FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Abnormalities Detection on Videos of Endoscopic Capsules (VEC)

Maria Teresa Gomes Silva Valério

DISSERTATION FOR THE DEGREE OF MASTER IN BIOENGINEERING

MSC IN BIOENGINEERING - BIOMEDICAL ENGINEERING

Supervisor: António Manuel Trigueiros da Silva Cunha

Co-supervisor: Hélder Filipe Pinto de Oliveira

July 24, 2019

Abnormalities Detection on Videos of Endoscopic Capsules (VEC)

Maria Teresa Gomes Silva Valério

MSc in Bioengineering - Biomedical Engineering

July 24, 2019

Resumo

A endoscopia capsular sem fio é uma técnica relativamente recente, usada em imagiologia do trato gastrointestinal. Ao contrário das abordagens tradicionais (endoscopia clássica, ultrassonografia e tomografia computadorizada), permite a visualização indolor de todo o trato gastrointestinal, incluindo o intestino delgado (uma região de difícil acesso). É utilizada uma cápsula pequena, leve, semelhante a um comprimido e não invasiva que é engolida pelo paciente e impulsionada por movimentos peristálticos, enquanto transmite imagens capturadas para um dispositivo externo. As cápsulas endoscópicas podem gravar durante cerca de 8 horas, produzindo cerca de $60,000$ imagens. A análise destes dados por um especialista é muito enfadonha e propensa a erros, incentivando assim o desenvolvimento de sistemas que analisem automaticamente estes dados. Tendo isto em consideração, o objetivo desta dissertação consistiu no desenvolvimento de um método para deteção e segmentação automáticas de lesões em vídeos de cápsulas endoscópicas. O método desenvolvido deverá classificar um *frame* relativamente à presença e tipo de lesão, em uma de três possíveis classes (normal, vascular ou inflamatória), além de providenciar uma localização ao nível do *pixel* da lesão presente na imagem.

Para responder ao problema da deteção de lesões, foi seguida um método baseado em *transfer learning*, enquanto que para a segmentação a abordagem implementada incluiu o uso de redes neuronais convolucionais direcionadas para esse propósito. Em ambos os casos, um passo de pré-processamento para melhoramento das imagens foi incluído no método. Quanto à tarefa de classificação de imagens, o modelo melhor sucedido foi o VGG16 com *batch normalisation*, que obteve 0.95 de *precision* e *recall*, e 0.90 de *area under curve*. Considerando a segmentação de lesões vasculares, foram obtidos um *Jaccard* de 0.821 e um *Dice* de 0.902 por parte de uma *U-Net*, enquanto que a segmentação de lesões inflamatórias alcançou apenas um *Jaccard* de 0.602 e um *Dice* de 0.751, conseguidos por uma *generative adversarial network*. O passo de pré-processamento implementado não mostrou ser vantajoso nem para a classificação nem para a segmentação de lesões vasculares, mas revelou ser crucial para a segmentação de lesões inflamatórias, onde o algoritmo de *multi-scale retinex with colour restoration* demonstrou ser bastante eficiente.

Tendo tudo isto em consideração, os resultados obtidos podem ser vistos como uma valiosa contribuição para a área, especialmente tendo em consideração que ainda não foram realizados trabalhos de classificação semelhantes ao aqui proposto, nem de segmentação de lesões inflamatórias. Apesar do sucesso alcançado em ambas as tarefas, algum trabalho pode ainda ser feito futuramente, de modo a melhorar os resultados obtidos.

Abstract

Wireless capsule endoscopy is a relatively novel technique used for imaging of the gastrointestinal tract. Unlike traditional approaches (classic endoscopy, ultrasound and computed tomography scan), it allows painless visualisation of the whole of the gastrointestinal tract, including the small bowel (a region of difficult access). A small, light, pill-like and non-invasive wireless capsule endoscope is swallowed by the patient and propelled by peristaltic movements, while wireless transmitting captured images to an external device. Endoscopic capsules can record for about 8 hours, producing around $60,000$ images. The analysis of these images by an expert is very tedious and prone to errors, thus encouraging the development of systems that automatically analyse this data. Taking this into consideration, this dissertation aimed for the development of a method for the automatic detection and segmentation of lesions in videos of endoscopic capsules (VEC). The developed method should classify a frame, concerning the presence and type of lesion, into one of three possible classes (normal, vascular or inflammatory), and also provide a pixel-wise localisation of the lesions within the VEC image.

To tackle the lesion detection problem a transfer learning based method was followed, while for the segmentation of lesions the implemented approach included the use of convolutional neural networks architectures for semantic segmentation. For both cases, a pre-processing step to enhance the VEC images was included in the framework. Regarding the image classification task, the best-succeeded model was the VGG16 with batch normalisation, that achieved 0.95 of precision and recall, and 0.90 of area under curve. Concerning the segmentation of vascular lesions, a Jaccard of 0.821 and a Dice of 0.902 where obtained by a U-Net model, while the segmentation of inflammatory lesions only reached a Jaccard of 0.602 and a Dice of 0.751, obtained by a generative adversarial network. The implemented pre-processing step was not advantageous for either the classification or vascular lesion segmentation tasks, but was crucial for the segmentation of inflammatory lesions, where the automated multi-scale retinex with colour restoration algorithm proved to be very effective.

All things considered, the obtained results can be recognised as a valuable contribution to the area, especially taking into consideration that no work addressing a classification problem like the one here proposed nor the segmentation of inflammatory lesions has been done. Despite the obtained success in both tasks, some work can still be performed hereafter in order to improve the achieved results.

Agradecimentos

Em primeiro lugar gostaria de deixar umas palavras de agradecimento aos Professores António Cunha e Hélder Oliveira por toda a disponibilidade e orientação ao longo deste ano. Agradeço também à Dr.^{a} Marta Salgado pela colaboração e ajuda durante a realização desta dissertação.

Quero deixar também um enorme obrigada à minha família por toda a exigência, confiança e apoio dos últimos 22 anos. Para vocês umas meras palavras nunca serão suficientes.

Por último, resta-me agradecer a todos aqueles que mais de perto me acompanharam e mais experiências partilharam comigo ao longo destes 5 anos que agora chegam ao fim. Aos que em setembro de 2014 também pisavam pela primeira vez o solo desta enorme faculdade à qual agora chamamos casa, sem sequer imaginar o que aí vinha. Aos que já por cá andavam, me fizeram crescer e ensinaram valores que não poderia ter aprendido em qualquer outro lado. Aos que vieram depois e com os quais também tive a oportunidade de aprender. Obrigada a todos vocês pelas longas noites de trabalho e pelas de boémia, pelas palavras de carinho mas também por aquelas que mais custaram ouvir, pelas gargalhadas e pelas lágrimas, pelas mitocôndrias e pelos complexos. Pelas melhores amizades que podia levar para a vida. Obrigada Metal&Bio por me dares a conhecer as melhores pessoas com quem poderia ter partilhado estes anos. Vemo-nos por aí.

Teresa Valério

"I tell young people: Do not think of yourself, think of others.
Think of the future that awaits you, think about what you can do and do not fear anything."

Rita Levi-Montalcini

Contents

List of Figures

List of Tables

Abbreviations

AUC	Area Under Curve
BIC	Bayesian Information Criterion
BoW	Bag of Words
CAD	Computer-Aided Diagnosis
CBI	Colour Balance Index
CLAHE	Contrast Limited Adaptive Histogram Equalisation
CNN	Convolutional Neural Network
CRF	Colour Restoration Function
CT	Computed Tomography
D	Discriminative model
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
EM	Expectation Maximisation
FN	False Negative
FP	False Positive
G	Generative model
GAN	Generative Adversarial Network
GI	Gastrointestinal
HE	Histogram Equalisation
IoU	Intersection over Union
LCP	Lower Clipping Point
MAP	Maximum a Posteriori
ML	Machine Learning
MLE	Maximum Likelihood Estimation
MRF	Markov Random Fields
MSR	Multi-Ccale Retinex
MSRCR	Multi-Scale Retinex with Colour Restoration
NN	Neural Network
OGIB	Obscure Gastrointestinal Bleeding
PM	Perona-Malik
ReLU	Rectified Linear Unit
RBF	Radial Basis Function
RF	Radio Frequency
ROC	Recever Operating Characteristic
ROI	Region of Interest
RoR	Residual Network of Residual Network
SIFT	Scale-Invariant Feature Transform
SSR	Single Scale Retinex
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
TP	True Positive
UCP	Upper Clipping Point
WCE	Wireless Capsule Endoscopy
SB	Small Bowel
VEC	Video of Endoscopic Capsule

Chapter 1

Introduction

1.1 Context

Traditional approaches for diagnosis of gastrointestinal (GI) tract diseases include endoscopy (esophagogastroduodenoscopy and colonoscopy), ultrasound, and computed tomography (CT) scan [35].

Endoscopy is the standardised method for GI tract visual analysis. This safe and effective method provides real-time images of the patient's GI tract, allowing the gastroenterologist to gather information for the diagnosis of possible lesions. The main disadvantages of this method include the discomfort that this invasive intervention causes to the patient, possibility of organ perforation and consequent haemorrhage, and impossibility in assessing the small bowel (SB) [7, 29].

Ultrasound gives GI tract images with high resolution of soft tissues (which can be very helpful for the detection of GI tract inflammation) without radiation exposure, provides information about intestinal wall and surroundings, and can also be used for intraluminal imaging. On the other hand, image quality may be lowered by intestinal gas, identification of artefacts can be hard, and visualisation of the whole intestine is not always possible [7].

CT provides information on the whole extension of the GI tract, including the SB and colon. This technique allows identification of tumours, inflammatory diseases and its complications, as well as other lesions. It is a non-invasive method unless contrast is used, as well as painless and relatively fast. This technique has the disadvantage of using radiation (X-Ray) to obtain the images of the organ and tissues, and the possibility of mistake intestinal loops or concentrations of fluid in the bowel as soft-tissue or tumours, leading to false positives or false negatives examinations [23].

Wireless capsule endoscopy (WCE) is a relatively recent technique (used since 2001, date of FDA approval) that allows painless endoscopic imaging of the whole of the GI tract, including the small bowel. This method uses a wireless capsule endoscope that is a small ($11 \times 26\,\text{mm}$), light (3.7 g), pill-like and non-invasive device. These characteristics allow the capsule to be swallowed by the patient and propelled by peristaltic movements. Along its route, the capsule transmits images through radiofrequency to a portable device attached to the patient's body [33]. Frame rate ranges from 2 to 6 frames per second in PillCam®SB3, but can go up to 35 frames per second in the PillCam$^{®}$COLON 2 capsule [40, 41]. When the capsule moves rapidly, the image capture rate is automatically increased (adaptive frame rate technology) to optimise a more complete and detailed tissue coverage [6, 41].

The device can record for about 8 hours, producing nearly $60,000$ images. The manual analysis of all this data by an expert gastroenterologist is very tedious, time-consuming ($40-60$ minutes) and susceptible to human error, since identifying lesions in this type of image is a very challenging task. This happens due to both the quality of the acquired images and the many different ways the same type of lesion can be presented [33].

Computer-aided diagnosis (CAD) tools are an ever-growing technique applied in many medical environments, that can help the doctor decide on diagnosis, treatment approach, and others. Regarding capsule endoscopy, some work for automatic lesion detection and segmentation has already been done; however, not with reliable enough performance to be suitable for medical use.

Therefore, there is a clear need to develop CAD based approaches for the automatic analysis of videos of endoscopic capsules (VEC), that allow the detection and segmentation of lesions in the GI tract, with comparable performances to a gastroenterologist.

1.2 Motivation

When performing an endoscopy, one of the main goals is to identify possible diseases and/or lesions along the GI tract. Those lesions can be organised into four main types: vascular, inflammatory, lymphangiectasias and polyps. This work will focus on the first two types of lesions. Vascular lesions are seriously taken into consideration by doctors, as they can be in the origin of many types of GI bleeding. Angiodysplasia (once called angiectasia), dieulafoy lesion, erythematous patch, red spot, phlebectasia and diminute angiectasia (also known as telangiectasia) are some of the vascular lesions that can be identified in the GI tract. All of them have specific characteristics that allow differentiation; however, this distinction is, sometimes, very hard to do, due to the difficulty associated to the visualisation of endoscopic images [54]. Inflammatory lesions also require some attention in the medical area, given that they can be linked to some conditions like ulcerative colitis or Crohn's disease, and can be of many different types as well: aphthae, ulcer, stenosis, mucosal erythema, mucosal cobblestone and mucosal edema [10].

Some of these lesions can be precancerous, which means that, over time, they are very likely to turn into GI cancer. For instance, inflammation, bleeding and some particular types of polyps can be in the origin of colon cancer, the third most common type of cancer in most countries [13, 58]. However, if these precancerous lesions are found in an early stage of development, the probability of survival of patients with this disease can be increased [59]. This fact reinforces the need for developing new methods that allow easy and early detection of gastrointestinal lesions.

WCE is now currently used for exploration of the small bowel, being considered a useful and safe method for the examination that, unlike most traditional procedures, allows complete examination of that portion of the GI tract. Capsule endoscopy has been used on analysis of SB diseases like obscure gastrointestinal bleeding (OGIB) and inflammatory bowel disease, as well as for esophageal and colonic diseases [33]. However, even for an expert in the area, the analysis of a VEC can take up to 60 minutes and is prone to human error, thus the urgent need of developing automatic systems for the analysis of this data [33]. Both lesion detection and segmentation should be addressed, in order to focus the doctor's attention in the less amount of data possible (the relevant one, only). Detecting lesions on the frames allows a shortened view of WCE videos that can drastically reduce the number of images that the doctor has to analyse, and lesion localisation in the frame directs the expert's attention for the relevant region of the picture, for a better analysis of the abnormality.

Based on published work on the area, it is possible to verify that many machine learning (ML) approaches have already been implemented for both lesion detection and segmentation in WCE frames. However, it is clear that vascular lesions have been much more studied than inflammatory ones, as most of the described work is related to the first ones.

Regarding lesion detection, image pre-processing followed by feature extraction and automatic classification has been implemented in many studies. Moreover, some recent works using CNN (Convolutional Neural Networks) and transfer learning have shown promising results [36]. Residual Networks have been applied in ImageNet and CIFAR-10 dataset for image classification, revealing very good results [14].

For the lesion segmentation problem, strategies like Expectation Maximisation Clustering, Maximum a Posteriori approach with Markov Random Fields and others have been described; however, semantic segmentation using CNN and GAN (Generative Adversarial Networks) are the ones that show the best results [50, 57].

Existing solutions for lesion detection and segmentation on VEC frames lack on suitable performance for acceptance in a clinical use, besides not being inclusive of all GI tract existing lesions. Taking this into consideration, it is of upmost importance the development of new techniques to fulfil these needs. Given the promising presented strategies, yet still with a margin for improvement, the main goal of the dissertation is to develop accurate methods for lesion detection and segmentation in frames of VEC, using ML approaches.

1.3 Objectives

Given the mentioned limitations offered by existing solutions on VEC frames classification and segmentation, the main goal of the dissertation is to develop algorithms that can accurately detect and segment lesions in WCE images. To do so, ML techniques, particularly deep learning, will be implemented. Both developed methods should achieve performance such that future clinical utilisation is possible.

1.4 Contributions

Achieved contributions by the present thesis for the detection and segmentation of lesions in VEC frames include:

- Creation of an algorithm for multi-class classification of VEC frames;
- Development of an algorithm for segmentation of vascular and inflammatory lesions in VEC frames;
- Assessment of the influence of image enhancement techniques in the classification and segmentation of lesions in VEC frames;
- Comparison of several neural network (NN) architectures commonly used for classification and segmentation tasks, applied in VEC images.

This thesis also resulted in the submission of a conference paper, accepted for publication:

- M. T. Valério, S. Gomes, M. Salgado, H. P. Oliveira, and A. Cunha. Lesions Multiclass Classification in Endoscopic Capsule Frames, 2019 (Submitted)

1.5 Document Structure

The remaining document of the present dissertation is composed by 4 more Chapters, organised as follows. Chapter 2 presents an overview of the anatomy of the GI tract, as well as some physiological aspects required to understand some concepts presented hereafter, followed by a description of the most common lesions of the GI tract. Then, some methods and techniques currently used for VEC images classification and segmentation tasks are introduced. This includes some background on NN architectures, methods that implement some of these networks, as well as enhancement techniques applied to VEC images.

Afterwards, Chapter 3 presents the explanation of the framework implemented to address the issues at hand, including the detailed exposition of each step of the pipeline.

The obtained results of the implemented framework are presented on Chapter 4. The outcome of each step of the pipeline is exposed, followed by a discussion of those results.

Finally, the main conclusions drawn from this thesis are presented in Chaper 5, as well as some suggestions for future improvement of the developed algorithms.

Chapter 2

Literature Review

In this Section, some background information for a better understanding of the problem at hand will be presented, followed by a review of some strategies applied to similar tasks. First, in Section 2.1, an overview of the anatomy and physiology of the GI tract is presented. Then, in Section 2.2 a characterisation of GI lesions can be found, followed by a description of endoscopic capsules and their operation mode in Section 2.3. Then follows the literature review on methods already implemented to solve problems similar to the ones at hand, that includes strategies for enhancement of VEC images (Section 2.4) and ML techniques for image classification and segmentation (Section 2.5). In this last case, some of the most used NN are introduced (Section 2.5.1) for a better understanding of the afterwards presented approaches for VEC images classification and segmentation (Section 2.5.2). Finally, in Section 2.6, a description of available datasets suitable for these tasks is presented.

2.1 Gastrointestinal Tract Anatomy and Physiology

Given the key role of the digestive system in maintaining life and the disruption of health when it malfunctions, it is of critical importance to be aware of its structure and function. Besides its acknowledge importance in nutrient digestion and absorption, the GI tract is also considered a very important immunological organ in the human body responsible for the protection against exogenous pathogens. Due to this exposure to the external environment, the GI tract is susceptible to the residence of organisms that can be the source of a wide variety of disorders [12].

In an adult human, the GI tract can measure up to 9 meters long, spanning all the way from the mouth to the anus. Two large groups of organs can be considered when concerning the digestive system: the GI tract organs and the accessory digestive organs (Figure 2.1). GI tract organs include the buccal cavity, pharynx, oesophagus, stomach, small intestine and large intestine, while accessory digestive organs include teeth, tongue, salivary glands, liver, gallbladder and pancreas [12].

The interior of the GI tract is surrounded by several layers (mucosa, submucosa, muscularis and serosa tunics), each one responsible for one specific function:

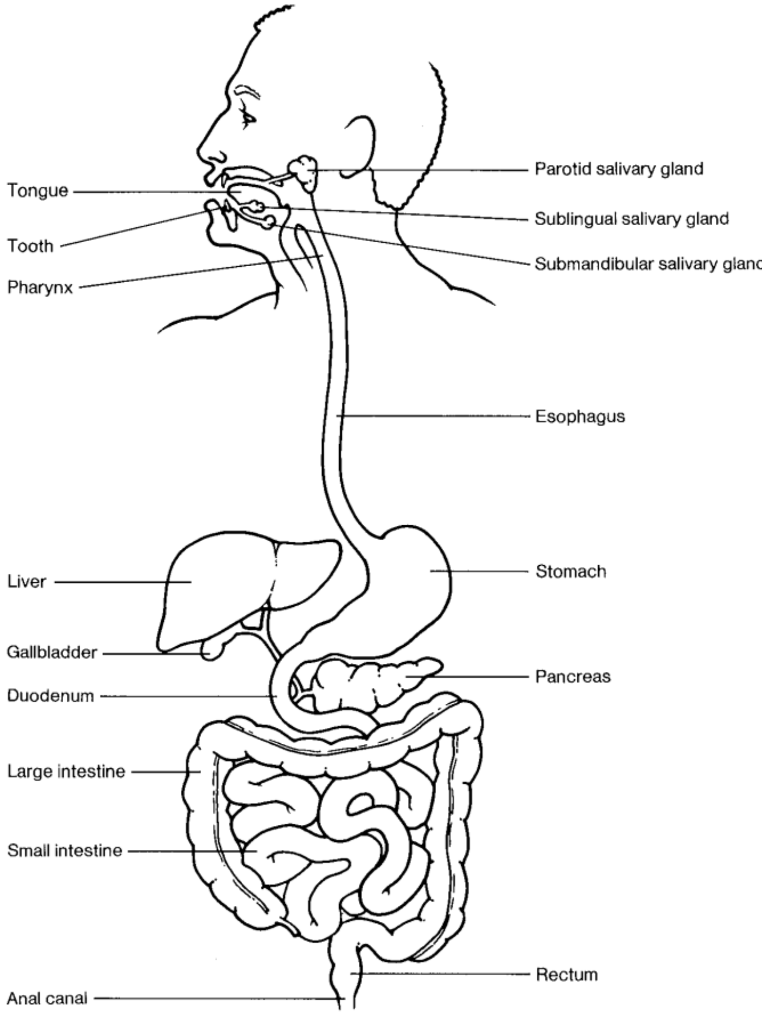


Figure 2.1: Scheme of the gastrointestinal tract.

- Mucosa is the layer that surrounds the lumen of the GI tract and has both absorptive and secretory functions. It includes a thin layer of smooth muscle that allows the folding of some portions of the intestine, leading to a great increase in the absorptive surface area. Specialisations of the mucosa include the ability for distension, absorption, water extraction and secretion of digestive enzymes, depending on the region of the GI tract [12, 62];

- Submucosa is the layer that immediately follows the mucosa. It is a crucial participant in the digestive process, given that it is not completed until the submucosa receives the food nutrients absorbed by the mucosa through capillary networks in the small intestine [12, 62];

- Contractions and peristaltic movements along the GI tract are possible thanks to the muscularis tunic. The contraction of this layer allows food to move peristaltically through the path, while being physically pulverised with digestive chemicals [12];

- The serosa tunic is a binding and protective layer that completes the wall of the GI tract [12].

The small intestine constitutes about 3 to 5 meters of the GI tract and is the most crucial component of the digestive system, being responsible for fluid and electrolyte haemostasis, digestion and absorption of nutrients, immunoregulation, and secretion of hormones. It takes part in the breakdown and absorption of important nutrients that allow the functioning of the human body at its peak performance. The absorptive capacity of this organ is greatly enhanced by the plicae circulares, villi and microvilli, that result in a total estimated surface area of around $30m^2$ [4, 62]. The small intestine can be divided into three sections: the duodenum is the first and shortest portion (measuring only about 25 cm long), extending from the stomach to the duodenojejunal junction, the beginning of the SB (constituted by the jejunum and the ileum). The jejunum connects the duodenum and the ileum, but the transition between these two portions of the small intestine is not very clear. Although there is no clear boundary separating both parts, a gradual transition in morphology can be identified. Finally, the ileum is the portion that connects to the large intestine at the ileocecal valve [62].

