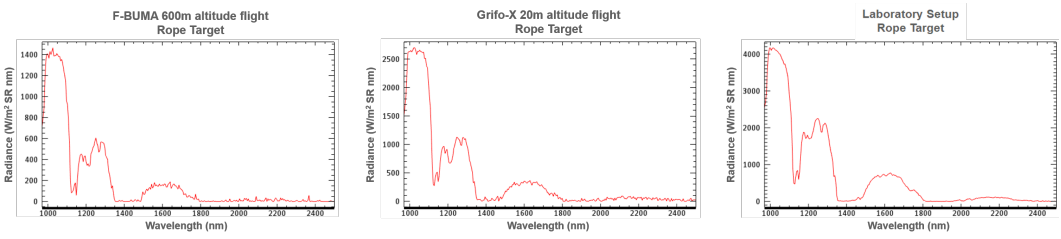


Figure 4.14: *Comparison between the rope target results for F-BUMA flight (600 m altitude), Grifo-X flight (20 m altitude) and laboratory setup, with a focus on HySpex Mjolnir S-620 bands (970 - 2500 nm).*



(a) (b)

Figure 4.15: *Flight trajectories: (**a**) shows the 20 September 2020 flight trajectory, performed by F-BUMA aircraft, while in (**b**) is possible to observe the trajectory of the 25 September 2020 flight performed with GRIFO-X aircraft.*

4.6 Summary

This chapter details the data acquisition performed in the framework of the ESAPlastics project. Afterwards, it presents the setup comprising two pushbroom hyperspectral cameras from 400 to 2500 nm. The first step was to analyse the marine litter samples by resorting to in-situ tests, using the samples wet and submerged in seawater to simulate a real test site. Then by using artificial targets with 10 $\times$ 10 m, it was possible to perform a marine litter dataset campaign to collect an extensive amount of data. This dataset campaign was performed in Porto Pim bay, Faial Island, Azores, for two weeks, one week for each platform. The F-BUMA aircraft flew four times, with an altitude speed ranging from 500 to 750 meters and 20 m/s to 30 m/s speeds. The UAV performed five flights, always with 1 m/s speed and flight altitudes from 25 to 35 meters. The flights resulted in approximately 2TB of data and allowed us to understand the effects of changing the altitude in the data collected by the system, especially in the spectral response and image resolution, and the effects it produces when compared to the in-situ analysis of a given target material.

5

Supervised Approaches for Hyperspectral Imaging Classification

5.1 Introduction

This chapter describes our work for detecting and classifying marine litter, it presents three supervised algorithms for marine litter detection and classification. The first two approaches, Random Forests and Support Vector Machine, are two state-of-the-art algorithms regularly used with hyperspectral imaging data. The third is an implementation of a CNN-3D which was developed to integrate spectral and spatial information for marine litter detection and classification.

5.2 Support Vector Machines and Random Forests - Implementation

The characterisation of the acquired hyperspectral imaging data is essential. However, to help solve the problem of marine litter detection and classification, automated means of detection must be used to detect and classify marine litter in vast amounts of data.

Therefore, in this work, we develop, train, and compare the obtained results of two Machine Learning (ML) algorithms to detect and classify marine litter samples, namely: RF [Breiman, 2001] and SVM [Boser et al., 1992].

The model's training was performed with data collected during the F-BUMA 18 September 2020 flight, which was divided to be used for training and fine-tuning

Table 5.1: *Targets used during the dataset collection and their division between the classes.*

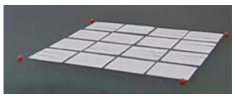
Class 0	**Class 1**	**Class 2**	**Class 3**
Water and Land (Houses, Trees, Streets, Cars, and Others Materials)	Orange Plastic Target	White Plastic Target	Ropes Target
			

Table 5.2: *Number of pixels (NP) present for each class in 18/9 F-BUMA flight: class 1 corresponds to the orange target data, class 2 to the white target and class 3 to the rope target data. Class 0 was created with all the remaining pixels containing land and water data.*

Flyby	**Class 0 (NP)**	**Class 1 (NP)**	**Class 2 (NP)**	**Class 3 (NP)**
0	57259	299	664	678
1	125732	399	857	732
2	86479	454	1203	1144
3	153180	462	1880	1338
4	51438	700	1182	0
5	138612	620	1658	1710
6	87266	696	1358	1200
7	117078	489	1373	1960

of the model parameters. The second flight, from 20 September 2020, was used for testing the results. This means the model was trained using data from a different day than the one used for the prediction. The data was measured in raw units and then transformed to radiance. By doing so, we are assessing the possibility of soon having a hyperspectral imaging system capable of detecting marine litter concentrations in near real-time, which needs to be performed with radiance data.

Given the data characteristics, the flights were divided into flybys over the target (only considering the data in the target vicinity). We only used flybys that had acquired the three targets for training and prediction. In Table 5.1 is possible to observe some examples of the class components. Tables 5.2 and 5.3 show each flyby's pixel distribution for each class. Class 1 represents the orange target data, class 2 corresponds to the white target data, and class 3 is the rope target data. Class 0 was created with all the remaining pixels containing land and water data. It was necessary to make this class to train the methods to recognise the characteristics of the non-marine litter pixels by grouping them all in one class.

To overcome some limitations in the training of the data, some training techniques were adopted in the experiments, namely:

Table 5.3: *Number of pixels (NP) present for each class in 20/9 F-BUMA flight: class 1 corresponds to the orange target data, class 2 to the white target and class 3 to the rope target data. Class 0 was created with all the remaining pixels containing land and water data.*

Flyby	Class 0 (NP)	Class 1 (NP)	Class 2 (NP)	Class 3 (NP)
0	15604	516	0	0
1	29522	858	0	0
2	47986	662	671	281
3	72742	715	1157	1026
4	52024	626	670	0
5	94414	1246	781	899
6	167177	1547	1682	2574
7	81713	813	864	930
8	183419	2640	1928	1733

- Class imbalance: As can be observed in Table 5.2 and Table 5.3, the number of pixels for each class has a considerable variation. Some flybys do not hold information about all the classes, and the difference between the class 0 pixels and the other classes is significant. Therefore, we randomly selected some samples so that the number of points was identical for each class. This was required so the class with the highest number of points does not become dominant, leading to overfitting;

- Feature normalisation: Given the characteristics of the hyperspectral data that is affected by the variability of atmospheric conditions, particularly direct sunlight, it is necessary to carry out its normalisation. To perform this step, we used the *sklearn* python library pre-processing module to normalise the feature data to unit variance.

- Given the characteristics of the flight (altitude), the targets will appear relatively small (resolution) in the datasets. In addition, there are targets that, due to their physical construction, will have water in the middle, such as the orange (class 1) and rope targets (class 3). Furthermore, due to the size of targets in the final dataset, marking the ground truth to exclude the water in the middle of the target completely is challenging. Therefore, it is still necessary to take this into account and try to balance this with the number of pixels provided to train the algorithm;

- The targets were always located in the same area. However, due to some flight variations mainly because of atmospheric conditions, there are flybys in which a considerable amount of land was observed. To allow the methods to learn the terrain characteristics, some land points were added to class 0. These land points contain data from houses, cars, ropes, trees, sand, and other objects. Due to this effect, the number of land pixels increased considerably to ensure that when selecting the points randomly, all these classes had some

representation in class 0. This class is not a pure class (it does not contain only one type of material), but it is a class with a combination of dozens of materials that are not intended to be identified per se.

Considering all the constraints mentioned above, the data were then processed for training. In this case, the number of pixels for ground truth in each class was the following:

- Class 0: 170773 randomly selected pixels;
- Class 1: All the pixels contained in the training dataset, 4119 pixels;
- Class 2: 4119 randomly selected pixels;
- Class 3: 7000 randomly selected pixels;

The selection of the number of pixels for training each class was based on the maximum number of pixels present in the training data for class 1 since this was the class with fewer available pixels to identify. During the training process, we also had to add more class 3 pixels since it was challenging to select pure rope pixels manually. Both RF and SVM methods implementation based on Python sklearn used the same training points from the 18/9 F-BUMA flight and the same test data from the 20/9 F-BUMA flight. We used a grid search method for both algorithms that allowed fine-tuning both RF and SVM to their best parameters. In the case of SVM, the parameters changed were: C-100, kernel-RBF and gamma-0.0001. For RF, the parameters changed were: number of trees-3500, maximum features-log2, maximum depth-10, the minimum number of samples in a leaf node-1 and minimum samples required to split an internal node-5.

5.2.1 Results

Both RF and SVM methods were trained with the training data from the 18 September 2020 flight. The aim was to create a model that would be evaluated using the data collected during the 20 September 2020 flight. Table 5.4 shows the prediction analysis for both algorithms (RF, SVM), using the following metrics: precision, recall, f1-score, and accuracy. For these tests, we only used the flybys with all targets present.

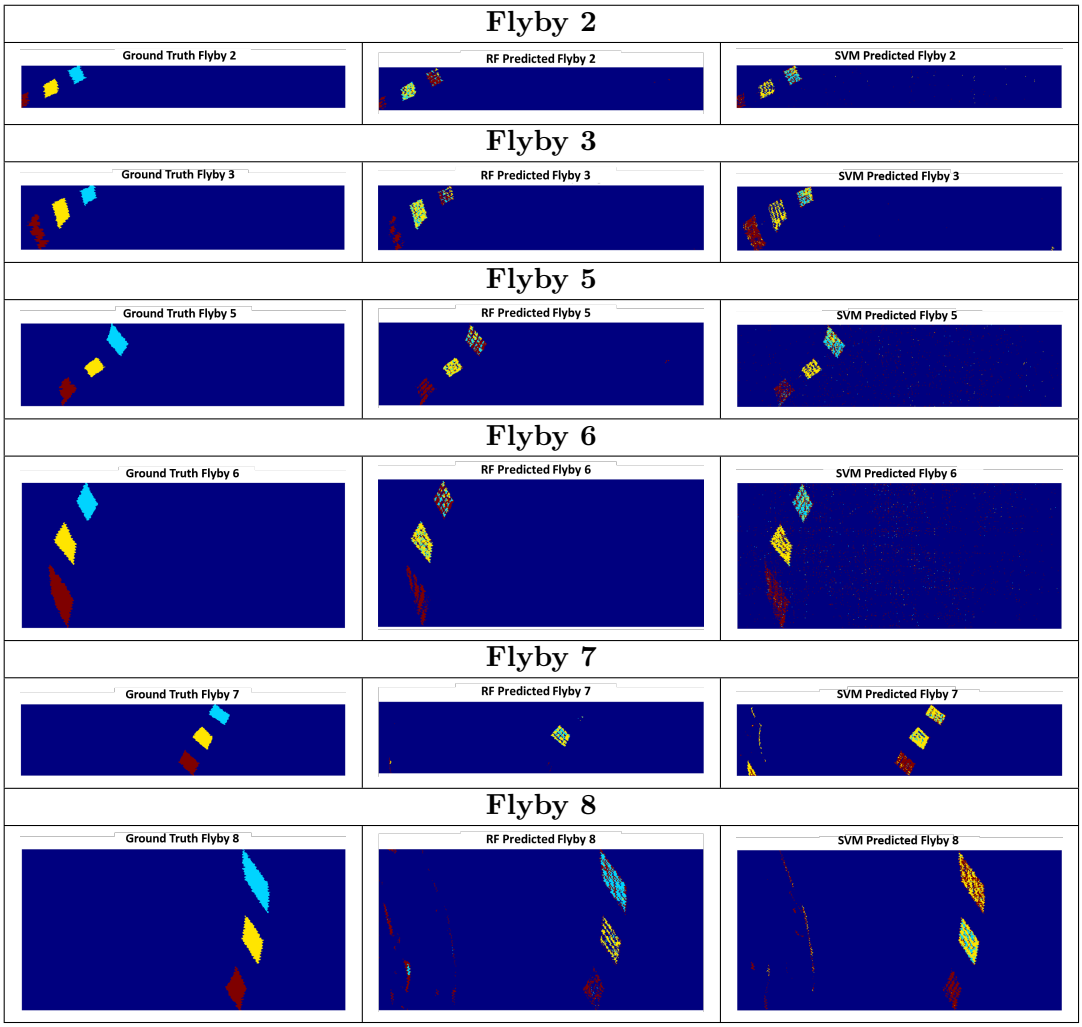
Table 5.5 illustrates the detection and classification results compared with ground-truth information.

The fact that the targets had 10 × 10 m dimensions composed of 4 × 4 squares, each one with 2.5 × 2.5 m, negatively influenced the results, which means that at 600 m average flight altitude is challenging to annotate ground-truth pixels as "pure" manually. Therefore, there is no absolute way to be sure that the target pixel is not over-dominated by water. Even though the targets are still able to be identified in most of the flybys, the precision rate and the recall rate are not uniform for all three targets indicating that on some of those flybys, the targets could be submerged in some areas, which is something to be expected in real environment scenarios.

Table 5.4: *Results obtained for the test data (20/09/2020 flight) for RF and SVM algorithms.*

Flyby 2 - F-BUMA								
	RF				**SVM**			
Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**
0	0.99	1	0.99	97.35%	0.99	0.99	0.99	97.47%
1	0.24	0.12	0.16		0.75	0.42	0.54	
2	0.85	0.42	0.56		0.66	0.47	0.55	
3	0.18	0.44	0.26		0.22	0.50	0.31	
Flyby 3 - F-BUMA								
	RF				**SVM**			
Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**
0	0.99	1	0.99	97.34%	0.99	0.99	0.99	97.27%
1	0.20	0.13	0.16		0.75	0.38	0.51	
2	0.85	0.53	0.65		0.56	0.43	0.49	
3	0.45	0.38	0.41		0.52	0.74	0.61	
Flyby 5 - F-BUMA								
	RF				**SVM**			
Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**
0	0.99	1	0.99	98.09%	0.99	0.99	0.99	97.06%
1	0.71	0.25	0.37		0.77	0.46	0.58	
2	0.74	0.62	0.67		0.57	0.59	0.58	
3	0.38	0.56	0.45		0.21	0.47	0.29	
Flyby 6 - F-BUMA								
	RF				**SVM**			
Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**
0	0.99	1	0.99	97.92%	0.99	0.98	0.98	96.56%
1	0.52	0.22	0.31		0.71	0.48	0.57	
2	0.85	0.59	0.70		0.78	0.67	0.72	
3	0.50	0.42	0.46		0.27	0.51	0.35	
Flyby 7 - F-BUMA								
	RF				**SVM**			
Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**
0	0.98	1	0.99	97.32%	1	0.99	0.99	98.01%
1	0.01	0	0.01		0.50	0.16	0.24	
2	0.95	0.45	0.61		0.44	0.72	0.54	
3	0.17	0.01	0.02		0.64	0.76	0.70	
Flyby 8 - F-BUMA								
	RF				**SVM**			
Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**
0	0.99	0.99	0.99	97.44%	0.99	0.99	0.99	97.28%
1	0.85	0.55	0.66		0.10	0.02	0.04	
2	0.94	0.36	0.52		0.46	0.55	0.50	
3	0.21	0.35	0.26		0.30	0.51	0.38	

Table 5.5: *Figures show the results obtained for the test data (20 September 2020 flight) for RF and SVM algorithms. The first column illustrates the ground truth for each flyby, the second column shows the prediction made by the random forest algorithm, and the last column represents the predicted values by the support vector machine algorithm.*



5.2.2 Discussion

There is an urgent need to understand and monitor marine litter concentrations. This work focused on developing a remote hyperspectral imaging system for detecting and classifying marine litter. The work was conducted to analyse marine litter samples from a known marine litter hotspot (Porto Pim Bay, Faial Island, Azores, Portugal). First, samples were collected and analysed in an outdoor environment to retrieve each sample's reference spectra. The second step was to develop the hyperspectral imaging setup and place it onboard two aircraft: the F-BUMA, a manned aircraft, and Grifo-X, an UAV. The payload was the same for both platforms, which allowed data acquisition at different altitudes. The final test trials took place in Porto Pim Bay, Faial Island, from the 16th to the 25th of September. Although, due to atmospheric conditions, it was not possible to compare the results with Sentinel-2 data, the results are promising. The results show similarities between the absorption bands obtained in the laboratory with the field trial results. It is also possible to observe that the difference in the flight altitude between the two aircraft does not affect the spectral response absorption bands of the plastics.

Concerning the implemented methods (RF, SVM), it is possible to see from the Tables mentioned above 5.3 and 5.4 that the SVM provides slightly more accurate results. The SVM can detect all three targets in all the flybys, but the precision and recall values vary between the flybys. RF provides more stable results when it detects a target but sometimes cannot detect the targets, see flybys 7 and 8.

However, even with the SVM method, the results show a non-uniform detection pattern, particularly in class 2. This is the case since the class 2 target is a more standard plastic type that can be found in structures of houses or agriculture, making it easier to be detected outside the target areas. In flybys 7 and 8, it is possible to identify anomalies in class 1 and class 2 results in the SVM method. Both flybys show the presence of "land" in identifying the targets, which are included in class 0. This class is not a water target-related class but a "non-marine litter" class. This fact suggests the appearance of unknown artefacts of class 0 that were wrongly classified as class 2 in flyby 7 and class 1 in flyby 8.

Overall, the RF and SVM methods show potential to be able to detect marine litter, i.e., plastic samples using hyperspectral imaging onboard an aircraft at a reasonable 600 meters flight altitude, with 0.70-0.80 precision values and few false positives. The model trained from the previous flight days allowed the classification of most of the targets correctly. Still, the presence of submerged target pixels made it more difficult for the methods to accurately classify all the target pixels, making the distinction between the different types of plastic that constituted each target extremely challenging.

5.3 CNN-3D Spectral/Spatial Marine Litter Detection and Classification - Implementation

In this section, we develop and train a 3D Convolutional Neural Network (CNN-3D) to assess the applicability of deep learning methods using hyperspectral imaging data to marine litter detection and classification. For hyperspectral imaging data, a CNN-3D means the neural network will use spatial and spectral information to classify the data. Therefore, the network's input will use small patches containing spectral and spatial information of the central pixel and their neighbours. For this evaluation, we use only one of the Grifo-X flights (24/09/2020), mainly due to the time-consuming manual annotation of all the pixels for ground truth and training purposes. Ground truth labelling is an important task, but very time-consuming and prone to errors. Considering that even with RGB image, this can influence the results when working with hyperspectral imaging data, this issue is even more prominent due to the need to make the ground truth for each pixel. Figure 5.1 shows the UAV flying over the targets in Porto Pim bay during the dataset campaign.



Figure 5.1: *Grifo-X flying over the targets during the dataset campaign.*

In this case, due to the significant difference between the number of pixels for each class, the division was made using the flybys (which contain the data collected over the target). In table 5.6 is possible to see the number of pixels present in each class for each flyby. Class 0 represents the water, class 1 the orange target, class 2 is the white target, and class 3 is the rope target.

The dataset on the day consisted of six flybys (0-5) and was divided into three parts: train, test and validation. For training the model, we used flybys 1, 2 and 3; for the test, we used flyby 0. While for validation purposes, we used the last two flybys, 4 and 5.

For conducting the CNN-3D training, we performed the following procedures:

- All the data were normalised to have zero mean and shuffled before providing it to the CNN-3D as input;
- Not all pixels available for training were used: due to each class characteristics, class 1, class 2, and class 3 had the number of points used balanced according

Table 5.6: *Number of pixels (NP) present in the Grifo-X dataset used for each class*

Flyby	Class 0 (NP) Water	Class 1 (NP) Orange target	Class 2 (NP) White target	Class 3 (NP) Rope target
0	174109	2197	57274	780
1	186348	24747	53104	6741
2	160437	29896	58406	43901
3	100532	42874	81495	59059
4	214882	21852	48505	20421
5	212794	37957	85579	41870

to the value of the class that had the least points (class 1). However, due to the class 0 variability, there was a need to use more points in this class. One of the main reasons for this is that some of the targets are partially submerged. To overcome this issue, we increased the number of pixels for class 0;

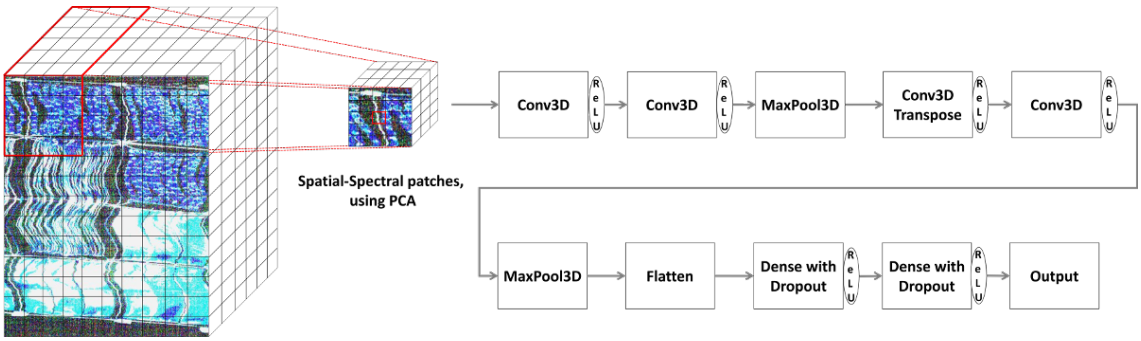


Figure 5.2: *Flowchart of the proposed CNN-3D architecture.*

To extract the most meaningful features from the data, before giving it as input to the CNN-3D, the data was pre-processed using the Principal Component Analysis (PCA) algorithm. In this case, the number of principal components used was 15. Another parameter to be tuned is the number of neighbours considered for each patch, which in this case was 11 $\times$ 11 pixels square. So, the input shape for the input layer of the CNN-3D has the following shape: $[p, p, n]$, where p corresponds to the patch size and n to the number of principal components used. After this layer, two 3D Convolutional layers are applied, followed by a 3D Max Pooling layer. After this, a 3D Convolutional layer is applied, followed by another 3D Convolutional layer and a 3D Max Pooling layer. Then, the data is flattened, followed by Dropout and Dense layers to avoid overfitting the training data. The CNN-3D was trained using an Adam optimiser, with 1^{-6} learning rate and 1^{-6} decay, with a batch size of 64 and 200 epochs. Figure 5.2 depicts the CNN-3D architecture.

