

A evolução das Redes Adversárias Generativas na Edição de Imagem e estudo do modelo *EDITGAN*

Diogo Filipe Araújo da Silva Rocha Pereira

Mestrado em Multimédia da Universidade do Porto;

Especialização em Tecnologias Interativas e Jogos Digitais

Orientador: João Manuel Ribeiro da Silva Tavares

Diogo Pereira, 2024

A evolução das Redes Adversárias Generativas na Edição de Imagem e estudo do modelo *EDITGAN*

Diogo Filipe Araújo da Silva Rocha Pereira

Mestrado em Multimédia da Universidade do Porto

Aprovado em provas públicas pelo Júri:

Presidente: Maria van Zeller de Macedo de Oliveira e Sousa

Vogal Externo: Fernando Joaquim Lopes Moreira

Orientador: João Manuel Ribeiro da Silva Tavares

Abstract

Automatic image editing has come a long way in recent times, with more and more editing methods being invented, bringing with them numerous advantages, but also numerous disadvantages. Some are easier to edit, others focus on greater precision and others on greater creativity. There is always a different problem with each different method and a general problem, which can be identified as a lack of privacy. This is due to the simple fact that generative adversarial networks use various images published on the internet for data training to generate or produce a final image, and there may be no consent from the person in question to use it.

This project was focused on analyzing and comparing in depth the different methods that use generative adversarial networks to produce images, as well as their main problems and possible ways to overcome them. It also included an in-depth study of the experimental model called EditGAN and an interactive prototype for the cell phone version to be used in its final phase. The developed interactive prototype used a user-centered design methodology, and usability tests and questionnaires were performed. The data obtained was used to identify flaws, interface problems and improvements to be taken into consideration at the design level. The results of the evaluation questionnaires showed great potential in an automatic editing application using the EditGAN method. Design guidelines that were derived serve as a valuable resource for future interface development. Future studies should pay attention to additional aspects and explore possible updates to improve the interface, also incorporating more feedback from diverse audiences, to improve and enrich the effectiveness of the interface.

Keywords: EditGAN, user-centered design, interactive prototype, future interface development, generative adversarial networks for image production, data training, experimental model, usability tests and questionnaires

Resumo

A edição de imagem automática tem trazido um enorme avanço nos últimos tempos, cada vez mais, existem diversos métodos de edição a serem inventados, trazendo consigo inúmeras vantagens, mas também inúmeras desvantagens. Uns consistem numa maior facilidade de edição, outros concentram-se numa maior precisão e outros numa maior criatividade. Existindo sempre um problema diferente em cada método distinto e um problema geral, que se identifica por falta de privacidade. Isto pelo simples facto de que as redes adversárias generativas utilizam diversas imagens publicadas na internet para treinamento de dados a fim de gerarem ou produzirem uma imagem final, podendo não existir um consentimento algum da sua utilização por parte da pessoa em questão.

Este projeto concentrou-se em analisar e comparar aprofundadamente os diferentes métodos que utilizam as redes adversárias generativas para a produção de imagens, bem como os seus principais problemas e possíveis formas de combatê-los. Incluiu também com um estudo aprofundado ao modelo em estado experimental com o nome de EditGAN e um protótipo interativo para a versão de telemóvel, a fim de ser utilizada na sua fase final. O protótipo interativo utilizou uma metodologia de design centrado no utilizador, abrangendo testes de usabilidade e questionários. Os dados obtidos foram utilizados para identificar falhas, problemas de interface e melhoramentos a ter em consideração a nível de design. Os resultados dos questionários de avaliação, demonstraram grande potencial numa aplicação de edição automática utilizando o método EditGAN. Diretrizes de design que foram derivadas a servirem como um recurso valioso para o desenvolvimento futuro da interface. Estudos futuros devem ter atenção aos aspetos adicionais e explorarem as possíveis atualizações, a fim de melhorar a interface, incorporando também mais feedback de público diversificado, com o objetivo de melhorar e enriquecer a eficácia da mesma.

Palavras-chave: EditGAN, design centrado no utilizador, protótipo interativo, desenvolvimento futuro da interface, redes adversárias generativas para a produção de imagem, modelo experimental, treino de dados, testes de usabilidade e questionários

Agradecimentos

Gostaria de expressar a minha sincera e formal gratidão às pessoas que desempenharam um papel significativo na realização e desenvolvimento desta Dissertação. Este estudo foi desenvolvido a pensar nas mais valias que a inteligência artificial pode trazer num futuro próximo em relação à edição de imagem.