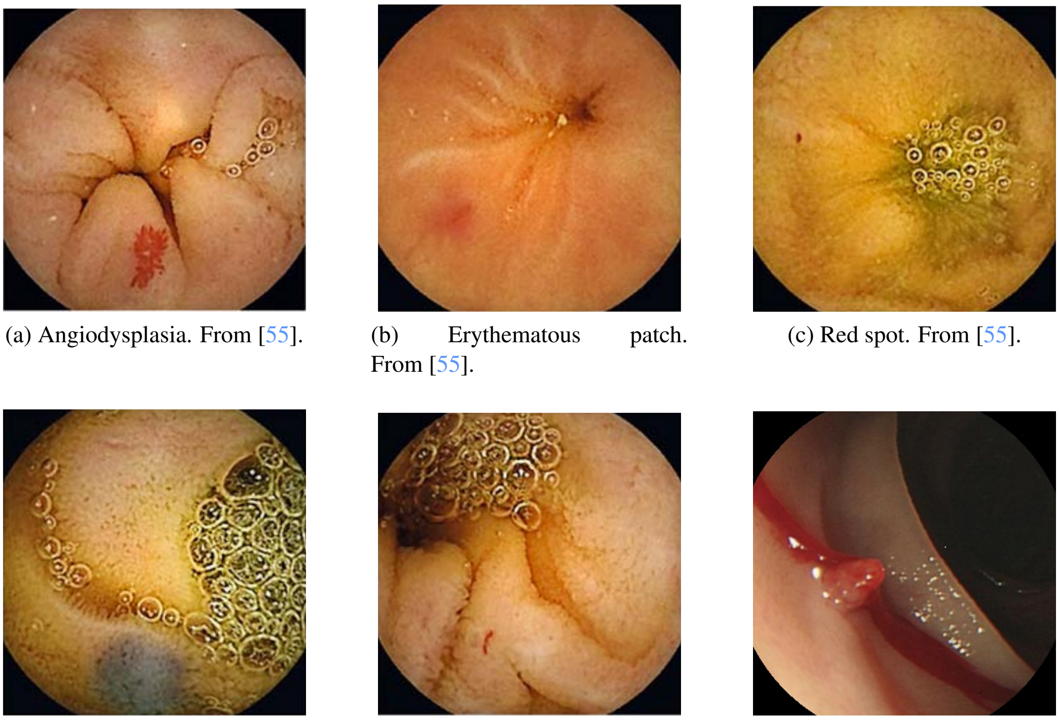
2.2 Gastrointestinal Tract Lesions Characterisation

It is of highest priority the study of diseases and abnormalities that can be found in the GI tract, given that many of them can be in the origin of serious health conditions. Many lesions can be found in the GI tract, being generally divided into four main types: vascular lesions, inflammatory lesions, lymphangiectasias and polyps. Although efforts have been made during the first years of clinical use of capsule endoscopy imaging, no clear definitions or full descriptions of the most common lesions have been widely accepted yet. In the medical literature, there are only a few semantic descriptions of angiodysplasia (one of the most common SB lesion), and these become even rarer when concerning other types of abnormalities [55]. Taking this into consideration, a brief description of some types of GI tract lesions will be presented, in order to provide some background knowledge regarding this issue.

Vascular lesions

From all four main types of GI abnormalities, vascular lesions are the ones that are mostly described in the literature. However, due to the inconsistency among descriptions, Romain Leenhardt et al. [55] conducted a study that established nomenclature and description for some of the most frequent vascular lesions, based on levels of agreement of several experts in the area. Some of the most important vascular lesions to take into consideration include angiodysplasias, erythematous patches, red spots, phlebectasias, diminute angiectasias and dieulafoy's lesions, and are described as follows:

- Angiodysplasia (Figure 2.2a), previously designated as angiectasia, is one of the most common SB vascular lesions, and also the leading cause of OGIB. These abnormalities are usually flat and clearly demarcated, presenting a bright-red colouration. They consist of tortuous and clustered capillary dilatations within the mucosal layer (surrounded by intestinal villi). Angiodysplasias size can vary from some mm to a few cm [55];



(a) Angiodysplasia. From [55].

(b) Erythematous patch. From [55].

(c) Red spot. From [55].

(d) Phlebectasia. From [55].

(e) Diminute angiectasia. From [55].

(f) Dieulafoy's lesion. From [22].

Figure 2.2: Gastrointestinal tract vascular lesions.

- Erythematous patch (Figure 2.2b) is also a flat and reddish vascular abnormality that usually measures only a few mm. However, unlike angiodysplasias, these lesions do not present any vessel appearance within the mucosal layer [55];

- Red spots (Figure 2.2c) are, as suggested by the nomenclature, a minuscule (less than 1 mm) and punctuate lesion with a bright-red area, also without linear or vessel appearance within the mucosal layer [55];

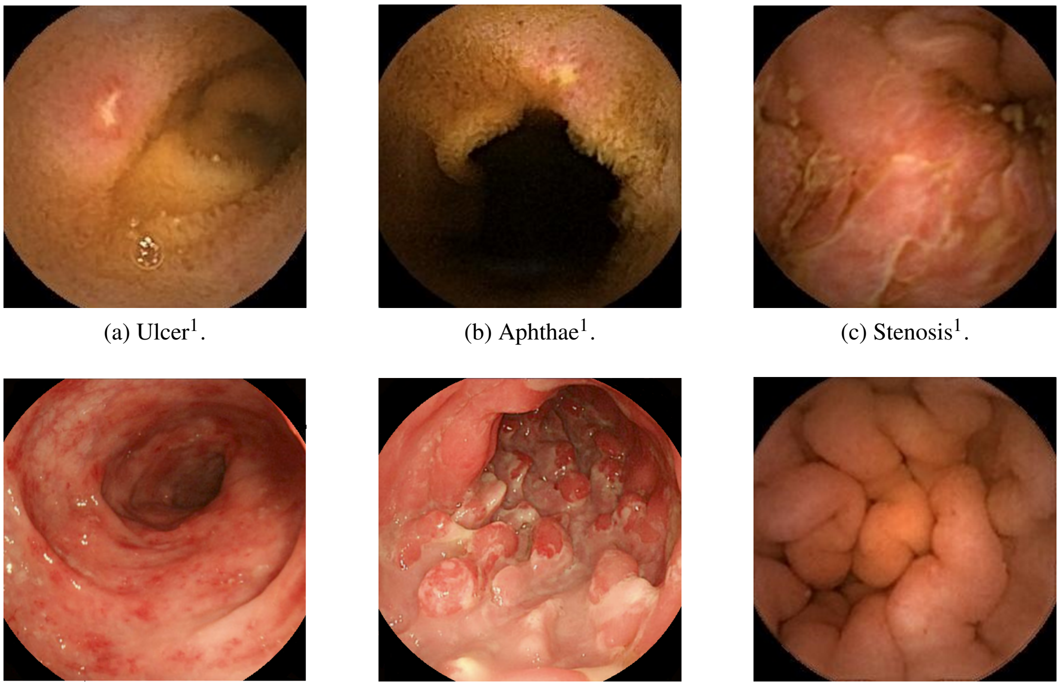
- Phlebectasias (Figure 2.2d) are flat to slightly-elevated vascular lesions that, unlike all the previous ones, present a bluish colouration derived from the venous dilatation running below the mucosa (covered by intestinal villi) [55];

- Diminutive angiectasia (Figure 2.2e), also known as telangiectasia, is more difficult to characterise, but some experts describe it as clearly demarcated, linear, bright-red lesion, consisting of tiny non-clustered capillary dilatations within the mucosal layer [55];

- Dieulafoy's lesion (Figure 2.2f) is one of the most rare causes of OGIB, however it can be potentially life-threatening. Unlike normal arteries of the GI tract that narrow progressively as they traverse the wall of the structure, dieulafoy's lesion arteries present an abnormally large diameter (1-3 mm) that is maintained constant. These abnormalities protrude through a small mucosal defect varying from 2-5 mm, appearing as reddish-brown protruding spots that have fibrinoid necrosis at its base [54, 1].

Inflammatory lesions

With respect to inflammatory abnormalities, ulcer, aphthae, stenosis, mucosal erythema, mucosal cobblestone and mucosal edema are the most common ones:



(a) Ulcer[1].

(b) Aphthae[1].

(c) Stenosis[1].

(d) Mucosal erythema. From [3]

(e) Mucosal cobblestone. From [3].

(f) Mucosal edema[1] .

Figure 2.3: Gastrointestinal tract inflammatory lesions.

- Ulcer (Figure 2.3a) is the most common cause of hospitalisation for upper GI tract bleeding and represents openings in the mucosa of any part of the GI tract that can occur with chronic inflammation [5];
- Aphthous (Figure 2.3b), sometimes also called aphthous ulcers, have been described as ulcers with diameter up to 3 mm, with a distinct white base and surrounding congestions or haemorrhage on a normal mucosal background. These lesions usually present red raised margins and central areas of yellow-white exudate with normal intervening mucosa [38].
- Stenosis (Figure 2.3c) is the name given to the narrowing or even obstruction of the intestinal lumen. This can be caused by many factors and when it happens due to thickening of the intestinal walls caused by inflammation of the mucosa, it can be designated inflammatory stenosis [39].

[1] Image from *PillCam* RAPID™ READER Atlas

- Mucosal erythema (Figure 2.3d) is the name given to the red colouration that the inner of the gastrointestinal tract acquires when the membrane incorporates dilated veins [3];
- Mucosal cobblestone (Figure 2.3e) is characterised by longitudinal and circumferential fissures that separate regions of the mucosa, giving it an appearance resembling cobblestones. This can result from the thickening and swallowing of the mucosa due to the intermittent pattern of diseased and healthy tissues [3];
- Mucosal edema (Figure 2.3f) is defined as thick and swollen mucosa, which is pooled with fluid, leading to thickened haustral folds. In severe edema, the mucosal narrowing can be found [3] .

Lymphangiectasias

Intestinal lymphangiectasias (Figure 2.4) are lesions characterised by dilated lymphatic capillary of the small intestine. These lesions are usually identified by the creamy yellow of jejunal villi, corresponding to marked dilation of the lymphatics within the intestinal mucosa [67].

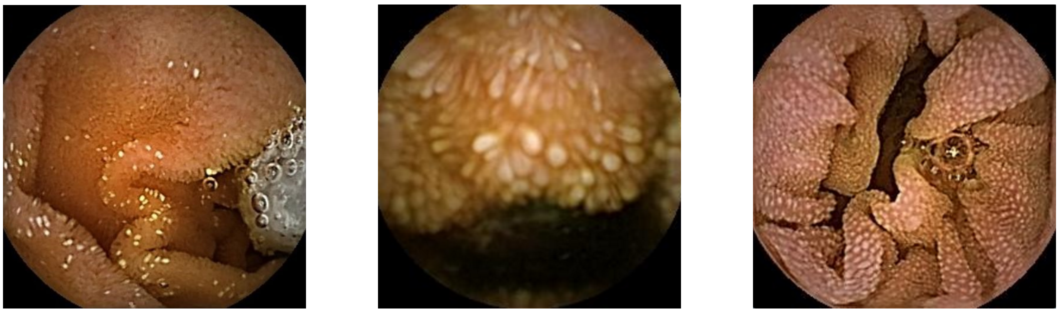


Figure 2.4: Gastrointestinal tract lymphangiectasias[2].

Polyps

Polyps (Figure 2.5) are luminal lesions projected above the plane of the mucosal surface, that can become inflamed and eroded. The appearance of these lesions ranges from slightly raised plaques to soft multilobed nodules to, more rarely, broad-based or sessile lesions [68].

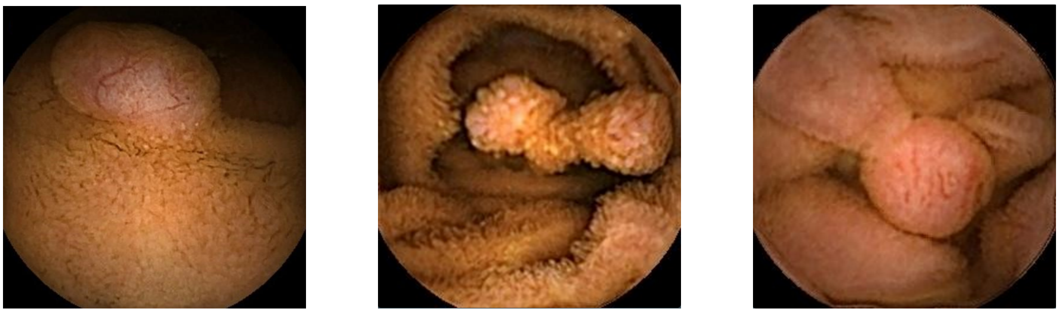


Figure 2.5: Gastrointestinal tract polyps[2].

[2] Images from *PillCam* RAPID™ READER Atlas

A quick comparison between the presented GI lesions shows some clear differences. While the majority of vascular and inflammatory lesions are flat, lymphangiectasias and polyps are projected above the plane of the intestine wall. These last two abnormalities have also very characteristic morphology, making them easier to identify and distinguish from others. Between vascular and inflammatory lesions, a clear difference stands out: the colour. While vascular lesions are darker than the background areas, presenting reddish colouration, most common inflammatory lesions are usually characterised by lighter colours, ranging from white to light yellow. However, these last ones can also include some reddish areas, which may erroneously lead one to relate them to vascular lesions.

2.3 Endoscopic Capsules

WCE was introduced in Iddan et al. [19] as a new form of endoscopy that came to revolutionise the way gastroenterologist view the GI tract. The fact that this method allows visualisation of the whole of the GI tract without pain, sedation, or air inflation (like previously used methods did), lead to fast acceptance of the device in the medical environment [35].

Swallowable electronic capsules have been used since the 1950s with the purpose of monitoring GI physiological parameters [19]. This was the starting point for the development of endoscopic capsules (Figure 2.6).

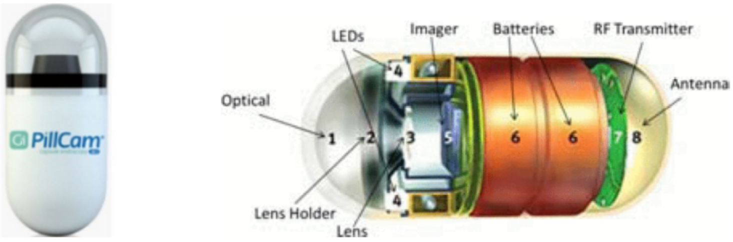


Figure 2.6: (left) Endoscopic capsule (PillCam®SB3). From [65]; (right) Components of the endoscopic capsule. From [53].

This small device (only 11 × 26 mm) weights less than 4 g and is made of a biocompatible and resistant to digestive fluids plastic. It uses a lens of short focal length and a miniature video camera to capture images of the GI tract, that are transmitted through radio frequency (RF) to a device attached to the patient's body [65]. The endoscopic capsule includes CMOS (complementary metal oxide silicon) image sensors that can capture images with comparable quality to those acquired with charge-coupled devices, ASIC (application-specific integrated circuit) RF transmitter that sends the recorded video to the external device, and white LEDs (light emitting diode) that illuminate the GI tract [19]. Synchronous switching of these three components minimises the power consumption of the two mercury-free silver oxide batteries so that the capsule can record from 8 to 12 hours.

Capsule features of the PillCam®SB3 include a $156-170°$ field of view, high resolution and sharpness, a more homogeneous light exposure in comparison to first capsules, and a depth of view of at least $20-30\,\mathrm{mm}$ [65]. A scheme of the endoscopic capsule components can be seen in Figure 2.6.

2.4 VEC Images Enhancement

Quality of images captured by endoscopic capsules is not ideal, due to several reasons. Owing to the need for saving the capsule battery data needs to be compressed, leading to not very clear images. Besides that, image resolution ranges from only 256×256 pixels to 576×576 pixels, depending on the capsule version, due to the power limitation of the device. Furthermore, the quality of these images can be even more degenerated by low illumination and complex circumstances found along the GI tract, which can result in dark frames where the contents are barely distinguishable. Finally, the camera range of focus is short, which can cause vignetting and effects of depth that are not desirable [35].

Given this, it is clear the need for developing approaches for image enhancement to improve the visibility of darker regions and remove noise, without loss of important information. Three proposed architectures to address this issue are presented next.

2.4.1 Adaptive Contrast Diffusion

Li and Meng [35] proposed a technique that improves the anisotropic diffusion suggested by Perona and Malik [48]. Image smoothing with a Gaussian or averaging filter is the equivalent to the application of the heat diffusion equation (Equation 2.1) [35].

$$\frac{\partial I(x,y,t)}{\partial t} = g\Delta I(x,y,t) \tag{2.1}$$

In the above equation, Δ and g are the Laplacian operator and the diffusion conductance constant, respectively, and $\Delta I = div(\triangledown I(x,y,t))$. This processing does not respect the natural boundaries of the objects, resulting in an overall blurred image [48]. To overcome this issue, Perona and Malik [48] proposed an improvement of this method. Equation 2.2 represents the anisotropic diffusion equation, where $\triangledown$ is the gradient operator.

$$\frac{\partial I(x,y,t)}{\partial t} = div(g(\|\triangledown I\|)\triangledown I) \tag{2.2}$$

When applying the model presented above to an image, the outcome is much sharper than if a simple Gaussian filter is applied. The content of the image is smoothed while edges are preserved, due to the fact that smoothing is applied only on objects on the same side of the boundary, i.e. intra-region smoothing is encouraged, while inter-region smoothing is inhibited. This behaviour is a function of g, the diffusion conductance constant [35].

The parameter $g(\|\triangledown I\|)$, also called the diffusion coefficient (represented as *(c(x,y,t)* in [48]), controls the diffusion rate and should be chosen as a function of the image gradient. The point is not only to preserve, but also sharpen the image edges [48]. Towards the solution of this problem, Li and Meng [35] proposed two functions for the diffusion coefficient (Equations 2.3 and 2.4).

$$g(\|\nabla I\|) = e^{-(\frac{\|\nabla I\|}{K})^2} \tag{2.3}$$

$$g(\|\nabla I\|) = \frac{1}{1+(\frac{\|\nabla I\|}{K})^2} \tag{2.4}$$

These functions have different effects on the image: while the first one favours high contrast edges over low contrast ones, the second one favours wide regions over smaller ones. K is a constant that controls the diffusion, determining if a certain region of the image should be smoothed or not. This value is either fixed by hand or calculated taking into account the noise of the image using a noise estimator described by Canny [48]. Due to the limitation that both these options offer, Li and Meng [35] came up with an adaptive contrast diffusion approach. The authors start by getting a contrast description of an image pixel by using the Hessian matrix. Equation 2.5 represents the Hessian matrix under a given scale σ for a given point of a grey-scale image, where I_{xx}, I_{yy} and I_{xy} represent the second-order derivatives of the image along the respective directions x, y, and xy.

$$H_\sigma(x,y) = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix} \tag{2.5}$$

Assuming that the Hessian matrix of a pixel has two eigenvalues, $\lambda_1(x,y)$ and $\lambda_2(x,y)$, the authors define a new concept of contrast (Equation 2.6) that characterises intensity variations.

$$c(x,y) = \lambda_1^2(x,y) + \lambda_2^2(x,y) \tag{2.6}$$

The use of the sum of the square eigenvalues may improve images by enhancing the details, which, in some cases, is something desirable for VEC images. This concept can then be applied to the whole image, obtaining a new expression for the anisotropic diffusion equation (Equation 2.7):

$$\frac{\partial c(x,y,t)}{\partial t} = div[g(\nabla c)\nabla c] \tag{2.7}$$

Finally, as diffusion is performed on contrast space, normalisation should be done for the transformation of the diffused result back into image space, as suggested in Equation 2.8.

$$I = \frac{c - c_{min}}{c_{max} - c_{min}} \times 255 \tag{2.8}$$

Nevertheless, a problem still stands. Different values of K lead to different consequences on the diffusion process: large K leads to smoothness, while small K leads to the sharpness of the processed region. So, although a fixed K can simplify implementation, problems may arise if the parameter is not suitable. Therefore, this parameter should be adaptive, so that it can fit each image and region within it. Li and Meng [35] designed a function to make the decision of the parameter K completely automatic, represented in Equation 2.9:

$$K(x,y) = \frac{1}{\lambda_1^2(x,y) + \lambda_2^2(x,y)} \tag{2.9}$$

This function allows enhancement in abnormal parts of VEC images (which is desirable, since these are the most relevant regions for analysis) and smoothing in regions where contrast is low (since these are probably background and have no interest). So, K should be low in regions where contrast is high, and vice versa.

Finally, this approach should be extended to colour space, since VEC images are colour images. Channels should be diffused simultaneously; otherwise, colour edge distortion can affect the image. Based on Equation 2.7, the authors extend it to RGB colour space, obtaining the expression in Equation 2.10:

$$\frac{\partial c_i(x,y,t)}{\partial t} = div[g(\phi)\triangledown c_i] \tag{2.10}$$

where $\phi = \sum c_i$ and i=R, G, B. Consequently, function for the conductance parameter is now given by Equation 2.11:

$$K(x,y) = \frac{1}{\sqrt{m_1^2(x,y) + m_2^2(x,y)}} \tag{2.11}$$

In the above equation, m_1 and m_1 are the eigenvalues of a certain pixel of the image, obtained from the Hessian matrix shown below (Equation 2.12).

$$H_\sigma(x,y) = \begin{bmatrix} \sum_{i=R,G,B} I_{xx}^i & \sum_{i=R,G,B} I_{xy}^i \\ \sum_{i=R,G,B} I_{xy}^i & \sum_{i=R,G,B} I_{yy}^i \end{bmatrix} \tag{2.12}$$

The proposed method based on eigenvalues and adaptive contrast diffusion has proven to be better for VEC image enhancement than other traditional methods like CLAHE (contrast limited adaptive histogram equalisation) and MSRCR (multi-scale retinex with colour restoration). Subject evaluation (by visual analysis only) of VEC images processed by several methods show that the proposed approach can lead to better visualisation of regions of interest while avoiding over-enhancement. Furthermore, images processed by this algorithm lead to better classification results than other methods, upon binary classification of VEC images into normal and abnormal classes [35].

2.4.2 Homomorphic Filtering

Homomorphic filtering techniques for contrast enhancement of VEC images based on Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT) are described by Ramaraj et al. [53].

An image *I(x,y)* can be defined as a function of the multiplication of illumination (*i(x,y)*) and reflectance (*r(x,y)*) components (Equation 2.13):

$$I(x,y) = i(x,y) * r(x,y) \tag{2.13}$$

i(x,y) has a lower frequency than *r(x,y)*, meaning that illumination intensity changes slower than reflectance. The goal of homomorphic filtering is to reduce the significance of illumination, which can be achieved by reducing low-frequency components of the VEC image when filtering the image in the frequency domain. So, firstly the image has to be transformed into the frequency domain, which is done in two steps: application of a logarithmic function to transform Equation 2.13 into an addition (Equation 2.14), and Fourier and wavelet transforms that transforms the image into the frequency domain (Equations 2.15 and 2.16).

$$\ln I(x,y) = \ln i(x,y) + \ln r(x,y) \tag{2.14}$$

$$F(u,v) = Fi(u,v) + Fr(u,v) \tag{2.15}$$

$$W(u,v) = Wi(u,v) + Wr(u,v) \tag{2.16}$$

Afterwards, the transformed image should go through a high pass filter *H(u,v)* to get lower frequencies filtered out (Equations 2.17 and 2.18).

$$H(u,v)F(u,v) = H(u,v)Fi(u,v) + H(u,v)Fr(u,v) \tag{2.17}$$

$$H(u,v)W(u,v) = H(u,v)Wi(u,v) + H(u,v)Wr(u,v) \tag{2.18}$$

$F(...)$ and $W(...)$ are the Fourier and wavelet functions, respectively. A Butterworth based high pass filter (BWF) can be applied, based on Equations 2.19 - 2.21 that express the filter response.

$$BWF = 1 - H(u,v) \tag{2.19}$$

$$H(u,v) = ((\alpha_H - \alpha_L) * h(u,v)) + \alpha_L \tag{2.20}$$

$$h(u,v) = \frac{1}{1 + (\frac{D_0}{D(u,v)})^{2n}} \tag{2.21}$$

Filter parameters should be chosen according to the goal of the processing. In this case, the filter should decrease illumination and return an image with enhanced contrast. Taking this into account, the authors chose the following parameter values: cutoff frequency $D_0 = 1$, low frequency gain $\alpha_L = 0.9$, high frequency gain $\alpha_H = 1.75$ and n=4. Using $\alpha_L < 1$ and $\alpha_H > 1$ drives the filter to occupy mostly the high frequency components, allowing the desirable contrast enhancement for VEC images.