5.3.1 Results

This section presents the results for the two flybys selected for evaluating the CNN-3D: flyby 4 and 5. Figure 5.3 shows the waterfall image for each flyby.

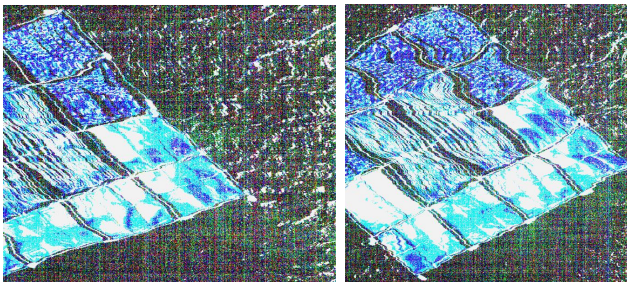


Figure 5.3: *Left image flyby 4 waterfall image, right image flyby 5 waterfall image.*

In Figure 5.4 5.5 is possible to observe the ground truth and the obtained CNN predictions for flybys 4 and 5.

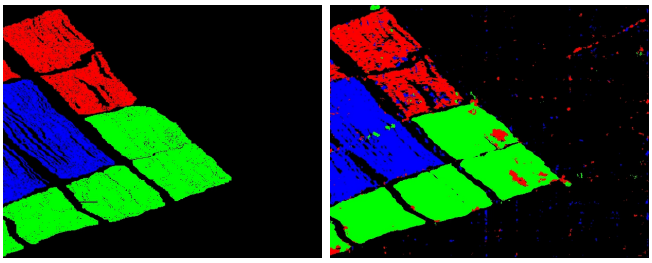


Figure 5.4: *Left image flyby 4 ground-truth image data, right image flyby 4 CNN-3D result.*

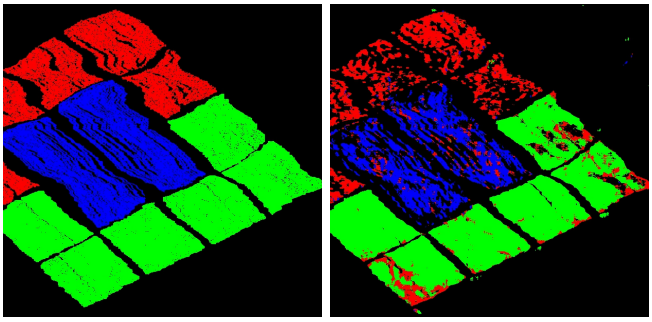


Figure 5.5: *Left image flyby 5 ground-truth image data, right image flyby 5 CNN-3D result.*

In Table 5.7 and Table 5.8 is possible to see the results metrics, including the obtained precision, recall, F1-score, and overall accuracy (OA) for each of the flybys.

5.3.2 Discussion

The overall accuracy varies between 91.67% for flyby 4 and 84.84% for flyby 5. In both flybys is possible to observe that class 1 (orange target) and class 3 (rope

Table 5.7: *Flyby 4 CNN-3D Overall Results*

	Precision	**Recall**	**F1-Score**	**Number of pixels**
Class 0	0.98	0.93	0.95	214882
Class 1	0.75	0.78	0.77	21852
Class 2	0.94	0.95	0.94	48505
Class 3	0.69	0.96	0.80	20421
OA	91.67%			

Table 5.8: *Flyby 5 CNN-3D Overall Results*

	Precision	**Recall**	**F1-Score**	**Number of pixels**
Class 0	0.83	0.97	0.90	212794
Class 1	0.62	0.50	0.55	37957
Class 2	0.98	0.84	0.91	85579
Class 3	0.88	0.54	0.67	41870
OA	84.84%			

target) are the ones that have lower accuracy results. The obtained results suffer from two deviations: First, to manually annotate the ground truth, see Figures 5.3, and make the process run faster. We resort to the K-Means algorithm to detect the majority of the points. After this first step, the selected pixels are verified. However, this process can induce some errors in the manual annotation of the ground-truth pixels, and the waterfall image also does not make it easy to distinguish the pixels from plastic and water classes, which can also induce errors. The second issue is the target water coverage. Taking into account the material's physical properties, there is the possibility that the materials are partially submerged, especially in the case of the orange target. The rope target also absorbs much water, making it difficult to distinguish between rope and water classes. This effect creates mixed pixels that could produce errors in some class pixels classifications. Spectral Unmixing techniques will need to be put in place for trying to achieve better classification accuracy. Overall and despite these effects, the CNN-3D showed the potential to detect and classify the different types of marine litter.

5.4 Summary

This chapter presented three supervised approaches for marine litter detection and classification using hyperspectral imaging. All these methods are state-of-the-art approaches used regularly with hyperspectral data.

The first two methods, Support Vector Machines and Random Forests, were applied to the F-BUMA aircraft data. This allows for establishing a comparison base with state-of-the-art methods and understanding the usability of these methods for

marine litter detection and classification. Both methods showed that they can detect and classify marine litter with at least 70% precision values and few false positives at 600 meters altitude. However, both methods are affected by the presence of some submerged target pixels, which decreases the method's precision.

To assess the applicability of deep learning methods for marine litter detection and classification, a CNN-3D was implemented, which uses spectral and spatial information combined as input. Due to the lack of spatial resolution in the F-BUMA dataset, this algorithm was implemented in the Grifo-X UAV dataset. The CNN-3D returned results with an overall accuracy of over 84% to distinguish the multiple types of marine litter.

6

Zero-Shot Learning Marine Litter Detection

6.1 Introduction

This chapter presents a ZSL approach for marine litter detection using hyperspectral imaging. By resorting to the spectral characteristics of the marine litter, this algorithm creates a model with only the seen classes. Still, during the classification stage, it can classify new class samples that were never seen during the training stage.

The ZSL algorithm is based on a generalised methodology, which means that it is trained with the known classes, and the classification is performed in all classes, known and unknown. The algorithm is adaptable, meaning it can classify the samples from unknown classes into different classes. It was trained and evaluated using a real hyperspectral imaging dataset, holding marine litter data acquired during field experiments in Porto Pim bay, Faial island, Azores.

6.2 Implementation

Several methods can be used for marine litter detection and classification, such as the ones presented in the previous chapter. However, these do not adapt to the continuous change in marine litter, whether due to sun and water erosion or the appearance of new types of marine litter. Another issue is the number of training samples required by these methods, which can be hard to fulfil. Given this, we implemented a novel ZSL method that can detect and classify samples from classes

not used during the training stage, decreasing the number of samples needed to train the model.

The data acquisition was completed in the radiance domain, from 400 to 2500 nm. The data pre-processing comprises two main steps: normalisation and dimensionality reduction. The first allowed us to normalise the data to have zero mean, and the second was needed due to the amount of data and because not all bands contribute equally to detecting the marine litter samples. In this case, the method used was Principal Component Analysis (PCA) [Rodarmel and Shan, 2002], reducing the number of utilised bands from 524 to 64 PCA bands.

GZSL solves the problem of classic ZSL approaches, where the prediction step uses unknown samples. In this case, our training dataset is composed of multiple spectra, each corresponding to one pixel of the pushbroom hyperspectral data cube, all belonging to known classes. Then, the prediction will also have the unknown classes of which the model has no information and has never seen. Let S_K denote the known classes, while S_U represents the unknown class labels, in which X and Y are the input and the output space, respectively. The input is composed of the S_K, while the output is composed of S_K and S_U. In the GZSL, $S = S_K \cup S_U$, with $S_k \cap S_U = \emptyset$. Given the training data, $(x_1, y_1, \ldots, (x_n, y_n))$ where x_n belongs to X and the label belongs to S_K, the goal is to extrapolate and recognise the spectra in Y. This algorithm can predict the label from sample $y \in Y$, where Y contains both known and unknown classes. The algorithm aims to extract semantic information about X, recognise K, and extrapolate to U, by mapping the training space X into a semantic space Z. Then, this semantic encoding is mapped to a class label.

This approach is an updated and enhanced version of the work presented in [Dong et al., 2020], and it is composed of four different components:

- Feature Extractor (F) - Convolutional Neural Network (CNN), with channel attention modules after each convolutional block [Woo et al., 2018];
- Classifier (C) - Fully Connected Neural Network;
- Decoder (D) - Autoencoder;
- Discriminator (P).

The first three components, F, C, and D, are used to build a spectral semantic space. Figure 6.1 shows a scheme for these blocks, while the following sections describe each one in detail.

6.2.1 Feature Extractor

The input signal consists of one row with 64 PCA bands. The Feature Extractor (F) module comprises a CNN architecture with four convolutional layers and three fully connected layers. After each convolutional block, a channel attention module improves the CNN architecture. This channel attention module is one part of the well-known Convolutional Block Attention Module (CBAM) [Woo et al., 2018], given that, in this case, there is no improvement in using both components of CBAM. To

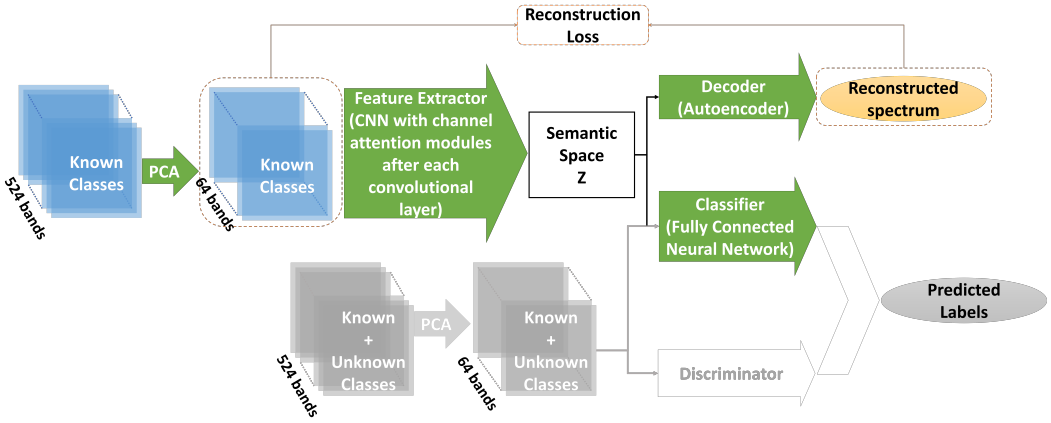


Figure 6.1: *Schematic of the algorithm used: first, the PCA method is used to reduce the data dimensionality for both the training and prediction phase. First, a feature extractor with channel attention modules after each convolutional layer is used to train the algorithm to create a semantic space Z. Then, this semantic space is used for the classifier, a fully connected neural network, and by the decoder to create a reconstructed spectrum. The classifier and the discriminator are used to perform the prediction, giving a predicted label to each pixel.*

minimise the intra-class variations in space Z and simultaneously maintain the inter-classes semantic features separated, this component resorts to centre loss [Wen et al., 2016]. Assuming that batch size is represented by B, and that $x_i \in X$, with label y_i, so $z_i = F(x_i) \in Z$, the centre loss is given by Equation (6.1), where sy_i corresponds to the semantic centre vector of the class y_i in Z, and s_{y_i} will be updated according to the class y_i semantic features changes.

$$L_{cl} = \frac{1}{2B} \sum_{i=1}^{B} ||F(x_i) - s_{y_i}||_2^2 \tag{6.1}$$

Given the limitations with the amount of hyperspectral data, the training dataset will be analysed in batches, which makes s_{y_i} updated according to $s_{y_i} \leftarrow s_{y_i} - \alpha \Delta_{s_{y_i}}$, where α denotes the learning rate and $\Delta_{s_{y_i}}$ can be obtained with Equation (6.2), where $\delta(\dot{})$ is equal to 1 if the condition is true and 0 otherwise.

$$\begin{cases} \Delta_{s_{y_i}} = 0, \ \ if \ \sum_{j=1}^{B} \delta(y_j = y_i) = 0, \\ \Delta_{s_{y_i}} = \frac{\sum_{j=1}^{B} \delta(y_j = y_i)(s_{y_i} - F(x_i))}{\sum_{j=1}^{B} \delta(y_j = y_i)} \ , \ otherwise \end{cases} \tag{6.2}$$

6.2.2 Classifier

The classifier, C, discriminates between the sample labels based on semantic features and comprises several fully connected layers. In this case, we used the cross-entropy loss defined in Equation (6.3), where $C(F(x_i))$ is the prediction for the sample x_i.

$$L_{cel} = -\frac{1}{B}\sum_{i=1}^{B} y_i \; log(C(F(x_i))) \tag{6.3}$$

6.2.3 Decoder

The decoder is composed of an autoencoder, using deconvolutions, unpolling, and fully connected layers. This component allows the reconstruction of the signal from the latent space Z, using the reconstruction loss to analyse the difference between the input and reconstructed signals. The loss is obtained with Equation (6.4), where $D(F(x_i))$ corresponds to the input signal x_i reconstructed.

$$L_{rl} = \frac{1}{2B}\sum_{i=1}^{B} ||D(F(x_i)) - x_i||_2^2 \tag{6.4}$$

The total loss function combining all the losses can be expressed as Equation (6.5).

$$L_t = L_{cel} + \lambda_{cl} L_{cl} + \lambda_{rl} L_{rl} \tag{6.5}$$

6.2.4 Discriminator

Given the latent space Z, the discriminator can distinguish between known and unknown classes by using Equation (6.6) with the feature extractor F for each known class k, obtaining the semantic centre vector for each known class, where m represents the number of training samples.

$$H_k = \frac{\sum_{j=1}^{m} \delta(y_j = k) F(x_j)}{\sum_{j=1}^{m} \delta(y_j = k)} \tag{6.6}$$

when a test spectrum Q is considered and $F(Q)$ is obtained, the difference between $F(Q)$ and H_k is measured for each known class k. This distance is obtained by using the Mahalanobis distance with Equation (6.7), where A_k corresponds to the covariance matrix Σ of semantic features of signals of class k, $d(\dot{;})$ transformation matrix associated to class k, and A_k^{-1} is the inverse matrix of A_k.

$$d(F(Q), H_k) = \sqrt{(F(Q) - H_k)^T \; A_k^{-1} \; (F(Q) - H_k)} \tag{6.7}$$

Given this distance, $d(F(Q), H_k)$, it is possible to distinguish between known and unknown classes. This is performed by comparing the threshold θ_1 with the minimal distance d_1 given by Equation (6.8), where H corresponds to the set of known semantic centre vectors.

$$d_1 = \min_{H_k \in S} \; d(F(Q), H_k) \tag{6.8}$$

Considering that y_Q represents the prediction of Q, if $d_1 < \theta_1,\ y_Q \in K$, otherwise $y_Q \in U$, where θ_1 is given by Equation (6.9), where λ_1 is the discrimination coefficient.

$$\theta_1 = \lambda_1 \times 3\sqrt{t} \tag{6.9}$$

However, if Q is classified as unknown in the previous step, it is necessary to understand if it will be part of a new class or must be combined with another class. If R refers to the recorded unknown classes, and S_R refers to the set of semantic centre vectors of R, if $R \subseteq \emptyset$, a new class label R_1 is added to R and $F(Q)$ is set to be the semantic centre vector H_{R_1}. The unknown spectrum Q of an unknown class is saved in set G_{R_1} and $y_Q = R_1$. If $R \nsubseteq \emptyset$, the threshold θ_2 defined in Equation (6.11), which is based on the distance between $F(x)$ and H_k, is compared to the minimal distance d_2 defined in Equation (6.10).

$$d_2 = \min_{H_{R_u} \in H_R} d(F(Q), R_u) \tag{6.10}$$

$$\theta_2 = \frac{d_1 + \lambda_2 \times d_{med}}{1 + \lambda_2} \tag{6.11}$$

Considering that n_R corresponds to the number of record spectrum labels registered in R, and if $d_2 > \theta_2$, a new spectrum label $R_{n_{R+1}}$ is added to R and set $y_Q = n_R + 1$. Given that n_R has no restriction, the model never knows the number of unknown classes left to be discriminated. If $d_2 \leqslant \theta_2$, then $y_Q = \underset{R_u}{\operatorname{argmin}}\, d(F(Q), H_{R_u})$ and the spectrum Q is saved in G_{y_Q}. The H_{y_Q} is updated using Equation (6.12), where $\#(G_{y_Q})$ corresponds to the number of spectra in the set G_{y_Q}.

$$H_{y_I} = \frac{\sum_{k \in G_{y_I}} F(k)}{\#(G_{y_I})} \tag{6.12}$$

Using this discriminator, the model can distinguish test samples into $n + n_R$ classes, where n_R is the number of unknown classes distinguished by the model.

6.3 Marine Litter Dataset Acquisition and Experimental Setup

The remote sensing of marine litter with hyperspectral systems from visible to the SWIR range is scarce, a series of issues related to the price of the systems and airborne campaign makes it difficult to find available open-source data, and most of the datasets available are from satellite data. Therefore, we resort to a hyperspectral imaging system setup composed of two pushbroom cameras: a Specim FX10e (1024 spatial pixels, spectral range from 400 to 1000 nm) and a HySpex Mjolnir S-620 (620 spatial pixels, spectral range from 970 to 2500 nm). For the system setup description in detail, see Chapter 4, Section 4.3.

The hyperspectral imaging system was then installed in a Cessna aircraft, F-BUMA, as shown in Figure 6.2, where it is possible to see the aeroplane, and the

payload is already mounted on the aeroplane landing gear. For the dataset acquisition, the flight altitude was around 600 m above sea level.



(a) (b)

Figure 6.2: *On the left-hand image, the F-BUMA aircraft, and on the right side of the image, the hyperspectral imaging payload installation on the aircraft landing gear.*

One of the goals of this campaign was also to assess Sentinel-2 (S2) satellite capability to detect marine litter and compare the results using satellite and airborne data. Due to the S2 spatial resolution, there was the need to have 10 $\times$ 10 metre artificial targets to be visible by the satellite. For this purpose, we used three different artificial targets composed of distinct plastic types: low-density polyethylene orange plastic (used oyster spat collectors), white plastic film, and ropes. Figure 6.3 shows the targets already in the water. In addition to the flybys over the targets, the dataset also holds information about other materials and objects present in the data: water, concrete pier, trees, and boats. These data will also be used in our algorithm to train the ZSL method.

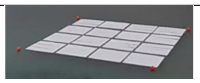
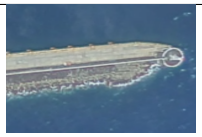
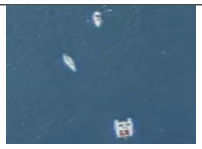


Figure 6.3: *Image acquired from the aircraft during a flight with the artificial targets placed in the water.*

Due to the hyperspectral image acquisition system (pushbroom) type, the dataset was collected during a single flight and contains information about other artefacts that are difficult to label manually, such as streets with moving cars and houses. Therefore, the solution used was to divide the dataset into flybys over the classes that can be manually labelled, such as water, orange target, white target, rope target, concrete pier, trees, and boats.

Table 6.1 shows the description of all the data present in the dataset and provides an example of each class.

Table 6.1: *Classes used for the ZSL algorithm, and main characteristics e.g., description, known and unknown information, image and number of pixels.*

ID	Description	Example	Number of Pixels
Class 0 (**Blue**) Known class	Water		1279132
Class 1 (**Red**) Unknown class	Orange target (low density polyethylene)		8118
Class 2 (**Green**) Known class	White target (White plastic film)		7913
Class 3 (**Magenta**) Unknown class	Rope target		8375
Class 4 (**White**) Known class	Concrete pier		16999
Class 5 (**Grey**) Known class	Trees		972160
Class 6 (**Yellow**) Known class	Boats		5023

Having defined the classes for the ZSL implementation, one needs to divide them into known and unknown classes. Since our objective was to detect marine litter (plastic), we chose two artificial plastic targets as unknown classes: class 1 - orange plastic target and class 3 - ropes target.