Primeiramente gostaria de dar o meu sincero agradecimento ao meu orientador João Manuel Tavares, por me ter ensinado a desenvolver esta Dissertação e me aconselhar a fazer os questionários e estudos necessários para produzir os testes de usabilidade. Os seus apoios inabaláveis durante todo o percurso da tese foram inestimáveis, estou verdadeiramente grato pela sua orientação e crença nas minhas capacidades.

Numa nota pessoal, gostaria de agradecer bastante à minha namorada Carolina por me ter feito uma revisão ortográfica, juntamente com a minha tia Diana que fez uma segunda revisão.

Um enorme obrigado também ao meu pai por todo o apoio e dedicação que me deu ao longo destes dois anos de estudo na FEUP e à minha mãe por ter confiado em mim e me prestar auxílio nas propinas.

Por fim, estou extremamente grato à FEUP por me aceitar no mestrado de Multimédia, proporcionando-me dois anos de imensa aprendizagem, rizadas, empatia e paixão. Estes momentos vão ser lembrados para a vida toda.

Índice

1. Introdução.....	1
1.1 Contexto.....	2
1.2 Objetivo do Projeto.....	4
1.3 Problemas e Hipóteses.....	4
2. Revisão Bibliográfica	6
2.1 Introdução	6
2.2 Processo	7
2.2.1 Algoritmo.....	8
2.3 Estudo de métodos com GANs em diversos contextos	9
2.3.1 Imagem	9
2.3.2 Vídeo	10
2.3.3 Detecção de Objetos	10
2.4 <i>EDITGAN</i>	11
2.4.1 <i>StyleGAN2</i>	11
2.4.2 Incorporação de Imagens no Espaço Latente	12
2.4.3 Encontrar a Semântica no Espaço Latente.....	12
2.4.4 Maneiras de Edição com a Interface.....	13
2.4.5 Código em estado BETA.....	13
2.5 Implementação e Resultados	13
2.5.1 Resultados Qualitativos	14
2.5.2 Resultados Quantitativos	16
2.6 Comparação de Métodos Generativos	17
3. Metodologia de Investigação	20
3.1 Plano de Pesquisa	20
3.2 Design centrado no utilizador	21
3.3 Métodos	22
3.4 Tipos de Aplicações de Edição de Imagem com IA	22
3.5 Requisitos de Design	25
4. Protótipo Interativo.....	26
4.1 Figma.....	27

4.2 Criação do Protótipo	28
4.3 Autentificação e Gestão do Utilizador.....	30
4.4 Módulo de Interação e Conteúdo.....	34
4.5 Editor do protótipo.....	40
5. Testes de Usabilidade e Inquéritos.....	45
5.1 Questionários	45
5.2 Testes de Usabilidade	48
5.3 Análise de Dados	49
6. Resultados	52
7. Conclusão e Oportunidades Futuras	54
7.1 Conclusões Finais	54
7.2 Oportunidades Futuras.....	55
8. Referências.....	57
9. Anexos	64

Lista de Figuras

Figura 1: Fatores que se destacam nas Redes Adversárias Generativas	3
Figura 2: O discriminador D e o gerador G são os dois modelos aprendidos durante o processo de treino de um GAN (Antonia Creswell <i>et al.</i> , 2018).	7
Figura 3: Esquema de rótulos do rosto de uma pessoa (Huan Ling <i>et al.</i> , 2021).	14
Figura 4: Exemplos de edições orientadas para a segmentação com <i>EditGAN</i>	15
Figura 5: Paleta de cores escolhidas para o protótipo desenvolvido.	28
Figura 6: Logótipo escolhido para representação do protótipo desenvolvido.	29
Figura 7: Representação dos ecrãs do Módulo de Autenticação e Gestão do Utilizador, início do protótipo desenvolvido.	30
Figura 8: Ecrãs do perfil do utilizador e a sua pesquisa do protótipo desenvolvido.	33
Figura 9: Representação das publicações no protótipo desenvolvido.	35
Figura 10: Ecrãs da página principal do protótipo desenvolvido.	36
Figura 11: Ecrã das definições do protótipo desenvolvido.	38
Figura 12: Ecrã das notificações do protótipo desenvolvido.	39
Figura 13: Ecrãs da galeria do protótipo desenvolvido.	40
Figura 14: Ecrãs do início do editor Editable-Gen desenvolvido.	41
Figura 15: Funcionalidades de edição do editor Editable-Gen desenvolvido.	42
Figura 16: Ecrãs finais do editor Editable-Gen desenvolvido.	43
Figura 17: Imagens dos testes de usabilidade da sessão de gravação.	49