After filtering out undesired frequencies, image should be converted back to spacial domain, by applying inverse Fourier (Equation 2.22) and wavelet (Equation 2.23) transforms:

$$F'[H(u,v)F(u,v)] = F'[H(u,v)Fi(u,v) + H(u,v)Fr(u,v)] \tag{2.22}$$

$$W'[H(u,v)W(u,v)] = W'[H(u,v)Wi(u,v) + H(u,v)Wr(u,v)] \tag{2.23}$$

Finally, to get the output of the enhancement processing, the exponential function is applied to the previous step output (Equations 2.24 and 2.25):

$$O(x,y) = exp[F'(H(u,v)F(u,v))] \tag{2.24}$$

$$O(x,y) = exp[W'(H(u,v)W(u,v))] \tag{2.25}$$

Subjective evaluation by simple visual analysis of the enhancement results by both DFT and DWT homomorphic filtering shows that the overall image quality is good. For an objective evaluation, image quality parameters were calculated for four different methods: HE (histogram equalisation), PM (Perona-Malik) diffusion and DFT and DWT homomorphic filtering. Results show that the proposed techniques perform better than both PM diffusion and HE methods [53].

2.4.3 Multi-Scale Retinex with Colour Restoration

Due to the fact that the dynamic range of a camera is much more narrow than the one of humans, the way our visual system perceives a scene when directly observed at a naked eye is different from the way that a camera captures it. While our vision allows us to perceive colours irrespective of the light source, the colours of the objects captured by a digital camera depend on the lighting conditions at the scene [45]. The retinex theory, described by Land and McCann [32], aims to model how the human visual system perceives scenes and enhance the quality of acquired images in order to make them resemble how a person would see it. The goal is to achieve the colour constancy feature, a property of the human colour perception [45, 49]. This can be achieved by an algorithm that fulfils the following three requirements: dynamic range compression (through the application of logarithmic transformations on the image), colour constancy (independence from the spectral distribution of the scene illuminant, that can be achieved by elimination of this component), and colour and lightness rendition.

2.4.3.1 Land's model

Originally, Land and McCann [32] proposed a random walk retinex algorithm to address this task. It takes into consideration every possible path that starts at random points and ends in the pixel where the lightness value is computed. This value is computed by taking the average of the products of ratios between the intensity values of consecutive edge pixels in the path. That is to say that the lightness *L(x)* (Equation 2.26) of a certain pixel x is obtained by the average of the relative lightness of the pixel x with respect to a pixel y_k, $L(x;y_k)$ (Equation 2.27), given all N possible paths between them, where $L(x;y_k)$ is a function of the ratio between values of two consecutive pixels in a given path γ_k with n_k pixels [49].

$$L(x) = \frac{\sum_{k=1}^{N} L(x;y_k)}{N} \tag{2.26}$$

$$L(x;y_k) = \sum_{t_k=1}^{n_k} \delta \left[\log \frac{I(x_{t_k})}{I(x_{t_k+1})} \right] \tag{2.27}$$

In the above equation, δ is defined as depicted in Equation 2.28. This is applied to remove the effect of uneven illumination over the image by considering the ratio unitary when the difference to one is smaller that a fixed threshold t.

$$\delta(s) = \begin{cases} s & \text{if } |s| > t \\ 0 & \text{if } |1| < t \end{cases} \tag{2.28}$$

Years later, Land [31] developed a centre/surround algorithm as an alternative to his random walk algorithm. This function determines the pixel lightness by computation of the ratio between a pixel value and the average value of the neighbours, considering that these surrounding pixels have density proportional to the inverse of the square distance (Equation 2.29). This operation can be seen as a sort of high pass filter [49].

$$L(x) = \frac{I(x)}{(I * G_\sigma)(x)} \tag{2.29}$$

The above equation can also be written as follows:

$$\log L(x) = \log I(x) - \log (I * G_\sigma)(x) \tag{2.30}$$

$$\log L(x) = \log I(x) - (\log I) * G_\sigma(x) \tag{2.31}$$

A subsequent analysis of the properties of Equation 2.31 lead to the single-scale retinex method, that has the ability to control a trade-off between rendition and dynamic range compression [49].

The same author also proposed a different approach, based on the image decomposition into its two components. Each image pixel can be defined as the product of its illumination and reflectance components, as already stated in Section 2.4.2:

$$I(x,y) = i(x,y) * r(x,y) \tag{2.13}$$

In order to satisfy the second property that the colour constancy algorithm must satisfy, the reflectance component *r(x,y)* must be eliminated. Knowing that illumination is the parcel with the lowest frequency (slower variance across the image, when compared to reflectance), it can be obtained by applying a low pass filtering to the image [45]. The use of a logarithmic function previous to this step allows the achievement of the dynamic range compression property. Given this, Equation 2.13 can be transformed into:

$$log(r(x,y)) = log(I(x,y)) - log(i(x,y)) \tag{2.32}$$

The illuminance component *r(x,y)* can then be obtained by a convolution with a low pass filter *F(x,y)*. One of the first suggestions was made by Land [30] that proposed to use $F(x,y) = 1/(x^2+y^2)$ as the low pass filter. Although this approach fulfils the first two requirements of the colour constancy algorithm, the last one is not accomplished.

2.4.3.2 Single Scale Retinex

The single scale retinex (SSR) equation derives from the centre/surround function (Equation 2.31), where the output is calculated by the difference between the centre (input) and the surround (average of the neighbourhood). The SSR equation can be obtained by extension of Equation 2.31 to all channels:

$$R_i(x,y) = \log\left(I_i(x,y)\right) - \log\left(I_i(x,y) * F(x,y)\right) \tag{2.33}$$

In the above equation $I_i(x,y)$ stands for the image distribution in the i^{th} colour channel, $F(x,y)$ is the surround function, and R_i is the associated retinex output. This retinex operation is performed on each colour channel [31]. Several surround functions were proposed, namely a single radial kernel proposed by Land [31]:

$$F(x,y) = \frac{C}{x^2+y^2} \tag{2.34}$$

The use of a Gaussian surround function as shown in Equation 2.35, proposed by Jobson et al. [25], overcomes the drawback imposed by others, offering a balanced trade-off between enhancement of the local dynamics and colour rendition.

$$F(x,y) = Ce^{-(x^2+y^2)/2\sigma^2} \tag{2.35}$$

In the above function, σ stands for the filter standard deviation that controls the amount of spatial detail that is retained, sacrificing dynamic range compression over the improvement of rendition, when varied from a small to a large value. For a value of $\sigma = 80$, a reasonable compromise between fairly compensated shadows and acceptable levels of image quality can be achieved [45, 49]. C is a normalisation factor such that $\int F(x,y)dxdy = 1$.

To overcome the disadvantages of only being capable of achieve either good dynamic range compression or good colour rendition, and the trouble associated with the choice of a proper scale σ for $F(x,y)$, the multi-scale retinex (MSR) method was developed.

2.4.3.3 Multi-Scale Retinex

The MSR output is a weighted sum of several single scale retinex outputs, as shown in Equation 2.36, and in opposition to SSR, it seems to grant a fair trade-off between good dynamic range and colour rendition.

$$R_{MSR_i} = \sum_{n=1}^{N} w_n R_{n_i} \tag{2.36}$$

In the MSR formula above, N is the number of scales, R_{ni} is the i^{th} component of the n^{th} scale, R_{MSRi} is the i^{th} spectral component of the MSR output, and w_n is the weight associated with the n^{th} scale [45]. Whereas in the SSR the surround function is given by Equation 2.35, in this case, it is given as follows:

$$F_n(x,y) = C_n e^{-(x^2+y^2)/2\sigma_n^2} \tag{2.37}$$

A study has proven that three scales is a proper value to fix, with values set to 15, 80 and 250, and identical weights for each scale [49].

2.4.3.4 Multi-Scale Retinex with Colour Restoration

The retinex algorithm takes into consideration the grey-world assumption, that considers that the average values of the red, green and blue components of an image with considerable colour variations should average out to a common grey value [49]. That is to say that an image obeys the grey-world assumption if the image reflectances in all three colour bands are the same on an average [45]. Thus, the retinex processing will have an adverse consequence over images that violate this assumption. In those cases where a certain colour may dominate, the use of the retinex algorithm may result in greyed-out images by decreasing the colour saturation, either globally or in precise areas [45, 49]. Taking this into consideration, it is clear the need for a step that provides colour restoration, without comprising colour constancy. The algorithm for multi-scale retinex with colour restoration (MSCRC) is given by:

$$R_{MSRCRi}(x,y) = G\left[C_i(x,y) R_{MSRi}(x,y) + b\right] \tag{2.38}$$

where $C_i(x,y) = f\left[I_i'(x,y)\right]$ is the i^{th} band of the colour restoration function (CRF), and G and b are the final gain and offset values. Jobson et al. [24] found that the CRF that provides the best overall colour restoration is:

$$C_i(x,y) = \beta \log \left[\alpha I_i'(x,y)\right] = \beta \log \left[\alpha \frac{I_i(x,y)}{\sum_{j=1}^{S} I_j(x,y)}\right] = \beta \log \left[\alpha I_i(x,y)\right] - \beta \log \left[\sum_{i=1}^{S} I_i(x,y)\right] \tag{2.39}$$

where β is a gain constant, and α controls the strength of the non-linearity [45, 49]. Suitable values for the constants are: $\beta = 46$, $\alpha = 125$, $G = 192$ and $b = -30$ [49].

Notwithstanding, it was verified that some images were still greyed-out after this colour restoration step. Given that the processing can generate negative and positive RGB values with arbitrary margins, the range of values has to be clipped into the display domain, i.e. in between $[0, 255]$. The final gain and offset values, G and b respectively, intend to address this issue, however not very successfully.

Later on, it was discovered that this degradation of the image happens at the SSR stage, so any kind of processing to prevent this issue should be applied in the single scale retinex enhanced image [45]. Many approaches were proposed, aiming for a solution to address this problem. Moore et al. [42] suggested a method where each colour channel is adjusted by taking into consideration the absolute minimum and maximum of the three colour bands. Jobson et al. [25] proposed the use of constant parameters to perform a linear transformation between the logarithmic and the display domain, after realising the peculiar shape of the retinex image output histogram (Figure 2.7), with quite extensive extremities that point out the need of clipping extreme colour values in order to achieve good image contrast. The canonical gain/offset method suggested in Jobson et al. [24] supports the use of an histogram based approach and proposes the gain and offset values previously presented, however resulting in images with poor appearance [45, 49].

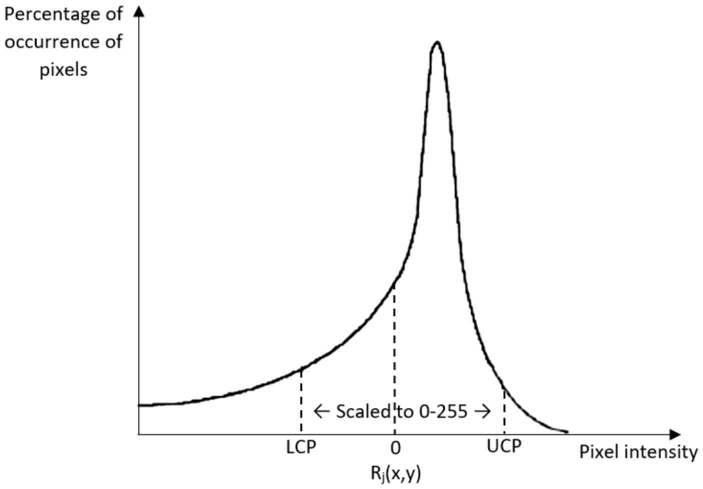


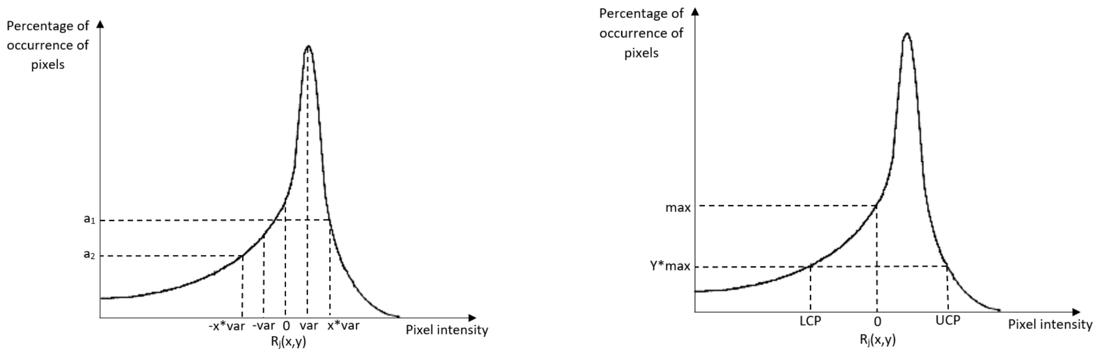
Figure 2.7: Histogram shape of the SSR enhanced image. From [45]. LCP - Lower Clipping Point; UCP - Upper Clipping Point.

2.4.3.5 Automated Multi-Scale Retinex with Colour Restoration

In Parthasarathy and Sankaran [45], the authors propose an automated and image independent method to choose the upper and lower clipping points of the SSR image histogram. This can be done by either using the variance of the histogram or the frequency of occurrence of pixels as a control measure.

In the case where the variance is used, the clipping point is chosen as x times the variance, where x can take any value from 1 to 5 (Figure 2.8a). The image histogram is then clipped and re-scaled to $[0, 255]$, as depicted in Figure 2.7. Nevertheless, after a proper test of this procedure across some images, the authors found that a unique x is not suitable for all images, concluding that the computation of the clipping points cannot be achieved with the use of the variance as the control measure. Automation of the process is not possible in this case due to the fact that the histogram is not a perfect Gaussian. Thus, equidistant pixels from either side of the mean do not have an equal percentage of occurrence of pixels, i.e. a_1 and a_2 in Figure 2.8a have different values.

In the method where the control measure depends on the frequency of occurrence of pixels, the first step is to find the number of pixels with value 0, *max*, as can be seen in Figure 2.8b. The lower and upper clipping points (LCP and UCP) are determined by finding the intersection of an horizontal line $y \times max$ with the histogram, where $y = 0.05$. This was found to be an optimal value after testing across many images. Unlike the previous approach, this method does not rely on the image, which is a desirable feature for methods with real time-applications.



(a) Clipping points using the variance. From [45].

(b) Clipping points using the frequency of occurrence of pixels. From [45].

Figure 2.8: Automated methods to choose upper and lower clipping points of SSR enhanced image histogram. LCP - Lower Clipping Point; UCP - Upper Clipping Point

Although this step was initially applied after the SSR stage, the authors found that better output images could be obtained when the procedure was applied after the MSR step instead. Besides that, the computational time also decreases if this small change is applied, since it is faster to perform one histogram processing in the MSR instead of three in the SSR.

2.5 Machine Learning for Image Classification and Segmentation

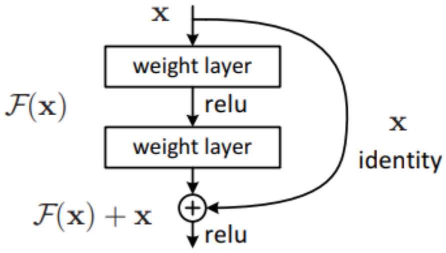
In this section, a wide range of ML techniques for image classification and segmentation tasks will be described. First, architectures of some of the most commonly used NN will be described (Section 2.5.1), followed by a summary of the application of these, as well as some other methods, for the specific case of VEC images (Section 2.5.2).

2.5.1 Background on Neural Networks

Deep learning using neural networks has been widely applied in several problems of image classification and segmentation in many different areas, including medicine. In this section, a quick overview of NN commonly used for these tasks is presented. Described architectures include Residual Network, Residual Network of Residual Network, Dense Convolutional Network, AlexNet, VGG, SqueezeNet, Inception V3, U-Net, TernausNet, AlbuNet, Fully Convolutional Dense Network and Generative Adversarial Network.

2.5.1.1 Residual Network

ResNets, described in He et al. [14] and widely used in many classification tasks, address the degradation problem of deep networks with the introduction of skip connections (also called shortcut connections) into the network architecture (Figure 2.9).



x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
x
weight layer
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Figure 2.9: Residual network building block. From [14].

Considering that $H(x)$ is the underlying map to be fit by a certain number of stacked layers, with x being the input of the first of them, the layers can take profit of the shortcut connection and learn a residual function $H(x) = F(x) + x$ instead. The network can more easily learn this residual mapping than the original one. Shortcut connections perform identity mapping, and their outputs are added to the output of the stacked layers (as outlined in Figure 2.9). So, a building block is defined as: $y = F(x, W_i) + x$, where $F(x, W_i)$ is the residual mapping to be learned. For the example in Figure 2.9, $F = W_2\sigma(W_1x)$, where σ is a rectified linear unit (ReLU) function. Dimensions of F and x should be the same, and when that does not happen, a linear projection is performed: $y = F(x, W_i) + W_sx$, where W_s can be a simple identity matrix with zero-padding.

Results show that, unlike what happens with plain networks, accuracy does not get saturated with the increase of the network depth. Skip connections of ResNet do not add any extra parameters or computational complexity to the model. Finally, residual networks exhibit considerably lower training error when compared to plain networks of the same depth, which confirms the effectiveness of residual learning on very deep networks.

2.5.1.2 Residual Networks of Residual Networks

As an improvement of ResNet, Zhang et al. [69] developed Residual Networks of Residual Networks (RoR). Compared to ResNets, these networks have extra shortcut connections on several levels (Figure 2.10).

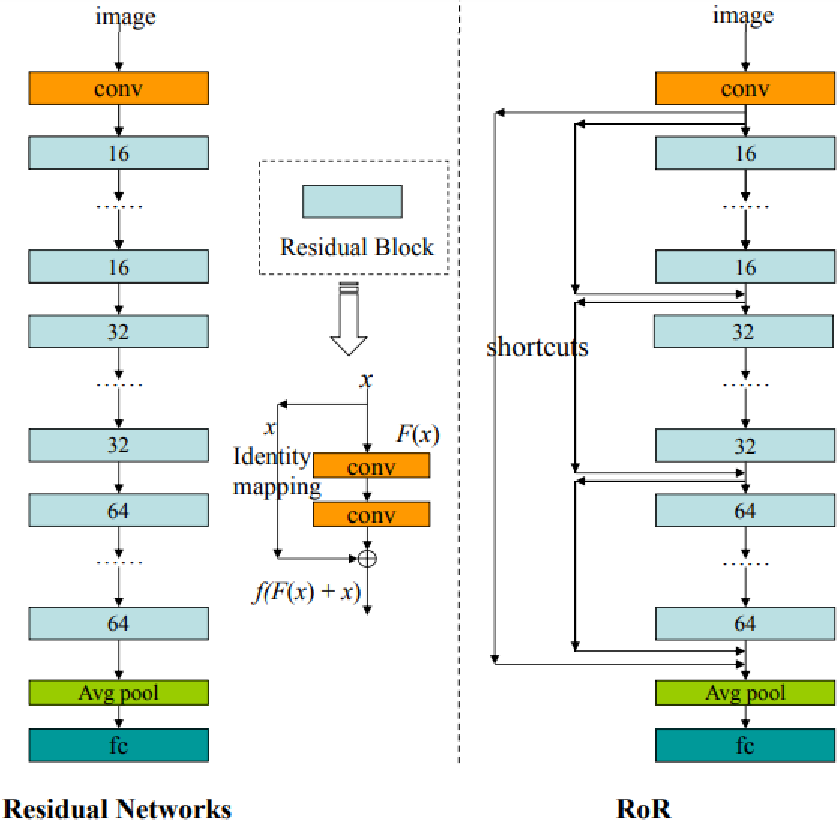


Figure 2.10: (left) Residual network architecture; (right) Residual network of residual network architecture. From [69].

Considering an original residual network with L residual blocks (final-level shortcuts), the authors first added a shortcut above all the residual blocks (root or first-level shortcut). Then, three residual blocks are formed (one for each type of filter that constitutes the convolutional layers), where each one contains $L/3$ residual blocks, and a shortcut connection is added above each one of them (second or middle-level shortcuts). More shortcuts can be added by dividing each residual block into equal parts.

RoR with three levels of shortcut connections was compared with other residual network architectures in image classification tasks, achieving better performance results than its basic residual network with the same depth [69].

2.5.1.3 Dense Convolutional Network

Huang et al. [15] proposed a novel architecture for convolutional networks, where each layer is connected to every other layer in a feed-forward fashion, thence the name Dense Convolutional Network (DenseNet). Traditional plain networks with L layers have L connections, whereas DenseNets with that same number of layers have $\frac{L(L+1)}{2}$ direct connections. In DenseNets, information flow along the network is maximised since each layer receives as input the feature maps off all the preceding layers, and its feature maps will also be used as inputs of all the following layers. This architecture is schematically illustrated in Figure 2.11.

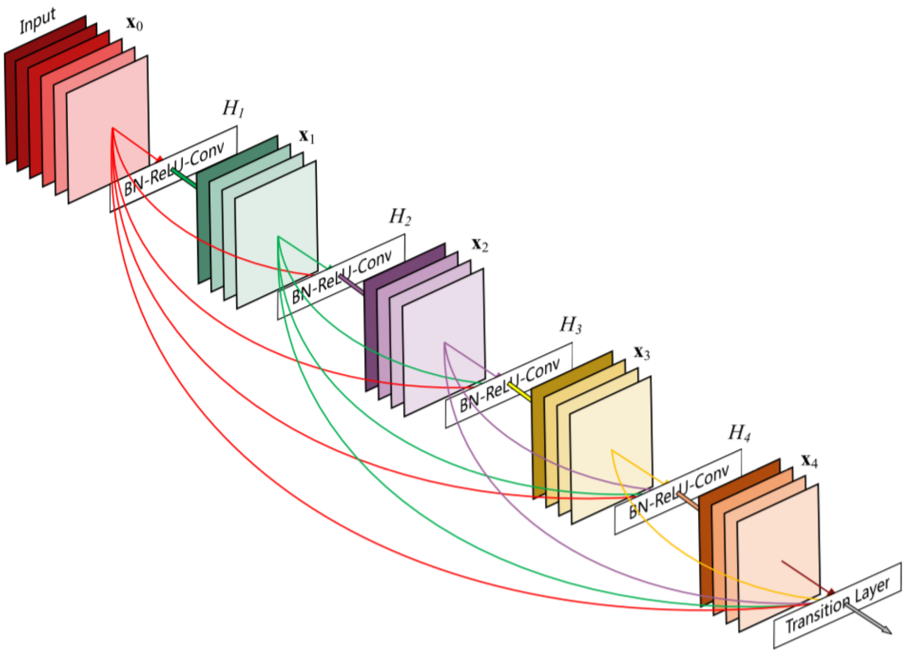


Figure 2.11: DenseNet block. From [15].

The proposed architecture has some advantages when compared with other convolutional architectures: alleviates the degradation problem associated with very deep networks, strengthens feature propagation while encouraging its reuse, and also reduces the total number of parameters. DenseNets have also proven to be able to reduce overfitting in problems with a low number of training samples [15].

2.5.1.4 AlexNet

AlexNet is a model descendant of LeNet architecture that contains 8 layers, 5 of which are convolutional (taking the role of feature extraction) and 3 are fully connected (working as a classifier).

The network architecture also includes max-pooling layers after the first, second and third convolutional layers. The output of the convolutional layers is then fed into a series of three fully connected layers, where the last one is passed through a softmax classifier with the number of class labels suitable for the classification problem. ReLU activation function is applied after all the convolutional and fully connected layers. The application of overlap pooling and dropout layers minimises overfitting in these networks [28].

2.5.1.5 VGG

The VGG network family, also descendant from the LeNet model, presents a similar, however deeper architecture in comparison to AlexNet. The increase of the network depth was possible due to the application of more convolution layers with very small sized filters (3×3). VGG family includes four main configurations with different depths, where the name is given according to the number of weight layers of the network. For example, VGG11 is constituted by 11 weight layers: 8 convolutional and 3 fully connected layers. VGG13, VGG16 and VGG19 are obtained by respectively adding two, five and eight convolution layers to the VGG11 architecture, thus totalling the correspondent number of total weight layers. ReLU activation function is always applied after each convolution. All VGG architectures include five max-pooling operations and a softmax layer after the three fully connected layers [60].