We then started the training procedure, where adjustments had to be performed to overcome the limitations of the training data, namely:

- Class imbalance: As shown in Table 6.1, the number of pixels for each class is different. To minimise the probability of overfitting during training, we use the lower number of pixels as a reference to randomly select samples from the other classes;

- Feature normalisation: The hyperspectral data acquisition is affected by the variability of atmospheric conditions, mainly direct sunlight. To minimise this effect, there is the need to perform data normalisation to normalise the feature data to unit variance.

In addition to the aforementioned pre-processing techniques, we also had to take into account some items concerning each of the known and unknown classes, such as:

- Due to the flight altitude, some classes will have small (resolution) in the datasets. In addition to this, there are classes with artefacts in the middle due to their physical characteristics and construction. This happens in class 1, orange target, and class 3, rope target. The first, given the physical design of the multiple components (used oyster spat collectors) and their orientation, can accumulate more water. In contrast, the second (rope target) has more gaps in the middle of the target. Additionally, this target also uses two different rope targets with variable flutuability.

- The water contains points with high exposure due to the wave effects and the camera adjustment.

- The concrete pier also contains rocks.

- It is necessary to consider that in class 6, boats have different hulls, and there are artefacts inside the boat that also have different materials. However, the flight altitude combined with the camera's spatial resolution does not allow a better distinction to facilitate the manual labelling process.

Another influential aspect is the camera exposure time. During the data collection, the exposure time was set to acquire the plastic targets in the best manner possible. Therefore, it can lead to overexposed zones that hold other classes and negatively influence the results.

6.4 Results

In this section, we will display the ZSL results for marine litter detection and classification. In this experiment, the algorithm was trained using seven classes in total: classes 0, 2, 4, 5, and 6 are the known classes, while the unknown classes (two of the plastic targets) were classes 1 and 3. Since one of the main objectives is to understand if the method can detect new plastic types that are not yet determined (unknown), we prefer to train with only one known plastic type. However, this will negatively influence the results due to the intrinsic physical characteristics of the plastic materials.

After performing the selection, we divided the dataset into training, validation, and test data. To promote the balance between the data, the number of points for each class used for training was 5023, and the number of pixels present in the class with fewer samples, class 6 - boats, was selected randomly. The data were then divided using 70% to train, 10% for validation, and all the remaining data to test, including data outside the first division.

Figures 6.4 show the reconstructed image of the dataset for both SPECIM FX10e (400 - 1000 nm) on the right side and the left side, the HySpex Mjolnir S-620 (970 - 2500 nm) systems, respectively, where it is possible to observe that there are multiple flybys over all the targets.

Figure 6.5 shows the manually annotated ground truth image on the left-hand side, while the right-hand side of the figure contains the ZSL results. Both images follow the same colour scheme to identify the different classes. There are some pixel outliers that represent four more classes, which are represented in orange, light pink, light grey and dark green. The colour scheme for each class is identified in the first column of Table 6.1. Figure 6.6 shows a zoom-enhanced version of one of the artificial target detections. It is possible to observe the two unknown classes (class 1 - orange target, red colour, and class 3 - rope target, pink colour) and some of the outliers around the edges.

Concerning the experimental setup, the tests were performed in a workstation with an Intel Core i7-8700k @ 3.70 GHz CPU, 32 GB RAM memory and one GeForce RTX 2080 Super 8 GB. The training stage took approximately 1052.71 s, while the prediction took approximately 243.44 s to finish.

To perform a comparison between the novel ZSL approach and already established state-of-the-art approaches, we compared the results (for the unknown samples) obtained by the ZSL with our previous work Random Forest (RF) and Support Vector Machines (SVM) implementation [Freitas et al., 2021b], and CNN-3D [Freitas et al., 2021a] for the same experimental data setup. We expect the ZSL performance metrics to be below that of the other methods, as in the ZSL RGB approaches, since the other methods are based on supervised approaches with all known samples. However, it is vital to understand the order of magnitude the ZSL can perform and if it is viable to use these semi-supervised or unsupervised approaches in remote airborne hyperspectral imaging systems for marine litter detection.

Table 6.2 shows the ZSL results, while Table 6.3 shows a comparison between these and the previous results, obtained and presented in [Freitas et al., 2021b,Freitas

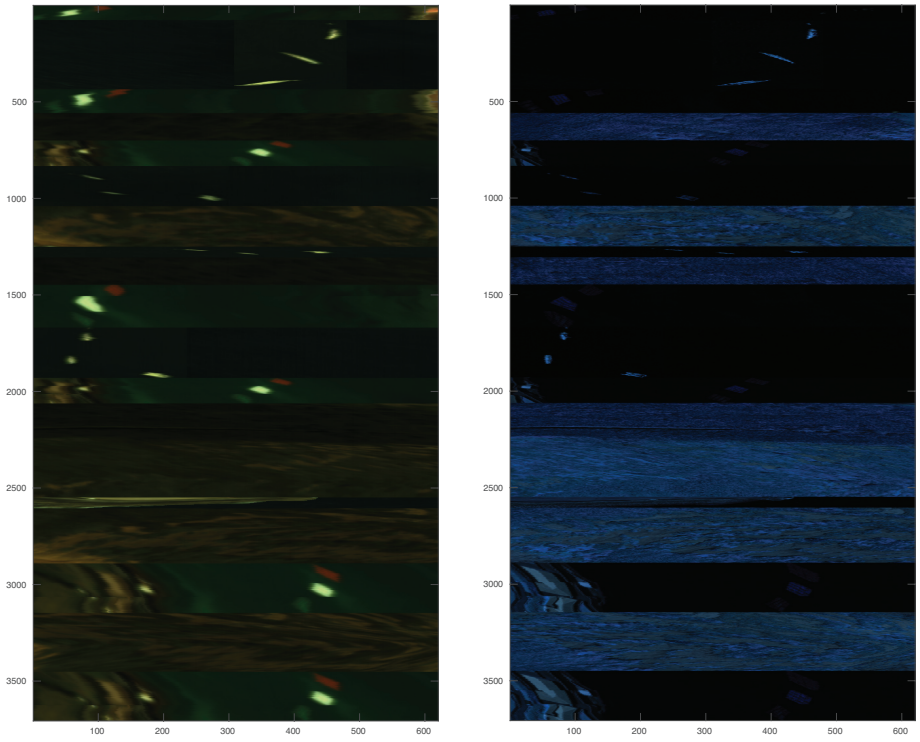


Figure 6.4: *Reconstructed image for each hyperspectral camera used. (Left image shows the Specim FX-10e data, while the right image shows the HySpex Mjolnir S-620 data.*

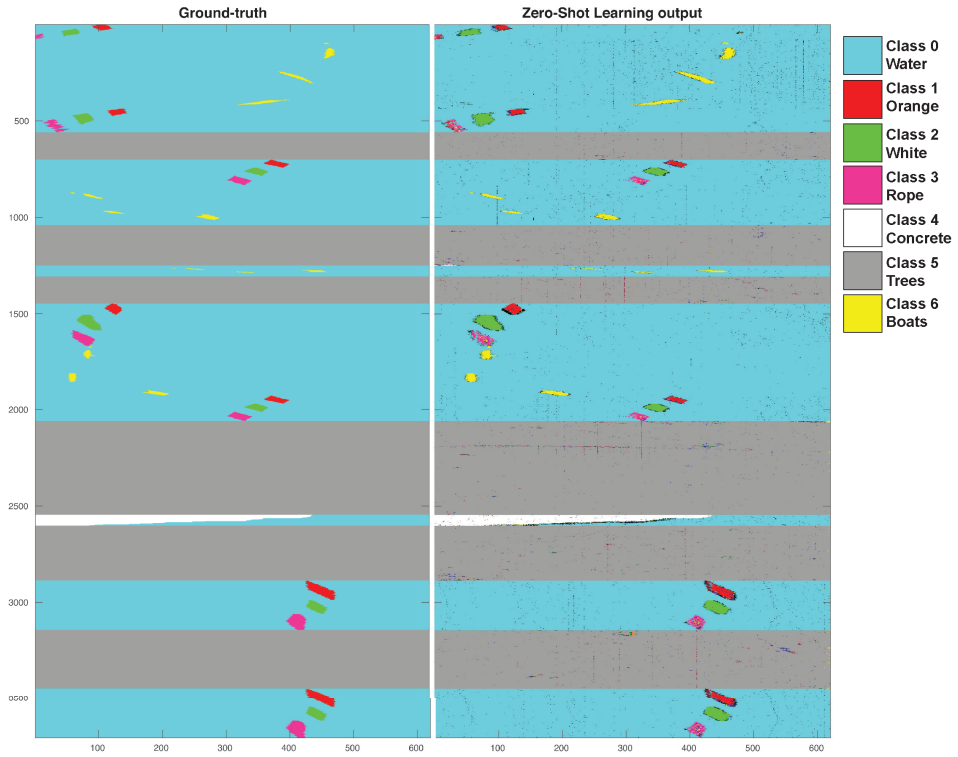


Figure 6.5: *Left image holds the ground truth while the right image shows the ZSL results.*

et al., 2021a].

Table 6.2: *ZSL Results.*

	Class 0 Water Known Class	**Class 1 Orange Target** Unknown class	**Class 2 White Target** Known Class	**Class 3 Rope Target** Unknown Class	**Class 4 Concrete Pier** Known class	**Class 5 Trees** Known Class	**Class 6 Boats** Known Class
Precision	0.9853	0.6659	0.7508	0.5632	0.9784	0.9927	0.5874
Recall	0.9872	0.6946	0.9960	0.5766	0.9885	0.9927	0.9962
F1-Score	0.9862	0.6800	0.8562	0.5698	0.9834	0.9927	0.7390
OA	98.71%						

6.5 Discussion

In this section, we will discuss the results obtained by the ZSL method, mainly focusing on the ability to detect unknown classes. We also compare and discuss the results obtained with our previous implementations (RF, SVM and CNN-3D). Table 6.2 depicts each class's precision, recall, f1-score, accuracy of each class, and overall accuracy (OA). It is possible to observe that unknown classes 1 and 3 both have

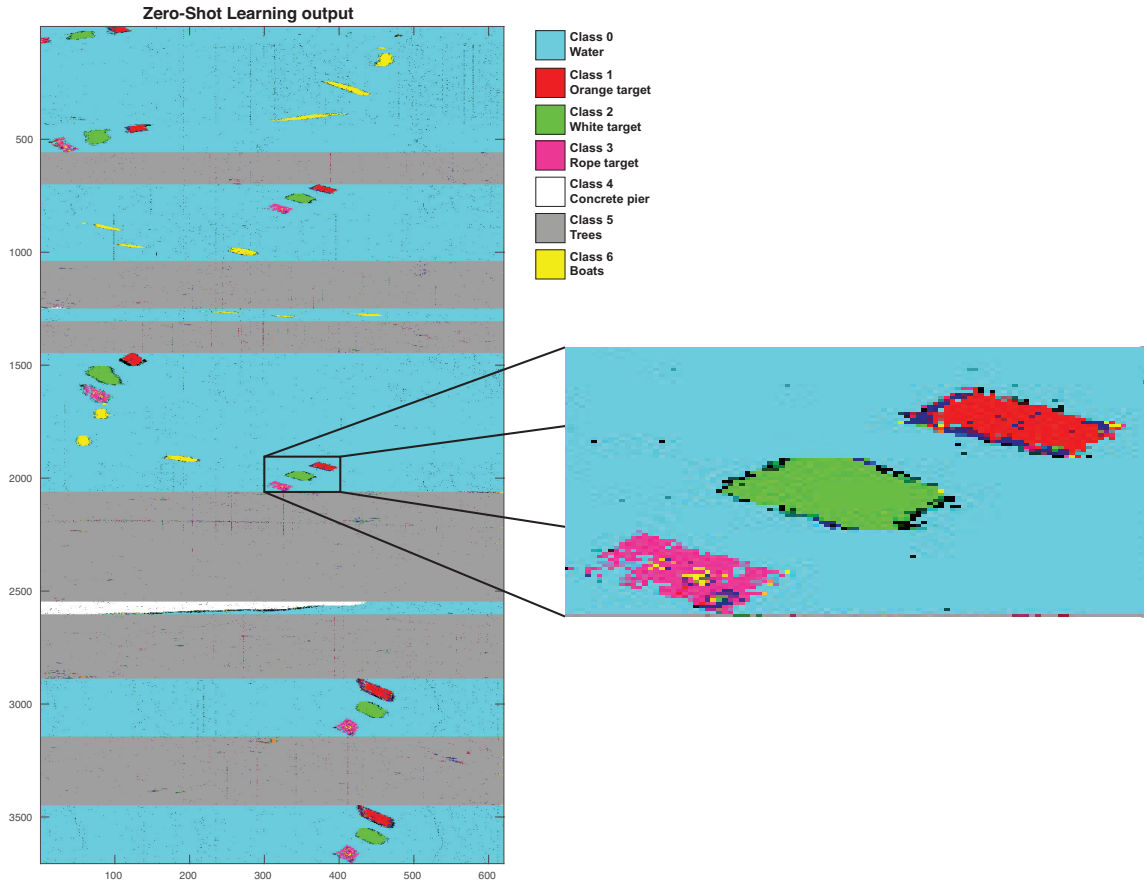


Figure 6.6: *ZSL results zooming on one of the artificial target's detections (including the two unknown classes (class 1, identified with the colour red, and class 3, identified with the colour pink).*

Table 6.3: *Results comparison using RF, SVM, CNN and ZSL methods.*

F-Buma flight - Flyby over the artificial targets								
Flyby 5								
	Random Forest (RF)				**Support Vector Machine (SVM)**			
Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**
0 Water and Land	0.99	1	0.99	98.09 %	0.99	0.99	0.99	97.06 %
1 Orange target	0.71	0.25	0.37		0.77	0.46	0.58	
2 White target	0.74	0.62	0.67		0.57	0.59	0.58	
3 Rope target	0.38	0.56	0.45		0.21	0.47	0.29	

F-BUMA Flight - ZSL								
Class	**0 Water Known class**	**1 Orange target Unknown class**	**2 White target Known class**	**3 Rope target Unknown class**	**4 Concrete pier Known class**	**5 Boats Known class**	**6 Boats Known class**	**Accuracy**
Precision	0.9853	0.6659	0.7508	0.5632	0.9784	0.9927	0.5874	98.71 %
Recall	0.9872	0.6946	0.9960	0.5766	0.9885	0.9927	0.9962	
F1-Score	0.9862	0.6800	0.8562	0.5698	0.9834	0.9927	0.7390	

Drone Flight - CNN3D				
Flyby 4				
Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**
0 Water	0.98	0.93	0.95	91.67 %
1 Orange target	0.75	0.78	0.77	
2 White target	0.94	0.95	0.94	
3 Rope target	0.69	0.96	0.80	
Flyby 5				
Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**
0 Water	0.83	0.97	0.90	84.84 %
1 Orange target	0.62	0.50	0.55	
2 White target	0.98	0.84	0.91	
3 Rope target	0.88	0.54	0.67	

recall values over 57%, which means that the true positive pixels were higher than 57% of the total samples in both cases. In terms of precision, class 1 has more than 66% while class 3 has more than 56%.

Given the average flight altitude of 600 m, it is difficult to manually annotate the pixels as being purely belonging to a single class. This makes it challenging to ensure that each pixel in the targets belonging to classes 1, 2, and 3 is not contaminated by water. In addition, the targets morphology and their constituents also make this task difficult. The targets had 10 $\times$ 10 m, composed of four squares, each one with 2.5 $\times$ 2.5 m, each target filled with the related material. However, for example, in the case of class 3 - ropes, the filling is not uniform due to the water effects, creating areas where the target is submerged, which is a phenomenon that also occurs with class 1 - orange target.

Regarding class 6 - boats, precision is over 58%, and the recall is over 99%, which indicates a high number of true positives (over 99% of the pixels are correctly identified as class 6). However, it also shows some false positives. In this case, the false positives are present in the water, which raises the issue of the ground truth annotation: due to the resolution, it is impossible to understand what is inside the boats. These were not placed in the water, especially for dataset purposes, so we do not have any control over what is inside them.

Figure 6.6 shows a zoom for one of the artificial targets ZSL results. As observed, the class 3 - rope target, represented by the pink points, exhibits some points marked as class 6 - boats (yellow). This shows that some of the boats can probably have ropes inside them. However, the spatial resolution makes it impossible to understand what is inside the boat. The pixel may not be pure (representing only one type of material) and instead be a combination of different materials inside the boat. It is also possible to observe some points marked as class 6 - boats (yellow) in the limits of the artificial targets, probably due to the water bottles used to help them float. Another issue is the camera configuration, and the camera gain set configured to acquire the artificial target data. Effects such as sun reflections can result in over-exposure pixels, mainly on pixels that do not represent the classes that the cameras were configured to acquire. This is especially noticeable in the waves of the water, which have several over-exposure pixels.

Analysing the left image in Figure 6.5, it is possible to observe the presence of different classes that do not belong to the ones identified by the ground truth near the concrete pier (class 5, colour grey). These are located in the rocks of the pier, most likely due to the foam created with the ocean motion due to oil residues, among others. However, it is impossible to identify this on the ground truth due to the camera's spatial resolution. This is also visible in Figure 6.6, in the contour of the orange and white artificial targets, and it can also be due to the foam that forms next to these targets due to the water ripple. This is not so visible in the rope target due to its characteristics: the ropes absorb much more water and offer less resistance to the water passage due to their shape.

Despite being a challenging dataset, due to the impossibility of controlling what is in the data that do not belong to the artificial targets, the ZSL algorithm acquired knowledge about the known classes and was able to detect and classify the unknown

classes to a reasonable extent.

Although this approach presents interesting results, using other algorithms to extract more meaningful information from the spectra can be an interesting solution. One approach that could extract more information about the characteristics of the plastics is the band selection algorithm, which will not project the data to another space, such as the method used for dimensionality reduction in this work (PCA algorithm).

Using spatial information combined with spectral one can be an interesting solution. However, this dataset was built from multiple flybys over points of interest (over the different classes: artificial targets, water, trees and boats) since the full dataset contained several areas whose ground truth was impossible to perform (roads with cars and people moving, for example), so there was a need to discard this data. Therefore, the spatial information was compromised by this dataset construction and could negatively influence the algorithm to learn some spatial representation that is not true. In addition to this, there is the fact that the artificial targets used for proof-of-concept are a perfectly square shape, which will not happen in a real scenario. The deformations suffered by the marine litter during their life are mainly due to sun exposure, water erosion and hitting other artefacts such as more marine litter and rocks, among others. These deformations will change the marine litter shape without affecting the spectral information.

In Table 6.3, it is possible to compare our state-of-the-art results and the ZSL approach. Some remarks should be addressed before analysing this table:

- Random Forest (RF), Support Vector Machine (SVM) and the Convolutional Neural Network 3D (CNN3D) were all implemented using only four classes:
  - Class 0: For the RF and SVM algorithm, it represents the water and land (houses, trees, streets, cars and other materials) to train the algorithms to learn the characteristics of the non-marine litter pixels. In the case of the CNN-3D algorithm, it represents the water;
  - Class 1: orange target;
  - Class 2: white target;
  - Class 3: rope target.
- RF, SVM, and ZSL approaches use the F-BUMA data, while the CNN-3D resorts to the drone data, using the same artificial targets and in the same test site. Due to the lack of resolution of the F-BUMA data, the results of the CNN-3D were unsatisfactory due to the lack of distinct spatial features. Although the drone data have much more spatial information, as can be observed in [Freitas et al., 2021a], it does not contain data from any other class besides the artificial targets and the water, which was not enough for the ZSL algorithm. Using RF and SVM with drone data would be possible. However, the algorithms will not benefit from the spatial information, and therefore, no novelty will be added to the results obtained with the F-BUMA data;
- The datasets used for RF, SVM and CNN-3D algorithms were organised by flyby over the artificial targets, giving each flyby its results. In the case of

the SVM and RF, Table 6.3 only presents the results for a single flyby (there were six flybys in total), while in the case of the CNN-3D, all the results are presented.