Lista de Tabelas

Tabela 1: Comparações quantitativas com várias linhas de base de edição de sorrisos baseada na de Huan Ling <i>et al.</i> , 2021.	18
Tabela 2: Plano de Pesquisa definido para esta Dissertação.	21
Tabela 3: <i>Feedback</i> dos participantes no teste de usabilidade	50
Tabela 4: Dados mais significativos do teste de satisfação do participante.	51

Abreviaturas

GAN	<i>Generative Adversarial Networks</i>
AI/IA	<i>Artificial Intelligence/Inteligência Artificial</i>
NeurIPS	<i>Neural Information Processing Systems</i>
ChatGPT	<i>Chat Generative Pre-Trained Transformer</i>
LPIPS	<i>Learned Perceptual Image Patch Similarity</i>
FFHQ	<i>Flickr-Faces-HQ</i>
FID	<i>Fréchet Inception Distance</i>
KID	<i>Kernel Inception Distance</i>
GPU	<i>Graphics Processing Unit</i>
3D	<i>Three Dimensional</i>
MIT	<i>Massachusetts Institute of Technology</i>
ICLR	<i>International Conference on Learning Representations</i>
ICML	<i>International Conference on Machine Learning</i>
CVPR	<i>Computer Vision and Pattern Recognition</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
ACM	<i>Association for Computing Machinery</i>
ICCV	<i>International Conference on Computer Vision</i>
API	<i>Application Programming Interface</i>
BETA	<i>Software Development Phase</i>
UEQ	<i>User Experience Questionnaire</i>
UCD	<i>User Centered Design</i>
UI	<i>User Interface</i>
UX	<i>User Experience</i>
OTP	<i>One Time Password</i>
CNN	<i>Convolutional Neural Network</i>

1. Introdução

A edição da imagem digital está constantemente a evoluir ao longo dos anos, nos dias de hoje um dos métodos mais rápidos e promissores, consta com a inteligência artificial e designa-se por Rede Adversária Generativa (GAN). Com o passar do tempo, estas redes generativas desenvolveram-se bastante no processamento de imagens e estão a ganhar cada vez mais interesse nas áreas académica e industrial (Ian Goodfellow *et al.*, 2020).

Atualmente, as GANs são utilizadas numa variedade de estudos e aplicações, incluindo gerenciamento de imagens, pinturas, texto, processamento de imagens médicas, segmentação semântica, colorização de imagens, tradução de imagem para imagem e geração de arte. Além disso, também são utilizados na síntese e edição facial. Usando o Scopus - uma das bases de referência de trabalhos científicos utilizadas para obter informação necessária para esta investigação, podemos obter de 100 mil resultados de trabalhos relacionados com Redes Adversárias Generativas (pesquisa realizada em 02/02/2024), o que dificulta bastante esta pesquisa. Por esse motivo, optou-se por orientar este estudo nas informações mais recentes de forma a ter uma boa estrutura de progressão para um futuro próximo.

Após comparar alguns métodos de edição via inteligência artificial, o mais relevante, foi o *EditGAN*. Isto devido à maioria dos métodos de edição de imagens baseados em GANs, que exigem frequentemente conjuntos de dados em grande escala com anotações de segmentação semântica para treino e apenas fornecem controle de alto nível ou apenas interpolam entre imagens diferentes. Já o *EditGAN*, permite que os utilizadores modifiquem as imagens alterando as suas máscaras de segmentação de partes extremamente detalhadas (Huan Ling *et al.*, 2021), diferenciando-se bastante com os métodos de edição automática atualmente.

O *EditGAN* foi introduzido pela empresa NVIDIA em 2022 e aparenta ser bastante eficaz na edição de alta precisão e semântica de alta qualidade. Um dos principais problemas do *EditGAN*, é o facto de se encontrar estado experimental, sem nenhum prazo definido para a sua finalização. Outro problema, é o facto de existir pouca informação sobre o método e haver apenas um

documento viável de informação online, com o título de “*EditGAN: High-Precision Semantic Image Editing*”. Por outro lado, é um método que pode ser implementado em cenários reais, trabalha com dados rotulados a fim de fornecer uma máscara de edição simples ao utilizador, permite também várias edições em simultâneo e funciona em imagens simples de um objeto em questão. Vantagens estas que levaram à escolha deste método para o protótipo da aplicação.

O planeamento deste protótipo foi realizado a pensar no utilizador, visando uma *interface* acessível e de fácil uso, fornecendo um resultado que supere as expectativas do utilizador com uma boa edição de imagem e precisão.