2.5.1.6 Inception V3

When designing a layer for a convolutional network, we may have to pick between 1×1, 3×3 and 5×5 convolutions, or even max-pooling layers. Inception modules solve this problem since they apply these options in parallel and concatenate the result of each one into one single output. This approach has a high computational cost, so bottleneck layers (1×1 convolution) are added in order to reduce output depth. These inception modules are depicted in Figure 2.12.

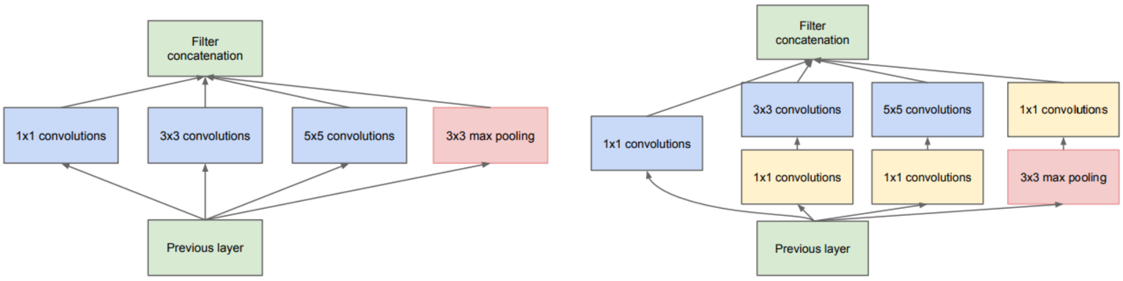


Figure 2.12: (left) Original inception module; (right) Inception module with dimension reductions. From [63].

Inception networks consist of inception modules stacked on top of each other, with occasional max-pooling layers. These modules should only be used later in the network architecture, keeping the first layers in the traditional convolutional fashion [63]. The Inception V3 model architecture includes traditional convolutional layers in the beginning and a sequence of inception modules of three different architectures later in the network [64].

2.5.1.7 SqueezeNet

SqueezeNet is a small CNN that is capable of achieving AlexNet-level performance on ImageNet, however 3 times faster and with 50 times fewer parameters. This architecture employs three main ideas in order to achieve an architecture with few parameters, while maintaining competitive performance:

- replace 3×3 filters by 1×1 filters, thus reducing the number of parameters to 9 times less;
- decrease number of input channels of the following 3×3 filters using squeeze layers (1×1 filters), in order to maintain a reduced total number of parameters in the network;
- perform the downsampling step late in the network, so that convolutional layers have large activation maps, which can lead to higher classification accuracy.

A building block of the SqueezeNet, usually called fire module (Figure 2.13), is constituted by two layers: a squeeze layer that contains only 1×1 filters, and an expand layer that is a mix of 1×1 and 3×3 filters. ReLU activation function follows each one of these layers. The network architecture is constituted by an initial convolutional layer, eight fire modules with gradually increased number of filters, and a final convolution layer followed by a softmax function. Max-pooling is performed after both convolution layers, and third and seventh fire modules. Dropout is applied after the last fire module [18].

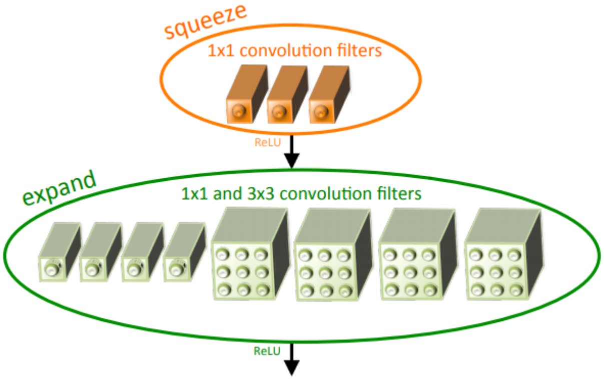


Figure 2.13: SqueezeNet fire module. From [18].

2.5.1.8 U-Net

U-Net is a novel convolutional network applied for image segmentation. Its architecture is outlined in Figure 2.14. We can clearly see that the network has a U-shape, which is due to the fact that it consists in two paths: a contracting one (left side) that captures context and, symmetrically, an expanding one that allows precise localisation. At the end, it has a final layer that produces the output segmentation map [56].

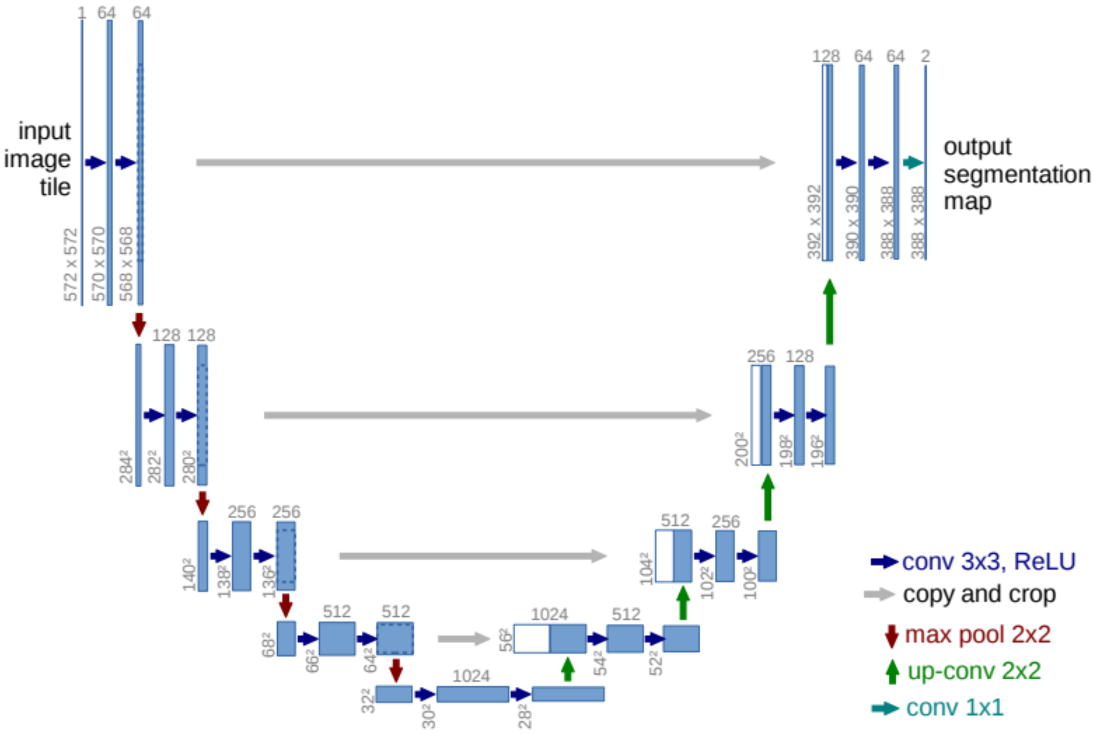


Figure 2.14: U-Net architecture. From [56]

Contracting path: gradually downsamples feature maps, while increasing the number of feature maps per layer. It resembles a typical convolutional network with interleaved convolutional and pooling operations, consisting of a sequence of the following operations:

- two 3x3 convolutions, each followed by a ReLU;
- 2x2 max pooling operation with stride 2 for downsampling (for each downsampling step, the number of feature channels is doubled).

Expansive path: increases the resolution of the output, with an architecture somewhat symmetric to the contracting path. Each step consists of an upsampling of the feature map followed by:

- 2x2 convolutions (halves the number of feature channels);
- concatenation with correspondingly cropped feature map from the contracting path;
- two 3x3 convolutions, each followed by a ReLU.

In the concatenation step, cropping needs to be applied due to the loss of pixels in the image border in each convolution.

Final layer: is required to map the resulting 64-component feature vector into the desired number of classes. The network output is a mask image where each pixel is assigned to a class.

Ronneberger et al. [56] described the use of U-Net for biomedical image segmentation, where it achieved very good performance. Moreover, those good results are reached without the need for a big amount of annotated images, thanks to data augmentation.

2.5.1.9 TernausNet

Given the popular application of classical U-Net architectures for pixel-wise object segmentation of many types of images, Iglovikov and Shvets [20] demonstrated that the performance of a U-Net like model can be improved by the use of a pre-trained encoder.

In this work, the authors propose the use of the VGG11 network as encoder, whose architecture consists of 11 sequential layers: eight 3×3 convolutional layers (each followed by a ReLU), five 2×2 max-pooling layers (each reducing feature map by 2) and three fully connected layers. In order to create the encoder deriving from this network, the fully connected layers are swapped by a single convolutional layer that works as the bottleneck central part of the model, constructing the transition between the encoder and the decoder. This one includes transposed convolutional layers that reduce the number of channels by half, while doubling the size of the feature map. Similarly to what happens in the U-Net, the output of a transposed convolution is concatenated with a copy of the corresponding feature map of the decoder. The feature map resultant of this operation suffers a convolution operation so that the number of channels is preserved. The overall architecture of the TernausNet11 network is depicted in Figure 2.15.

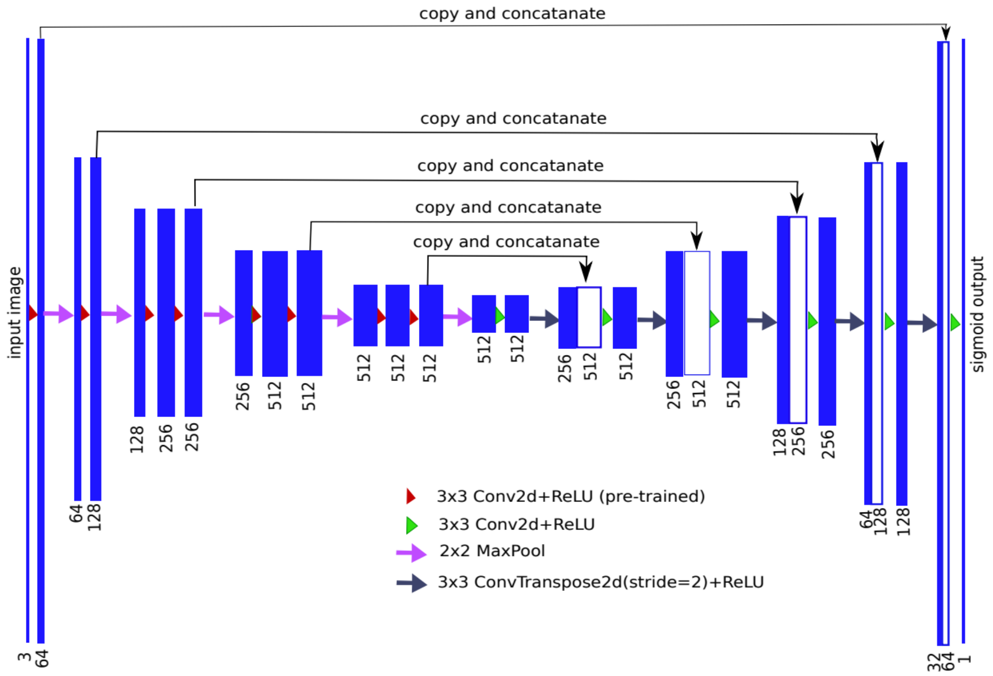


Figure 2.15: TernausNet11 architecture. From [20].

Iglovikov and Shvets [20] tested the TernausNet11 in a segmentation task of urban settlements (Inria Aerial Image Labeling Dataset), proving that the performance of the widely used U-Net model can be improved by the use of a pre-trained encoder in the network. Shvets et al. [57] applied both TernausNet11 and Ternaus16 (similar to TernausNet11, but uses VGG16 network as encoder instead of VGG11) to an angiodysplasia segmentation problem, confirming the benefit of using pre-trained encoders in a U-Net like architecture.

2.5.1.10 AlbuNet

The AlbuNet network is also an improvement of the standard U-Net architecture that uses a ResNet-type pre-trained encoder. In Shvets et al. [57] the authors used the pre-trained ResNet34 as the encoder. The architecture of this model consists of 34 sequential layers: an initial 7 × 7 convolution with stride 2 followed by a max-pooling also with stride 2 and a sequence of residual blocks. Each one of them includes an initial 3 × 3 convolution with stride 2, in order to provide downsampling, and 3 × 3 convolutions with stride 1. Each pair of 3 × 3 filters of the encoder has a typical ResNet shortcut connection. The network decoder is constructed by a set of decoder blocks, each one including a 1 × 1 convolution that reduced the number of filters by a factor of 4, followed by batch normalisation and transposed convolution to upsample the feature map. Besides this, each encoder block is connected to the correspondent decoder block through a skip connection that computes the summation of both. The architecture of the AlbuNet34 network is depicted in Figure 2.16.

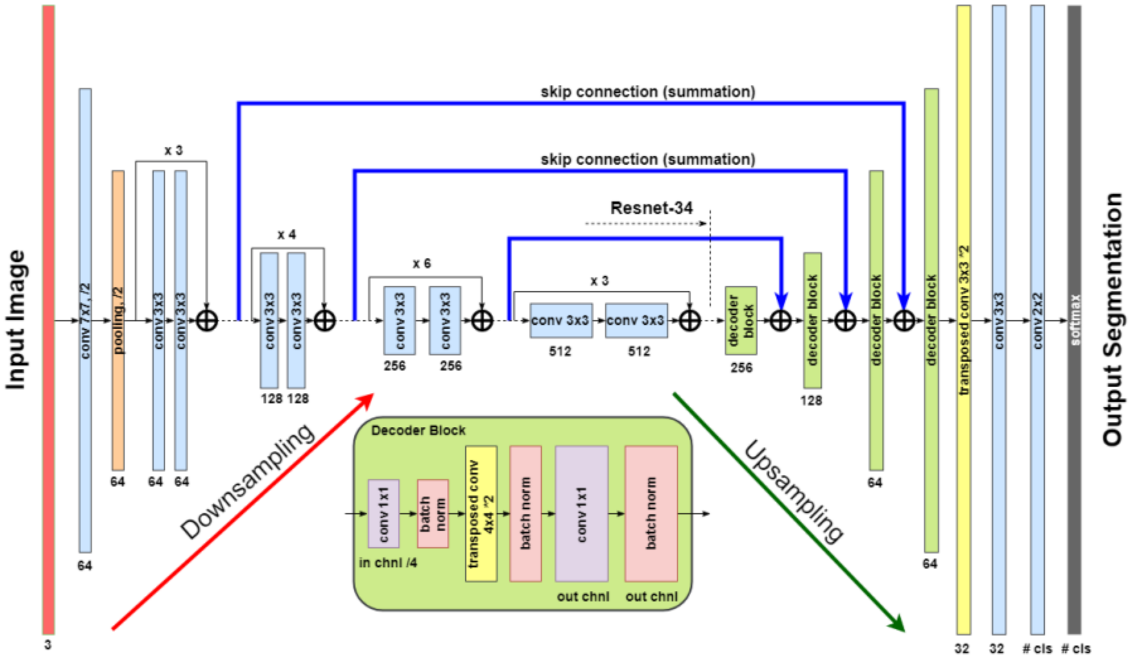


Figure 2.16: AlbuNet34 architecture. From [57].

2.5.1.11 Fully Convolutional DenseNet

Jégou et al. [21] extended the previously presented DenseNets to address semantic segmentation problems. To make DenseNets suitable for this task, an upsampling path needs to be added, to recover full input resolution, constructing a network called Fully Convolutional DenseNet (FC-DenseNet). The proposed architecture of the network is schematised in Figure 2.17.

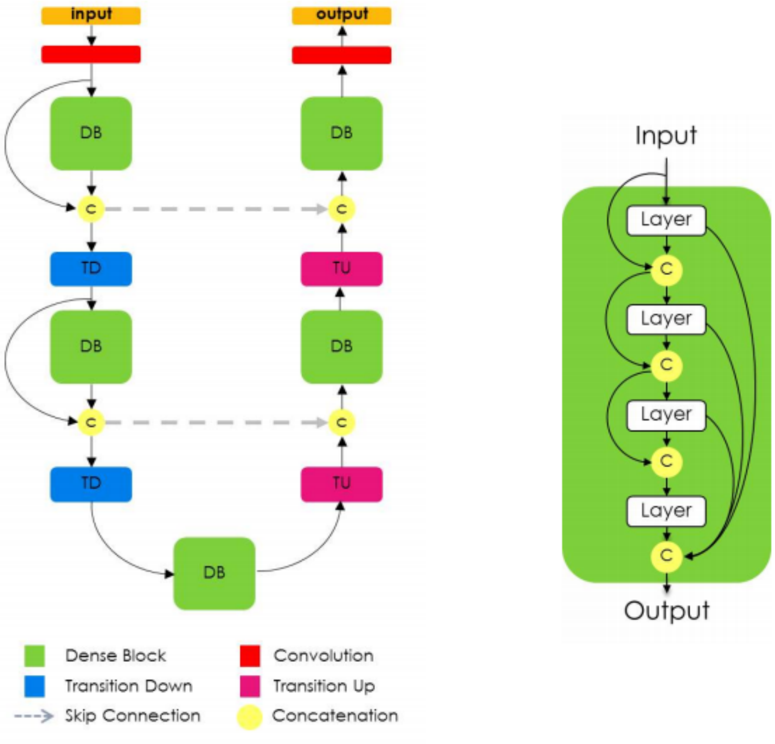


Figure 2.17: (left) Fully Convolutional DenseNet architecture; (right) 4-layer dense block diagram. From [21].

Each layer of the dense block is composed by batch normalisation, ReLU, 3 × 3 convolution and dropout. Transitions down are blocks of batch normalisation, ReLU, 1x1 convolution, dropout and max-pooling, while transitions up are just 3 × 3 transposed convolutions that upsample the previous future maps. The final layer of the network is a 1 × 1 convolution followed by softmax, so that it returns an image with each pixel class as output.

The network is composed of a downsampling and an upsampling path, just like what happens in the U-Net architecture. Also, feature maps from the downsampling path are concatenated with the corresponding ones in the upsampling path. However there is a difference between the two paths: while in the downsampling path the output of a dense block is concatenated with the input, in the upsampling path that does not happen since it would be too memory-demanding due to the growth of the number of features caused by the increase of the feature maps spacial resolution.

FC-DenseNets have few parameters in comparison to other networks used for the same purpose and, just like DenseNets, allow reuse of features. Moreover, this architecture achieved state of the art results on urban scene understanding datasets, outperforming other implemented methods [21].

2.5.1.12 Generative Adversarial Network

GAN were described for the first time by Goodfellow et al. [11], as deep neural network architectures where two models, a generative and a discriminative one, are trained simultaneously.

The generative model (G) captures the data distribution (of the training dataset) and generates new instances by passing noise through the model. The goal of G is to create new images that will be deemed as authentic by the discriminative model (D), even though they are fake, maximising the probability of D making a mistake (i.e. classifying an image as real, when in fact it is fake).

On the other hand, the discriminative model aims to, as already stated, classify images regarding its origin, being able to identify images created by G as fake. So, D learns how to distinguish a sample that comes from the data distribution from one created by G.

GAN can be seen as two NN competing with each other, with the goal of making the other fail. Competition between both models drives them to improve their methods until real data is distinguishable from data generated by G. If multilayer perceptrons are used for both models, backpropagation and dropout algorithms can be implemented to train the entire system [11].

These networks can be used for image semantic segmentation, as described by Luc et al. [37]. In this case, a convolutional semantic segmentation network along with an adversarial network should be used. The adversarial model discriminates segmentation masks coming either from the ground truth or from the generative model.

For GAN training, Luc et al. [37] propose the use of a hybrid objective function, that is the weighted sum of two terms: multi-class cross-entropy term and adversarial term. The first one encourages the segmentation model to predict the right class label for each pixel, while the second one encourages the segmentation model to produce masks that cannot be distinguishable from ground-truth ones by an adversarial binary classification model.

The training of the segmentation model minimises the multi-class cross-entropy loss, while at the same time degrading the performance of the adversarial model, which encourages the segmentation model to produce masks more similar to ground truth ones.

2.5.2 Methods for VEC Images Classification and Segmentation

In this section, a review of some of the most relevant work on lesion detection (for image classification) and segmentation of VEC frames is presented. Section 2.5.2.1 compiles a set of methods applied for the detection of lesions in VEC images, while in Section 2.5.2.2 some approaches for lesion segmentation are presented.

2.5.2.1 Detection of Lesions in the GI Tract

Many different approaches for lesion detection in images of VEC can be found in the literature. These include some works that aim to simultaneously detect different types of abnormalities and others that focus on the detection of one specific type of lesion, namely vascular or inflammatory. Due to a large number of studies done in the area, only the most relevant ones for this work will be discussed. A summary of the reviewed literature can be seen in Table 2.1 (page 36).

Simultaneous detection of many types of lesions

Regarding the automatic detection of several types of lesions, both Gueye et al. [13] and Iakovidis and Koulaouzidis [17] describe works focused on this problem.

In Gueye et al. [13], the proposed algorithm aims to an automatic video analysis for binary classification (normal/abnormal) of each frame, taking into consideration the presence (or absence) of lesions in the image. Abnormalities present in the videos include polyps, inflammations, cancer and bleeding areas. The proposed workflow can be divided into three main stages: pre-processing, feature extraction and classification. The authors justify the need of a pre-processing step saying that VEC images can suffer from vignetting, which is described as an irregular distribution of the pixel intensities from the centre of the image. Taking this into consideration, there is the need to reduce image degradation due to illumination changes and high reflectance factors, which can be achieved by a normalisation of the intensity values of the pixels. For the feature extraction step, a SIFT (Scale-Invariant Feature Transform) descriptor algorithm with BoW (Bag of Words) was the followed approach. SIFT is a well-known robust descriptor and BoW is a popular method for visually classify data. Finally, for the classification step, a linear SVM (Support Vector Machine) is used to distinguish normal from abnormal frames. Two different classifiers are also trained to identify images with polyps or inflammation separately. Classification rates of 98.25% and 92.67% are achieved for separate classification of polyps and inflammation, respectively, whereas that metric decreases to 88.50% for identification of both lesions simultaneously.

A different method for automatic detection of some lesions in VEC images is described in Iakovidis and Koulaouzidis [17]. In this case, the four main types of GI lesions are addressed: vascular and inflammatory lesions, lymphangiectasias and polyps. The proposed methodology is based on the fact that colour is an important factor to take into consideration for the discrimination of lesions from normal surrounding tissue and luminal content. The algorithm aims to classify pixels of the image into two classes (normal and abnormal tissues), and can be divided into four main steps that are succinctly described next. The first one is the colour transformation of the image from RGB to other colour space like CIELab. Then the SURF (Speeded Up Robust Features) algorithm is applied for automatic detection of points of interest (based solely on colour information), followed by feature extraction from each one of these points. Finally, the classification of the obtained feature vectors is performed by an SVM with RBF (Radial Basis Function) kernel, achieving a mean AUC (Area Under Curve) of 89%.

Detection of vascular lesions

Bleeding and angiodysplasia lesions are the ones that have been addressed the most when considering vascular lesions detection. An analysis of some proposed methodologies found in the literature that focus on this issues is presented next.

Bleeding detection: Assessment of WCE videos or images by medical specialists has a performance of only 17% for the detection of bleeding [70]. To automatically detect bleeding regions (OGIB and other), a software tool called Suspected Blood Indicator is provided by Given Image Company. However, this software has a sensitivity of 72% and a specificity of 85%, which is still not enough for medical use alone, as both metrics should be at least 85% [16, 50]. Thus, many studies have been done in this area in order to provide automatic detection of bleeding areas.

As an answer to this problem, Fu et al. [8] described a process based on super-pixel segmentation, that tries to balance the conflicting goals of reducing image complexity and avoid under-segmentation. The first step of the algorithm is the pre-processing, where a canny detector is used to find edge pixels, that will be later on removed by masking. Then, close pixels with similar colour are grouped by semantic segmentation. The resulting super-pixels are then classified by an SVM with RBF, based on RGB colour features. Finally, the image is classified as bleeding if it contains at least one super-pixel classified as bleeding, or non-bleeding otherwise. The proposed method achieves sensitivity, specificity and accuracy of 99%, 94% and 95%, respectively. Additional tests also show that these metrics are improved by the edge removal step.