Due to the difference in the number of classes, it is different to compare the accuracy directly. However, it is possible to analyse the precision, recall and F1-score for the artificial targets classes (classes 1, 2 and 3). However, the analysis will focus on the precision and recall results, given that the F1-score combines the precision and recall values. Observing the results for class 1, orange target, in terms of precision, the values are similar. In the case of the ZSL approach, this algorithm was not trained with data from this class and can still obtain comparable results in terms of precision compared to the other three algorithms. When comparing the recall values, it is possible to observe that both the RF and SVM results are lower than the ZSL approach, and when comparing ZSL and CNN-3D, the results are similar. Class 2, white target, a known class for the ZSL algorithm, i.e., the model was trained with data from this class, also presents similar results for all the approaches. SVM shows lower performance, while CNN-3D display better results. However, ZSL performs better than the RF algorithm in terms of precision and recall. The last comparable class is the number 3, rope target, which is the most challenging for all the methods due to their characteristics. This is evident in the results presented for all the algorithms, with CNN-3D performing slightly better. ZSL overcomes both RF and SVM approaches, even considering that it was an unknown class whose data was hidden during the training phase of the ZSL model.

Overall, the ZSL algorithm shows potential to detect unknown marine litter, i.e., plastic samples using a hyperspectral imaging system installed in an aircraft at approximately 600 m altitude, with precision values over 56% for all the classes, including the two unknown classes, which are not used to train the model. This is in line with the results obtained by ZSL methods using only visible images, such as [Li et al., 2022a], which shows that ZSL methods can also have potential applications in hyperspectral imaging data.

6.6 Summary

The ZSL approach for marine litter detection using hyperspectral imaging data described in this chapter has been proven to classify unknown marine litter. These plastic samples were not available to the model during the training stage. To the best of our knowledge, this is the first implementation of a ZSL model to detect marine litter using hyperspectral imaging. The results show that the model can achieve precision values over 56% for all classes, including the unknown classes. It can be improved by collecting new data with more resolution, which will also allow to combine spatial information. This algorithm reduces the data needed to train the model, mainly because some classes will not need to provide information during the training stage. Another advantage is the ability to detect and classify new samples that belong to different classes than the ones used to train the model.

7

Conclusions and Future Work

In this chapter, we will enumerate the thesis advances and contributions, draw some conclusions about the obtained results during the duration of the thesis, and also point out the main future research paths, some already ongoing and others that will be the subject of research soon.

7.1 Conclusions

The thesis focuses on developing a hyperspectral imaging system for Remote Marine litter detection and classification using learning-based approaches. To accomplish this work, we performed the following contributions and developments:

- **Dataset and Acquisition Setup**
  The dataset campaign aimed to collect information about marine litter, mainly plastic types. Given the studies performed and presented in state of the art, it was possible to conclude that to distinguish between plastic types, there is the need to have a system capable of collecting information in the 400 to 2500 nm spectral range due to the specific characteristics of the plastic. So, one of the first steps developed in this thesis was to mount and develop a hyperspectral imaging system capable of collecting information from 400 to 2500 nm. This was accomplished using two hyperspectral cameras: one for the visible range (Specim FX10e) and another for the SWIR spectral range (HySpex Mjolnir S-620). However, using two independent systems brings issues during the acquisition, such as the trigger for both cameras simultaneously. Considering that the HySpex Mjolnir is a complete hyperspectral imaging system composed of the hyperspectral camera, GPS/INS and CPU, it is already pre-

pared to work as a standalone system, and the trigger is already synchronised with GPS, it is possible to use this trigger to synchronise both cameras. To first test the system and understand the plastic spectral response, the hyperspectral imaging system was assembled in a laboratory setup and used to analyse the marine litter samples collected at the test site. This study was performed with the samples placed in a container with seawater, with direct sunlight, at different times of the day. After these preliminary tests, it was possible to perform the dataset campaign in the marine litter hotspot using the hyperspectral imaging system installed in two aircraft: manned aircraft and unmanned aerial vehicles. The datasets took place in Porto Pim bay, Faial island, Azores, for two weeks in September 2020. Given the possibility of comparing the data collected with both aircraft with the data from Sentinel-2 satellites, we placed at the test site three artificial plastic targets consisting of different types of plastic. These targets had 10×10 m dimensions to meet the spatial resolution of the Sentinel-2 satellites multispectral sensors. During the first week, four flights were carried out with the F-BUMA aeroplane due to the limitations imposed by the atmospheric conditions. The second week was dedicated to performing the UAV flights. Although the airspace and atmospheric restrictions, the UAV did five flights over the artificial targets. However, it was not possible to process the Sentinel-2 satellite due to adverse atmospheric conditions.

- **Spectral/Spatial Classification (Random Forest, Support Vector Machine)**
  Two state-of-the-art algorithms were implemented to establish a baseline for comparison purposes: Random Forest and Support Vector Machine. These algorithms were implemented as a first approach to evaluate the potential of machine learning techniques to distinguish between different types of plastics. The results show that SVM has slightly more accurate results. Overall, both methods show potential to distinguish between the multiple types of marine litter, i.e. plastic samples using hyperspectral imaging data, with 0.70 - 0.80 precision values and few false positives. Both methods were implemented with the F-BUMA data, and the models were trained using data from the previous flight days and allowed to classify most targets correctly. Still, the presence of submerged target pixels decreases the accuracy of the methods.

- **CNN-3D**
  The previous approaches performed classification with only spectral information. However, the spatial information can also be combined with the spectral data to extract more meaningful features from the dataset. Given the accurate results obtained with the application of CNN in hyperspectral data, it is interesting to expand the application of these algorithms to combine both spectral and spatial information, resulting in a CNN-3D. This algorithm benefits from more spatial resolution, which can be obtained if the dataset is performed at a lower altitude, such as the dataset performed with the UAV at 25 meters altitude. Given the finest spatial resolution of these datasets, this

was the data selected to implement the CNN-3D algorithm, which shows good performance at classifying the different types of plastic with overall accuracies between 91.67% and 84.84%. However, some of the plastic types presented lower accuracy results, mainly due to errors in the manual annotation of the ground truth and also due to the target water coverage, given the possibility of some of the materials being partially submerged.

- **Zero-Shot Learning for Hyperspectral Imaging**
  Hyperspectral imaging data is scarce and difficult to obtain. We implemented a Zero-Shot Learning approach for hyperspectral imaging data to overcome this issue and decrease the data needed to train a hyperspectral classification model. This algorithm was used to detect and classify marine litter samples and can distinguish between different plastic types. By resorting to the features collected from the seen classes during the training stage, the algorithm can create an intermediate layer to transfer this knowledge to the unseen classes during the classification phase and classify these samples into new classes that the model never saw during the training stage.

  To the best of our knowledge, this is the first implementation of a Zero-Shot Learning approach for hyperspectral imaging and the first to be applied to marine litter classification. Despite the challenging dataset, due to the low resolution and the water coverage of some targets, the ZSL algorithm shows promising results. The results show that the two unseen classes selected have both recall values over 57%, which means that the true positive pixels were higher than 57% of the total samples in both cases. In terms of precision, both classes have a result higher than 56%, in line with the existing ZSL approaches for RGB images.

7.2 Future Work

Despite the relevance of the developed work and the obtained results during the scope of this thesis, there are secondary aspects that can produce relevant scientific work that forms the core objectives to pursue in our ongoing and future work. The relevant ones are enumerated as follows:

- **New Dataset**

  Due to some equipment restrictions, there were some issues related to the geo-referencing of the data. The HySpex system has demanding requirements to collect accurate information from the GPS/INS system, which sometimes could not be met due to flight requirements. In addition, it would also be interesting to change the disposition of the cameras. Due to platform requirements, there is no alternative to using another mounting scheme. The lack of satellite data due to adverse atmospheric conditions when the satellite flew over the test site is another point that must be improved in a new dataset. Although it is not something we can control, we can try to make the new dataset at a time with

better atmospheric conditions and in an area with a lower probability of this issue happening.

- **Zero-Shot Learning algorithm improvements**

  In future work, we plan to introduce also spatial information in a combination of the spectral data to complement the information. This will allow the creation of new semantic representations with more meaningful information to help the algorithm achieve better results. Using band selection techniques to extract more meaningful information from a hyperspectral imaging dataset could also help improve the results from the ZSL algorithm.

- **Edge Computing - Real-Time performance**

  Real-time solutions could be implemented onboard the aircraft using data in the radiance domain. One of these solutions would be to implement the ZSL algorithm in a Graphics Processing Unit (GPU), such as the NVIDIA Jetson AGX Xavier, due to its computational capacity and compact size, allows it to be integrated onboard.

- **Spectral Unmixing**

  In future work, we plan to introduce spectral unmixing techniques to preprocess the data before the classification stage. Due to the resolution of the cameras and the altitude of the flights, sometimes the pixels are not pure but instead a mixture of multiple materials spectra. This can induce errors in the classification stage due to the high spectra variability inside the sample class. A spectral unmixing technique would allow the extraction of multiple spectra separated, classifying a single pixel as multiple classes.

A

Hyperspectral Imaging Datasets

Unfortunately, publicly available hyperspectral data for the scientific community is scarce. The most used and known hyperspectral remote sensing scenes are available on the *Grupo de Inteligencia Computacional (GIS)* website[1]. Table A.1 briefly describes some of the most used hyperspectral and multispectral imaging datasets. Figure A.1 shows the image obtained using only one wavelength and the ground truth of the following datasets: Indian Pines, Pavia, Pavia_U and Salinas.

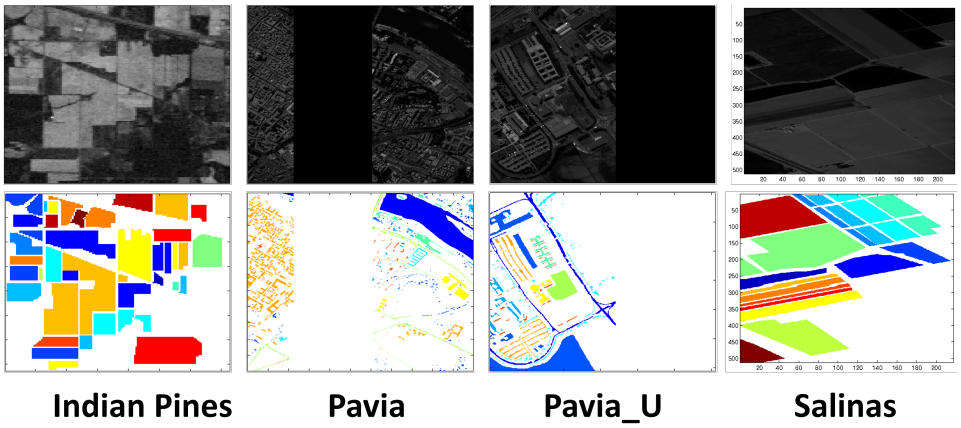


Figure A.1: *Hyperspectral imaging datasets: the images on the first line are obtained using only one wavelength, while the images on the second line represent the ground truth of each dataset.*

[1] `http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes`, accessed on 7 February 2022

Table A.1: *Brief description of the most used hyperspectral and multispectral imaging datasets, adapted from [Shafaey et al., 2021].*

Dataset	**Sensor/ Platform**	**Test Site**	**Scene Dimensions**	**# Spectral Bands**	**Spectral Range (nm)**
Indian Pines	AVIRIS	North--western India	145 x 145	200	400 - 2500
Salinas	AVIRIS	Salinas Valley, California	512 x 217	224	400 - 2500
Pavia_C	ROSIS	Pavia Center, North Italy	1096 x 1096	102	430 - 860
Pavia_U	ROSIS	Pavia University	610 x 610	103	430 - 860
Kennedy Space Center	AVIRIS	Kennedy Space Center, Florida	614 x 512	176	400 - 2500
Washington DC Mall	Airborne hyperspectral data flight	Washington DC Mall	1208 x 307	191	400 - 2400
Botswana	EO-1	Okavango Delta, Botswana	1496 x 256	145	400 - 2500
Cooke City	HyMap Airborne Hyperspectral Imaging Sensor	Cooke City, USA	280 x 280	126	450 - 2480
Urban	-	-	307 x 307	162	400 - 2500
Samson	-	-	952 x 952	156	401 - 889
Jasper Ridge	-	-	512 x 614	198	380 - 2500
Cuprite	AVIRIS	Las Vegas, USA	250 x 190	188	400 - 2500
MARIDA	Sentinel 2	Coastal areas, all over the world	1381 patches, consisting of 837357 pixels	12	442 - 2202

B

Dataset Campaign Logbook

The dataset campaign took place for two weeks on a marine litter hotspot - Porto Pim bay, Faial island, Azores. The first week was used to perform flights with the F-BUMA aircraft, while the second was dedicated to Grifo-X flights. These flights were always limited due to flight authorization restrictions and atmospheric conditions. Figure B.1 shows pictures of the Grifo-X and F-BUMA platforms during the dataset campaign and the targets being prepared and placed in the marine litter hotspot for the dataset campaign. Tables B.1 and B.2 present the details of the F-BUMA and Grifo-X flights performed.



Figure B.1: *The first line of images shows the Grifo-X and F-BUMA platforms during the dataset campaign. The second line shows the targets being prepared and placed in the marine litter hotspot for the dataset campaign.*

Table B.1: *F-BUMA aeroplane dataset campaign details*

Flight	**1**	**2**	**3**	**4**
Date	17/9/2020	18/9/2020	18/9/2020	20/9/2020
Start time	15:15h	11:10h	12:53h	13:04h
Stop time	15:48h	11:38h	13:27h	13:39h
Flight duration	Aprox. 33 min	Aprox. 28 min	Aprox. 34 min	Aprox. 35 min
Altitude	600 meters	600 meters	500 meters in the beginning and 500 meters in the last flyby	750 meters in the first flyby and 650 meters in the last flyby
Speed	30 m/s	25 m/s	25 m/s	20 m/s
Targets	Only the orange plastic target (10x10 meters) on the water, moored to a boat, without GPS position	Two targets on the water (one with ropes and other with the white plastic), moored to a boat, with GPS. The two targets were connected by ropes with 10 meters.	Three targets on the water (white plastic, ropes and orange plastic). The first two had 10x10 meters, and the last one only had 5x7.5 meters due to a problem with the structure. The targets were moored to a boat, and connected by ropes with 10 meters.	Three 10x10 meters targets on the water (white plastic, ropes and orange plastic). The targets were adrift, anchored to the beach (parallel to the beach).
Observations	Cloudy	Cloud	Cloudy	Cloudy

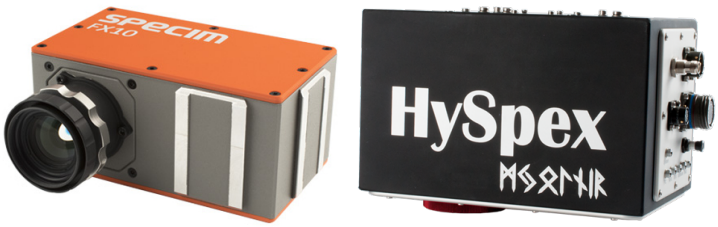
Flight	**1**	**2**	**3**	**4**	**5**
Date	22/9/2022	24/9/2022	24/9/2022	25/9/2022	25/9/2022
Start time	14:25h	12:39h	15:48h	11:00h	12:35h
Stop time	14:33h	12:51h	15:55h	11:11h	12:42h
Flight duration	Aprox. 8 min	Aprox. 12 min	Aprox. 7 min	Aprox. 11 min	Aprox. 7 min
Altitude	25 meters	25 meters	In the first flyby, 25 meters, and in the last flybys 35 meters	In the first flyby, 25 meters, and in the last flybys 35 meters	In the first flyby, 25 meters, and in the last flybys 35 meters
Speed	1 m/s	1 m/s	1 m/s	1 m/s	1 m/s
Targets	10 x 10 meters rope target adrift, anchored to the beach	10 x 10 meters mixed plastic target (one single target with the three plastic types)	10 x 10 meters mixed plastic target (one single target with the three plastic types)	Three 10 x 10 meters targets on the water (white plastic, ropes and orange plastic), moored to a boat, with GPS. The targets were connected by ropes with 10 meters	Three 10 x 10 meters targets on the water (white plastic, ropes and orange plastic), moored to a boat, with GPS. The targets were connected by ropes with 10 meters
Observations	Cloudy	Cloudy	Partially sunny	Cloudy	Cloudy

Table B.2: *Grifo-X aircraft dataset campaign details*

C

Hyperspectral Cameras Characteristics

The hyperspectral imaging systems setup was composed of two pushbroom hyperspectral cameras: one in the visible range, from 400 to 1000 nm - Specim FX10e and another in the SWIR range, from 970 to 2500 - HySpex Mjolnir S-620. The first one, Specim FX10E, is a GigE camera with 1024 pixels of spatial resolution and needs an external data acquisition system to collect the data. In the case of the HySpex Mjolnir S-620, this is a hyperspectral imaging system composed of the hyperspectral camera, a data acquisition unit, GPS/INS, RGB camera integrated into a single closed system. Table C.1 and Table C.2 show the main features of each Specim FX10e and HySpex Mjolnir S-620 hyperspectral camera. Figure C.1(a) shows the Specim FX10e while Figure C.1(b) illustrates the HySpex Mjolnir S-620 system.



(a) (b)

Figure C.1: *Specim FX10 (400-1000 nm) and HySpex Mjolnir S-620 (970-2500 nm), respectively (not to scale).*

Table C.1: *Specim FX10e Characteristics*

Specifications	**Specim FX10e**		
Spectral range	400 - 1000 nm		
Detector type	CMOS		
Slit width	Physical width 42 μm. Projection on sensor 32 μm (M$\sim$1/1.3)		
Pixel size	16 x 8 μm		
# Spatial pixels	1024		
Binning (spectral x spatial)	2 x 1		
Spectral binning options	2x	4x	8x
# Spectral bands covering the specified range	224	112	56
Spectral sampling/pixel	2.7 nm	5.4 nm	10.8 nm
Spectral resolution FWHM	5.5 nm (mean)		
SNR @ max. signal	420:1		
Frame rate (fps), full range (220 bands) max.	327 FPS		
Frame rate (fps), MROI examples	20 bands = 2800 FPS 5 bands = 6500 FPS		
Integration time	Adjustable, within frame time		
Sensor coding	Passive		
Shutter	Electromechanical shutter for dark background registration		
Order blocking filter	Yes		

Table C.2: *HySpex Mjolnir S-620 Specifications*

Specifications	**HySpex Mjolnir S-620**
Spectral range	970 - 2500 nm
Spatial pixels	620
Spectral channels and sampling	300 bands @ 5.1 nm
F-number	F1.9
FOV	20°
Pixel FOV accross/along	0.54/0.54 mrad
Bit resolution	16 bit
Noise floor	80 e^-
Dynamic range	10000
Peak SNR (at full resolution)	>900
Max speed (at full resolution)	100 fps
Power consumption	70 W
Dimensions (l-w-h)	374 - 202 - 178 mm
Weight	6 kg (inc. IMU/GPS and DAU) 6.5 kg (inc. also standard battery)

Bibliography

[Abdulridha et al., 2020] Abdulridha, J., Ampatzidis, Y., Roberts, P., and Kakarla, S. C. (2020). Detecting powdery mildew disease in squash at different stages using uav-based hyperspectral imaging and artificial intelligence. *Biosystems Engineering*, 197:135–148.

[Abdulridha et al., 2019] Abdulridha, J., Batuman, O., and Ampatzidis, Y. (2019). Uav-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning. *Remote Sensing*, 11(11).

[Acito et al., 2003] Acito, N., Corsini, G., and Diani, M. (2003). An unsupervised algorithm for hyperspectral image segmentation based on the gaussian mixture model. In *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*, volume 6, pages 3745–3747.

[Acquarelli et al., 2018] Acquarelli, J., Marchiori, E., Buydens, L. M., Tran, T., and Van Laarhoven, T. (2018). Spectral-spatial classification of hyperspectral images: Three tricks and a new learning setting. *Remote Sensing*, 10(7).

[Acuña-Ruz et al., 2018] Acuña-Ruz, T., Uribe, D., Taylor, R., Amézquita, L., Guzmán, M. C., Merrill, J., Martínez, P., Voisin, L., and Mattar B., C. (2018). Anthropogenic marine debris over beaches: Spectral characterization for remote sensing applications. *Remote Sensing of Environment*, 217(62):309–322.

[Ahn et al., 1999] Ahn, C.-W., Baumgardner, M. F., and Biehl, L. L. (1999). Delineation of soil variability using geostatistics and fuzzy clustering analyses of hyperspectral data. *Soil Science Society of America Journal*, 63(1):142–150.

[Akata et al., 2016] Akata, Z., Perronnin, F., Harchaoui, Z., and Schmid, C. (2016). Label-embedding for image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(7):1425–1438.

[Andriolo et al., 2021] Andriolo, U., Gonçalves, G., Rangel-Buitrago, N., Paterni, M., Bessa, F., Gonçalves, L., Sobral, P., Bini, M., Duarte, D., Fontán-Bouzas, Á., Gonçalves, D., Kataoka, T., Luppichini, M., Pinto, L., Topouzelis, K., Vélez-Mendoza, A., and Merlino, S. (2021). Drones for litter mapping: An inter-operator concordance test in marking beached items on aerial images. *Marine Pollution Bulletin*, 169.