Após a conclusão do protótipo, foi feita uma análise geral de aplicações concorrentes no mercado e uma comparação entre as mais utilizadas. Foram feitos vários testes de usabilidade com o objetivo de melhorar o protótipo e perceber a perspectiva de cada possível utilizador.

1.1 Contexto

As redes adversárias generativas (GAN) foram introduzidas por Ian Goodfellow e os seus parceiros em 2014, numa conferência com o nome de “*Neural Information Processing Systems*” (NeurIPS). Estas redes tiveram um grande avanço na aprendizagem automática, especialmente nos modelos generativos.

O termo “adversarial” deriva do fato de as GANs se aproveitarem de duas redes neuronais num ambiente competitivo. Derivadas de gerador e discriminador. Sendo o gerador responsável por produzir dados sintéticos, enquanto o discriminador é responsável por distinguir entre dados gerados e dados verdadeiros. O processo de formação resulta do facto que o gerador melhora a capacidade de produzir dados confiáveis e o discriminador melhorar a capacidade de distinguir entre amostras geradas e as reais.

A capacidade impressionante das GANs para produzir dados realistas e de alta qualidade, contribuiu para avanços na visão computacional, gráficos e outras aplicações de inteligência artificial, contribuiu também para a geração das “*deepfakes*” um tema bastante polémico nos últimos anos. O campo da aprendizagem profunda foi significativamente afetado pelas GANs, que ainda é um campo de pesquisa e inovação ativo.

Estas redes introduziram uma abordagem inovadora à modelação generativa e foram motivadas por vários fatores como os identificados na Figura 1.

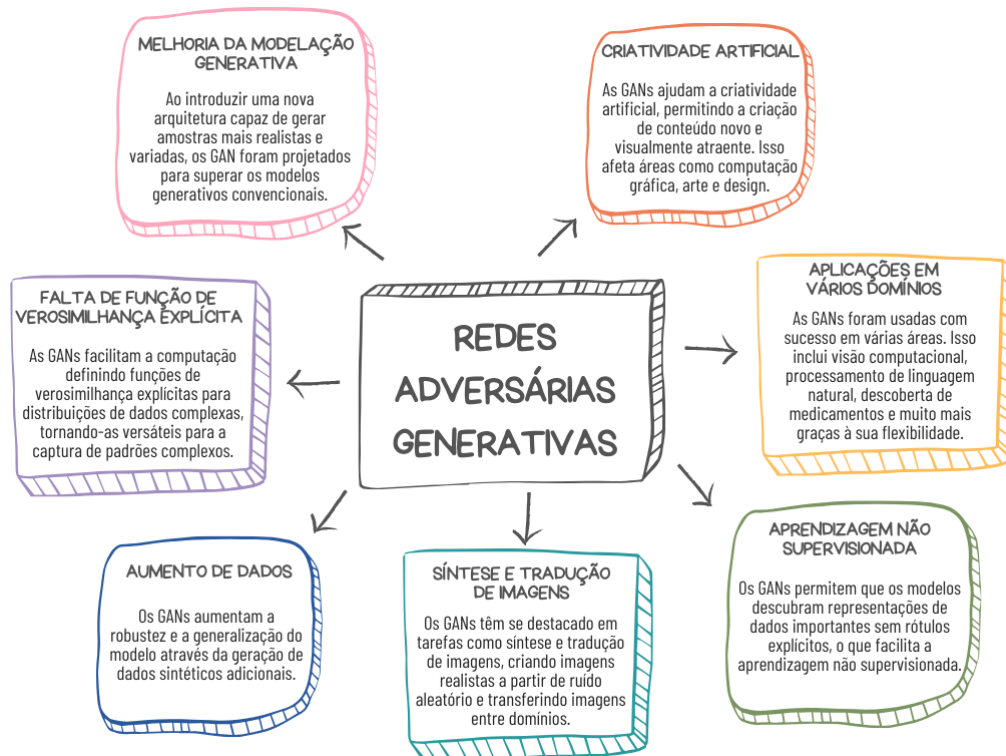


Figura 1: Fatores que se destacam nas Redes Adversárias Generativas

Após o marco histórico de Ian Goodfellow *et al.*, em 2014, na conferência NeurIPS, as Redes Adversárias Generativas foram-se desenvolvendo ao longo dos anos. Sendo que em 2015, os investigadores exploraram variações da arquitetura original e foi desenvolvido o modelo *DCGAN* (*Deep Convolution GAN*) (Radford *et al.*, 2015), introduzindo camadas convulsionais, melhorando e estabilizando o treinamento das GANs. Mais tarde, foi desenvolvida uma nova métrica para medir a diferença entre as distribuições geradas e as reais, chamada de *Wasserstein GAN* (WGAN) que foi introduzida por Arjovsky *et al.*, na conferência ICML em 2017, esta variante consegue resolver problemas comuns nas redes adversarias generativas como o colapso entre elas e a dificuldade de estas serem treinadas, devido a existir extrema instabilidade.