In Pan et al. [44] a bleeding detection method using improved Euler distance in CIELab colour space is described. The goal is to improve the Euler distance with the covariance matrix of the image, computed using the mathematical expectation of each bleeding pattern pixels. Euler distance is a metric commonly used to measure colour similarity. By adding the covariance matrix to the Euler distance, a new distance to represent the colour difference that widens the difference between bleeding and non-bleeding patterns can be obtained. For the bleeding detection step, a threshold value is defined, setting the maximal distance between pixels for the one that is being tested to be considered bleeding. This technique obtains sensitivity and specificity values of 91.94% and 94.92%, respectively.

A method for bleeding detection based on the Colour Balance Index (CBI) to compensate for irregular image condition is described in Jung et al. [26]. CBI is determined by intrinsic colours of the background tissues, as well as by illumination conditions, describing the overall colour tone of the image. CBI is computed for a high number of samples ($50,000$ images), and a mean value is obtained. Then, the proposed workflow for the detection of GI bleedings is divided into three main steps. First of all, bad illuminated (under and over-illuminated) regions are masked, meaning that too bright or too dark pixels are set to zero. Afterwards, colour spectrum transformation method is applied (using the previously obtained CBI value), in order to divide the VEC frame into bleeding and non-bleeding region, according to the set threshold value. Finally, the output mask is filtered to have noise (detected small bleeding regions) filtered out. The obtained results show that the use of CBI is beneficial in the bleeding detection problem, leading to an increase of sensitivity from 80.87% to 94.87% and specificity from 74.25% to 96.12%.

Hwang et al. [16] described a bleeding detection method using Expectation Maximisation (EM) clustering algorithm and Bayesian Information Criterion (BIC). EM clustering algorithm trains probability models for both blood and non-blood pixels, dividing each group into several sub-groups and modelling each one of these by a set of parameters of Gaussian mixture density.

The EM algorithm is used to find appropriate clusters by iteratively estimating the Maximum Likelihood Estimation (MLE) of the parameters. In opposition to this method, BIC is used to automatically obtain the number of clusters. The methodology is finalised with the blood regions detection, where three main steps can be highlighted: dark colour pixel removal, application of a conditional probability restriction rule (to select blood pixel candidates) and region size filtering (to filter out small blood regions). This approach can reach sensitivity and specificity values of 92.55% and 98.10%, respectively, for blood VEC image detection.

More recently, Li et al. [36] suggested the use of CNN and transfer learning for the bleeding detection in VEC images problem. Transfer learning is especially useful when the amount of available data to train the CNN is not enough. Due to this problem, the authors applied data augmentation (image rotation and flip, colour distribution perturbation and application of SMOTE algorithm for minority class oversampling) to the training dataset. The proposed strategy transfers the pre-trained Inception V3 model trained on ImageNet dataset, that works as a generic feature extractor of mid-level image representation. It can then be used as a starting point to train another network for a different problem, like lesion detection in endoscopic images. The model is then fine-tuned and its weights are updated by training with labelled VEC images. This work surpasses other methods that do not use transfer learning, getting 99.1% of AUC and 91.9%, 87.2% and 98.62% of precision, recall and accuracy, respectively.

Angiodysplasia detection: VEC images and videos analysis performed by a medical specialist has a detection performance of about 69% for angiodysplasia [70]. This value is not reliable enough, so automatic detection of these lesions has also been addressed by many investigators, whose work is presented next.

For the detection of these lesions in VEC images, Noya et al. [43] proposed a solution based on a colour thresholding approach. First, the image is pre-processed by HE, CLAHE and RGB decorrelation stretch. After this, a potential region of interest (ROI) is selected, taking into consideration RGB channels information (especially the green channel), as well as geometric data like area, perimeter and extent of the ROI. Feature extraction is the next step, where statistical, texture and geometrical features are extracted, followed by feature selection. Finally, a RUSBoosted trees classifier is used, getting 93.15% AUC, 96.63% accuracy and 96.8% specificity.

More recent publications addressing this problem show that the use of deep learning techniques is very promising in the area. A CNN-based semantic segmentation algorithm was recently used for deep-feature extraction and classification of VEC images, regarding the presence of angiodysplasia lesions [34]. The CNN will first extract local features by training the network on a set of labelled images, and then classify new images, based on the learner features in the previous step. This algorithm yielded a 100% sensitivity and 96% specificity in the detection of angiodysplasias in VEC frames.

Another approach using deep learning is presented in Pogorelov et al. [50]. In this paper, both angiodysplasia detection at frame-level and segmentation at pixel-level are presented, but only the first one will be addressed by now. So, in order to detect if there is an angiodysplasia lesion in the VEC frame, the authors propose different methods: global features (handcrafted extraction of global features that describe the image on a global level), deep features (extracted by a neural network like ResNet 50, VGG 19 and Inception V3), and a variation of GAN (that takes as input the number of positively marked pixels in the image, and classifies image taking into account the number of pixels classified as positive). For the classification step, Random Tree, Random Forrest and Logistic Model Tree classifiers were adopted. The approach where GAN was used outperformed all of the others, with a sensitivity of 98% and specificity of 100%.

Detection of inflammatory lesions

Concerning inflammatory lesions, the detection of some types of these lesions by specialists has a correct detection rate of only 38%, which justifies the need of developing tools for their automatic detection [70].

A proposed algorithm based on CNN is presented in Georgakopoulos et al. [10]. The authors define the method as weakly-supervised, once it only requires image-level labelling, instead of pixel-level annotations. For comparison, a BoW-based classifier method was also implemented. Regarding the CNN based strategy, the implemented NN has five convolutional layers, and the training images suffer data augmentation through rotation and flip. In the BoW method, SURF algorithm is used for automatic key-point detection and description of the images, and the resulting feature set is clustered using k-means algorithm. The CNN-based method outperformed the BoW-based strategy, with accuracy, sensitivity and specificity values of 90.2%, 92.6% and 88.9%, respectively, against 72.2%, 88.9% and 55.6%. A different approach where the images were divided into patches was also tested, however with worst performance when compared to either previous results.

For an easier comparative analysis of all the presented reviewed works on lesion detection on VEC images, a summary can be found in Table 2.1. The presented methods cannot be directly compared, since the lesion that is being detected is not the same for all cases.

Given all reviewed approaches for lesion detection in VEC frames, only two of them propose algorithms for a multi-class classification problem, although none of them addresses the classification of images into normal, vascular and inflammatory. Both approaches suggest the application of a pre-processing step to the images and the use of SVM classifier. Results cannot be compared, since a number of labels of the classification problem is different, and classification metrics computed to evaluate the algorithms are not the same.

Regarding bleeding detection, a wider range of approaches was reviewed. Applied pre-processing includes removal of edge and bad-illuminated regions. One of the approaches also performs data augmentation. Classification algorithms comprise SVM, threshold value setting and, the one that achieved the best performance concerning the accuracy, Inception V3 pre-trained model.

Table 2.1: Summary of the most relevant literature reviewed on lesion detection in VEC images.

Lesion	Data augmentation and pre-processing	Feature extraction	Classification and Pos-processing	Accuracy	Sensitivity	Specificity	AUC	Source
Polyps ^{1} Inflammatory ^{2}	Normalisation of intensity values of the pixels	SIFT algorithm with BoW	Linear SVM	98.25% ^{1} 92.67% ^{2} 88.50% ^{12}	-	-	-	[13]
Vascular Inflammatory Lymphangiectasias Polyps	Color transformation from RGB to CIELab	SURF algorithm	SVM with RBF	-	-	-	89%	[17]
Bleeding	Canny detector to find edge pixels Masking of edge regions Creation of super-pixels	RGB colour features	SVM with RBF	95%	99%	94%	-	[8]
Bleeding	Euler distance improvement (with covariance matrix)		Threshold value is set for maximal Euler distance	93.23%	91.94%	94.92%	-	[44]
Bleeding	Bad illuminated regions masking Application of the colour spectrum transformation method		Threshold value set on CST method determines pixel classification Small bleeding regions are filtered out	95.75%	94.87%	96.12%	-	[26]
Bleeding	Dark colour pixel removal		EM clustering algorithm and BIC Application of conditional probability restriction rule Small bleeding regions are filtered out	-	92.55%	98.10%	-	[16]
Bleeding	Data augmentation (image flipping and rotation) SMOTE algorithm	Transfer learning (Inception V3 model trained on ImageNet and fine-tuned)		98.62%	87.2%	-	99.1%	[36]
Angiodysplasia	HE, CLAHE and RGB decorrelation stretch	Statistical, texture and geometrical features Feature selection	RUSBoosted trees	96.63%	89.51%	96.8%	93.12%	[43]
Angiodysplasia		CNN for semantic segmentation		98%	100%	96%	-	[34]
Angiodysplasia		GAN (classifies image according to number of positively labeled pixels)		99%	98%	100%	-	[50]
Inflammatory	Data augmentation (image flipping and rotation)	CNN		90.2%	92.6%	88.9%	-	[10]
Inflammatory		SURF algorithm	K-means algorithm	72.2%	88.9%	55.6%	-	

For angiodysplasia detection, one of the presented algorithms proposes a workflow where the images are pre-processed for a following feature extraction step. Before classification (performed by RUSBoosted trees classifier), a feature selection step is applied. However, two deep learning based approaches (CNN and GAN) also applied for this task, were the ones that best performed.

Finally, regarding inflammatory lesions, the best performing algorithm was the one that implements a CNN (alongside a previous data augmentation step), outperforming an approach of feature extraction and classification by a k-means algorithm.

All in all, for each lesion (bleeding, angiodysplasia and inflammatory), especially the last two, it is clear that the deep learning approaches (CNN and GAN) show the best results.

2.5.2.2 Segmentation of Lesions in the GI Tract

Concerning lesion segmentation in VEC images, the number of proposed strategies to address this problem is much lower, when compared to the lesion detection task. Moreover, no article with the aim of inflammatory lesion segmentation was found. Thus, only vascular lesions will be approached in this section. A summary of the reviewed literature can be seen in Table 2.2 (page 38).

The workflow proposed by Pogorelov et al. [50], already described in the previous section, applied a GAN for semantic segmentation of angiodysplasia lesions. The authors were able to use an implementation initially developed for retinal vessel segmentation, described in Son et al. [61], by applying a small modification to it: addition of an output layer to the generator network that will allow the creation of a binary segmentation as output, given that it implements an activation layer with a step function. The GAN segmentation algorithm achieved 99.9% accuracy and specificity, and 88% sensitivity.

A different approach is proposed by Vieira et al. [66], where segmentation of angiodysplasias is accomplished by using a MAP (Maximum a Posteriori) approach with Markov Random Fields (MRF). The pre-processing step includes removal of not lesion related regions highlighted in the *a* component of the CIELab colour space. The MAP approach is used to divide the image into two regions, with MRF theory to include spatial information. Segmentation is performed using statistical classification based on Bayes rule. Based on that, for each pixel, the MAP estimate is calculated for all classes, and the pixel is assigned to the class with the maximum MAP. Class conditional probability is modelled by a Gaussian function, where the observations are modelled by a Gaussian mixture that has his parameters iteratively estimated by EM algorithm (with Maximum Likelihood criterion). MRF can be used to improve *a priori* probabilities by including neighbourhood information. For performance comparison, a standard segmentation using Otsu's method was also implemented. The MAP algorithm alone achieved a mean dice of 32.91%, whereas the MAP with MRF approach increased this value to 55.11%. Otsu's method had a dice performance of only 31.26%.

Hwang et al. [16] described a method that was used for both blood detection and segmentation. The implementation was already described in Section 2.5.2.1 and, regarding lesion detection at pixel-level (i.e. segmentation), the algorithm got 83.96% sensitivity and 94.93% specificity.

The winning solution for Angiodysplasia Detection and Localisation SubChallenge of MICCAI 2017 Endoscopic Vision Challenge applied deep convolutional NN for the segmentation of those lesions [57]. Here, four different architectures are presented: U-Net, TernausNet11, TernausNet16, and AlbuNet34. Regarding pre-processing, the authors apply random affine transformation and colour augmentations in HSV space to the training images. After training the models, the masks resulting from the test suffer post-processing in order to remove small detected lesions that are considered noise. The network with the best results was AlbuNet34, that got IoU (intersection over union, also called Jaccard index) of 75.35% and Dice of 84.98%.

For an easier comparative analysis of all the presented reviewed works on lesion detection on VEC images, a summary can be found in Table 2.2.

Table 2.2: Summary of the most relevant literature reviewed on lesion segmentation in VEC images.

Lesion	**Algorithm**	**Sensitivity**	**Specificity**	**Jaccard**	**Dice**	**Source**
Angiodysplasia	GAN (adaptation of GAN used for retinal vessel segmentation)	88%	99.9%	-	-	[50]
Angiodysplasia	MAP	-	-	-	32.91%	[66]
	MAP with RMF	-	-	-	55.11%	
	Otsu's method	-	-	-	31.26%	
Angiodysplasia	U-Net	-	-	73.18%	83.06%	[57]
	TernausNet11	-	-	74.94%	84.43%	
	TernausNet16	-	-	73.83%	83.05%	
	AlbuNet34	-	-	75.35%	84.98%	
Bleeding	EM clustering algorithm and BIC Application of conditional probability restriction rule	83.96%	94.93%	-	-	[16]

Literature review on segmentation of lesions in VEC fell short on the expectations, since any published work addresses inflammatory lesions, and regarding vascular ones, only studies concerning segmentation of a specific type were found.

For segmentation of angiodysplasia lesions, MAP and Otsu's method based approaches were implemented; however, the obtained results are not satisfactory. On the other hand, U-Net and GAN neural networks achieved very good results, especially the second one. Regarding bleeding detection, a strategy based on EM clustering algorithm was reviewed. This approach obtained good results; however, the lack of another algorithm for results comparison does not allow a fair draw of conclusions.

Despite the lower number of studies performed on this subject, deep learning approaches (namely GANs, U-Nets and U-Net like architectures) show promising results. However, these networks were still not applied for segmentation of several types of vascular lesions, neither for any kind of inflammatory lesion. Thus, the success of these approaches to new problems is not assured.

2.6 Datasets

In order address both VEC images classification and segmentation tasks, available endoscopic datasets should be considered.

GIANA Endoscopic Vision Challenge provides a dataset that can be accessed by any participant registered in the challenge. It has a total of 1812 WCE frames, divided into three classes, regarding the presence or absence of lesions: 600 normal, 607 inflammatory and 605 vascular. For the frames where either vascular or inflammatory abnormalities are present, binary masks of the lesions segmentation are also available. All frames were captured with the PillCam®SB3 endoscopic capsule, that produces JPEG images with 576×576 pixels.

Additionally, we have an anonymised private database. It consists of a total of 449 VEC (338 recorded with PillCam®SB2, and the remaining 116 with PillCam®SB3). These have the correspondent medical report with information regarding relevant medial findings in the video, which can be specially helpful in the cases where annotations regarding the presence of vascular and inflammatory lesions are provided.

2.7 Summary

The reviewed literature described in this section allowed for a better and deeper understanding of already developed work in VEC image classification, segmentation and enhancement, as well as promising techniques that have not yet been applied for the particular problems proposed for this dissertation.

As expected, the use of ML techniques has proven to be advantageous over other methods, for both lesion detection and localisation in VEC frames. Although none of the presented methodologies was applied for either image classification into three classes (normal, vascular and inflammatory), or inflammatory lesion segmentation, similar deep learning approaches should be suitable for these problems.

The importance of a pre-processing stage should also not go unnoticed, as was evidenced by some studies that this step can boost image quality, and consequently improve image classification and segmentation methods.

Chapter 3

Lesion Detection and Segmentation in VEC Images

In this chapter, the approach followed to address the classification and segmentation tasks at hand in this dissertation will be presented. The developed methodology took into consideration several issues pointed out in Chapter 2:

- The need of implementing a pre-processing step on the VEC images, that arises from the fact that the insufficient quality of these frames can hamper the analysis of the information contained within the captured image;

- The reviewed literature pointed out the advantage of deep learning based methods over others in lesion detection and segmentation works. Therefore, both tasks will be performed with resort to deep learning techniques, where the use of large datasets is almost mandatory to achieve a model with good performance and without overfitting. Data augmentation can be applied to the original dataset in order to create new samples originated from the real ones, thus increasing the number of available images;

- Given the ever-growing acceptance of the use of deep learning techniques in a wide range of applications, many neural networks architectures have been created. Thus, a selection of the most suitable ones for the problems at hand has to be made, in order to elect those that should perform better;

- Towards the analysis of the obtained results and their validation, the choice of the right metrics of evaluation is a very important step. These should be elected taking into consideration the task and be representative of the results, as well as allow comparison with other works.

With all this in mind, a methodology to implement was delineated (Figure 3.1). The pre-processing and data augmentation steps are equivalent for both VEC image classification and segmentation tasks, whereas concerning model training and evaluation, the pipeline diverges. Each step of the framework will be described in detail in the following sections.

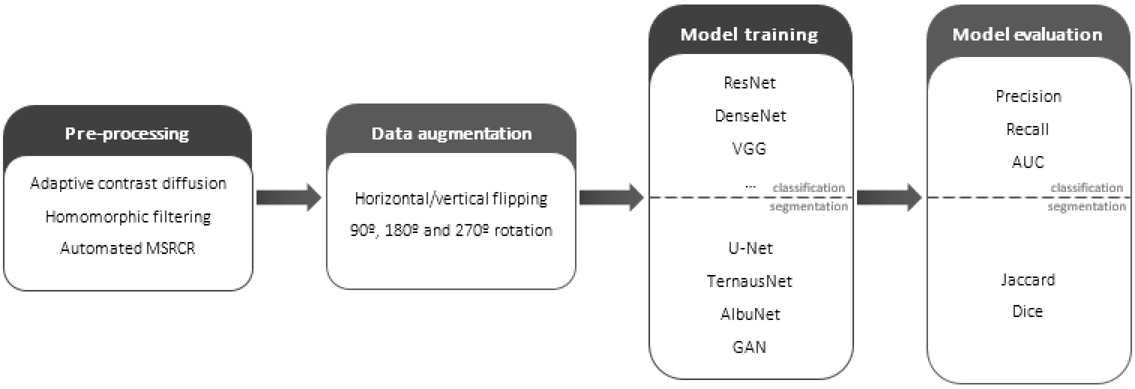


Figure 3.1: Pipeline of the implemented framework for VEC images classification and segmentation.

3.1 Image Pre-processing

It is known that the quality of VEC frames is not always the best. This happens due to the conjugation of a few aspects including data compression, low resolution of the frames, low illumination, and other complex circumstances that the capsule has to go through along its passage. With this in mind, several approaches that aim to attenuate the effect of these conditions where analysed in Section 2.4. Given the promising results of the described techniques in VEC images enhancement, adaptive contrast diffusion (Section 2.4.1), homomorphic filtering (Section 2.4.2), and multi-scale retinex with colour restoration (Section 2.4.3) algorithms were applied.

Adaptive Contrast Diffusion

Following the procedure described in Section 2.4.1, the first step of the adaptive contrast diffusion algorithm for an RGB colour space image is the computation of the image contrast description (Equation 2.6), with resort to the eigenvalues of each pixel Hessian matrix (Equation 2.12). Then, for each image pixel, the value of the parameter that controls the diffusion process, K, that is also a function of the eigenvalues of the aforementioned Hessian matrix of the respective pixel (Equation 2.11) can be computed. Finally, the anisotropic diffusion equation (Equation 2.10) can be applied to the previously obtained contrast description of the image, followed by normalisation of the diffused result back to image space (Equation 2.8). An example of the effect of the adaptive contrast diffusion enhancement in an image can bee seen in Figure 3.2.

$$\frac{\partial c_i(x,y,t)}{\partial t} = div[g(\phi)\nabla c_i] \tag{2.10}$$

$$K(x,y) = \frac{1}{\sqrt{m_1^2(x,y) + m_2^2(x,y)}} \tag{2.11}$$

$$H_\sigma(x,y) = \begin{bmatrix} \sum_{i=R,G,B} I_{xx}^i & \sum_{i=R,G,B} I_{xy}^i \\ \sum_{i=R,G,B} I_{xy}^i & \sum_{i=R,G,B} I_{yy}^i \end{bmatrix} \tag{2.12}$$

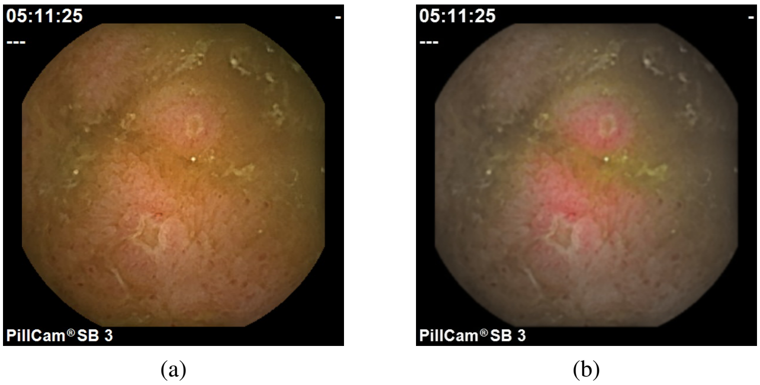


Figure 3.2: Example of the effect of the adaptive contrast diffusion enhancement: (a) original image; (b) enhanced image.

Homomorphic Filtering

The homomorphic filtering technique (described in detail in Section 2.4.2) is based on the representation of an image as the sum of two components, illumination ($i(x,y)$) and reflectance ($r(x,y)$). In order to convert the images to the frequency domain, the Fourier transform was applied. Towards the elimination of the illumination parcel (the lower frequency component), the previously transformed image goes through a high pass filter. The image is then converted back to image space by application of the inverse DFT, finalising the homomorphic filtering process. Figure 3.3 presents the result of an image enhanced by the homomorphic filtering technique.

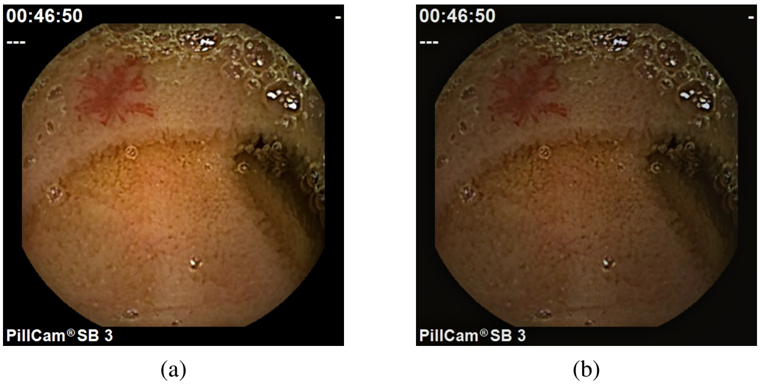
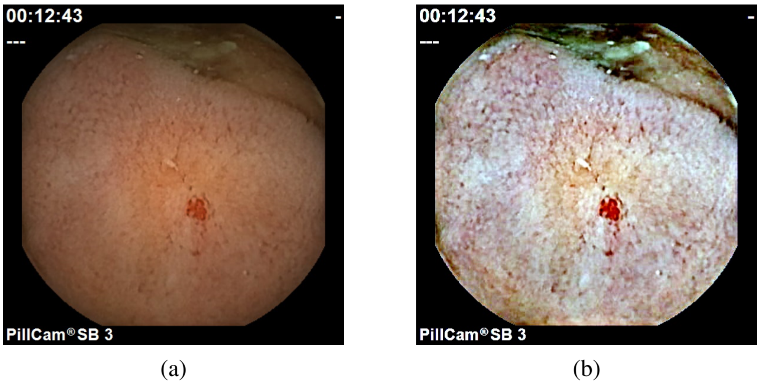


Figure 3.3: Example of the effect of the homomorphic filtering enhancement: (a) original image; (b) enhanced image.