[Arjovsky et al., 2017] Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. *Proceedings of the 34th International Conference on Machine Learning*, 70:214–223.

[Arroyo-Mora et al., 2019] Arroyo-Mora, J. P., Kalacska, M., Inamdar, D., Soffer, R., Lucanus, O., Gorman, J., Naprstek, T., Schaaf, E. S., Ifimov, G., Elmer, K., and Leblanc, G. (2019). Implementation of a uav–hyperspectral pushbroom imager for ecological monitoring. *Drones*, 3(1).

[Arroyo-Mora et al., 2021] Arroyo-Mora, J. P., Kalacska, M., Løke, T., Schläpfer, D., Coops, N. C., Lucanus, O., and Leblanc, G. (2021). Assessing the impact of illumination on uav pushbroom hyperspectral imagery collected under various cloud cover conditions. *Remote Sensing of Environment*, 258:112396.

[Asner and Heidebrecht, 2003] Asner, G. and Heidebrecht, K. (2003). Imaging spectroscopy for desertification studies: comparing aviris and eo-1 hyperion in argentina drylands. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6):1283–1296.

[Attila et al., 2013] Attila, J., Koponen, S., Kallio, K., Lindfors, A., Kaitala, S., and Ylöstalo, P. (2013). Meris case ii water processor comparison on coastal sites of the northern baltic sea. *Remote Sensing of Environment*, 128:138–149.

[Atzmon and Chechik, 2019] Atzmon, Y. and Chechik, G. (2019). Adaptive confidence smoothing for generalized zero-shot learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11663–11672.

[Atzmon et al., 2020] Atzmon, Y., Kreuk, F., Shalit, U., and Chechik, G. (2020). A causal view of compositional zero-shot recognition. *Advances in Neural Information Processing Systems*, 33:1462–1473.

[Ba et al., 2015] Ba, J. L., Swersky, K., Fidler, S., and Salakhutdinov, R. (2015). Predicting deep zero-shot convolutional neural networks using textual descriptions. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 4247–4255.

[Bäcklund et al., 2011] Bäcklund, H., Hedblom, A., and Neijman, N. (2011). A density-based spatial clustering of application with noise. *Data Mining TNM 033*, 33:11–30.

[Balsi et al., 2018] Balsi, M., Esposito, S., and Moroni, M. (2018). Hyperspectral characterization of marine plastic litters. In *2018 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea)*, pages 28–32.

[Balsi et al., 2021] Balsi, M., Moroni, M., Chiarabini, V., and Tanda, G. (2021). High-resolution aerial detection of marine plastic litter by hyperspectral sensing. *Remote Sensing*, 13(8).

[Banerjee et al., 2020] Banerjee, B. P., Raval, S., and Cullen, P. J. (2020). Uav-hyperspectral imaging of spectrally complex environments. *International Journal of Remote Sensing*, 41(11):4136–4159.

[Baraldi and Parmiggiani, 1995] Baraldi, A. and Parmiggiani, F. (1995). A neural network for unsupervised categorization of multivalued input patterns: an application to satellite imaee clustering. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2):305–316.

[Barnsley et al., 2004] Barnsley, M., Settle, J., Cutter, M., Lobb, D., and Teston, F. (2004). The proba/chris mission: a low-cost smallsat for hyperspectral multiangle observations of the earth surface and atmosphere. *IEEE Transactions on Geoscience and Remote Sensing*, 42(7):1512–1520.

[Basu et al., 2021] Basu, B., Sannigrahi, S., Sarkar Basu, A., and Pilla, F. (2021). Development of novel classification algorithms for detection of floating plastic debris in coastal waterbodies using multispectral sentinel-2 remote sensing imagery. *Remote Sensing*, 13(8).

[Baumgardner et al., 2015] Baumgardner, M. F., Biehl, L. L., and Landgrebe, D. A. (2015). 220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3. Purdue University Research Repository.

[Beck et al., 2016] Beck, R., Zhan, S., Liu, H., Tong, S., Yang, B., Xu, M., Ye, Z., Huang, Y., Shu, S., Wu, Q., Wang, S., Berling, K., Murray, A., Emery, E., Reif, M., Harwood, J., Young, J., Nietch, C., Macke, D., Martin, M., Stillings, G., Stump, R., and Su, H. (2016). Comparison of satellite reflectance algorithms for estimating chlorophyll-a in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations. *Remote Sensing of Environment*, 178:15–30.

[Ben-Arfa et al., 2015] Ben-Arfa, J., Bergès, J. C., Beltrando, G., Rim, K., and Zargouni, F. (2015). Mapping the land cover in coastal gabes oases using the eo-1 hyperion hyperspectral sensor. In *EGU General Assembly Conference Abstracts*, EGU General Assembly Conference Abstracts, page 4314.

[Benediktsson et al., 1990] Benediktsson, J., Swain, P., and Ersoy, O. (1990). Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):540–552.

[Biederman, 1987] Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2):115–147.

[Biermann et al., 2020] Biermann, L., Clewley, D., Martinez-Vicente, V., and Topouzelis, K. (2020). Finding plastic patches in coastal waters using optical satellite data. *Scientific Reports*, 10(5364).

[Bioucas-Dias and Figueiredo, 2009] Bioucas-Dias, J. and Figueiredo, M. (2009). Logistic regression via variable splitting and augmented lagrangian tools. *Instituto Superior Técnico, TULisbon, Portugal, Tech. Rep.*

[Biswas and Annadani, 2018] Biswas, S. and Annadani, Y. (2018). Preserving semantic relations for zero-shot learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7603–7612.

[Blitzer et al., 2009] Blitzer, J., Foster, D. P., and Kakade, S. M. (2009). Zero-shot domain adaptation: A multi-view approach. *Technical Report, TTI-TR-2009-1.*

[Borges et al., 2011] Borges, J. S., Bioucas-Dias, J. M., and Marcal, A. R. S. (2011). Bayesian hyperspectral image segmentation with discriminative class learning. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6):2151–2164.

[Boser et al., 1992] Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, page 144–152, New York, NY, USA. Association for Computing Machinery.

[Breiman, 1996] Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(1):123–140.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

[Brekke and Solberg, 2005] Brekke, C. and Solberg, A. H. (2005). Oil spill detection by satellite remote sensing. *Remote Sensing of Environment*, (95):1–13.

[Bresciani et al., 2022] Bresciani, M., Giardino, C., Fabbretto, A., Pellegrino, A., Mangano, S., Free, G., and Pinardi, M. (2022). Application of new hyperspectral sensors in the remote sensing of aquatic ecosystem health: Exploiting prisma and desis for four italian lakes. *Resources*, 11(2).

[Bézy et al., 2015] Bézy, J.-L., Huot, J.-P. M., Delwart, S. M., Bourg, L., Bessudo, R., and Delclaud, Y. (2015). *Medium Resolution Imaging Spectrometer for Ocean Colour onboard ENVISAT*, chapter 4, pages 91–120. John Wiley &Sons, Ltd.

[Campbell et al., 2011] Campbell, G., Phinn, S., Dekker, A. G., and Brando, V. E. (2011). Remote sensing of water quality in an australian tropical freshwater impoundment using matrix inversion and meris images. *Remote Sensing of Environment*, 115(9):2402–2414.

[Canters, 1997] Canters, F. (1997). Evaluating the uncertainty of area estimates derived from fuuy land-cover classification. *Photogrammetric Engineering &Remote Sensing*, 63(4):403–414.

[Carbary et al., 1994] Carbary, J. F., Darlington, E. H., Harris, T. J., McEvaddy, P. J., Mayr, M. J., Peacock, K., and Meng, C. I. (1994). Ultraviolet and visible imaging and spectrographic imaging instrument. *Applied Optics*, 33(19):4201–4213.

[Carpenter and Carpenter, 1983] Carpenter, D. and Carpenter, S. (1983). Modeling inland water quality using landsat data. *Remote Sensing of Environment*, 13(4):345–352.

[Carreira-Perpiñán and Zemel, 2004] Carreira-Perpiñán, M. A. and Zemel, R. S. (2004). Proximity graphs for clustering and manifold learning. In *Proceedings of the 17th International Conference on Neural Information Processing Systems*, NIPS'04, page 225–232, Cambridge, MA, USA. MIT Press.

[Carter et al., 2009] Carter, G. A., Lucas, K. L., Blossom, G. A., Lassitter, C. L., Holiday, D. M., Mooneyhan, D. S., Fastring, D. R., Holcombe, T. R., and Griffith, J. A. (2009). Remote sensing and mapping of tamarisk along the colorado river, usa: A comparative use of summer-acquired hyperion, thematic mapper and quickbird data. *Remote Sensing*, 1(3):318–329.

[Casal et al., 2011] Casal, G., Kutser, T., Domínguez-Gómez, J., Sánchez-Carnero, N., and Freire, J. (2011). Mapping benthic macroalgal communities in the coastal zone using chris-proba mode 2 images. *Estuarine, Coastal and Shelf Science*, 94(3):281–290.

[Castaldi et al., 2016] Castaldi, F., Palombo, A., Santini, F., Pascucci, S., Pignatti, S., and Casa, R. (2016). Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. *Remote Sensing of Environment*, 179:54–65.

[Celeux and Diebolt, 1985] Celeux, G. and Diebolt, J. (1985). The sem algorithm: A probabilistic teacher algorithm derived from the em algorithm for the mixture problem. *Computational Statistics Quarterly*, 2:73–82.

[Chan and Paelinckx, 2008] Chan, J. C.-W. and Paelinckx, D. (2008). Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112(6):2999–3011.

[Chandhok and Balasubramanian, 2021] Chandhok, S. and Balasubramanian, V. N. (2021). Two-level adversarial visual-semantic coupling for generalized zero-shot learning. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 3099–3107.

[Changpinyo et al., 2016] Changpinyo, S., Chao, W.-L., Gong, B., and Sha, F. (2016). Synthesized classifiers for zero-shot learning. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5327–5336.

[Chao et al., 2016] Chao, W.-L., Changpinyo, S., Gong, B., and Sha, F. (2016). An empirical study and analysis of generalized zero-shot learning for object recognition in the wild. In *Computer Vision – ECCV 2016*, pages 52–68, Cham. Springer International Publishing.

[Chen et al., 2020a] Chen, X., Lan, X., Sun, F., and Zheng, N. (2020a). A boundary based out-of-distribution classifier for generalized zero-shot learning. In *Computer Vision - ECCV 2020*, pages 572 – 588, Cham. Springer International Publishing.

[Chen et al., 2013] Chen, X., Shrivastava, A., and Gupta, A. (2013). Neil: Extracting visual knowledge from web data. In *2013 IEEE International Conference on Computer Vision*, pages 1409–1416.

[Chen et al., 2016] Chen, Y., Jiang, H., Li, C., Jia, X., and Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):6232–6251.

[Chen et al., 2017] Chen, Y., Ma, S., Chen, X., and Ghamisi, P. (2017). Hyperspectral data clustering based on density analysis ensemble. *Remote Sensing Letters*, 8(2):194–203.

[Chen et al., 2015] Chen, Y., Zhao, X., and Jia, X. (2015). Spectral–spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2381–2392.

[Chen et al., 2020b] Chen, Z., Li, J., Luo, Y., Huang, Z., and Yangyang, Y. (2020b). Canzsl: Cycle-consistent adversarial networks for zero-shot learning from natural language. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 863–872.

[Cheng et al., 2006] Cheng, Y.-B., Zarco-Tejada, P. J., Riaño, D., Rueda, C. A., and Ustin, S. L. (2006). Estimating vegetation water content with hyperspectral data for different canopy scenarios: Relationships between aviris and modis indexes. *Remote Sensing of Environment*, 105(4):354–366.

[Clark et al., 1989] Clark, R. B., Frid, C., and Attrill, M. (1989). Marine pollution. *Oxford: Clarendon Press*, 4.

[Cocking et al., 2022] Cocking, J., Narayanaswamy, B. E., Waluda, C. M., and Williamson, B. J. (2022). Aerial detection of beached marine plastic using a novel, hyperspectral short-wave infrared (swir) camera. *ICES Journal of Marine Science*, 79(3):648–660.

[Cocks et al., 1998] Cocks, T., Jenssen, R., Stewart, A., Wilson, I., and Shields, T. (1998). The hymap airborne hyperspectral sensor: The system, calibration and performance. *EARSEL Workshop on Imaging Spectroscopy.*

[Cudahy et al., 2001] Cudahy, T., Hewson, R., Huntington, J., Quigley, M., and Barry, P. (2001). The performance of the satellite-borne hyperion hyperspectral vnir-swir imaging system for mineral mapping at mount fitton, south australia. In *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, volume 1, pages 314–316 vol.1.

[Das et al., 2008] Das, S., Abraham, A., and Konar, A. (2008). Automatic clustering using an improved differential evolution algorithm. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 38(1):218–237.

[Datta et al., 2011] Datta, A., Feris, R., and Vaquero, D. (2011). Hierarchical ranking of facial attributes. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pages 36–42.

[de Vries et al., 2021] de Vries, R., Egger, M., Mani, T., and Lebreton, L. (2021). Quantifying floating plastic debris at sea using vessel-based optical data and artificial intelligence. *Remote Sensing*, 13(17).

[Dekker et al., 2001a] Dekker, A., Brando, V., Anstee, J., Pinnel, N., and Held, A. (2001a). Preliminary assessment of the performance of hyperion in coastal waters. cal/val activities in moreton bay, queensland, australia. In *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, volume 6, pages 2665–2667 vol.6.

[Dekker et al., 2001b] Dekker, A. G., Brando, V. E., Anstee, J. M., Pinnel, N., Kutser, T., Hoogenboom, E. J., Peters, S., Pasterkamp, R., Vos, R., Olbert, C., and Malthus, T. J. (2001b). *Imaging Spectrometry of Water*, pages 307–359. Springer Netherlands, Dordrecht.

[Del Frate et al., 2007] Del Frate, F., Pacifici, F., Schiavon, G., and Solimini, C. (2007). Use of neural networks for automatic classification from high-resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(4):800–809.

[Deng et al., 2020] Deng, X., Zhu, Z., Yang, J., Zheng, Z., Huang, Z., Yin, X., Wei, S., and Lan, Y. (2020). Detection of citrus huanglongbing based on multi-input neural network model of uav hyperspectral remote sensing. *Remote Sensing*, 12(17).

[Derraik, 2002] Derraik, J. G. (2002). The pollution of the marine environment by plastic debris: a review. *Marine Pollution Bulletin*, 44(9):842–852.

[Ding et al., 2021a] Ding, C., Li, Y., Wen, Y., Zheng, M., Zhang, L., Wei, W., and Zhang, Y. (2021a). Boosting few-shot hyperspectral image classification using pseudo-label learning. *Remote Sensing*, 13(17).

[Ding et al., 2021b] Ding, J., Hu, X., and Zhong, X. (2021b). A semantic encoding out-of-distribution classifier for generalized zero-shot learning. *IEEE Signal Processing Letters*, 28:1395–1399.

[Dong et al., 2020] Dong, Y., Jiang, X., Zhou, H., Lin, Y., and Shi, Q. (2020). Sr2cnn: Zero-shot learning for signal recognition. *CoRR*, abs/2004.04892.

[Donlon et al., 2012] Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.-H., Féménias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C.,

Nieke, J., Rebhan, H., Seitz, B., Stroede, J., and Sciarra, R. (2012). The global monitoring for environment and security (gmes) sentinel-3 mission. *Remote Sensing of Environment*, 120:37–57. The Sentinel Missions - New Opportunities for Science.

[Du et al., 2016] Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., and Li, X. (2016). Water bodies' mapping from sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the swir band. *Remote Sensing*, 8(4).

[Dungan, 2002] Dungan, J. L. (2002). *Toward a Comprehensive View of Uncertainty in Remote Sensing Analysis*, chapter 3, pages 25–35. John Wiley &Sons, Ltd.

[Elhamifar and Vidal, 2009] Elhamifar, E. and Vidal, R. (2009). Sparse subspace clustering. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2790–2797.

[Elhamifar and Vidal, 2013] Elhamifar, E. and Vidal, R. (2013). Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781.

[Elhoseiny and Elfeki, 2019] Elhoseiny, M. and Elfeki, M. (2019). Creativity inspired zero-shot learning. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5783–5792.

[Elhoseiny et al., 2013] Elhoseiny, M., Saleh, B., and Elgammal, A. (2013). Write a classifier: Zero-shot learning using purely textual descriptions. In *2013 IEEE International Conference on Computer Vision*, pages 2584–2591.

[European Space Agency and Plymouth Marine Laboratory (PML), 2020] European Space Agency and Plymouth Marine Laboratory (PML) (2020). Remote sensing of marine litter optimal-optical methods for marine litter detection final report (d7). (2).

[Falcone and Gomez, 2005] Falcone, J. A. and Gomez, R. (2005). Mapping impervious surface type and sub-pixel abundance using hyperion hyperspectral imagery. *Geocarto International*, 20(4):3–10.

[Farhadi et al., 2009] Farhadi, A., Endres, I., Hoiem, D., and Forsyth, D. (2009). Describing objects by their attributes. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1778–1785.

[Fei-Fei et al., 2006] Fei-Fei, L., Fergus, R., and Perona, P. (2006). One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):594–611.

[Felix et al., 2018] Felix, R., Kumar, B. G. V., Reid, I., and Carneiro, G. (2018). Multi-modal cycle-consistent generalized zero-shot learning. *European Conference on Computer Vision (ECCV)*.

[Felix et al., 2019] Felix, R., Sasdelli, M., Reid, I. D., and Carneiro, G. (2019). Multi-modal ensemble classification for generalized zero shot learning. *CoRR*, abs/1901.04623.

[Fergus et al., 2010] Fergus, R., Bernal, H., Weiss, Y., and Torralba, A. (2010). Semantic label sharing for learning with many categories. In *Computer Vision – ECCV 2010*, pages 762–775, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Fernández-Manso et al., 2016] Fernández-Manso, A., Fernández-Manso, O., and Quintano, C. (2016). Sentinel-2a red-edge spectral indices suitability for discriminating burn severity. *International Journal of Applied Earth Observation and Geoinformation*, 50:170–175.

[Feyisa et al., 2014] Feyisa, G. L., Meilby, H., Fensholt, R., and Proud, S. R. (2014). Automated water extraction index: A new technique for surface water mapping using landsat imagery. *Remote Sensing of Environment*, 140:23–35.

[Firat et al., 2022] Firat, H., Asker, M. E., İlyas Bayindir, M., and Hanbay, D. (2022). 3d residual spatial–spectral convolution network for hyperspectral remote sensing image classification. *Neural Computing and Applications*.

[Fouhey et al., 2016] Fouhey, D. F., Gupta, A. K., and Zisserman, A. (2016). Understanding higher-order shape via 3d shape attributes. *ArXiv*, abs/1612.06836.

[Freitas et al., 2021a] Freitas, S., Silva, H., Almeida, C., Viegas, D., Amaral, A., Santos, T., Dias, A., Jorge, P. A. S., Pham, C. K., Moutinho, J., and Silva, E. (2021a). Hyperspectral imaging system for marine litter detection. In *OCEANS 2021: San Diego – Porto*, pages 1–6.

[Freitas et al., 2018] Freitas, S., Silva, H., Almeida, J., and Silva, E. (2018). Hyperspectral imaging for real-time unmanned aerial vehicle maritime target detection. *Journal of Intelligent &Robotic Systems*.

[Freitas et al., 2019] Freitas, S., Silva, H., Almeida, J. M., and Silva, E. (2019). Convolutional neural network target detection in hyperspectral imaging for maritime surveillance. *International Journal of Advanced Robotic Systems*, 16(3):1–13.

[Freitas et al., 2021b] Freitas, S., Silva, H., and Silva, E. (2021b). Remote hyperspectral imaging acquisition and characterization for marine litter detection. *Remote Sensing*, 13(13).

[Freitas et al., 2022] Freitas, S., Silva, H., and Silva, E. (2022). Hyperspectral imaging zero-shot learning for remote marine litter detection and classification. *Remote Sensing*, 14(21).

[Friedl et al., 2001] Friedl, M. A., McGwire, K. C., and McIver, D. K. (2001). *An Overview of Uncertainty in Optical Remotely Sensed Data for Ecological Applications*, pages 258–283. Springer New York, New York, NY.

[Frome et al., 2013] Frome, A., Corrado, G. S., Shlens, J., Bengio, S., Dean, J., Ranzato, M., and Mikolov, T. (2013). Devise: A deep visual-semantic embedding model. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, page 2121–2129, Red Hook, NY, USA. Curran Associates Inc.

[Fu et al., 2014a] Fu, Y., Hospedales, T. M., Xiang, T., Fu, Z., and Gong, S. (2014a). Transductive multi-view embedding for zero-shot recognition and annotation. In *Computer Vision – ECCV 2014*, pages 584–599, Cham. Springer International Publishing.