Um ano depois, Karras *et al.*, na conferência ICLR em 2018, introduziu os GANs progressivos, que revolucionaram o processo de treino ao aumentar gradualmente a complexidade do gerador e do discriminador. Esta técnica permitiu a criação de imagens de alta resolução e ajudou na criação de rostos realistas. Já em 2019, um gerador baseado em estilos com o nome de *StyleGAN* (modelo desenvolvido em parceria com a empresa NVIDIA), permitiu um controlo maior sobre as imagens geradas. Este estilo é ainda bastante utilizado devido a ser bem-sucedido na geração de imagens diversas e de boa qualidade, o mesmo gerador foi atualizado para *StyleGAN2* e está em estudo no método *EditGAN* (Karras *et al.*, 2019). Nesse mesmo ano, foi

apresentada a expansão da arquitetura das GAN para lidar com a geração de imagens em grande resolução, dando assim o nome de *BigGAN* (*Big Generative Adversarial Network*) ao modelo. Contextualizando, este foi o ano que o tema das “*deepfakes*” começou a ser viral e houve bastante polêmica em volta das GANs, pois esta variante foi introduzida para gerar imagens de alta qualidade e resolução de maior fidelidade que os modelos anteriores (Brock *et al.*, 2019).

Contudo, as Redes Adversárias Generativas têm sido usadas numa variedade de setores. Isso inclui processamento de linguagem natural, descoberta de medicamentos e cuidados de saúde. Um dos maiores exemplos que podemos observar neste momento, é o *ChatGPT*.

1.2 Objetivo do Projeto

Esta Dissertação, teve como objetivo realizar um estudo nas Redes Adversárias Generativas, mais concretamente no modelo de edição *EditGAN*, como também desenvolver a arquitetura de uma aplicação para ser utilizada com este modelo na sua fase final. Por conseguinte, contou com o desenvolvimento de *mockups* de alta qualidade e um estudo a aplicações equiparáveis no mercado.

Este estudo teve por base duas questões de investigação:

Q1- Considerando o método *EDITGAN*, um gerador de imagem editado ao pormenor, irá ter grande impacto em relação aos outros editores de imagem por inteligência artificial já existentes?

Q2- Que estratégias podem ser utilizadas para diminuir problemas relacionados com o colapso de modos em GANs, e como é que essas estratégias têm influência na diversidade das amostras geradas?

1.3 Problemas e Hipóteses

As GANs são uma ferramenta útil para a edição de imagem, mas existem também questões éticas e problemas a ter em conta:

No que diz respeito à privacidade, as GAN podem ser usadas para criar imagens falsas realistas ou até “*fake news*”, o que levanta dúvidas sobre a utilização indevida da tecnologia para fins enganadores ou prejudiciais. Sem o consentimento de uma pessoa a fotografia não pode ser utilizada, por esse motivo, levanta questões éticas uma vez que viola o direito do indivíduo à autoimagem. Por outro lado, para combater este problema, os programadores devem criar regras claras sobre a utilização de conteúdo criado pelo utilizador e garantir que o consentimento do mesmo seja obtido antes de manipular ou compartilhar imagens. Existe sempre a opção de produzir ferramentas que ajudem os utilizadores a verificar a autenticidade das imagens, facilitando assim a deteção de conteúdo manipulado. Por fim, educar os utilizadores sobre a literacia mediática e ajudá-los a entender a existência dos possíveis efeitos, informando-os sobre as vantagens e desvantagens da tecnologia GAN durante a edição de imagens.

Quanto ao envio de dados para treino, as GANs podem produzir resultados injustos e discriminatórios se forem treinadas em conjuntos de dados enviesados. Isto porque as GANs têm a capacidade de se adaptar demasiado aos dados de treino, tornando-os sensíveis a pequenas mudanças. Por outro lado, para reduzir os enviesamentos, os conjuntos de dados de formação têm de ser diversificados e representativos. Uma atualização regularmente nos conjuntos de dados, pode evitar enviesamentos. Na fase de treino, o utilizador pode optar por usar técnicas de regularização a fim de evitar ajustes e garantir que o modelo se adapte bem a novos dados.