Automated Multi-Scale Retinex with Colour Restoration

The first stage of the processing is the computation of the SSR of the image. This step relies on a surround function that was defined as a Gaussian filter. The standard deviation of the filter, σ, represents a trade-off between fairly compensated shadows and acceptable levels of image quality and will be set in the next stage of the processing. Thereupon follows the MSR, that is merely a weighted sum of several single scale retinex outputs. The outcome of an image to which the automated MSRCR algorithm was applied can be seen in Figure 3.4.



(a) (b)

Figure 3.4: Example of the effect of the automated MSRCR enhancement: (a) original image; (b) enhanced image.

3.2 Model Training

3.2.1 Data Augmentation

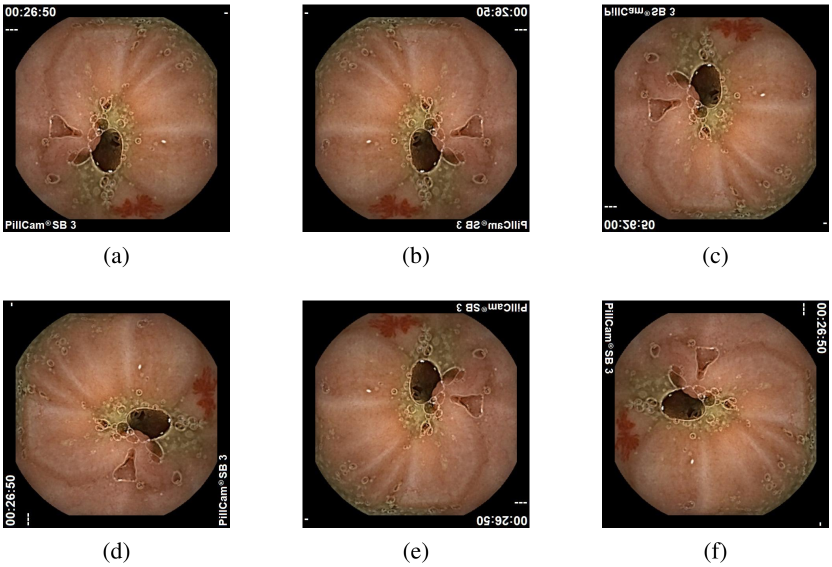
It is known that the more data an algorithm has access to, the more effective it can be. However, image and video classification problems frequently have insufficient data publicly available. This is particularly an issue for medical tasks, where access to data is heavily protected as a result of privacy issues. Models trained on small datasets usually do not generalise data well enough, overfitting to the training data and leading to poor performance on the validation and test sets [47]. Training the model with additional synthetic generated data can make it invariant to translation, rotation, size, and so on, increasing the robustness of the algorithm for a real-world application, where data may be found in a much wider variety of conditions. Thus, data augmentation can prevent NN from learning irrelevant feature patterns from a small dataset, boosting overall model performance.

Taking this into consideration, traditional data augmentation transformations where applied to the training dataset: vertical and horizontal flipping, as well as 90°, 180° and 270° rotations. This means that, for a dataset of size N, a new one of $6 \times N$ size is generated, which represents a significant increase in the number of training samples. Examples of this transformations can be seen in Figure 3.5.

3.2.2 VEC Images Classification

3.2.2.1 Implemented Networks for VEC Images Classification

The need for using large datasets in deep learning methods arises from the large number of parameters associated with NN. The deeper the model, the higher amount of data required to train it.



(a) (b) (c) (d) (e) (f)

Figure 3.5: Transformations applied to the VEC images for data augmentation: (a) Original VEC frame; (b) Horizontal flipping; (c) Vertical flipping; (d) 90° rotation; (e) 180° rotation; (f) 270° rotation.

Models need to be deep enough to be able to capture resemblance like texture, colour and shape features between image samples: while initial layers of the model gather those high level features, later ones capture information that relates those features with the outputs, learning to discriminate between them. Unfortunately, in many cases, the amount of available data is not sufficient to adopt an approach like this. Transfer learning solves this problem by applying models previously trained on a large available dataset (usually trained for a completely different task, using the same input but returning different output), in a novel task. This is achieved by a previous capture of the relations in the features of that data that can then be reused for other problems.

With this in mind and due to the previously mentioned success of transfer learning in image classification (Section 2.5), namely in the medical area, an approach using this technique was followed for the lesion detection task. In transfer learning approaches, reusable features returned by some layers will be used as input features that allow the training of a new model that only needs to learn the relations of those features for the new problem, given that it has already learnt about data patterns in the data that was used to pre-train the model. Besides requiring much less parameters to learn, transfer learning also has the advantage of better generalisation of the models, given that they underlay the phenomenon more that they model the data, owing to the fact that the model has access to different types of data.

PyTorch provides several models [51] previously trained on ImageNet, a dataset of more than 15 million labelled high-resolution images belonging to around 22,000 classes [28]. Available networks include AlexNet, VGG, ResNet, SqueezeNet, DenseNet and Inception V3. All these architectures were already described in Section 2.5.1.

Testing such a wide range of models allows for a comparison of very different architectures, which can be useful to understand which ones are more suitable for this problem. Afterwards, we can focus on the best-succeeded approaches and how to improve their performance.

3.2.2.2 Experiments

In order to investigate the performance of the previously presented networks, the models were trained and evaluated in a balanced dataset that has to be split into three sets: training (50%), validation (25%) and test (25%).

Raw VEC images usually have an outer border with text information that includes the time of the frame in the video from where the frame was extracted and endoscopic capsule version. Therefore, in order to remove these text annotations without losing relevant information, images have to be cropped before further processing. Shvets et al. [57] suggests a crop to 512×512 pixels, for images of 576×576 pixels of original size.

Besides cropping, all images of the dataset should be normalised and resized. This is demanded by the used pre-trained models, that require either 299×299 (Inception model) or 224×224-sized (all other networks) input RGB images [51].

Each pre-trained model is then trained using two different approaches: only the last layer is updated, i.e. the pre-trained model works as an feature extractor and only the weights of the classification layers are changed, or the whole model is finetuned. Each model was trained for 30 epochs, with a batch size of 8, and using stochastic gradient descent as the optimiser. Momentum was set to 0.9 and learning rate to 0.001, with a decay by a factor of 0.1 every 7 epochs. The loss is computed for every training and validation step of each epoch, using the cross-entropy loss function (Equation 3.1) that applies a logarithmic softmax followed by the negative log-likelihood loss [52].

$$loss(x, class) = -log\left(\frac{exp(x[class])}{\sum_j exp(x[j])}\right) = -x[class] + log\left(\sum_j exp(x[j])\right) \tag{3.1}$$

3.2.3 VEC Images Segmentation

3.2.3.1 Implemented Networks for VEC Images Segmentation

The rise of many scenarios that require accurate and efficient segmentation tools has been leading to the ever-growing interest of image semantic segmentation for computer vision and machine learning researches. Consequently, this demand motivated the growth of deep learning techniques in this area, usually CNN, that are surpassing other methods by a large margin regarding accuracy and sometimes even efficiency. The key advantage of CNN applied in pixel-level labelling problems over traditional methods is the ability to learn appropriate feature representations for the task at hand [9].

With this in mind, for lesion segmentation within the VCE frame, four neural network architectures were tested: U-Net, TernausNet, AlbuNet and GAN, already described in Section 2.5.1. These were the selected models due to the produced promising results in some reviewed works that addressed issues similar to the one at hand. Each model was separately trained for each type of lesion (vascular and inflammatory). Concerning the GAN architecture, the implementation described in Son et al. [61] was followed.

3.2.3.2 Experiments

Equivalently to the lesion detection task, for the lesion segmentation the dataset of each type of lesion is also split into training (50%), validation (25%) and test (25%) sets.

In a similar way to what was done in the lesion classification task, images are cropped in order to remove the canvas and text annotations around the region of interest. Binary masks of the segmented lesions existing in these images are also cropped, so that images and masks have direct pixel-by-pixel matching. Unlike the NN employed for the lesion detection, all models implemented for the segmentation task support 512 × 512-sized images as input, so no further resize is required.

Training and validation images, as well as matching masks are then fed into the network that trains for 500 epochs, with learning rate set to 0.0001. Binary cross entropy with logits loss (Equation 3.2) is used to compute loss at both training and validation steps of every epoch.

$$\ell(x,y) = L = \{l_1,\dots,l_N\}^\top, \quad l_n = -w_n\left[y_n \cdot \log\sigma(x_n) + (1-y_n)\cdot\log(1-\sigma(x_n))\right] \tag{3.2}$$

In the above equation N is the batch size. This loss function combines a sigmoid layer with the binary cross entropy loss in one single class, which is more numerically stable since the aggregation of both operations into one single layer takes the advantage of the log-sum-exp trick [52]. Adam optimiser is the used to update model weights. This algorithm grants efficient stochastic optimisation, requiring only first-order gradients. It is an adaptive learning rate optimisation algorithm that computes individual values of that hyper-parameter for different parameters from estimates of first and second moments of the gradient. Adam is well suited for problems that employ deep learning techniques [27]. After the training step is finished, the model with the lowest loss in the validation step is saved.

3.3 Model Evaluation

3.3.1 VEC Images Classification

After the training and validation steps are completed, the model weights of the epoch with the lowest validation loss are saved for further evaluation in the test set. Given the ground truth classification and the output prediction of the model, a confusion matrix can be obtained. Metrics commonly used for classification problems can then be computed with resort to the number of true positives (TP), false positives (FP) and false negatives (FN). Recall and precision are some of the most widely known classification metrics and are defined by equations 3.4 and 3.3, respectively. The area under the Receiver Operating Characteristic (ROC) curve is another evaluation metric commonly used to evaluate the performance of classifiers, and will also be computed in this work. The area under the ROC curve can be computed using the trapezoidal integration (Equation 3.5) [2].

$$precision = \frac{TP}{TP+FP} \tag{3.3}$$

$$recall = \frac{TP}{TP+FN} \tag{3.4}$$

$$AUC = \sum_{i} \left\{ (1-\beta_i \cdot \Delta\alpha) + \frac{1}{2} [\Delta(1-\beta) \cdot \Delta\alpha] \right\} \tag{3.5}$$

In Equation 3.5 i is the number of decision thresholds, i.e., the number of points for which the ROC was plotted; $\Delta(1-\beta) = (1-\beta_i) - (1-\beta_{i-1})$; $\Delta\alpha = \alpha_i - \alpha_{i-1}$.

Precision represents the fraction of cases where the algorithm correctly predicts a class, from all instances where it predicted that class, either correctly or incorrectly. On the other hand, recall is the fraction of instances where a class is correctly predicted, from all of the samples labelled as belonging to that class. So, recall and precision give us information regarding the classifier's performance with respect to FN and FP, respectively. Thus, the use of these two metrics allows us to have a better overview of the model regarding miss-classifications, which can be helpful when we are focused in reducing the number of FN and/or FP. On the other hand, the AUC is a measurement of model performance at various threshold values, representing the capability of the model in distinguishing between classes. The curve is a plot of the false positive rate against the true positive rate. In multi-class problems like the one we stand before in this work, one ROC curve can be plotted for each class, using the one-against-all classification rule.

3.3.2 VEC Images Segmentation

Similarly to what was described previously for the classification task, the model of the validation epoch with lowest loss is saved for evaluation in the test set.

To evaluate the algorithm performance, Jaccard index, also known as IoU(Equation 3.6), and Dice coefficient (Equation 3.7) were used.

$$IoU(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{3.6}$$

$$Dice(A,B) = \frac{2|A \cap B|}{|A| + |B|} \tag{3.7}$$

IoU quantifies the percentage of overlap between the ground truth mask and the prediction output of the model, relatively to the union of the target and predicted masks. Dice coefficient between two binary masks measures their intersection, scaled by their size.

These metrics will be computed for every mask resultant from the setting of each threshold value between 0 and 1 to the segmentation map, in order to determine the one that creates binary masks more similar to the ground truth ones, resulting in higher IoU and dice values. This threshold value defines if a given pixel of a segmentation map is set to either 0 or 1, given the pixel probability value. For object segmentation tasks, threshold value is many times set to 0.5. However, this approach is not suitable for all cases, thus the need to test several values in order to understand which is the value that gives the best results.

3.4 Summary

In this Section, an overview of the devised framework for the couple of tasks of this work was presented, taking into consideration the previously reviewed literature on the area (Chapter 2).

The implemented pipeline includes two main steps that are common to both VEC image classification and segmentation: pre-processing and data augmentation. The pre-processing step is of extremely importance given the need of enhancing the quality of VEC images and already proven aid in lesion detection tasks. Although pre-processing techniques applied to lesion segmentation have not been review, the implementation of this step should also lead to a higher performance of the models in this task. Furthermore, due to the need of having large training datasets when using deep learning approaches, a step of data augmentation was also included in the pipeline.

To address the lesion detection problem, neural networks like ResNet, DenseNet, VGG and others will be applied, employing transfer learning technique. For the segmentation of lesions within the VEC frame, U-Net, two variations of this network (TernausNet and AlbuNet) and a GAN will be implemented.

All in all, the presented approach tries to address the issues at hand through implementation of novel deep learning techniques which past reviews indicate promising results in this tasks.

Chapter 4

Results and Discussion

4.1 Dataset Description

The dataset used for both lesion detection and segmentation tasks was the one from GIANA Endoscopic Vision Challenge, described in detail in Section 2.6. It includes images of three balanced classes (normal, inflammatory, and vascular), regarding the presence and type of lesion within the frame, adding up to a total of 1812 VEC images. Concerning the abnormal frames, i.e. those classified as vascular or inflammatory, segmentation masks of the lesions are also provided. This data was randomly split into three balanced sets (training, validation and test), following the previously mentioned proportions (50%-25%-25%).

4.2 Image Pre-processing

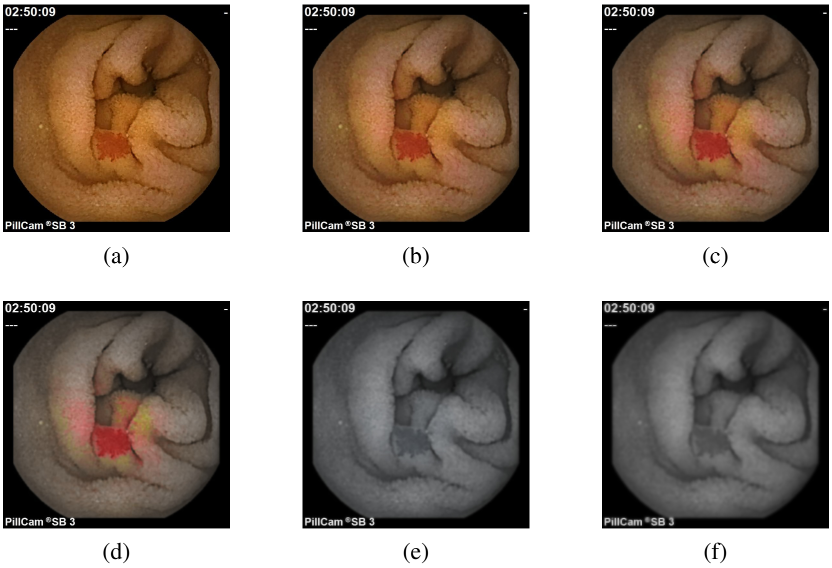
The evaluation of the results of the employment of the aforementioned image enhancement techniques was addressed into two stages: subject evaluation by visual analysis only and objective evaluation by analysis of the impact that the employment of the enhancement techniques has on the performance of the classification and segmentation tasks.

Adaptive Contrast Diffusion

The whole algorithm is independent from user-defined parameters, except for the Hessian matrix scale parameter σ and the number of iterations. Given that Li and Meng [35] proved that σ does not have any influence in the performance of the algorithm, this parameter was set to 1, in order to reduce the computational cost of the processing. Regarding the number of iterations, several values were tested in a small subset of the dataset. The algorithm was tested for 10, 20, 40, 60 and 80. After visual analysis of the results (Figure 4.1), 20 was defined as the most suitable number of iterations.

Homomorphic Filtering

In order to achieve the desired outcome (elimination of the illumination effect in the images), the



(a) (b) (c) (d) (e) (f)

Figure 4.1: Enhancement results of a VEC frame for different number of iterations of the adaptive contrast diffusion algorithm: (a) Original image; (b) 10 iterations; (c) 20 iterations; (d) 40 iterations; (e) 60 iterations; (f) 80 iteraions.

parameters in Equations 2.20 and 2.21 where defined as in Ramaraj et al. [53]. Some other values where tested for both high and low frequency gains, however $\alpha_L = 0.9$ and $\alpha_H = 1.75$ seemed to be the most suitable set of values.

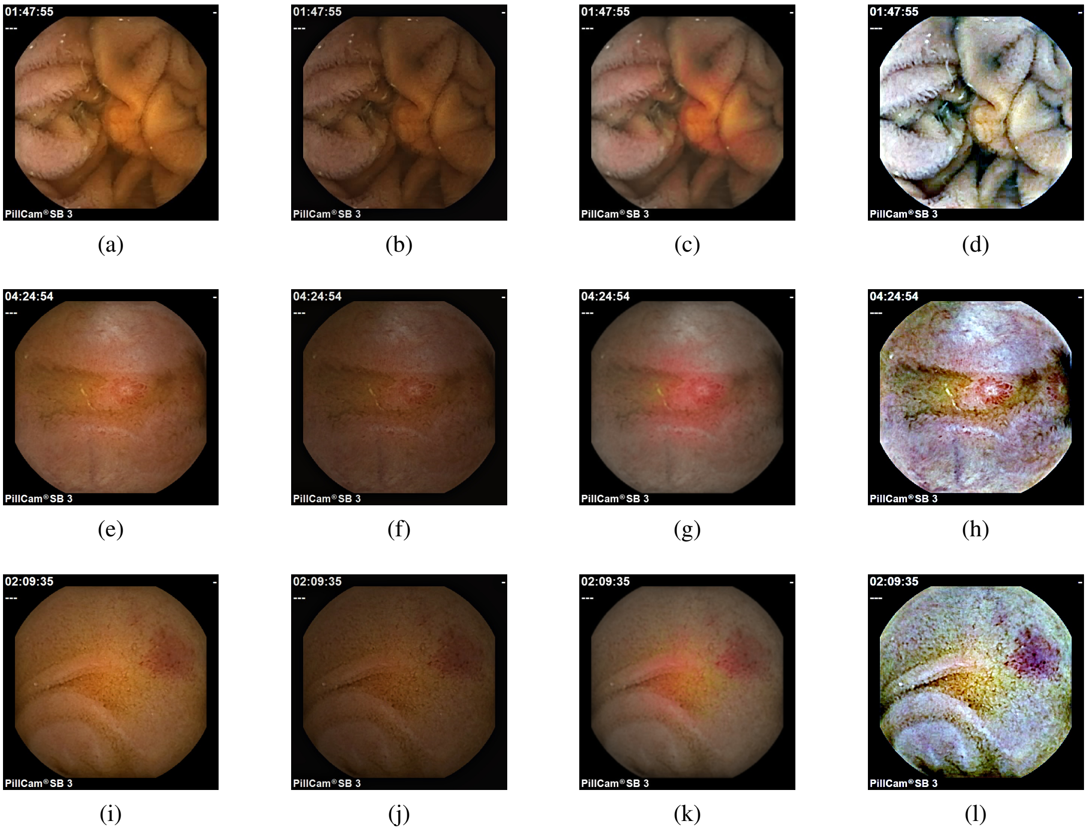
Automated Multi Scale Retinex with Colour Restoration

In similar fashion to what happens in the anisotropic diffusion filtering, the automated MSRCR method described in Section 2.4.3 does not rely on many fixed values set by the user since most parameters are computed by resorting to image information alone.

Similarly to what Parthasarathy and Sankaran [45] and Petro et al. [49] proposed, the scales σ were set to 15, 80 and 250. Finally, for the automatic settlement of the LCP and UCP of the MSR image histogram, the approach followed was the one that recurs to the frequency of occurrence of pixels rather that the one that resorts to the variance, given the already presented flaws of this second procedure.

Some examples of the influence of image enhancement by these methods can be seen in Figure 4.2. In order to have a better idea of the outcome in each type of images, examples of all three types of frames (normal, inflammatory and vascular) are presented, each along with the result of each one of the three processing techniques.

Regarding the homomorphic filtering, we can clearly see that the goal of reducing the effect of the illumination component of the image is well achieved. The overall image is much more uniform when concerning light distribution. However, the darkening of the frame can be an adverse effect of the processing since it can hinder the detection of potential lesions present in the frame.



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure 4.2: Effect of the applied pre-processing techniques on dataset images: (left to right) normal frame and pre-processed frames with homomorphic, adaptive contrast diffusion and automated MSRCR filtering; (top to bottom) normal, inflammatory and vascular frame.

For example, the light homogenisation provided by the homomorphic filtering may be beneficial in the normalisation of the image colouration; however, the darkening of the overall image may conceal some lesions. This can be verified by taking a close look at Figure 4.2j, where it is evident that the global darkening of the image makes it harder to identify the vascular lesion present in the frame.

On the other hand, the anisotropic diffusion filtering appears to have a contrasting effect of the one produced by the homomorphic technique. In this case, the regions where the light intensity is major (usually close to the centre of the frame) appear quite pinkish, whereas the surrounding area seems reasonably uniform concerning both light distribution and colour variation. Besides that, the darkening effect is not as severe as in the homomorphic filtering, which can represent an advantage over the previous method. The effect of the anisotropic diffusion filtering may be uncertain and dependant of the type and location of the lesions within the frame, if there is any. The frame presented in Figure 4.2g contains an inflammatory lesion in the centre region of the frame. The lesion clearly stands out from the surrounding background area, which can be favourable for the identification of the abnormality. However, in cases where the abnormal regions is not the enhanced one (Figure 4.2k), this processing may be harmful to the correct identification of the lesion.

The automated MSRCR processing is the one that most distorts the original colouration of the frame. The so distinctive colour of the intestine walls becomes much more bright and white, while pink/red areas are quite enhanced. This leads to a higher contrast between the background and the lesions, which can be very beneficial, given that the point of this pre-processing step is to obtain images where the lesions can be more easily identified. Furthermore, in Figure 4.2h, we can notice that the applied processing was able to enhance a small inflammatory lesion on the right side of the frame that was barely visible in the original image.

All in all, the implemented pre-processing techniques applied to VEC frames provided quite distinctive outputs, that can lead to contrasting outcomes in the classification and segmentation tasks. The effect of these techniques in the performance of the developed algorithms will be evaluated in the following sections.

4.3 Lesion Detection in VCE frames

As described in Section 3.2.2, the implemented methodology for the development of a VEC image classification algorithm combined the adoption of several pre-trained models to be applied to this task: ResNet, DenseNet, VGG, Inception V3, AlexNet and SqueezeNet.

In order to have an overview of the overall performance of the implemented architectures in this task, a prior evaluation was made before further progression in the pipeline. This step will allow a previous selection of the models most suitable to perform this task and reduce the time of computational implementation in the subsequent steps.

The aforementioned pre-trained models where finetuned and evaluated in the selected dataset, with no pre-processing applied. The computed evaluation metrics obtained by each model in the test set lead to the exclusion of AlexNet and SqueezeNet from the set of models suitable for this problem, given the poor performance that was achieved by these two architectures, in comparison to the remaining ones (see Table 4.1). AlexNet was the first CNN to be successfully applied to a big dataset in an image classification task by winning by a large margin the ImageNet LSVRC-2010 challenge Krizhevsky et al. [28]. The achieved performance derived from the addition of dropout, softmax and max-pooling layers in the architecture, along with the replacement of the sigmoid activation function for the ReLU. At the time, this network was much larger than previous developed CNN used in machine learning tasks. However, when compared to modern architectures the AlexNet model is relatively simple, thence the low performance obtained when compared against recent models. The low performance of the SqueezeNet may be a consequence of the use of 1×1 filters instead of 3×3 filters. The fact that the squeeze layer of the fire module only contains 1×1 filters hinders the ability of spatial abstract, which does not happen when 3×3 filters are used, since these are capable of capturing spatial information of pixels in the neighbourhood. Moreover, the decreased number of feature maps in the 3×3 filters of the expand layer can limit the information flow along with the network.