[Fu et al., 2012] Fu, Y., Hospedales, T. M., Xiang, T., and Gong, S. (2012). Attribute learning for understanding unstructured social activity. In *Computer Vision – ECCV 2012*, pages 530–543, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Fu et al., 2014b] Fu, Y., Hospedales, T. M., Xiang, T., and Gong, S. (2014b). Learning multimodal latent attributes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(2):303–316.

[Fu et al., 2016] Fu, Y., Hospedales, T. M., Xiang, T., Xiong, J., Gong, S., Wang, Y., and Yao, Y. (2016). Robust subjective visual property prediction from crowdsourced pairwise labels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(3):563–577.

[Gao et al., 2018] Gao, Q., Lim, S., and Jia, X. (2018). Hyperspectral image classification using convolutional neural networks and multiple feature learning. *Remote Sensing*, 10(2).

[Gao et al., 2019] Gao, R., Hou, X., Qin, J., Liu, L., Zhu, F., and Zhang, Z. (2019). A joint generative model for zero-shot learning. In *Computer Vision – ECCV 2018 Workshops*, pages 631–646, Cham. Springer International Publishing.

[Garaba and Dierssen, 2018] Garaba, S. P. and Dierssen, H. M. (2018). An airborne remote sensing case study of synthetic hydrocarbon detection using short wave infrared absorption features identified from marine-harvested macro- and microplastics. *Remote Sensing of Environment*, 205(November 2017):224–235.

[Garcia-Garin et al., 2021] Garcia-Garin, O., Monleón-Getino, T., López-Brosa, P., Borrell, A., Aguilar, A., Borja-Robalino, R., Cardona, L., and Vighi, M. (2021). Automatic detection and quantification of floating marine macro-litter in aerial images: Introducing a novel deep learning approach connected to a web application in r. *Environmental Pollution*, 273:116490.

[Garrity and Levings, 1993] Garrity, S. D. and Levings, S. C. (1993). Marine debris along the caribbean coast of panama. *Marine Pollution Bulletin*, (26, 317–324):317–324.

[Ghamisi et al., 2016] Ghamisi, P., Chen, Y., and Zhu, X. X. (2016). A self-improving convolution neural network for the classification of hyperspectral data. *IEEE Geoscience and Remote Sensing Letters*, 13(10):1537–1541.

[Ghamisi et al., 2017] Ghamisi, P., Plaza, J., Chen, Y., Li, J., and Plaza, A. J. (2017). Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geoscience and Remote Sensing Magazine*, 5(1):8–32.

[Giardino et al., 2007] Giardino, C., Brando, V. E., Dekker, A. G., Strömbeck, N., and Candiani, G. (2007). Assessment of water quality in lake garda (italy) using hyperion. *Remote Sensing of Environment*, 109(2):183–195.

[Glowienka and Zembol, 2022] Glowienka, E. and Zembol, N. (2022). Forest community mapping using hyperspectral (chris/proba) and sentinel-2 multispectral images. *Geomatics and Environmental Engineering*, 16(4):103–117.

[Goddijn-Murphy et al., 2018] Goddijn-Murphy, L., Peters, S., Van Sebille, E., James, N. A., and Gibb, S. (2018). Concept for a hyperspectral remote sensing algorithm for floating marine macro plastics. *Marine Pollution Bulletin*, 126(November 2017):255–262.

[Goel et al., 2003] Goel, P., Prasher, S., Patel, R., Landry, J., Bonnell, R., and Viau, A. (2003). Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. *Computers and Electronics in Agriculture*, 39(2):67–93.

[Goetz, 1995] Goetz, A. F. H. (1995). Imaging spectrometry for remote sensing: vision to reality in 15 years. In *Imaging Spectrometry*, volume 2480, pages 2 – 13. International Society for Optics and Photonics, SPIE.

[Gonçalves and Andriolo, 2022] Gonçalves, G. and Andriolo, U. (2022). Operational use of multispectral images for macro-litter mapping and categorization by unmanned aerial vehicle. *Marine Pollution Bulletin*, 176(113431).

[Gonçalves et al., 2020a] Gonçalves, G., Andriolo, U., Pinto, L., and Bessa, F. (2020a). Mapping marine litter using uas on a beach-dune system: a multidisciplinary approach. *Science of The Total Environment*, 706(135742).

[Gonçalves et al., 2020b] Gonçalves, G., Andriolo, U., Pinto, L., and Duarte, D. (2020b). Mapping marine litter with unmanned aerial systems: A showcase comparison among manual image screening and machine learning techniques. *Marine Pollution Bulletin*, 155(111158).

[Gross and Klemas, 1986] Gross, M. F. and Klemas, V. (1986). The use of airborne imaging spectrometer (ais) data to differentiate marsh vegetation. *Remote Sensing of Environment*, 19(1):97–103.

[Guanter et al., 2015] Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., Van der Linden, S., Leitão, P. J., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., and

Sang, B. (2015). The enmap spaceborne imaging spectroscopy mission for earth observation. *Remote Sensing*, 7(7):8830–8857.

[Gulrajani et al., 2017] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. (2017). Improved training of wasserstein gans. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 5769–5779, Red Hook, NY, USA. Curran Associates Inc.

[Guo et al., 2019] Guo, Y., Ding, G., Han, J., Ding, X., Zhao, S., Wang, Z., Yan, C., and Dai, Q. (2019). Dual-view ranking with hardness assessment for zero-shot learning. *Proceedings of the Association for the Advancement of Artificial Intelligence*, 33:8360–8367.

[Guo et al., 2016] Guo, Y., Ding, G., Jin, X., and Wang, J. (2016). Transductive zero-shot recognition via shared model space learning. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, page 3494–3500. AAAI Press.

[Harris, 1981] Harris, Z. S. (1981). Distributional structure. *The Netherlands: Springer*, pages 3–22.

[Haut et al., 2017] Haut, J. M., Paoletti, M., Plaza, J., and Plaza, A. (2017). Cloud implementation of the k-means algorithm for hyperspectral image analysis. *The Journal of Supercomputing*, 73(1):514–529.

[Hofmann, 2006] Hofmann, M. (2006). Support vector machines - kernels and the kernel trick. *An elaboration for the Hauptseminar "Reading Club: Support Vector Machines"*.

[Hsu et al., 2003] Hsu, C.-W., Chang, C.-C., and Lin, C.-J. (2003). A practical guide to support vector classification.

[Hu et al., 2015a] Hu, F., Xia, G.-S., Hu, J., and Zhang, L. (2015a). Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, 7(11):14680–14707.

[Hu et al., 2015b] Hu, W., Huang, Y., Wei, L., Zhang, F., and Li, H. (2015b). Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors - Hindawi Publishing Corporation*, 2015.

[Huang et al., 2012] Huang, E. H., Socher, R., Manning, C. D., , and Ng, A. Y. (2012). Improving word representationsvia global context and multiple word prototypes. *Proceedings Association Computational Linguistics Conference*, pages 873–882.

[Huang et al., 2019a] Huang, H., Wang, C., Yu, P. S., and Wang, C.-D. (2019a). Generative dual adversarial network for generalized zero-shot learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, page 801–810.

[Huang et al., 2019b] Huang, N., Xiao, L., and Xu, Y. (2019b). Bipartite graph partition based coclustering with joint sparsity for hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(12):4698–4711.

[Huang et al., 2017] Huang, S., Zhang, H., and Pižurica, A. (2017). A robust sparse representation model for hyperspectral image classification. *Sensors*, 17(9).

[Hubbard and Crowley, 2005] Hubbard, B. E. and Crowley, J. K. (2005). Mineral mapping on the chilean-bolivian altiplano using co-orbital ali, aster and hyperion imagery: Data dimensionality issues and solutions. *Remote Sensing of Environment*, 99(1-2):173–186.

[Hung et al., 2011] Hung, C.-C., Kulkarni, S., and Kuo, B.-C. (2011). A new weighted fuzzy c-means clustering algorithm for remotely sensed image classification. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):543–553.

[Hunger et al., 2016] Hunger, S., Karrasch, P., and Wessollek, C. (2016). Evaluating the potential of image fusion of multispectral and radar remote sensing data for the assessment of water body structure. In *Remote Sensing for Agriculture, Ecosystems, and Hydrology XVIII*, volume 9998 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*.

[Hutton et al., 2016] Hutton, J. J., Gopaul, N., Zhang, X., Wang, J., Menon, V., Rieck, D., Kipka, A., and Pastor, F. (2016). Centimeter-level robust gnss-aided inertial post-processing for mobile mapping without local reference stations. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41B3:819–826.

[Hutton et al., 2007] Hutton, J. J. S., Bourke, T., and Scherzinger, B. M. (2007). New developments of inertial navigation systems at applanix.

[Huynh and Elhamifar, 2020] Huynh, D. and Elhamifar, E. (2020). Fine-grained generalized zero-shot learning via dense attribute-based attention. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4482–4492.

[Jakovels et al., 2016] Jakovels, D., Filipovs, J., Brauns, A., Taskovs, J., and Erins, G. (2016). Land cover mapping in latvia using hyperspectral airborne and simulated sentinel-2 data. In *Fourth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2016)*, volume 9688 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 96881Q.

[Ji et al., 2020] Ji, Z., Wang, H., Pang, Y., and Shao, L. (2020). Dual triplet network for image zero-shot learning. *Neurocomputing*, 373:90 – 97.

[Ji et al., 2022] Ji, Z., Yu, X., Yu, Y., Pang, Y., and Zhang, Z. (2022). Semantic-guided class-imbalance learning model for zero-shot image classification. *IEEE Transactions on Cybernetics*, 52(7):6543–6554.

[Jimenez and Landgrebe, 1998] Jimenez, L. and Landgrebe, D. (1998). Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 28(1):39–54.

[Jin et al., 2012] Jin, H., Li, P., Cheng, T., and Song, B. (2012). Land cover classification using chris/proba images and multi-temporal texture. *International Journal of Remote Sensing*, 33(1):101–119.

[Joelsson et al., 2007] Joelsson, S. R., Benediktsson, J. A., and Sveinsson, J. R. (2007). Random forest classification of remote sensing data. *Signal and Image Processing for Remote Sensing*, pages 327–344.

[Kalisperakis et al., 2015] Kalisperakis, I., Stentoumis, C., Grammatikopoulos, L., and Karantzalos, K. (2015). Leaf area index estimation in vineyards from uav hyperspectral data, 2d image mosaics and 3d canopy surface models. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-1/W4:299–303.

[Karlson et al., 2020] Karlson, M., Ostwald, M., Bayala, J., Bazié, H. R., Ouedraogo, A. S., Soro, B., Sanou, J., and Reese, H. (2020). The potential of sentinel-2 for crop production estimation in a smallholder agroforestry landscape, burkina faso. *Frontiers in Environmental Science*, 8.

[Kato et al., 2018] Kato, K., Li, Y., and Gupta, A. (2018). Compositional learning for human object interaction. In *Computer Vision – ECCV 2018*, pages 247–264, Cham. Springer International Publishing.

[Keshari et al., 2020] Keshari, R., Singh, R., and Vatsa, M. (2020). Generalized zero-shot learning via over-complete distribution. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13297–13305.

[Kikaki et al., 2022] Kikaki, K., Kakogeorgiou, I., Mikeli, P., Raitsos, D. E., and Karantzalos, K. (2022). Marida: A benchmark for marine debris detection from sentinel-2 remote sensing data. *PLoS ONE 17*.

[Kodirov et al., 2015] Kodirov, E., Xiang, T., Fu, Z., and Gong, S. (2015). Unsupervised domain adaptation for zero-shot learning. *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 2452–2460.

[Kravitz et al., 2020] Kravitz, J., Matthews, M., Bernard, S., and Griffith, D. (2020). Application of sentinel 3 olci for chl-a retrieval over small inland water targets: Successes and challenges. *Remote Sensing of Environment*, 237:111562.

[Kremezi et al., 2021] Kremezi, M., Kristollari, V., Karathanassi, V., Topouzelis, K., Kolokoussis, P., Taggio, N., Aiello, A., Ceriola, G., Barbone, E., and Corradi, P. (2021). Pansharpening prisma data for marine plastic litter detection using plastic indexes. *IEEE Access*, 9:61955–61971.

[Krinidis and Chatzis, 2010] Krinidis, S. and Chatzis, V. (2010). A robust fuzzy local information c-means clustering algorithm. *IEEE Transactions on Image Processing*, 22(3):1328–37.

[Kruse, 2002] Kruse, F. A. (2002). Comparison of aviris and hyperion for hyperspectral mineral mapping. *11th JPL Airborne Geoscience Workshop*.

[Kumar et al., 2020] Kumar, C., Chatterjee, S., and Oommen, T. (2020). Mapping hydrothermal alteration minerals using high-resolution aviris-ng hyperspectral data in the hutti-maski gold deposit area, india. *International Journal of Remote Sensing*, 41(2):794–812.

[Kumar et al., 2009] Kumar, N., Berg, A. C., Belhumeur, P. N., and Nayar, S. K. (2009). Attribute and simile classifiers for face verification. In *2009 IEEE 12th International Conference on Computer Vision*, pages 365–372.

[Kuo and Landgrebe, 2004] Kuo, B.-C. and Landgrebe, D. (2004). Nonparametric weighted feature extraction for classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42(5):1096–1105.

[Lake et al., 2013] Lake, B. M., Salakhutdinov, R. R., and Tenenbaum, J. (2013). One-shot learning by inverting a compositional causal process. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

[Lampert et al., 2009] Lampert, C. H., Nickisch, H., and Harmeling, S. (2009). Learning to detect unseen object classes by between-class attribute transfer. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 951–958.

[Lampert et al., 2014] Lampert, C. H., Nickisch, H., and Harmeling, S. (2014). Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):453–465.

[Larsen et al., 2016] Larsen, A. B. L., Sønderby, S. K., Larochelle, H., and Winther, O. (2016). Autoencoding beyond pixels using a learned similarity metric. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1558–1566, New York, New York, USA.

[Lavigne and Ruddick, 2021] Lavigne, H. and Ruddick, K. (2021). Inter-band calibration for hyperspectral water remote sensing: Demonstration for chris-proba. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 7771–7774.

[Lebreton et al., 2018] Lebreton, L., Slat, B., Ferrari, F., Sainte-Rose, B., Aitken, J., Marthouse, R., Hajbane, S., Cunsolo, S., Schwarz, A., Levivier, A., Noble, K., Debeljak, P., Maral, H., Schoeneich-Argent, R., Brambini, R., and Reisser, J. (2018). Evidence that the great pacific garbage patch is rapidly accumulating plastic. *Scientific Reports*, 8(4666).

[Lee et al., 2015] Lee, J., Lee, J. S., Jang, Y. C., Hong, S. Y., Shim, W. J., Song, Y. K., Hong, S. H., Jang, M., Han, G. M., Kang, D., and Hong, S. (2015). Distribution and size relationships of plastic marine debris on beaches in south korea. *Arch Environ Contam Toxicol.*, 69(3).

[Lee et al., 2018] Lee, K., Lee, K., Min, K., Zhang, Y., Shin, J., and Lee, H. (2018). Hierarchical novelty detection for visual object recognition. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1034–1042.

[Li et al., 2019a] Li, A., Qin, A., Shang, Z., and Tang, Y. Y. (2019a). Spectral-spatial sparse subspace clustering based on three-dimensional edge-preserving filtering for hyperspectral image. *International Journal of Pattern Recognition and Artificial Intelligence*, 33(03):1955003.

[Li et al., 2017] Li, C.-F., Liu, L., Lei, Y.-M., Yin, J.-Y., Zhao, J.-j., and Sun, X.-K. (2017). Clustering for hsi hyperspectral image with weighted pca and ica. *Journal of Intelligent &Fuzzy Systems*, 32(5):3729–3737.

[Li et al., 2020] Li, F., Zhu, Z., Zhang, X., Cheng, J., and Zhao, Y. (2020). From anchor generation to distribution alignment: Learning a discriminative embedding space for zero-shot recognition.

[Li et al., 2010] Li, J., Bioucas-Dias, J. M., and Plaza, A. (2010). Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4085–4098.

[Li et al., 2011] Li, J., Bioucas-Dias, J. M., and Plaza, A. (2011). Hyperspectral image segmentation using a new bayesian approach with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10):3947–3960.

[Li et al., 2012] Li, J., Bioucas-Dias, J. M., and Plaza, A. (2012). Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 50(3):809–823.

[Li et al., 2019b] Li, J., Jing, M., Lu, K., Ding, Z., Zhu, L., and Huang, Z. (2019b). Leveraging the invariant side of generative zero-shot learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7394–7403.

[Li et al., 2022a] Li, J., Jing, M., Lu, K., Zhu, L., and Shen, H. T. (2022a). Investigating the bilateral connections in generative zero-shot learning. *IEEE Transactions on Cybernetics*, 52(8):8167–8178.

[Li et al., 2019c] Li, J., Jing, M., Lu, K., Zhu, L., Yang, Y., and Huang, Z. (2019c). Alleviating feature confusion for generative zero-shot learning. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, page 1587–1595, New York, NY, USA. Association for Computing Machinery.

[Li et al., 2019d] Li, J., Lan, X., Liu, Y., Wang, L., and Zheng, N. (2019d). Compressing unknown images with product quantizer for efficient zero-shot classification. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5458–5467.

[Li et al., 2014] Li, M., Zang, S., Zhang, B., Li, S., and Wu, C. (2014). A review of remote sensing image classification techniques: the role of spatio-contextual information. *European Journal of Remote Sensing*, 47(1):389–411.

[Li et al., 2013] Li, S., Zhang, B., Li, A., Jia, X., Gao, L., and Peng, M. (2013). Hyperspectral imagery clustering with neighborhood constraints. *IEEE Geoscience and Remote Sensing Letters*, 10(3):588–592.

[Li et al., 2015] Li, Y., Swersky, K., and Zemel, R. (2015). Generative moment matching networks. *International Conference on Machine Learning*, page 1718–1727.

[Li et al., 2022b] Li, Z., Liu, M., Chen, Y., Xu, Y., Li, W., and Du, Q. (2022b). Deep cross-domain few-shot learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–18.

[Lin et al., 2019] Lin, G., Chen, W., Liao, K., Kang, X., and Fan, C. (2019). Transfer feature generating networks with semantic classes structure for zero-shot learning. *IEEE Access*, 7:176470–176483.

[Liu et al., 2019] Liu, G., Guan, J., Zhang, M., Zhang, J., Wang, Z., and Lu, Z. (2019). Joint projection and subspace learning for zero-shot recognition. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1228–1233.

[Liu et al., 2011] Liu, J., Kuipers, B., and Savarese, S. (2011). Recognizing human actions by attributes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3337–3344.

[Liu et al., 2017] Liu, S., Mello, S. D., Gu, J., Zhong, G., Yang, M., and Kautz, J. (2017). Learning affinity via spatial propagation networks. *Procedings in Advanced Neural Information Processing Systems (ANIPS)*.

[Liu et al., 2021a] Liu, S., Shi, Q., and Zhang, L. (2021a). Few-shot hyperspectral image classification with unknown classes using multitask deep learning. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6):5085–5102.

[Liu et al., 2016] Liu, W., Li, S., Lin, X., Wu, Y., and Ji, R. (2016). Spectral–spatial co-clustering of hyperspectral image data based on bipartite graph. *Multimedia Systems*, 22:355–366.

[Liu et al., 2020] Liu, Y., Gao, X., Gao, Q., Han, J., and Shao, L. (2020). Label-activating framework for zero-shot learning. *Neural Networks*, 121:1–9.

[Liu et al., 2022] Liu, Y., Gao, X., Han, J., Liu, L., and Shao, L. (2022). Zero-shot learning via a specific rank-controlled semantic autoencoder. *Pattern Recognition*, 122(108237).

[Liu et al., 2021b] Liu, Y., Zhou, L., Bai, X., Huang, Y., Gu, L., Zhou, J., and Harada, T. (2021b). Goal-oriented gaze estimation for zero-shot learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3793–3802.

[Liu et al., 2021c] Liu, Z., Li, Y., Yao, L., Wang, X., and Long, G. (2021c). Task aligned generative meta-learning for zero-shot learning. *CoRR*, abs/2103.02185.

[Lloyd, 1982] Lloyd, S. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137.

[Loizzo et al., 2018] Loizzo, R., Guarini, R., Longo, F., Scopa, T., Formaro, R., Facchinetti, C., and Varacalli, G. (2018). Prisma: The italian hyperspectral mission. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 175–178.

[Malila, 1980] Malila, W. A. (1980). Change vector analysis: An approach for detecting forest changes with landsat. *Symposium on Machine Processing of Remotely Sensed Data and Soil Information Systems and Remote Sensing and Soil Survey*, 80:326–335.