Os requisitos computacionais podem ser um problema limitador, isto porque as GANs são computacionalmente intensivas, o que significa que necessitam de um bom *hardware* para a formação e a inferência, limitando assim utilizadores com dispositivos fracos. Para contornar esta situação, pode ser fornecido ao utilizador serviços de edição de imagens utilizando soluções baseadas na *cloud* (um espaço de armazenamento de dados *online*), a fim de aliviar a carga computacional dos utilizadores e tornar a tecnologia mais acessível.

As GAN têm o potencial de causar resultados imprevistos e inesperados, principalmente quando são utilizadas em conteúdos sensíveis ou possíveis efeitos que não foram totalmente compreendidos pelas respetivas entidades. De certo modo, se avaliarmos regularmente os resultados dos sistemas de edição de imagens baseados em GAN, conseguimos corrigir rapidamente qualquer resultado desfavorável. Contudo, é necessário ter atenção ao *feedback* transmitido pelos utilizadores, pois estes podem mais facilmente notar alguma perda ou falha existente.

As GANs sofrem de problemas de avaliação, pois não há uma métrica clara para medir seu desempenho e qualidade. Contudo, uma partilha social da arte gerada, a fim de o ser humano avaliar a sua qualidade e desempenho é melhor forma métrica de avaliação (Zhaoqing Pan *et al.*, 2019).

2. Revisão Bibliográfica

2.1 Introdução

Os modelos generativos tornaram-se um tema popular na pesquisa de aprendizagem profunda devido à grande quantidade de imagens usadas nas redes sociais. As Redes Adversárias Generativas (GANs) desenvolvem-se bastante no processamento de imagens e estão a ganhar cada vez mais interesse nas áreas acadêmica e industrial. Atualmente, as GANs são usadas numa variedade de estudos e aplicações no domínio do processamento de imagens, incluindo pinturas, geração de texto, processamento de imagens médicas, segmentação semântica de imagem, colorização de imagens, tradução de imagem para imagem e geração de arte. Além disso, também são utilizadas na síntese e edição facial (Wei Chen *et al.*, 2020).

É normal considerar que a pesquisa sobre GANs está dividida em duas áreas (Wei Chen *et al.*, 2020):

1) Uma investigação teórica sobre GANs é baseada na teoria da informação ou modelos baseados na energia e concentra-se nos problemas não resolvidos dos GANs durante o treino, como o colapso do modo instável e difícil de avaliar.

2) O uso das GAN em várias tarefas de visão computacional. Embora ainda haja alguns problemas para resolver, várias variantes de GAN melhoraram o desempenho das GANs.

Introduzindo o *EditGAN*, é um novo método de edição de imagens semânticas de alta precisão e qualidade. Baseia-se em um algoritmo GAN que modela imagens RGB e suas segmentações semânticas por pixel. O treino é feito com muito poucos dados anotados. A otimização no espaço latente e o condicionamento de máscaras de segmentação editadas permitem a edição. Pode-se economizar essa otimização criando vetores de edição no espaço latente, que podem ser aplicados diretamente em outras imagens. Isso permite a edição interativa

em tempo real sem qualquer otimização ou apenas com um pouco mais de otimização. Ficaram demonstradas uma ampla gama de técnicas de edição para vários tipos de imagens, oferecendo uma flexibilidade e liberdade de edição sem precedentes (Huan Ling *et al.*, 2021).

2.2 Processo

Antonia Creswell *et al.*, em 2018, descreveram como funciona o processo das GANs. O falsificador, também conhecido como gerador G na literatura sobre GAN, cria falsificações com o objetivo de criar imagens realistas. O perito D, também conhecido como discriminador, tenta distinguir imagens falsas de imagens reais. Ambos são treinados simultaneamente e competem entre si. O gerador não tem acesso direto às imagens reais, a única maneira de ele aprender é interagindo com o discriminador. O discriminador pode ter acesso tanto a falsificações como a imagens reais, como mostra na Figura 2. O simples fato de saber se a imagem veio do gerador ou da pilha, dá ao discriminador um sinal de erro.