With respect to all models, the ones that achieved the best results were those where all weights were updated during the training phase, instead of only the final layers. These results were quite predictable, since finetuning the whole model is beneficial as it allows the network to improve the feature extraction step as well. After evaluation on the test set, precision, recall and AUC metrics were computed for these best-succeeded algorithms. The obtained results can be seen in Table 4.1.

Table 4.1: Results on lesion classification in VEC frames from GIANA dataset obtained by the best-succeeded models on the test set without pre-processing.

Model	**Precision**	**Recall**	**AUC**
AlexNet	0.73	0.73	0.77
SqueezeNet	0.78	0.78	0.61
ResNet-18	0.92	0.90	0.81
ResNet-34	0.92	0.92	0.86
ResNet-50	0.93	0.92	0.84
ResNet-101	0.94	0.93	0.86
ResNet-152	0.93	0.93	0.88
DenseNet-121	0.94	0.93	0.88
DenseNet-161	0.94	0.94	0.88
DenseNet-169	0.94	0.93	0.85
DenseNet-201	0.93	0.92	0.84
Inception V3	0.93	0.92	0.86
VGG11	0.90	0.90	0.83
VGG11 [3]	0.94	0.94	0.89
VGG13	0.76	0.75	0.68
VGG13 [3]	0.94	0.93	0.86
VGG16	0.72	0.72	0.54
VGG16 [3]	**0.95**	**0.95**	**0.90**
VGG19	0.79	0.79	0.63
VGG19 [3]	0.94	0.93	0.87

Best performing algorithms include Inception V3 model, as well as ResNet, DenseNet and VGG families of networks. Afterwards, with the ambition of improving the performance of the VEC classification algorithm, the above-presented models (with exception to AlexNet and SqueezeNet) were also trained and evaluated in datasets where the aforementioned pre-processing techniques were employed. Equivalently to what was done previously, classification metrics were computed in order to assess the performance of the models in enhanced VEC datasets. The results can be found in Table 4.2.

A careful evaluation of Tables 4.1 and 4.2 allows us to draw some conclusions regarding the implemented methodology towards VEC images classification:

- One phenomenon that can be noticed is that for model families where architectures of different depths were tested (ResNet, DenseNet and VGG), the ones with the highest number of layers do not always achieve the best performance, which is suggestive of model overfitting

[3] VGG model with batch normalisation

Table 4.2: Results on lesion classification in VEC frames from GIANA dataset obtained by the best-succeeded models on the test set with the three pre-processing techniques employed.

Model	Homomorphic			Anisotropic diffusion			Automated MSRCR		
	Precision	Recall	AUC	Precision	Recall	AUC	Precision	Recall	AUC
ResNet-18	0.92	0.91	0.82	0.88	0.87	0.80	0.90	0.89	0.78
ResNet-34	0.92	0.91	0.81	0.91	0.91	0.84	0.90	0.89	0.79
ResNet-50	0.94	0.94	0.88	0.89	0.89	0.82	0.92	0.92	0.84
ResNet-101	0.94	0.94	0.89	0.90	0.88	0.75	0.93	0.92	0.82
ResNet-152	0.90	0.89	0.79	0.92	0.91	0.82	**0.94**	**0.94**	**0.88**
DenseNet-121	0.93	0.93	0.87	0.90	0.87	0.75	0.91	0.90	0.79
DenseNet-161	**0.95**	**0.94**	**0.88**	**0.92**	**0.91**	**0.83**	0.90	0.89	0.80
DenseNet-169	0.93	0.93	0.88	0.91	0.89	0.79	0.92	0.92	0.84
DenseNet-201	0.92	0.90	0.80	0.91	0.90	0.82	0.92	0.91	0.81
Inception V3	0.92	0.91	0.83	0.90	0.89	0.80	0.93	0.93	0.87
VGG11	0.77	0.77	0.64	0.84	0.83	0.79	0.88	0.87	0.76
VGG11 [4]	0.93	0.92	0.84	0.89	0.87	0.76	0.92	0.91	0.81
VGG13	0.79	0.79	0.66	0.91	0.90	0.81	0.88	0.87	0.75
VGG13 [4]	0.89	0.87	0.74	0.91	0.90	0.84	0.92	0.91	0.82
VGG16	0.73	0.74	0.58	0.90	0.88	0.79	0.90	0.89	0.77
VGG16 [4]	0.92	0.91	0.81	0.90	0.88	0.78	0.92	0.92	0.84
VGG19	0.71	0.70	0.51	0.90	0.72	0.54	0.88	0.87	0.78
VGG19 [4]	0.93	0.92	0.85	0.90	0.90	0.81	0.91	0.91	0.83

to the training data. To address this issue, dropout layers and weight decay were applied to the models. Despite this approach, the effect of the overfitting was not hindered;

- Focusing our attention in the VGG results, it becomes clear that the incorporation of a batch normalisation step into the models architectures has a major positive influence in the algorithm performance;

- Considering each one of the four situations in which the algorithms were tested (without any pre-processing and with each one of the three implemented VEC enhancement techniques), VGG16 applied in the raw dataset was the best-succeeded one, achieving 0.95 of precision and recall, and 0.90 of AUC. Nevertheless, if we take a close look at the best performances achieved with resource to the pre-processed data, we can conclude that no significant discrepancies can be observed. This is especially true for the homomorphic and automated MSRCR techniques for which the difference between the performance metrics is insignificant (maximum variation of 0.01 for precision and recall and 0.02 for AUC). Taking these commentaries into consideration, we may assume that the pre-processing step of VEC frames enhancement does not have a noticeable impact, and can even be harmful in the classification of endoscopic images;

- We can also notice that the obtained results for the classification metrics are significantly similar between the four types of networks, if the VGG models without batch normalisation are not taken into consideration, i.e. despite presenting quite distinctive architectures, the several models are able to achieve similar performances. Besides that, among the four

[4] VGG model with batch normalisation

different conditions in which the models were tested, there is not any model that stands out from the remaining ones. In fact, while without any kind of pre-processing the best-succeeded model is a VGG, when the images suffer homomorphic or anisotropic diffusion filtering the model that performs the better is a DenseNet, and for the automated MSRCR enhancement, a ResNet is the model that returns the best results.

Since there are no published works that take into consideration a 3-class classification like the one here presented, we can consider the binary classification results of the best-succeeded model presented in Table 4.3, allowing a comparison with reviewed literature presented in Section 2.5.2.1 and summarised in Table 2.1.

Table 4.3: Binary lesion classification results obtained by the best-succeeded models.

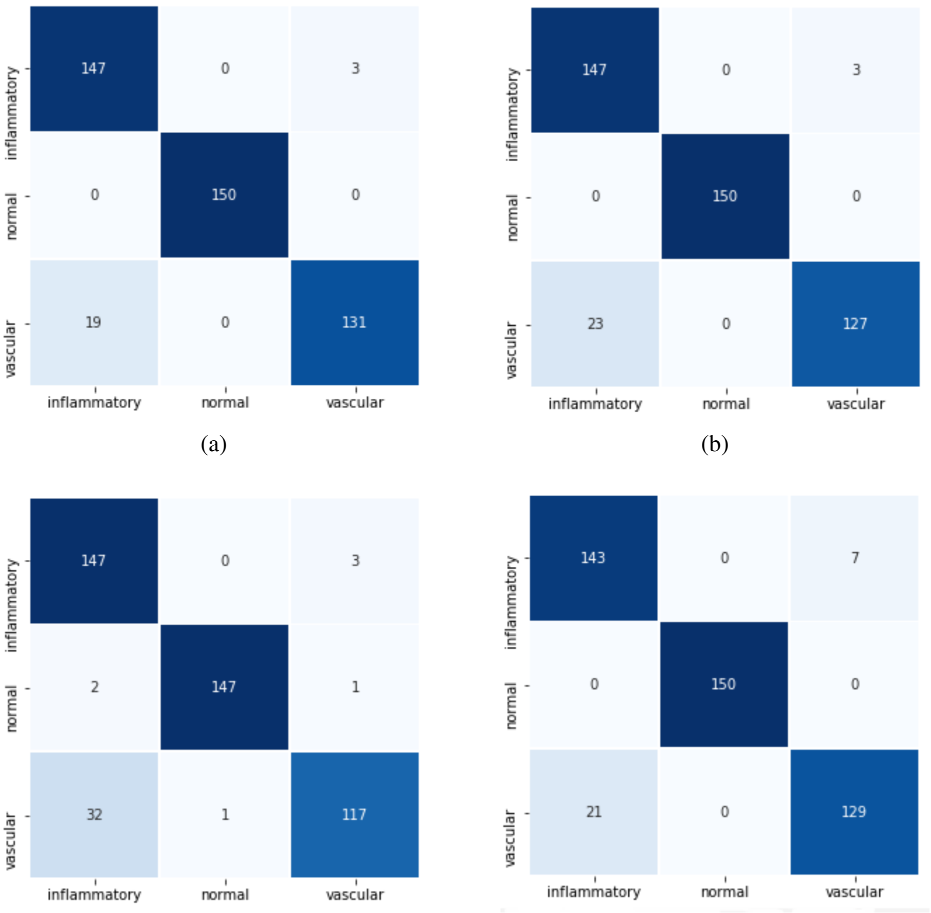
Classification	Precision	Recall	Accuracy	Specificity	AUC
Abnormal	1.00	1.00	1.00	1.00	1.00
Vascular	0.98	0.87	0.95	0.69	0.99
Inflammatory	0.89	0.98	0.94	0.66	0.99

Regarding a classification of VEC frames into normal and abnormal, the proposed method is able to achieve results without any FP or FN, leading to an ideal performance that surpasses other works on the literature. Concerning the detection of vascular lesions in endoscopic frames, the obtained results are quite satisfactory; however not as good as others like the ones described in [36, 50]. Notwithstanding, the works described in [36] and [50] take into account only one specific type of lesion (bleeding and angiodysplasia, respectively), whereas in the proposed methodology several types of vascular lesions are taken into consideration. Finally, concerning the detection of inflammatory lesions, the presented work surpasses others like the one described in [10], also with the advantage of considering many types of lesions.

Recall, precision and AUC are classification metric that only provides a general overview of the algorithm's performance. In order to better understand the model's behaviours in the dataset, we can resort to the confusion matrix. In Figure 4.3 the confusion matrix of the best-succeeded model is presented.

A close analysis of the presented confusion matrices shows that images belonging to the normal class ate not erroneously classified, except in the anisotropic diffusion filtering where the DenseNet161 classified two normal frames as inflammatory and one as vascular. Moreover, this model also classifies vascular frame as normal. This means that practically all images labelled as normal are classified by the algorithms as belonging to that same class and, besides that, the models do not miss-classify other images into this class.

Regarding images with inflammatory lesions, VGG16 (applied on the raw dataset) and DenseNet161 (tested on images enhanced by homomorphic and anisotropic diffusion filterings) only miss-classify 3 frames as vascular, whereas the ResNet152 implemented on frames filtered with automated MSRCR, 7 inflammatory frames are classified as belonging to the vascular class. This



(a) (b)

(c) (d)

Figure 4.3: Confusion matrices obtained by the best-succeeded models of each pre-processing technique: (a) VGG16 with batch normalisation in the dataset without pre-processing; (b) DenseNet161 in the dataset enhanced by homomorphic filtering; (c) DenseNet161 in the dataset enhanced by anisotropic diffusion filtering; (d) ResNet152 in the dataset enhanced by automated MSRCR filtering. Left labels indicate the ground truth of the classification and bottom labels represent the model predictions.

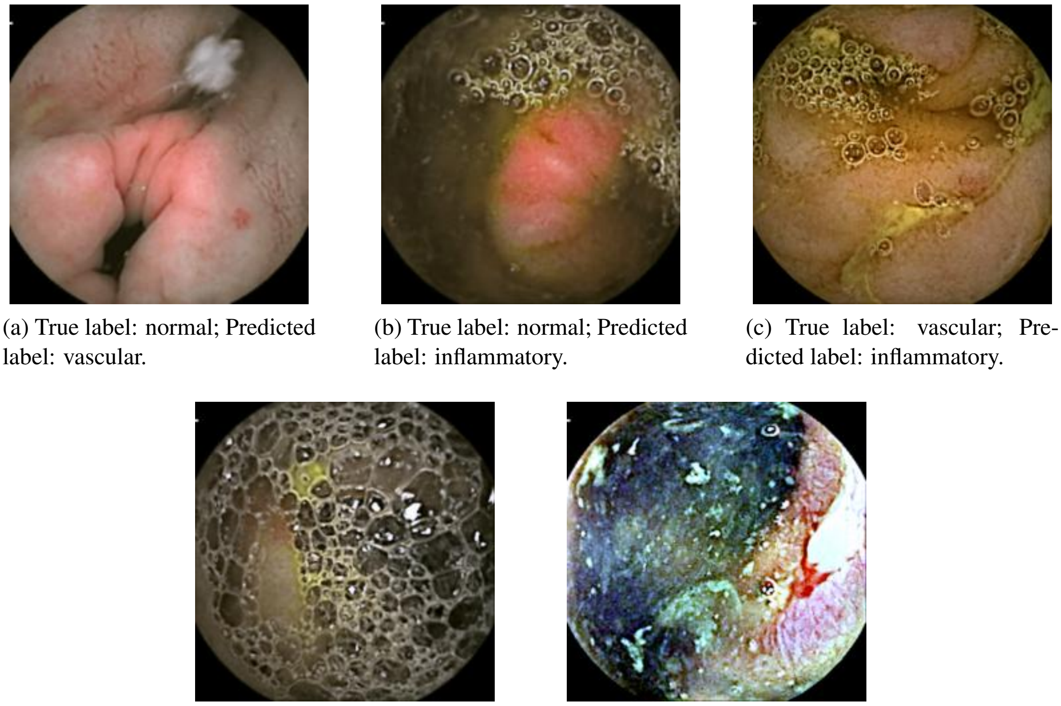
can mean that the automated MSRCR has a negative impact on the inflammatory frames, causing the model to incorrectly classify some of them as vascular.

Miss-classifications the other way around, i.e. vascular frames classified by the models as inflammatory, are more common than the previous: VGG16 trained in images without pre-processing outputs 19 of those miss-classifications; ResNet152 on the automated MSRCR pre-processing presents 21 similar cases; DenseNet161 applied on the datasets enhanced by homomorphic and anisotropic diffusion filterings respectively classifies 23 and 32 vascular frames as inflammatory.

These results prove to be extremely positive, given the intended application. With practically no miss-classifications concerning the normal images, the employment of this algorithm in a real situation would allow experts to analyse only the images classified as having a lesion, without the risk of overlooking an important lesion.

In Figure 4.4 some examples of miss-classificated VEC frames are presented. Figure 4.4a is the only normal frame that was mistakenly classified as vascular. Despite the assigned label of this frame is normal, the model prediction can be justified by the presence of a reddish area in the bottom right quarter of the image, that resembles a vascular abnormality. The normal VEC frame presented in Figure 4.4b was erroneously classified as an inflammatory one. Images similar to the one here presented represent a drawback for algorithms that aim the automatic classification of images, due to the high amount of GI content in the image, which minimises the visible area of the GI tract walls. Figures 4.4c and 4.4d are both vascular but were miss-classified as inflammatory and normal, respectively. In both cases, the visualisation of a vascular lesion is hardly possible. Besides that, Figure 4.4c presents some artefacts that resemble inflammatory lesions, namely ulcers. Finally, Figure 4.4e is an inflammatory lesion that contains blood as well, which may be a reasonable explanation for the classification of the frame as vascular. All miss-classified frames were shown to a gastroenterologist in order to get an experts opinion on the matter. This doctor confirmed that among the 22 erroneously classified images by the model, 5 of them have both lesion present, and 7 others should not be taken into consideration for diagnosis since they present low quality and obstructions that hamper the visualisation of the GI walls.

All in all, the trained models are able to correctly distinguish normal frames from images with lesions, but discrimination between vascular and inflammatory abnormalities still has margin for improvement.



(a) True label: normal; Predicted label: vascular.

(b) True label: normal; Predicted label: inflammatory.

(c) True label: vascular; Predicted label: inflammatory.

(d) True label: vascular; Predicted label: normal.

(e) True label: inflammatory; Predicted label: vascular.

Figure 4.4: Examples of misclassified VEC frames.

4.4 Lesion Segmentation in VCE frames

As described in Section 3.2.3, the followed pipeline for the development of an algorithm for lesion segmentation in VEC image included the implementation of several networks for semantic segmentation (U-Net, AlbuNet, TernausNet and GAN). These models were separately trained for vascular and inflammatory lesions, following the same approach, that was applied either in the original dataset, either in the pre-processed images. Obtained results of vascular and inflammatory lesions segmentation can be seen in Tables 4.4 and 4.5, respectively, where the best Jaccard index and Dice coefficient are presented, along with the threshold value that outputs that results.

Table 4.4: Results on vascular lesions segmentation in VEC frames from GIANA dataset.

Model	Pre-processing	Threshold	Jaccard	Dice
U-Net	**No pre-processing**	**0.55**	**0.821**	**0.902**
	Homomorphic	0.73	0.713	0.832
	Anisotropic diffusion	0.51	0.782	0.878
	Automated MSRCR	0.51	0.785	0.880
AlbuNet34	No pre-processing	0.78	0.701	0.824
	Homomorphic	0.82	0.767	0.868
	Anisotropic diffusion	0.76	0.698	0.822
	Automated MSRCR	0.69	0.795	0.886
TernausNet11	No pre-processing	0.86	0.777	0.875
	Homomorphic	0.88	0.767	0.868
	Anisotropic diffusion	0.56	0.693	0.818
	Automated MSRCR	0.70	0.790	0.883
GAN	No pre-processing	0.51	0.806	0.893
	Homomorphic	0.53	0.758	0.862
	Anisotropic diffusion	0.51	0.799	0.888
	Automated MSRCR	0.51	0.768	0.869

Several conclusions can be drawn from the careful analysis of Tables 4.4 and 4.5:

- Jaccard index and Dice coefficient obtained for the vascular lesions are much higher than the ones resulting from the segmentation of inflammatory ones. In fact, the segmentation of inflammatory lesions was not possible for most models and pre-processing techniques. Only two decent results were obtained by the employment of a GAN in the datasets enhanced by anisotropic diffusion and automated MSRCR techniques. This results from the fact that inflammatory lesions are much harder to identify in VEC frames due to the lighter colouration they present, sometimes quite similar to the surrounding intestinal wall, as well as faded contours which can hamper the precise outline of the lesion;

- Overall, U-Net and GAN models surpass both Albunet and TernausNet. The better performance of U-Net when compared to the two modifications of this network shows that the use of pre-trained encoders in U-Net like architectures is not benefic for this task, despite proven advantages demonstrated in other works;

Table 4.5: Results on inflammatory lesion segmentation in VEC frames from GIANA dataset[5] .

Model	Pre-processing	Threshold	Jaccard	Dice
U-Net	No pre-processing	-	NR	NR
	Homomorphic	-	NR	NR
	Anisotropic diffusion	0.51	0.239	0.386
	Automated MSRCR	0.51	0.192	0.322
AlbuNet34	No pre-processing	-	NR	NR
	Homomorphic	-	NR	NR
	Anisotropic diffusion	-	NR	NR
	Automated MSRCR	-	NR	NR
TernausNet11	No pre-processing	-	NR	NR
	Homomorphic	-	NR	NR
	Anisotropic diffusion	-	NR	NR
	Automated MSRCR	0.22	0.275	0.432
GAN	No pre-processing	-	NR	NR
	Homomorphic	0.51	0.166	0.284
	Anisotropic diffusion	0.51	0.386	0.557
	Automated MSRCR	**0.51**	**0.602**	**0.751**

- The threshold value that outputs the most accurate segmentation masks is much more variable in vascular lesions (between 0.51 and 0.88) than in the inflammatory ones (0.51 for most cases where reasonable Jaccard and Dice results are obtained). The variation of values that occurs in the vascular lesions may derive from the fact that these abnormalities are usually darker than the background area, thus the higher threshold value. Furthermore, we can see that regarding the vascular lesions there is not any pattern in the threshold values for the best segmentation metrics, concerning either the employed method or the processing technique applied to the images;

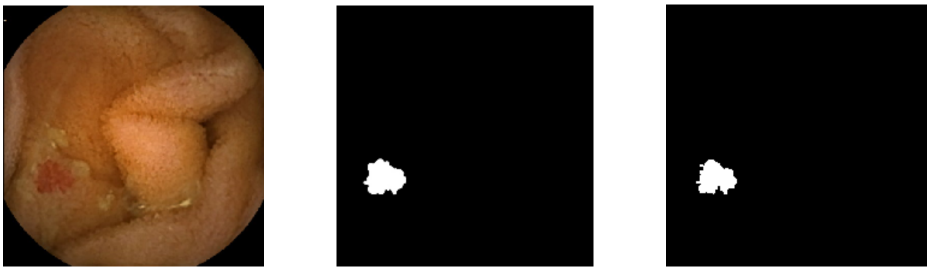
- Regarding the impact of the pre-processing techniques in the models efficiency in lesion segmentation, distinctive outcomes can be observed:

  – When concerning vascular lesions, the best-succeeded model was the U-Net applied to the raw dataset. For both TernausNet and GAN, the best segmentations were also obtained in the images without any pre-processing, meaning that the applied enhancement techniques did not prove to be effective in the improvement of methods for the segmentation of vascular lesions. This may be due to the fact that vascular lesions are generally not very difficult to identify given the usual reddish colouration, and the employment of enhancement techniques may distort the real appearance of the abnormalities in relation to the background, in such way that the recognition of the abnormality becomes harder to perform. Figures 4.2i-4.2l depict an example of an image containing a vascular lesion that may be harmed by the employment of pre-processing techniques;

[5] Classification metrics for cases where lesion segmentation was not possible were not considered relevant and presented in the table as NR (Not Relevant).

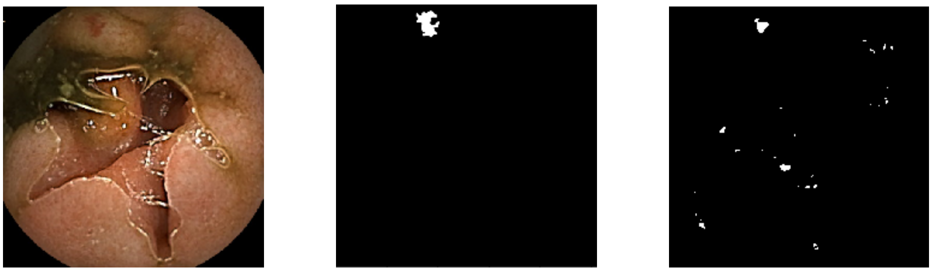
- Regarding inflammatory lesions, the conclusions drawn are quite the opposite. Given the few decent segmentation results obtained, none of them was obtained with the employment of a model to the images without any enhancement technique applied. The best segmentations obtained were the ones that used the GAN and the automated MSRCR pre-processing outperformed any other results. This proves that the employment of a pre-processing step is crucial for the segmentation of inflammatory lesions, unlike what happens for the vascular ones.

Figures 4.5-4.8 show some result segmentations obtained with the respective best-succeeded approach (implemented model and pre-processing technique) for each type of lesion.



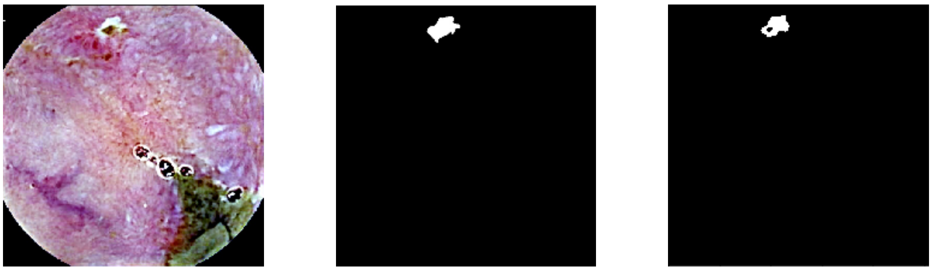
(a) Original frame. (b) Ground truth mask. (c) Predicted mask.