[Mandal et al., 2019] Mandal, D., Narayan, S., Dwivedi, S. K., Gupta, V., Ahmed, S., Khan, F. S., and Shao, L. (2019). Out-of-distribution detection for generalized zero-shot action recognition. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9977–9985.

[Mannheim et al., 2004] Mannheim, S., Segl, K., Heim, B., and Kaufmann, H. (2004). Monitoring of lake water quality using hyperspectral chris-proba data. ESA special publication ; SP-578, Frascati, Italy.

[Martin et al., 2018] Martin, C., Parkes, S., Zhang, Q., Zhang, X., McCabe, M. F., and Duarte, C. M. (2018). Use of unmanned aerial vehicles for efficient beach litter monitoring. *Marine Pollution Bulletin*, 131:662–673.

[Maulik and Saha, 2010] Maulik, U. and Saha, I. (2010). Automatic fuzzy clustering using modified differential evolution for image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(9):3503–3510.

[McFEETERS, 1996] McFEETERS, S. K. (1996). The use of the normalized difference water index (ndwi) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7):1425–1432.

[McIlgorm et al., 2011] McIlgorm, A., Campbell, H. F., and Rule, M. J. (2011). The economic cost and control of marine debris damage in the asia-pacific region. *Ocean &Coastal Management*, 54(9):643–651.

[Melgani and Bruzzone, 2004] Melgani, F. and Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790.

[Mensink et al., 2014] Mensink, T., Gavves, E., and Snoek, C. G. M. (2014). Costa: Co-occurrence statistics for zero-shot classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2441–2448.

[Mensink et al., 2012] Mensink, T., Verbeek, J., Perronnin, F., and Csurka, G. (2012). Metric learning for large scale image classification: Generalizing to new classes at near-zero cost. In *Computer Vision – ECCV 2012*, pages 488–501, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Merényi et al., 2014] Merényi, E., Farrand, W. H., Taranik, J. V., and Minor, T. B. (2014). Classification of hyperspectral imagery with neural networks: comparison to conventional tools. *EURASIP Journal on Advances in Signal Processing*, 2014(71):1–19.

[Middleton et al., 2013] Middleton, E. M., Ungar, S. G., Mandl, D. J., Ong, L., Frye, S. W., Campbell, P. E., Landis, D. R., Young, J. P., and Pollack, N. H. (2013). The earth observing one (eo-1) satellite mission: Over a decade in space. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2):243–256.

[Mishra et al., 2018] Mishra, A., Reddy, S. K., Mittal, A., and Murthy, H. A. (2018). A generative model for zero shot learning using conditional variational autoencoders. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2269–2277.

[Miura et al., 2003] Miura, T., Huete, A. R., Ferreira, L. G., and Sano, E. E. (2003). Discrimination and biophysical characterization of cerrado physiognomies with eo-1 hyperspectral hyperion. *Proceedings of the Simpósio Brasileiro de Sensoriamento Remoto*, pages 1077–1082.

[Moller, 1993] Moller, M. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4):525–533.

[Moore et al., 1999] Moore, G. F., Aiken, J., and Lavender, S. J. (1999). The atmospheric correction of water colour and the quantitative retrieval of suspended particulate matter in case ii waters: Application to meris. *International Journal of Remote Sensing*, 20(9):1713–1733.

[Mountrakis et al., 2011] Mountrakis, G., Im, J., and Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3):247–259.

[Murphy and Maggioni, 2020] Murphy, J. M. and Maggioni, M. (2020). Spectral–spatial diffusion geometry for hyperspectral image clustering. *IEEE Geoscience and Remote Sensing Letters*, 17(7):1243–1247.

[Neagoe and Chirila-Berbentea, 2016] Neagoe, V.-E. and Chirila-Berbentea, V. (2016). Improved gaussian mixture model with expectation-maximization for clustering of remote sensing imagery. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3063–3065.

[Nelson, 1983] Nelson, R. F. (1983). Detecting forest canopy change due to insect activity using a 3 landsat mss. *Photogrammetric Engineering and Remote Sensing*, 49(9):1303–1314.

[Neuenschwander et al., 1998] Neuenschwander, A., Crawford, M., and Provancha, M. (1998). Mapping of coastal wetlands via hyperspectral aviris data. In *IGARSS '98. Sensing and Managing the Environment. 1998 IEEE International Geoscience and Remote Sensing. Symposium Proceedings. (Cat. No.98CH36174)*, volume 1, pages 189–191.

[Ni et al., 2019] Ni, J., Zhang, S., and Xie, H. (2019). *Dual Adversarial Semantics-Consistent Network for Generalized Zero-Shot Learning.* Curran Associates Inc., Red Hook, NY, USA.

[Norouzi et al., 2013] Norouzi, M., Mikolov, T., Bengio, S., Singer, Y., Shlens, J., Frome, A., Corrado, G., and Dean, J. (2013). Zero-shot learning by convex combination of semantic embeddings. *CoRR*, abs/1312.5650.

[Odena et al., 2017] Odena, A., Olah, C., and Shlens, J. (2017). Conditional image synthesis with auxiliary classifier gans. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 2642–2651.

[Pal et al., 2022] Pal, D., Bundele, V., Sharma, R., Banerjee, B., and Jeppu, Y. (2022). Few-shot open-set recognition of hyperspectral images with outlier calibration network. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2091–2100.

[Pal and Foody, 2010] Pal, M. and Foody, G. M. (2010). Feature selection for classification of hyperspectral data by svm. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5):2297–2307.

[Palatucci et al., 2009] Palatucci, M., Pomerleau, D., Hinton, G. E., and Mitchell, T. M. (2009). Zero-shot learning with semantic output codes. In *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc.

[Paoletti et al., 2018] Paoletti, M., Haut, J., Plaza, J., and Plaza, A. (2018). A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145:120–147. Deep Learning RS Data.

[Paoli et al., 2009] Paoli, A., Melgani, F., and Pasolli, E. (2009). Clustering of hyperspectral images based on multiobjective particle swarm optimization. *IEEE Transactions on Geoscience and Remote Sensing*, 47(12):4175–4188.

[Papakonstantinou et al., 2021] Papakonstantinou, A., Batsaris, M., Spondylidis, S., and Topouzelis, K. (2021). A citizen science unmanned aerial system data acquisition protocol and deep learning techniques for the automatic detection and mapping of marine litter concentrations in the coastal zone. *Drones*, 5(1).

[Parikh and Grauman, 2011] Parikh, D. and Grauman, K. (2011). Relative attributes. In *2011 International Conference on Computer Vision*, pages 503–510.

[Paz-Argaman et al., 2020] Paz-Argaman, T., Atzmon, Y., Chechik, G., and Tsarfaty, R. (2020). Zest: Zero-shot learning from text descriptions using textual similarity and visual summarization. *CoRR*, abs/2010.03276.

[Pedrycz, 1996] Pedrycz, W. (1996). Conditional fuzzy c-means. *Pattern Recognition Letters*, 17(6):625–631.

[Pengra et al., 2007] Pengra, B. W., Johnston, C. A., and Loveland, T. R. (2007). Mapping an invasive plant, phragmites australis, in coastal wetlands using the eo-1 hyperion hyperspectral sensor. *Remote Sensing of Environment*, 108(1):74–81.

[Peschoud et al., 2017] Peschoud, C., Minghelli, A., Mathieu, S., Lei, M., Pairaud, I., and Pinazo, C. (2017). Fusion of sun-synchronous and geostationary images for coastal and ocean color survey application to olci (sentinel-3) and fci (mtg). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(1):45–56.

[Petropoulos et al., 2012a] Petropoulos, G. P., Arvanitis, K., and Sigrimis, N. (2012a). Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping. *Expert Systems with Applications*, 39(3):3800–3809.

[Petropoulos et al., 2012b] Petropoulos, G. P., Kontoes, C. C., and Keramitsoglou, I. (2012b). Land cover mapping with emphasis to burnt area delineation using co-orbital ali and landsat tm imagery. *International Journal of Applied Earth Observation and Geoinformation*, 18:344–355.

[Pham, 2001] Pham, D. L. (2001). Spatial models for fuzzy clustering. *Computer Vision and Image Understanding*, 84(2):285–297.

[Pignatti et al., 2013] Pignatti, S., Palombo, A., Pascucci, S., Romano, F., Santini, F., Simoniello, T., Umberto, A., Vincenzo, C., Acito, N., Diani, M., Matteoli, S., Corsini, G., Casa, R., De Bonis, R., Laneve, G., and Ananasso, C. (2013). The prisma hyperspectral mission: Science activities and opportunities for agriculture and land monitoring. In *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS*, pages 4558–4561.

[Pourpanah et al., 2022] Pourpanah, F., Abdar, M., Luo, Y., Zhou, X., Wang, R., Lim, C. P., Wang, X.-Z., and Wu, Q. M. J. (2022). A review of generalized zero-shot learning methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–20.

[Pu et al., 2008] Pu, R., Gong, P., and Yu, Q. (2008). Comparative analysis of eo-1 ali and hyperion, and landsat etm+ data for mapping forest crown closure and leaf area index. *Sensors*, 8(6):3744–3766.

[Pu et al., 2005] Pu, R., Yu, Q., Gong, P., and Biging, G. S. (2005). Eo-1 hyperion, ali and landsat 7 etm+ data comparison for estimating forest crown closure and leaf area index. *International Journal of Remote Sensing*, 26(3):457–474.

[Purushwalkam et al., 2019] Purushwalkam, S., Nickel, M., Gupta, A., and Ranzato, M. (2019). Task-driven modular networks for zero-shot compositional learning. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3592–3601.

[Qu et al., 2019] Qu, Y., Baghbaderani, R. K., and Qi, H. (2019). Few-shot hyperspectral image classification through multitask transfer learning. In *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–5.

[Rahman, 2015] Rahman, S. A. E. (2015). Hyperspectral imaging classification using isodata algorithm: Big data challenge. In *2015 Fifth International Conference on e-Learning (econf)*, pages 247–250.

[Ratle et al., 2010] Ratle, F., Camps-Valls, G., and Weston, J. (2010). Semisupervised neural networks for efficient hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5):2271–2282.

[Reed et al., 2016] Reed, S., Akata, Z., Lee, H., and Schiele, B. (2016). Learning deep representations of fine-grained visual descriptions. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 49–58.

[Ren et al., 2019] Ren, Z., Sun, L., Zhai, Q., and Liu, X. (2019). Mineral mapping with hyperspectral image based on an improved k-means clustering algorithm. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 2989–2992.

[Roberts et al., 2003] Roberts, D., Dennison, P., Gardner, M., Hetzel, Y., Ustin, S., and Lee, C. (2003). Evaluation of the potential of hyperion for fire danger assessment by comparison to the airborne visible/infrared imaging spectrometer. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6):1297–1310.

[Rodarmel and Shan, 2002] Rodarmel, C. and Shan, J. (2002). Principal component analysis for hyperspectral image classification. *Surv Land inf Syst*, 62.

[Rodriguez and Laio, 2014] Rodriguez, A. and Laio, A. (2014). Clustering by fast search and find of density peaks. *Science*, 344(6191):1492–1496.

[Rodriguez-Galiano et al., 2012] Rodriguez-Galiano, V., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67:93–104.

[Rohrbach et al., 2011] Rohrbach, M., Stark, M., and Schiele, B. (2011). Evaluating knowledge transfer and zero-shot learning in a large-scale setting. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1641–1648.

[Romera-Paredes and Torr, 2015] Romera-Paredes, B. and Torr, P. (2015). An embarrassingly simple approach to zero-shot learning. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2152–2161, Lille, France. PMLR.

[Romero et al., 2016] Romero, A., Gatta, C., and Camps-Valls, G. (2016). Unsupervised deep feature extraction for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3):1349–1362.

[Rossiter et al., 2020] Rossiter, T., Furey, T., McCarthy, T., and Stengel, D. B. (2020). Uav-mounted hyperspectral mapping of intertidal macroalgae. *Estuarine, Coastal and Shelf Science*, 242:106789.

[Rouse et al., 1973] Rouse, J. W., J., Haas, R. H., Schell, J. A., and Deering, D. W. (1973). Monitoring vegetation systems in the great plains with erts. In *Proceedings of the Third Earth Resources Technology Satellite-1 Symposium*, pages 309–317.

[Rudd and Boult, 2016] Rudd, Ethan M.and Günther, M. and Boult, T. E. (2016). Moon: A mixed objective optimization network for the recognition of facial attributes. In *Computer Vision – ECCV 2016*, pages 19–35, Cham. Springer International Publishing.

[Ruiz et al., 2020] Ruiz, I., Basurko, O. C., Rubio, A., Delpey, M., Granado, I., Declerck, A., Mader, J., and Cózar, A. (2020). Litter windrows in the south-east coast of the bay of biscay: An ocean process enabling effective active fishing for litter. *Frontiers in Marine Science*, 7:308.

[Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

[Ryan, 2015a] Ryan, P. G. (2015a). *A Brief History of Marine Litter Research*, pages 1–25. Springer International Publishing, Cham.

[Ryan, 2015b] Ryan, P. G. (2015b). Does size and buoyancy affect the long-distance transport of floating debris? *Environmental Research Letters*, (10).

[Ryan et al., 2009] Ryan, P. G., Moore, C. J., Franeker, V., A., J., and Moloney, C. L. (2009). Monitoring the abundance of plastic debris in the marine environment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1526):1999–2012.

[Salakhutdinov et al., 2013] Salakhutdinov, R., Tenenbaum, J. B., and Torralba, A. (2013). Learning with hierarchical-deep models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1958–1971.

[Salimans et al., 2016] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., and Chen, X. (2016). Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

[Sankey et al., 2018] Sankey, T. T., McVay, J., Swetnam, T. L., McClaran, M. P., Heilman, P., and Nichols, M. (2018). Uav hyperspectral and lidar data and their fusion for arid and semi-arid land vegetation monitoring. *Remote Sensing in Ecology and Conservation*, 4(1):20–33.

[Schiller and Doerffer, 2005] Schiller, H. and Doerffer, R. (2005). Improved determination of coastal water constituent concentrations from meris data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(7):1585–1591.

[Schroeder et al., 2007] Schroeder, T., Schaale, M., and Fischer, J. (2007). Retrieval of atmospheric and oceanic properties from meris measurements: A new case-2 water processor for beam. *International Journal of Remote Sensing*, 28(24):5627–5632.

[Schönfeld et al., 2019] Schönfeld, E., Ebrahimi, S., Sinha, S., Darrell, T., and Akata, Z. (2019). Generalized zero- and few-shot learning via aligned variational autoencoders. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8239–8247.

[Scott, 1992] Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.*

[Shafaey et al., 2021] Shafaey, M. A., Salem, M. A.-M., Al-Berry, M. N., Ebied, H. M., and Tolba, M. F. (2021). Review on supervised and unsupervised deep learning techniques for hyperspectral images classification. In *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2021)*, pages 66–74, Cham. Springer International Publishing.

[Shah et al., 2004] Shah, C. A., Arora, M. K., and author, P. K. V. C. (2004). Unsupervised classification of hyperspectral data: an ica mixture model based approach. *International Journal of Remote Sensing*, 25(2):481–487.

[Shah et al., 2007] Shah, C. A., Varshney, P. K., and Arora, M. K. (2007). Ica mixture model algorithm for unsupervised classification of remote sensing imagery. *International Journal of Remote Sensing*, 28(8):1711–1731.

[Shahidul Islam and Tanaka, 2004] Shahidul Islam, M. and Tanaka, M. (2004). Impacts of pollution on coastal and marine ecosystems including coastal and marine fisheries and approach for management: a review and synthesis. *Marine Pollution Bulletin*, 48(7):624–649.

[Shao and Li, 2020] Shao, J. and Li, X. (2020). Generalized zero-shot learning with multi-channel gaussian mixture vae. *IEEE Signal Processing Letters*, 27:456–460.

[Shen and Li, 2010] Shen, L. and Li, C. (2010). Water body extraction from landsat etm+ imagery using adaboost algorithm. In *2010 18th International Conference on Geoinformatics*, pages 1–4.

[Shi and Malik, 2000] Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905.

[Singh and Lee, 2016] Singh, K. K. and Lee, Y. J. (2016). End-to-end localization and ranking for relative attributes. *European Conference on Computer Vision (ECCV)*.

[Smith et al., 2018] Smith, M., Robertson Lain, L., and Bernard, S. (2018). An optimized chlorophyll a switching algorithm for meris and olci in phytoplankton-dominated waters. *Remote Sensing of Environment*, 215:217–227.

[Smith et al., 2003] Smith, M.-L., Martin, M., Plourde, L., and Ollinger, S. (2003). Analysis of hyperspectral data for estimation of temperate forest canopy nitrogen concentration: comparison between an airborne (aviris) and a spaceborne (hyperion) sensor. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6):1332–1337.

[Sohn et al., 2015] Sohn, K., Yan, X., and Lee, H. (2015). Learning structured output representation using deep conditional generative models. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'15, page 3483–3491, Cambridge, MA, USA. MIT Press.

[Song et al., 2018] Song, J., Shen, C., Yang, Y., Liu, Y., and Song, M. (2018). Transductive unbiased embedding for zero-shot learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1024–1033.

[Spoto et al., 2012] Spoto, F., Sy, O., Laberinti, P., Martimort, P., Fernandez, V., Colin, O., Hoersch, B., and Meygret, A. (2012). Overview of sentinel-2. In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pages 1707–1710.

[Stagakis et al., 2010] Stagakis, S., Markos, N., Sykioti, O., and Kyparissis, A. (2010). Monitoring canopy biophysical and biochemical parameters in ecosystem scale using satellite hyperspectral imagery: An application on a phlomis fruticosa mediterranean ecosystem using multiangular chris/proba observations. *Remote Sensing of Environment*, 114(5):977–994.

[Sung et al., 2018] Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P. H. S., and Hospedales, T. M. (2018). Learning to compare: Relation network for few-shot learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, page 1199–1208.

[Sylvain et al., 2019] Sylvain, T., Petrini, L., and Hjelm, D. (2019). Locality and compositionality in zero-shot learning. *Eighth International Conference on Learning Representations (ICLR)*.

[Taddia et al., 2021] Taddia, Y., Corbau, C., Buoninsegni, J., Simeoni, U., and Pellegrinelli, A. (2021). Uav approach for detecting plastic marine debris on the beach: A case study in the po river delta (italy). *Drones*, 5(4).

[Taggio et al., 2022] Taggio, N., Aiello, A., Ceriola, G., Kremezi, M., Kristollari, V., Kolokoussis, P., Karathanassi, V., and Barbone, E. (2022). A combination of machine learning algorithms for marine plastic litter detection exploiting hyperspectral prisma data. *Remote Sensing*, 14(15).

[Tasseron et al., 2021] Tasseron, P., van Emmerik, T., Peller, J., Schreyers, L., and Biermann, L. (2021). Advancing floating macroplastic detection from space using experimental hyperspectral imagery. *Remote Sensing*, 13(12).

[Themistocleous et al., 2020] Themistocleous, K., Papoutsa, C., Michaelides, S., and Hadjimitsis, D. (2020). Investigating detection of floating plastic litter from space using sentinel-2 imagery. *Remote Sensing*, 12(16).

[Thenkabail et al., 2004] Thenkabail, P. S., Enclona, E. A., Ashton, M. S., Legg, C., and De Dieu, M. J. (2004). Hyperion, ikonos, ali, and etm+ sensors in the study of african rainforests. *Remote Sensing of Environment*, 90(1):23–43.

[Tian et al., 2018a] Tian, L., Du, Q., Kopriva, I., and Younan, N. (2018a). Kernel spatial-spectral based multi-view low-rank sparse sbuspace clustering for hyperspectral imagery. In *2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4.

[Tian et al., 2018b] Tian, L., Du, Q., Kopriva, I., and Younan, N. (2018b). Spatial-spectral based multi-view low-rank sparse sbuspace clustering for hyperspectral imagery. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 8488–8491.

[Toming et al., 2016] Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., and Nõges, T. (2016). First experiences in mapping lake water quality parameters with sentinel-2 msi imagery. *Remote Sensing*, 8(8).

[Tong et al., 2020] Tong, X., Yin, J., Han, B., and Qv, H. (2020). Few-shot learning with attention-weighted graph convolutional networks for hyperspectral image classification. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 1686–1690.

[Topouzelis et al., 2019] Topouzelis, K., Papakonstantinou, A., and Garaba, S. P. (2019). Detection of floating plastics from satellite and unmanned aerial systems (plastic litter project 2018). *International Journal of Applied Earth Observation and Geoinformation*, 79(January):175–183.

[Torma et al., 2012a] Torma, M., Hatunen, S., Harma, P., and Jarvenpaa, E. (2012a). Sentinel-2 images and finnish corine land cover classification. In *First Sentinel-2 Preparatory Symposium*, volume 707 of *ESA Special Publication*, page 20.