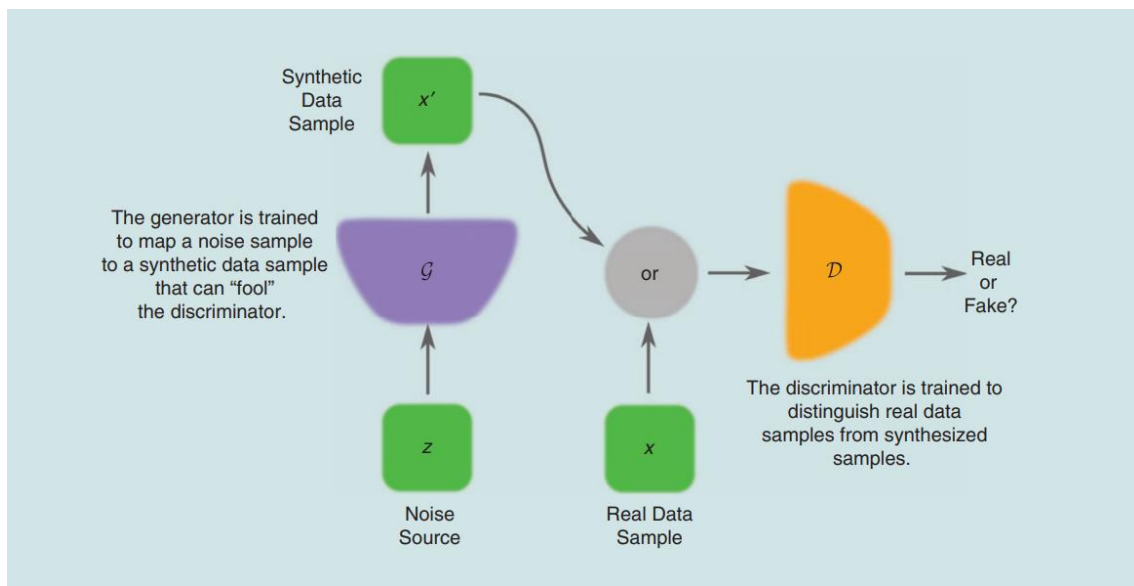


Figura 2: O discriminador D e o gerador G são os dois modelos aprendidos durante o processo de treino de um GAN (Antonia Creswell *et al.*, 2018).

O discriminador D e o gerador G são tipicamente implementadas com redes neurais, mas podem ser implementados por qualquer forma de sistema diferenciável que mapeia dados de um espaço para outro.

Normalmente, as redes multicamadas com camadas convulsionais estão totalmente ligadas e são usadas para representar o gerador e o discriminador. Embora não seja necessário que as redes geradoras e discriminatórias sejam diretamente invertíveis, ambas devem ser diferenciáveis. Se considerarmos a rede geradora como um mapeamento do espaço latente, para o espaço dos dados (neste caso, concentrando-nos nas imagens), podemos definir isso formalmente como $G: G(z) \rightarrow R^x$, onde $z \in R^z$ é uma amostra do espaço latente, $x \in R^x$ é uma imagem e o ponto (.) representa o número de dimensões.

2.2.1 Algoritmo

Ian Goodfellow *et al.*, em 2020, referiram que podemos pensar nas GANs como uma analogia entre os falsificadores e a polícia. Os falsificadores fazem dinheiro falso enquanto a polícia tenta capturar os falsificadores e continuar a permitir o uso de dinheiro legítimo. O dinheiro torna-se realista devido ao conflito entre os falsificadores e a polícia, até que os falsificadores produzem falsificações perfeitas e a polícia não consegue distinguir entre dinheiro real e falso. Uma complicação desta analogia é que o gerador aprende pelo discriminador, mas a operação *argmin* é difícil de trabalhar dessa forma. A função de cálculo pode ser dada como:

$$J^{(G)}\left(\theta^{(G)}, \underset{\theta^{(D)}}{\operatorname{argmin}} J^{(D)}(\theta^{(G)}, \theta^{(D)})\right),$$

onde G é o gerador e D o discriminador, $J^{(G)}$ foi definido como a probabilidade logarítmica negativa que o discriminador atribui à relação verdadeiro contra falso, logo $J^{(G)} = -J^{(D)}$ e $\theta^{(G)}$ é considerado que o gerador está a gerar uma amostra (x) (Radford *et al.*, 2015).

Considerando esse cenário como um jogo de dois jogadores é a maneira mais comum de pensar. A maior parte da literatura sobre teoria dos jogos concentra-se em jogos com espaços de ação discretos e finitos, perdas convexas ou outras características que os tornam mais simples. Os GAN precisam de usar a teoria dos jogos em ambientes que ainda não foram bem explorados, onde os custos não são convexas e as políticas e as ações são contínuas e de grande dimensão.