Figure 4.5: Example of a well-achieved segmentation of a vascular lesion.



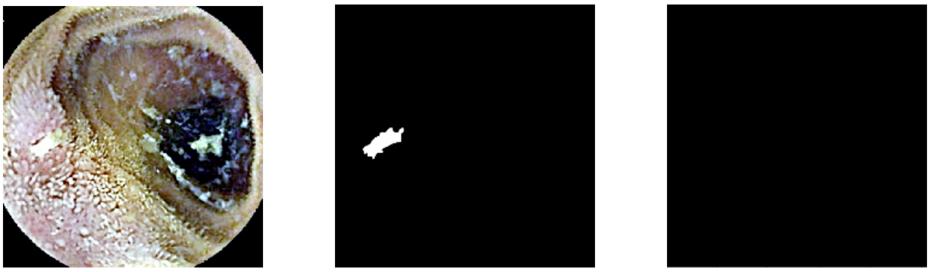
(a) Original frame. (b) Ground truth mask. (c) Predicted mask.

Figure 4.6: Example of a poor-achieved segmentation of a vascular lesion.



(a) Original frame. (b) Ground truth mask. (c) Predicted mask.

Figure 4.7: Example of a well-achieved segmentation of an inflammatory lesion.



(a) Original frame. (b) Ground truth mask. (c) Predicted segmentation.

Figure 4.8: Example of a poor-achieved segmentation of an inflammatory lesion.

Some quite satisfactory segmentation results were obtained, like the ones depicted in Figures 4.5 (vascular lesion) and 4.7 (inflammatory lesion), that achieved a Jaccard index of 0.882 and 0.754, and a dice coefficient of 0.937 and 0.860, respectively. In both cases the overall contour shape of the lesion is quite well-accomplished and there are no other regions besides the abnormal one being classified as such. However, the predicted segmentation of the inflammatory lesion presents a small area within the region identified as lesion that is considered normal. In the VEC frame this region clearly has an appearance quite distinguishable from the lesion, which may mean one of tho things: either the ground truth segmentation mask was performed in a somewhat coarse manner and did not take into account that small region or it actually belongs to the lesion and the algorithm was not able to identify it as such.

On the other hand, some segmentation results fell short on the expectations. Figures 4.6 and 4.8 represent some not so well-achieved vascular and inflammatory lesions segmentations. Concerning the vascular one, two main issues stand out: the under segmentation of the lesion area and the recognition of some healthy regions as abnormal. The imprecise identification of the lesion contour may be owed to the defective illumination conditions of that region of the image. Furthermore, the regions erroneously identified as abnormal correspond to light glimmer present in the VEC frame.

All things considered, quite satisfactory results were obtained for the segmentation of lesions in VEC frames. Nonetheless, there is still margin for improvement, especially for the inflammatory lesions.

4.5 Models Validation in Private Dataset Frames

In order to get a further evaluation of the previously presented algorithms for both abnormalities detection and segmentation, these were tested on a different set of images apart from the GIANA dataset. To do so, the private dataset described in Section 2.6 was used. To allow a fair comparison with the previous resutls, from all 449 VEC available, only the ones recorded with PillCam$^{®}$SB3 were taken into consideration. From 17 of those videos of different patients, a set of set of 643 frames (250 normal, 154 vascular and 239 inflammatory) was selected. The labelling of each image was done with resort to the medical report of each exam, that has annotations regarding the presence of lesions in the video, and later validated by a gasteroenterologist. Furthermore, for the frames where lesions are present, a manual annotation of the localisation of the abnormalities was performed using Sensarea software tool [46]. These segmentations where also made under the supervision of the same gastroenterologist.

The newly formed dataset was then used for a new validation of the proposed models for both tasks at hand. Regarding the lesion detection task, the obtained results are presented on Table 4.6, whereas the results for abnormalities segmentation can be seen in Table 4.7.

Table 4.6: Results on lesion classification in VEC frames from private dataset.

Model	Precision	Recall	AUC
ResNet-50	0.75	0.75	0.63
ResNet-101	**0.76**	**0.75**	**0.64**
DenseNet-169	0.58	0.50	0.47
DenseNet-201	0.55	0.57	0.59

From all previously trained models for the classification task, only a few of them (the ones presented in Table 4.6) were able to get considerable results (metric values above 0.5). The previously best-succeeded model (VGG16 with batch normalisation) is not among the ones that achieved the best results in the private dataset, neither is any architecture besides ResNet or DenseNet. This can be indicative of VGG16 and other models overfitting to the GIANA dataset. On the other hand, ResNet and DenseNet architectures are somewhat capable of surpassing this problem, with special highlight to ResNet-101 that is capable of achieving precision, recall and AUC of 0.76, 0.75 and 0.64, respectively. Although there is a meaningful decrease when in comparison to the previous results, we must take into consideration that these images belong to VEC exams from a private dataset, instead of a carefully designed dataset for lesions classification like GIANA.

Table 4.7: Results on vascular lesion segmentation in VEC frames from private dataset.

Model	Threshold	Jaccard	Dice
U-Net	**0.55**	**0.724**	**0.840**
AlbuNet34	0.78	0.709	0.794
TernausNet11	0.86	0.712	0.832
GAN	0.51	0.600	0.750

Regarding the obtained results on the private dataset images for lesion segmentation, only considerable outcomes were achieved for the vascular lesions. In this case there is a decrease in both Jaccard and Dice metrics; however the results are still quite satisfactory. Without discarding the hypothesis of overfitting to the GIANA dataset, we should also take into consideration that the segmentation masks where obtained in different conditions, which can justify some variations of the metrics.

4.6 Summary

This work focused on two main activities: detection and segmentation of lesions in VCE images, using deep learning techniques. The methods implemented for lesion detection proved to be very successful, although there is no term of comparison in the literature, since we are using a 3 class classification, not seen before. Notwithstanding, the best-tested model was able to achieve precision and recall of 0.95 and AUC of 0.90. However, there is still room for improvement, given that the system is not fully capable of distinguishing between the two types of lesions considered and that some overfitting on the training data was observed.

The segmentation step also has some margin for progress, especially when concerning the inflammatory lesions. Although mild success was obtained when analysing the vascular lesions (0.821 of Jaccard and 0.902 of Dice), the inflammatory lesions proved much harder to segment in the dataset provided (obtaining only 0.602 of Jaccard and 0.751 of Dice).

The validation of models for both tasks in a different dataset showed a decrease in the performance in both cases. This can be indicative of overfitting to the GIANA dataset, leading to decreased performance in others.

Chapter 5

Conclusions and Future Work

Since its approval by the FDA in 2001, WCE has been used as a GI tract imaging method that, unlike traditional approaches, allows the visual analysis of the whole of the GI tract, including the SB. The capsule itself resembles a pill, which makes it easy for the patient to swallow it, not causing any discomfort during the whole procedure. When performing an endoscopy, one of the procedures of upmost importance is the identification of lesions along the GI tract. This already challenging task, associated to the enormous quantity of data produced by a single WCE exam, makes the analysis of this data by an expert very tedious and prone to errors.

With this in mind, this dissertation aimed to the development of a CAD based approach for the automatic detection and segmentation of lesions of the GI tract in VEC, using deep learning methods. The resultant method should be able to classify a given VEC image into one of three classes (normal, vascular, inflammatory), regarding the presence and type of lesion in the frame, and localise the lesions at the pixel-level.

The reviewed literature on these subjects proved to be very enlightening regarding the understanding of emerging techniques applied in this area. The ever-growing application of machine learning, and especially deep learning techniques, to a wide range of problems, has proven to be effective in medical applications, including tasks for lesion detection and segmentation in images of VEC. These approaches outperformed other non deep learning based methods, thus encouraging the employment of deep learning algorithms.

Particularly for the lesion detection problem, transfer learning has shown very promising results in this subject, making it the selected approach to be followed for this task. For biomedical image segmentation, CNN architectures like U-Net and GAN algorithms have been widely used, including in segmentation of lesions in VEC frames tasks. Given the not ideal quality of VEC images and low illumination conditions, the implementation of a pre-processing step was included in the implemented pipeline.

5.1 Final Remarks

One of the goals of the proposed dissertation seeked to address the detection of lesions in VEC frames. This task was approached by the implementation of a deep learning based approach, preceded by an image enhancement step, that classifies a given VEC image into one of three classes: normal if no lesion is detected in the frame, or vascular/inflammatory if a lesion of that type is detected. Any of the three implemented pre-processing techniques proved to favour the classification task, given that the best results were obtained on the raw dataset. The model that achieved those results was the VGG16 with batch normalisation, reaching 0.95 of precision and recall, and 0.90 of AUC. Although this model is able to correctly discriminate normal from abnormal frames, the classification of the type of lesion present in the abnormal ones can still be improved. However, it should be taken into consideration that some of the erroneous classified images are obscured by intestinal content, hindering the visualisation of the GI tract walls, and others seem to have present both types of lesions. Furthermore, comparison between studies is not possible, since no work has addressed a classification problem similar to this. Notwithstanding, the obtained results seem very promising but still have room for improvement.

Moreover, the work proposed in this dissertation also aimed to assess the segmentation of abnormalities in VEC frames. To tackle this issue, a pre-processing step for enhancement of the VEC images followed by implementation of neural networks for semantic segmentation was the followed approach. While for vascular lesions the implementation of the pre-processing step was not advantageous, for the segmentation of the inflammatory lesions this step was crucial. Regarding vascular lesions, a Jaccard of 0.821 and a Dice of 0.902 were obtained with the U-Net model applied to the raw dataset, while for the inflammatory lesions the best-succeeded architecture was a GAN implemented in the dataset enhanced by automated MSRCR, achieving a Jaccard of 0.602 and a Dice of 0.751. Concerning the segmentation of vascular lesions, this results can be compared with works that address the segmentation of one specific kind of vascular lesion, but no work has addressed the simultaneous segmentation of many types of vascular lesions. Although this comparison is difficult to perform, the proposed framework can surpass some similar works in the area. On the other hand, segmentation of inflammatory lesions has still not been addressed, which makes it impossible to make a comparison with state-of-the-art results.

All in all, the obtained results can be considered a great contribution to the field, especially taking into account the multi-class classification of VEC frames and segmentation of inflammatory lesions, two tasks that have still not been addressed by others. Nevertheless, some future work can still be performed in order to improve the results here presented.

5.2 Future Work

Work developed hereafter should not only aim for the improvement of the obtained results, but also seek for a further evaluation of the implemented methodologies. This is of upmost importance since the good performance of a model in a given dataset does not always mean that it will be equally successful in a real application, i.e. the developed methodologies for lesion detection and segmentation seem to have reasonably performance on the given dataset, but when applied to a real endoscopy video, the same may not happen. Given that we have access to a private dataset containing 449 endoscopic capsule videos, a further evaluation of the models could be performed with resort to some of this data. Moreover, given that some of these videos have annotations performed by an expert regarding the presence of abnormalities, these data could also be used to train the models with a higher amount of data in the future.

Concerning the classification task, two main issues should be addressed: given the possibility of having both vascular and inflammatory lesions present in the same VEC frame, the algorithm should be able to attribute two labels to the same image; the overfitting should be reduced by introducing some other changes in the NN architecture and model training in a larger dataset.

Regarding the segmentation of lesions, the need to set a threshold value to obtain the correspondent mask of the image is a major drawback of the implemented model. This problem can be overcome by the use of an adaptive threshold, i.e. the implementation of a step that finds the most suitable threshold value for a given image, depending only on the VEC frame. Finally, taking into consideration that the GAN was one of the models that showed the most promising results and that is the one that can suffer more modifications, future work should also focus on the improvement of this model performance. To tackle this issue, different generator and discriminator models should be implemented.

References

[1] Baxter, M. and Aly, E. H. (2010). Dieulafoy's lesion: current trends in diagnosis and management. *Annals of the Royal College of Surgeons of England*, 92(7):548–54.

[2] Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159.

[3] Chan, H. C.-h., Kim, D. H., Lung, P. F. C., Cheon, J. H., and Ng, S. C. (2015). Complications of Inflammatory Bowel Disease. In *Atlas of Inflammatory Bowel Diseases*, pages 175–186. Springer Berlin Heidelberg, Berlin, Heidelberg.

[4] Collins, J. T. and Badireddy, M. (2019). *Anatomy, Abdomen and Pelvis, Small Intestine*. StatPearls Publishing.

[5] Cryer, B. and Mahaffey, K. W. (2014). Gastrointestinal ulcers, role of aspirin, and clinical outcomes: pathobiology, diagnosis, and treatment. *Journal of multidisciplinary healthcare*, 7:137–46.

[6] Friedel, D., Modayil, R., and Stavropoulos, S. (2016). Colon Capsule Endoscopy: Review and Perspectives. *Gastroenterology Research and Practice*, 2016:1–6.

[7] Frøkjaer, J. B., Drewes, A. M., and Gregersen, H. (2009). Imaging of the gastrointestinal tract-novel technologies. *World journal of gastroenterology*, 15(2):160–8.

[8] Fu, Y., Zhang, W., Mandal, M., and Meng, M. Q.-H. (2014). Computer-Aided Bleeding Detection in WCE Video. *IEEE Journal of Biomedical and Health Informatics*, 18(2):636–642.

[9] Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., and Garcia-Rodriguez, J. (2017). A Review on Deep Learning Techniques Applied to Semantic Segmentation.

[10] Georgakopoulos, S. V., Iakovidis, D. K., Vasilakakis, M., Plagianakos, V. P., and Koulaouzidis, A. (2016). Weakly-supervised Convolutional learning for detection of inflammatory gastrointestinal lesions. In *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, pages 510–514. IEEE.

[11] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Nets. Technical report.

[12] Graaff, K. M. V. d. (1986). Anatomy and physiology of the gastrointestinal tract. *The Pediatric Infectious Disease Journal*, 5(1):11–16.

[13] Gueye, L., Yildirim-Yayilgan, S., Cheikh, F. A., and Balasingham, I. (2015). Automatic detection of colonoscopic anomalies using capsule endoscopy. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 1061–1064. IEEE.

[14] He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep Residual Learning for Image Recognition.

[15] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2016). Densely Connected Convolutional Networks. Technical report.

[16] Hwang, S., Oh, J., Cox, J., Tang, S. J., and Tibbals, H. F. (2006). Blood detection in wireless capsule endoscopy using expectation maximization clustering. volume 6144, page 61441P. International Society for Optics and Photonics.

[17] Iakovidis, D. K. and Koulaouzidis, A. (2014). Automatic lesion detection in wireless capsule endoscopy &amp;#x2014; A simple solution for a complex problem. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 2236–2240. IEEE.

[18] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and.

[19] Iddan, G., Meron, G., Glukhovsky, A., and Swain, P. (2000). Wireless capsule endoscopy. *Nature*, 405(6785):417–417.

[20] Iglovikov, V. and Shvets, A. (2018). TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation. Technical report.

[21] Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., and Bengio, Y. (2016). The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. Technical report.

[22] Jeon, H. K. and Kim, G. H. (2015). Endoscopic Management of Dieulafoy's Lesion. *Clinical endoscopy*, 48(2):112–20.

[23] Jhaveri, K., McSweeney, S., and ODonoghue, P. (2010). Current and emerging techniques in gastrointestinal imaging. *Journal of Postgraduate Medicine*, 56(2):109.

[24] Jobson, D., Rahman, Z., and Woodell, G. (1997a). A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7):965–976.

[25] Jobson, D., Rahman, Z., and Woodell, G. (1997b). Properties and performance of a center/surround retinex. *IEEE Transactions on Image Processing*, 6(3):451–462.

[26] Jung, Y. S., Kim, Y. H., Lee, D. H., Lee, S. H., Song, J. J., and Kim, J. H. (2009). Automatic patient-adaptive bleeding detection in a capsule endoscopy. volume 7260, page 72603T. International Society for Optics and Photonics.

[27] Kingma, D. P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization.

[28] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90.

[29] Laing, C. J., Tobias, T., Rosenblum, D. I., Banker, W. L., Tseng, L., and Tamarkin, S. W. (2007). Acute Gastrointestinal Bleeding: Emerging Role of Multidetector CT Angiography and Review of Current Imaging Techniques 1. *RadioGraphics*, 27.

[30] Land, E. H. (1986a). An alternative technique for the computation of the designator in the retinex theory of color vision (Mach bands). Technical report.

[31] Land, E. H. (1986b). Recent advances in retinex theory. *Vision Research*, 26(1):7–21.

[32] Land, E. H. and McCann, J. J. (1971). Lightness and Retinex Theory. *Journal of the Optical Society of America*, 61(1):1.

[33] Lee, N. M. and Eisen, G. M. (2010). 10 Years of Capsule Endoscopy: an Update.

[34] Leenhardt, R., Vasseur, P., Li, C., Saurin, J. C., Rahmi, G., Cholet, F., Becq, A., Marteau, P., Histace, A., Dray, X., Sacher-Huvelin, S., Mesli, F., Leandri, C., Nion-Larmurier, I., Lecleire, S., Gerard, R., Duburque, C., Vanbiervliet, G., Amiot, X., Philippe Le Mouel, J., Delvaux, M., Jacob, P., Simon-Shane, C., and Romain, O. (2018). A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointestinal Endoscopy*.

[35] Li, B. and Meng, M. Q.-H. (2012). Wireless capsule endoscopy images enhancement via adaptive contrast diffusion. *Journal of Visual Communication and Image Representation*, 23(1):222–228.

[36] Li, X., Zhang, H., Zhang, X., Liu, H., and Xie, G. (2017). Exploring transfer learning for gastrointestinal bleeding detection on small-size imbalanced endoscopy images. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1994–1997. IEEE.

[37] Luc, P., Couprie, C., Chintala, S., and Verbeek, J. (2016). Semantic Segmentation using Adversarial Networks. Technical report.

[38] Lusk, L. B., Reichen, J., and Levine, J. S. (1984). Aphthous Ulceration in Diversion Colitis: Clinical Implications. *Gastroenterology*, 87(5):1171–1173.

[39] Maconi, G., Carsana, L., Fociani, P., Sampietro, G. M., Ardizzone, S., Cristaldi, M., Parente, F., Vago, G. L., Taschieri, A. M., and Bianchi Porro, G. (2003). Small bowel stenosis in Crohn's disease: clinical, biochemical and ultrasonographic evaluation of histological features. *Alimentary Pharmacology and Therapeutics*, 18(7):749–756.

[40] Medtronic (2016). PillCam ™ COLON 2 System.

[41] Medtronic (2018). PillCam ™ SB 3 System - Advanced imaging and visualization of the small bowel.

[42] Moore, A., Allman, J., and Goodman, R. (1991). A real-time neural system for color constancy. *IEEE Transactions on Neural Networks*, 2(2):237–247.

[43] Noya, F., Alvarez-Gonzalez, M. A., and Benitez, R. (2017). Automated angiodysplasia detection from wireless capsule endoscopy. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3158–3161. IEEE.

[44] Pan, G.-b., Yan, G.-z., Song, X.-s., and Qiu, X.-l. (2010). Bleeding detection from wireless capsule endoscopy images using improved euler distance in CIELab. *Journal of Shanghai Jiaotong University (Science)*, 15(2):218–223.

[45] Parthasarathy, S. and Sankaran, P. (2012). An automated multi Scale Retinex with Color Restoration for image enhancement. In *2012 National Conference on Communications (NCC)*, pages 1–5. IEEE.

[46] Pascal Bertolino (2013). Sensarea | Pascal BERTOLINO software.

[47] Perez, L. and Wang, J. (2017). The Effectiveness of Data Augmentation in Image Classification using Deep Learning.

[48] Perona, P. and Malik, J. (1990). Scale-Space and Edge Detection Using Anisotropic Diffusion. Technical Report 7.

[49] Petro, A. B., Sbert, C., and Morel, J.-M. (2014). Multiscale Retinex. *Image Processing On Line*, 4:71–88.

[50] Pogorelov, K., Ostroukhova, O., Petlund, A., Halvorsen, P., de Lange, T., Espeland, H. N., Kupka, T., Griwodz, C., and Riegler, M. (2018). Deep learning and handcrafted feature based approaches for automatic detection of angiectasia. In *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 365–368. IEEE.

[51] PyTorch (2018). PyTorch master documentation — torchvision.models.

[52] PyTorch (2019). PyTorch master documentation — torch.nn.

[53] Ramaraj, M., Raghavan, S., and Khan, W. A. (2013). Homomorphic filtering techniques for WCE image enhancement. In *2013 IEEE International Conference on Computational Intelligence and Computing Research*, pages 1–5. IEEE.

[54] Regula, J., Wronska, E., and Pachlewski, J. (2008). Vascular lesions of the gastrointestinal tract. *Best Practice & Research Clinical Gastroenterology*, 22(2):313–328.

[55] Romain Leenhardt, A., Li, C., Koulaouzidis, A., Cavallaro, F., Cholet, F., Eliakim, R., Fernandez-Urien, I., Kopylov, U., McAlindon, M., Németh, A., Plevris, J. N., Rahmi, G., Rondonotti, E., Saurin, J.-C., Eugenio Tontini, G., Toth, E., Yung, D., Marteau, P., Dray, X., Romain, L., and Saint-Antoine, H. (2017). Nomenclature and semantic description of vascular lesions in small bowel capsule endoscopy: an international Delphi consensus statement *.

[56] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation.

[57] Shvets, A., Iglovikov, V., Rakhlin, A., and Kalinin, A. A. (2018). Angiodysplasia Detection and Localization Using Deep Convolutional Neural Networks.

[58] Siegel, R. L., Miller, K. D., and Jemal, A. (2017). Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians*, 67(1):7–30.

[59] Simadibrata, M. and Adiwinata, R. (2017). Precancerous Lesions in Gastrointestinal Tract. *The Indonesian Journal of Gastroenterology, Hepatology, and Digestive Endoscopy*, 18(2):112.

[60] Simonyan, K. and Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition.

[61] Son, J., Park, S. J., and Jung, K.-H. (2017). Retinal Vessel Segmentation in Fundoscopic Images with Generative Adversarial Networks. Technical report.

[62] Standring, S. (2016). *Gray's anatomy : the anatomical basis of clinical practice*. 41 edition.

[63] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2014). Going Deeper with Convolutions.

[64] Szegedy, C., Vanhoucke, V., Ioffe, S., and Shlens, J. (2015). Rethinking the Inception Architecture for Computer Vision. Technical report.

[65] Van de Bruaene, C., De Looze, D., and Hindryckx, P. (2015). Small bowel capsule endoscopy: Where are we after almost 15 years of use? *World journal of gastrointestinal endoscopy*, 7(1):13–36.

[66] Vieira, P. M., Goncalves, B., Goncalves, C. R., and Lima, C. S. (2016). Segmentation of angiodysplasia lesions in WCE images using a MAP approach with Markov Random Fields. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1184–1187. IEEE.

[67] Vignes, S. and Bellanger, J. (2008). Primary intestinal lymphangiectasia (Waldmann's disease). *Orphanet Journal of Rare Diseases*, 3(1):5.

[68] Youn Park, D. and Lauwers, G. Y. (2008). Gastric Polyps Classification and Management. Technical report.

[69] Zhang, K., Miao, S., Han, T. X., Yuan, X., Guo, L., and Liu, T. (2016). Residual Networks of Residual Networks: Multilevel Residual Networks. Technical report.

[70] Zheng, Y., Hawkins, L., Wolff, J., Goloubeva, O., and Goldberg, E. (2012). Detection of Lesions During Capsule Endoscopy: Physician Performance Is Disappointing. *The American Journal of Gastroenterology*, 107(4):554–560.