[Torma et al., 2012b] Torma, M., Lewinski, S., Aleksandrowicz, S., Esch, T., Metz, A., Smith, G., Lamb, A., and Turlej, K. (2012b). Seasonality of land cover types as basis for improved land cover classification within pan-euopean area frame sampling scheme. In *First Sentinel-2 Preparatory Symposium*, volume 707 of *ESA Special Publication*, page 47.

[Tripathi and Garg, 2021] Tripathi, P. and Garg, R. D. (2021). Feature extraction of desis and prisma hyperspectral remote sensing datasets for geological applications. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 44M3:169–173.

[Vaglio Laurin et al., 2016] Vaglio Laurin, G., Puletti, N., Hawthorne, W., Liesenberg, V., Corona, P., Papale, D., Chen, Q., and Valentini, R. (2016). Discrimination of tropical forest types, dominant species, and mapping of functional guilds by hyperspectral and simulated multispectral sentinel-2 data. *Remote Sensing of Environment*, 176:163–176.

[van Sebille et al., 2015] van Sebille, E., Wilcox, C., Lebreton, L., Maximenko, N., Hardesty, B. D., van Franeker, J. A., Eriksen, M., Siegel, D., Galgani, F., and Law, K. L. (2015). A global inventory of small floating plastic debris. *Environmental Research Letters*, 10(12):124006.

[Vane et al., 1993] Vane, G., Green, R. O., Chrien, T. G., Enmark, H. T., Hansen, E. G., and Porter, W. M. (1993). The airborne visible/infrared imaging spectrometer (aviris). *Remote Sensing of Environment*, 44(2):127–143. Airbone Imaging Spectrometry.

[Vanegas et al., 2018] Vanegas, F., Bratanov, D., Powell, K., Weiss, J., and Gonzalez, F. (2018). A novel methodology for improving plant pest surveillance in vineyards and crops using uav-based hyperspectral and spatial data. *Sensors*, 18(1).

[Vanhellmont and Ruddick, 2016] Vanhellmont, Q. and Ruddick, K. (2016). Acolite for sentinel-2: Aquatic applications of msi imagery. In *Living Planet Symposium*, volume 740 of *ESA Special Publication*, page 55.

[Veiga et al., 2016] Veiga, J., Fleet, D., Kinsey, S., Nilsson, P., Vlachogianni, T., Werner, S., Galgani, F., Thompson, R., Dagevos, J., Gago, J., Sobral, P., and Cronin, R. (2016). *Identifying sources of marine litter.*

[Verma et al., 2018] Verma, V. K., Arora, G., Mishra, A., and Rai, P. (2018). Generalized zero-shot learning via synthesized examples. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4281–4289.

[Verma et al., 2020] Verma, V. K., Brahma, D., and Rai, P. (2020). Meta-learning for generalized zero-shot learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, page 6062–6069.

[Verma et al., 2021a] Verma, V. K., Liang, K. J., Mehta, N., and Carin, L. (2021a). Meta-learned attribute self-gating for continual generalized zero-shot learning. *CoRR*, abs/2102.11856.

[Verma et al., 2021b] Verma, V. K., Mishra, A., Pandey, A., Murthy, H. A., and Rai, P. (2021b). Towards zero-shot learning with fewer seen class examples. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 2240–2250.

[Verma and Rai, 2017] Verma, V. K. and Rai, P. (2017). A simple exponential family framework for zero-shot learning. In *Machine Learning and Knowledge Discovery in Databases*, pages 792–808. Springer International Publishing.

[von Luxburg, 2007] von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416.

[Wang et al., 2016] Wang, J., Cheng, Y., and Feris, R. S. (2016). Walk and learn: Facial attribute representation learning from egocentric video and contextual data. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2295–2304.

[Wang and Shi, 2013] Wang, Q. and Shi, W. (2013). Unsupervised classification based on fuzzy c-means with uncertainty analysis. *Remote Sensing Letters*, 4(11):1087–1096.

[Wang et al., 2017a] Wang, R., Nie, F., and Yu, W. (2017a). Fast spectral clustering with anchor graph for large hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, 14(11):2003–2007.

[Wang et al., 2018] Wang, W., Pu, Y., Verma, V. K., Fan, K., Zhang, Y., Chen, C., Rai, P., and Carin, L. (2018). Zero-shot learning via class-conditioned deep generative models. *AAAI Conference on Artificial Intelligence*.

[Wang et al., 2021] Wang, W., Xu, H., Wang, G., Wang, W., and Carin, L. (2021). Zero-shot recognition via optimal transport. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 3470–3480.

[Wang, 2015] Wang, X. (2015). Learning from big data with uncertainty – editorial. *J. Intell. Fuzzy Syst.*, 28(5):2329–2330.

[Wang and Zhang, 2011] Wang, X. and Zhang, T. (2011). Clothes search in consumer photos via color matching and attribute learning. In *Proceedings of the 19th ACM International Conference on Multimedia*, MM '11, pages 1353–1356, New York, NY, USA. ACM.

[Wang and Bu, 2010] Wang, X.-Y. and Bu, J. (2010). A fast and robust image segmentation using fcm with spatial information. *Digital Signal Processing*, 20(4):1173–1182.

[Wang et al., 2022] Wang, Y., Liu, M., Yang, Y., Li, Z., Du, Q., Chen, Y., Li, F., and Yang, H. (2022). Heterogeneous few-shot learning for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5.

[Wang et al., 2020] Wang, Y., Zhang, H., Zhang, Z., and Long, Y. (2020). Asymmetric graph based zero shot learning. *Multimedia Tools and Applications*, 79:33689–33710.

[Wang et al., 2017b] Wang, Z., He, K., Fu, Y., Feng, R., Jiang, Y.-G., and Xue, X. (2017b). Multi-task deep neural network for joint face recognition and facial attribute prediction. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, ICMR '17, page 365–374, New York, NY, USA. Association for Computing Machinery.

[Waske et al., 2012] Waske, B., Benediktsson, J. A., and Sveinsson, J. R. (2012). Random forest classification of remote sensing data. *Signal and Image Processing for Remote Sensing*, pages 363–374.

[Waske et al., 2009] Waske, B., Benediktsson, J. A., Árnason, K., and Sveinsson, J. R. (2009). Mapping of hyperspectral aviris data using machine-learning algorithms. *Canadian Journal of Remote Sensing*, 35(sup1):S106–S116.

[Wegman, 1990] Wegman, E. J. (1990). Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association*, 85(411):664–675.

[Wei et al., 2019] Wei, L., Huang, C., Zhong, Y., Wang, Z., Hu, X., and Lin, L. (2019). Inland waters suspended solids concentration retrieval based on pso-lssvm for uav-borne hyperspectral remote sensing imagery. *Remote Sensing*, 11(12).

[Wei et al., 2020] Wei, L., Wang, Z., Huang, C., Zhang, Y., Wang, Z., Xia, H., and Cao, L. (2020). Transparency estimation of narrow rivers by uav-borne hyperspectral remote sensing imagery. *IEEE Access*, 8:168137–168153.

[Wen et al., 2016] Wen, Y., Zhang, K., Li, Z., and Qiao, Y. (2016). A discriminative feature learning approach for deep face recognition. In *Computer Vision – ECCV 2016*, pages 499–515, Cham. Springer International Publishing.

[Weng et al., 2008] Weng, Q., Hu, X., and Lu, D. (2008). Extracting impervious surfaces from medium spatial resolution multispectral and hyperspectral imagery: a comparison. *International Journal of Remote Sensing*, 29(11):3209–3232.

[WENG et al., 2010] WENG, Y.-L., GONG, P., and ZHU, Z.-L. (2010). A spectral index for estimating soil salinity in the yellow river delta region of china using eo-1 hyperion data. *Pedosphere*, 20(3):378–388.

[Willoughby et al., 1997] Willoughby, N., Sangkoyo, H., and Lakaseru, B. O. (1997). Beach litter: an increasing and changing problem for indonesia. *Marine Pollution Bulletin*, 34(6):469–478.

[Wilson and Sader, 2002] Wilson, E. H. and Sader, S. A. (2002). Detection of forest harvest type using multiple dates of landsat tm imagery. *Remote Sensing of Environment*, 80(3):385–396.

[Windrim et al., 2019] Windrim, L., Ramakrishnan, R., Melkumyan, A., Murphy, R. J., and Chlingaryan, A. (2019). Unsupervised feature-learning for hyperspectral data with autoencoders. *Remote Sensing*, 11(7).

[Woerd and Wernand, 2015] Woerd, H. J. v. d. and Wernand, M. R. (2015). True colour classification of natural waters with medium-spectral resolution satellites: Seawifs, modis, meris and olci. *Sensors*, 15(10):25663–25680.

[Wolf et al., 2020] Wolf, M., van den Berg, K., Garaba, S. P., Gnann, N., Sattler, K., Stahl, F., and Zielinski, O. (2020). Machine learning for aquatic plastic litter detection, classification and quantification (APLASTIC-q). *Environmental Research Letters*, 15(11):114042.

[Woo et al., 2018] Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). Cbam: Convolutional block attention module.

[Woodcock and Strahler, 1987] Woodcock, C. E. and Strahler, A. H. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment*, 21(3):311–332.

[Xian et al., 2016] Xian, Y., Akata, Z., Sharma, G., Nguyen, Q., Hein, M., and Schiele, B. (2016). Latent embeddings for zero-shot classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, page 69–77.

[Xian et al., 2018] Xian, Y., Lorenz, T., Schiele, B., and Akata, Z. (2018). Feature generating networks for zero-shot learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5542–5551.

[Xian et al., 2017] Xian, Y., Schiele, B., and Akata, Z. (2017). Zero-shot learning - the good, the bad and the ugly. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3077–3086.

[Xiang et al., 2021] Xiang, H., Xie, C., Zeng, T., and Yang, Y. (2021). Multi-knowledge fusion for new feature generation in generalized zero-shot learning. *CoRR*, abs/2102.11566.

[Xie et al., 2019] Xie, G.-S., Liu, L., Jin, X., Zhu, F., Zhang, Z., Qin, J., Yao, Y., and Shao, L. (2019). Attentive region embedding network for zero-shot learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9376–9385.

[Xie et al., 2020] Xie, G.-S., Liu, L., Zhu, F., Zhao, F., Zhang, Z., Yao, Y., Qin, J., and Shao, L. (2020). Region graph embedding network for zero-shot learning. *European Conference on Computer Vision*, page 562–580.

[Xie et al., 2022a] Xie, G.-S., Zhang, Z., Liu, G., Zhu, F., Liu, L., Shao, L., and Li, X. (2022a). Generalized zero-shot learning with multiple graph adaptive generative networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33(7):2903–2915.

[Xie et al., 2022b] Xie, G.-S., Zhang, Z., Xiong, H., Shao, L., and Li, X. (2022b). Towards zero-shot learning: A brief review and an attention-based embedding network. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–17.

[Xie et al., 2018] Xie, H., Zhao, A., Huang, S., Han, J., Liu, S., Xu, X., Luo, X., Pan, H., Du, Q., and Tong, X. (2018). Unsupervised hyperspectral remote sensing image clustering based on adaptive density. *IEEE Geoscience and Remote Sensing Letters*, 15(4):632–636.

[Xu et al., 2016] Xu, B., Fu, Y., Jiang, Y.-G., Li, B., and Sigal, L. (2016). Video emotion recognition with transferred deep feature encodings. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, ICMR '16, page 15–22, New York, NY, USA. Association for Computing Machinery.

[Xu, 2006] Xu, H. (2006). Modification of normalised difference water index (ndwi) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27(14):3025–3033.

[Xu et al., 2018] Xu, J., Li, H., Liu, P., and Xiao, L. (2018). A novel hyperspectral image clustering method with context-aware unsupervised discriminative extreme learning machine. *IEEE Access*, 6:16176–16188.

[Yang et al., 2020a] Yang, G., Huang, K., Zhang, R., Goulermas, J. Y., and Hussain, A. (2020a). Self-focus deep embedding model for coarse-grained zero-shot classification. In *Advances in Brain Inspired Cognitive Systems*, pages 12–22, Cham. Springer International Publishing.

[Yang et al., 2007] Yang, G.-P., Liu, H.-Y., and Yu, X.-C. (2007). Hyperspectral remote sensing image classification based on kernel fisher discriminant analysis. In *2007 International Conference on Wavelet Analysis and Pattern Recognition*, volume 3, pages 1139–1143.

[Yang et al., 2020b] Yang, S., Wang, K., Herranz, L., and van de Weijer, J. (2020b). Simple and effective localized attribute representations for zero-shot learning. *CoRR*, abs/2006.05938.

[Yang et al., 2017] Yang, W., Hou, K., Liu, B., Yu, F., and Lin, L. (2017). Two-stage clustering technique based on the neighboring union histogram for hyperspectral remote sensing images. *IEEE Access*, 5:5640–5647.

[Yang Long and Shao, 2016] Yang Long, L. L. and Shao, L. (2016). Attribute embedding with visual-semantic ambiguity removal for zero-shot learning. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 40.1–40.11. BMVA Press.

[Ye et al., 2019] Ye, Z., Lyu, F., Li, L., Fu, Q., Ren, J., and Hu, F. (2019). Sr-gan: Semantic rectifying generative adversarial network for zero-shot learning. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 85–90.

[Yeh and Li, 2020] Yeh, M. and Li, F. (2020). Zero-shot recognition through image-guided semantic classification. *CoRR*, abs/2007.11814.

[Yoon and Kim, 2007] Yoon, Y.-S. and Kim, Y.-S. (2007). Application of hyperion hyperspectral remote sensing data for wildfire fuel mapping. *Korean Journal of Remote Sensing*, 23(1):21–32.

[Yu et al., 2018] Yu, Y., Ji, Z., Fu, Y., Guo, J., Pang, Y., and Zhang, Z. M. (2018). Stacked semantics-guided attention model for fine-grained zero-shot learning. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.

[Yu et al., 2019] Yu, Y., Ji, Z., Guo, J., and Zhang, Z. (2019). Zero-shot learning via latent space encoding. *IEEE Transactions on Cybernetics*, 49(10):3755–3766.

[Yu et al., 2020] Yu, Y., Ji, Z., Han, J., and Zhang, Z. (2020). Episode-based prototype generating network for zero-shot learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14032–14041.

[Yue et al., 2015] Yue, J., Zhao, W., Mao, S., and Liu, H. (2015). Spectral–spatial classification of hyperspectral images using deep convolutional neural networks. *Remote Sensing Letters*, 6(6):468–477.

[Zhai et al., 2021] Zhai, H., Zhang, H., Li, P., and Zhang, L. (2021). Hyperspectral image clustering: Current achievements and future lines. *IEEE Geoscience and Remote Sensing Magazine*, 9(4):35–67.

[Zhai et al., 2017] Zhai, H., Zhang, H., Xu, X., Zhang, L., and Li, P. (2017). Kernel sparse subspace clustering with a spatial max pooling operation for hyperspectral remote sensing data interpretation. *Remote Sensing*, 9(4).

[Zhai et al., 2019] Zhai, H., Zhang, H., Zhang, L., and Li, P. (2019). Total variation regularized collaborative representation clustering with a locally adaptive dictionary for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1):166–180.

[Zhang et al., 2013] Zhang, B., Li, S., Wu, C., Gao, L., Zhang, W., and Peng, M. (2013). A neighbourhood-constrained k-means approach to classify very high spatial resolution hyperspectral imagery. *Remote Sensing Letters*, 4(2):161–170.

[Zhang et al., 2018] Zhang, H., Bruzzone, L., Shi, W., Hao, M., and Wang, Y. (2018). Enhanced spatially constrained remotely sensed imagery classification using a fuzzy local double neighborhood information c-means clustering algorithm. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(8):2896–2910.

[Zhang et al., 2017] Zhang, H., Li, Y., Zhang, Y., and Shen, Q. (2017). Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network. *Remote Sensing Letters*, 8(5):438–447.

[Zhang et al., 2019a] Zhang, H., Long, Y., Guan, Y., and Shao, L. (2019a). Triple verification network for generalized zero-shot learning. *IEEE Transactions on Image Processing*, 28(1):506–517.

[Zhang et al., 2016a] Zhang, H., Zhai, H., Zhang, L., and Li, P. (2016a). Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(6):3672–3684.

[Zhang et al., 2016b] Zhang, L., Zhang, L., and Du, B. (2016b). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):22–40.

[Zhang et al., 2022] Zhang, S., Xu, M., Zhou, J., and Jia, S. (2022). Unsupervised spatial-spectral cnn-based feature learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17.

[Zhang et al., 2019b] Zhang, X., Han, L., Dong, Y., Shi, Y., Huang, W., Han, L., González-Moreno, P., Ma, H., Ye, H., and Sobeih, T. (2019b). A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral uav images. *Remote Sensing*, 11(13).

[Zhang and Saligrama, 2015] Zhang, Z. and Saligrama, V. (2015). Zero-shot learning via semantic similarity embedding. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 4166–4174.

[Zhang and Saligrama, 2016a] Zhang, Z. and Saligrama, V. (2016a). Learning joint feature adaptation for zero-shot recognition. *CoRR*, abs/1611.07593.

[Zhang and Saligrama, 2016b] Zhang, Z. and Saligrama, V. (2016b). Zero-shot learning via joint latent similarity embedding. pages 6034–6042.

[Zhang and Saligrama, 2016c] Zhang, Z. and Saligrama, V. (2016c). Zero-shot recognition via structured prediction. *European Conference on Computer Vision (ECCV)*, page 533–548.

[Zhao et al., 2017a] Zhao, B., Gao, L., Liao, W., and Zhang, B. (2017a). A new kernel method for hyperspectral image feature extraction. *Geo-spatial Information Science*, 20(4):309–318.

[Zhao et al., 2017b] Zhao, B., Sun, X., Yao, Y., and Wang, Y. (2017b). Zero-shot learning via shared-reconstruction-graph pursuit. *CoRR*, abs/1711.07302.

[Zhao et al., 2017c] Zhao, B., Wu, B., Wu, T., and Wang, Y. (2017c). Zero-shot learning posed as a missing data problem. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 2616–2622.

[Zhao and Du, 2016] Zhao, W. and Du, S. (2016). Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8):4544–4554.

[Zhong et al., 2016] Zhong, P., Gong, Z. Q., and Schönlieb, C. (2016). A diversified deep belief network for hyperspectral image classification. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41B7:443–449.

[Zhong et al., 2008] Zhong, P., Zhang, P., and Wang, R. (2008). Dynamic learning of smlr for feature selection and classification of hyperspectral data. *IEEE Geoscience and Remote Sensing Letters*, 5(2):280–284.

[Zhong and Zhang, 2012] Zhong, Y. and Zhang, L. (2012). An adaptive artificial immune network for supervised classification of multi-hyperspectral remote sensing imagery. *IEEE Transactions Geoscience Remote Sensing*, 50(3):894–909.

[Zhong et al., 2011] Zhong, Y., Zhang, L., and Gong, W. (2011). Unsupervised remote sensing image classification using an artificial immune network. *International Journal of Remote Sensing*, 32(19):5461–5483.

[Zhong et al., 2006] Zhong, Y., Zhang, L., Huang, B., and Li, P. (2006). An unsupervised artificial immune classifier for multi/hyperspectral remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 44(2):420–431.

[Zhong et al., 2013] Zhong, Y., Zhang, S., and Zhang, L. (2013). Automatic fuzzy clustering based on adaptive multi-objective differential evolution for remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(5):2290–2301.

[Zhu et al., 2019a] Zhu, P., Wang, H., and Saligrama, V. (2019a). Generalized zero-shot recognition based on visually semantic embedding. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2990–2998.

[Zhu et al., 2017] Zhu, W., Chayes, V., Tiard, A., Sanchez, S., Dahlberg, D., Bertozzi, A. L., Osher, S., Zosso, D., and Kuang, D. (2017). Unsupervised classification in hyperspectral imagery with nonlocal total variation and primal-dual hybrid gradient algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5):2786–2798.

[Zhu et al., 2019b] Zhu, Y., Xie, J., Tang, Z., Peng, X., and Elgammal, A. (2019b). Semantic-guided multi-attention localization for zero-shot learning. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, page 14943–14953.

[Zielinski et al., 2009] Zielinski, O., Busch, J. A., Cembella, A. D., Daly, K. L., Engelbrektsson, J., Hannides, A. K., and Schmidt, H. (2009). Detecting marine hazardous substances and organisms: sensors for pollutants, toxins, and pathogens. *Ocean Science*, 5(3):329–349.

[Zielinski et al., 2006] Zielinski, O., Hengstermann, T., and Robbe, N. (2006). *Detection of oil spills by airborne sensors*, pages 255–271. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Özdemir et al., 2014] Özdemir, A. O. B., Gedik, B. E., and Çetin, C. Y. Y. (2014). Hyperspectral classification using stacked autoencoders with deep learning. In *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4.