O objetivo do algoritmo de aprendizagem automática neste contexto é descobrir o equilíbrio local de *Nash*, um ponto mínimo local do custo de cada jogador em relação aos parâmetros do jogador. Com movimentos locais, nenhum jogador pode reduzir o preço se os parâmetros do outro jogador não mudarem. O algoritmo de treinamento mais usado consiste simplesmente em usar um otimizador baseado em gradientes para dar repetidamente passos simultâneos a ambos os jogadores, reduzindo gradualmente o custo para cada jogador em relação aos parâmetros desse

jogador. Mesmo para conjuntos de dados extremamente complexos com imagens de alta resolução, as GAN geralmente produzem amostras realistas no final do processo de formação (Ian Goodfellow *et al.*, 2020).

2.3 Estudo de métodos com GANs em diversos contextos

2.3.1 Imagem

As aplicações de geração de rostos são relativamente fáceis pois o problema envolve apenas uma classe de objetos. Cada objeto é caro, e a maioria dos conjuntos de dados de rostos é composta por pessoas para a câmara. A maioria das pessoas regista os seus pontos de referência, como o nariz, os olhos e outros, em locais consistentes (Jie Gui *et al.*, 2021). É um pouco mais difícil fazer com que as GANs funcionem em conjuntos de dados diferentes, que contém mil classes de objetos diferentes. No entanto, nos últimos anos, houve um progresso rápido no trabalho, e a qualidade das imagens produzidas tem vindo a melhorar continuamente.

Muitas aplicações incluem a comunicação entre pessoas e o processo de criação de imagens. A manipulação realista de imagens em tais circunstâncias, por outro lado, é difícil porque exige que o utilizador controle as alterações feitas na imagem, mantendo a aparência realista. Durante a edição, a imagem pode-se desviar facilmente do conjunto de imagens naturais se o utilizador não tiver habilidades artísticas suficientes. O GAN interativo (*IGAN*) cria uma classe de operações de edição de imagens e garante que os seus resultados estejam sempre localizados num coletor aprendido. Além disso, as redes introspectivas adversárias oferecem a oportunidade de realizar uma edição interativa de fotos (Li *et al.*, 2020). Contudo o *IGAN*, destaca-se por permitir que utilizadores editem e criem imagens de forma intuitiva e rápida, gerando imagens mais realistas.

Os resultados foram apresentados principalmente para edição de rostos. *GauGAN* (uma demonstração de IA para geração de imagens foto realistas) tem a capacidade de transformar rabiscos em paisagens maravilhosas e reais (T. Park *et al.*, 2019). Teve um marco histórico bastante relevante, pois oferecia uma ferramenta para artistas e *designers* que precisavam de uma criação rápida de conteúdo visual.

2.3.2 Vídeo

Mais tarde, um novo modelo, representou a rede desembaraçada (*DRNET*). A *DRNET* pode aprender representações de imagens desembaraçadas de vídeos baseados em GAN, pois foi uma nova abordagem de síntese de vídeo para vídeo, sugerido para um quadro de aprendizagem adversário generativo, tem como objetivo capturar e aprender as mudanças dinâmicas em sequências de imagens á medida do tempo, conforme o movimento dos objetos ou da alteração luminosa, separando os aspetos estáticos da cena (Dento *et al.*, 2017).

2.3.3 Detecção de Objetos

Como podemos treinar um detetor de objetos que não sofre de oclusões ou deformações? Uma estratégia baseada em dados consiste em recolher conjuntos de dados em grande escala que contenham vários exemplos de objetos que aparecem em diferentes situações.

Podemos esperar que o classificador final possa usar numerosas instâncias para aprender invariâncias utilizando esta abordagem. Algumas deformações e oclusões são tão incomuns que quase nunca acontecem em condições reais, mas queremos que nosso método seja invariável nesses casos. Para resolver este problema, foram criadas instâncias com oclusões e deformações usando GANs. O propósito do gerador era produzir instâncias, já o objetivo do mesmo, era gerar instâncias que o detetor de objetos não pudesse classificar (Wang *et al.*, 2017).

K. Ehsani *et al.*, na conferência CVPR em 2018, estudaram objetos que estavam a ser interrompidos por outros objetos numa imagem utilizando um modelo de segmentação e as GANs, chamando-lhe de *SeGAN* (*Segmentation Generative Adversarial Network*), que tinha como objetivo de melhorar a segmentação semântica utilizando uma abordagem baseada em GANs, tendo como função classificar cada pixel de uma imagem para uma determinada categoria específica. Já nesse mesmo ano, foi inventado um GAN multitarefa de ponta a ponta (*MTGAN*), que foi desenvolvida para lidar com mais do que uma tarefa de uma vez só. É geralmente utilizado para melhorar a eficácia e precisão em gerir dados sintéticos ou transformar os mesmos para cumprir múltiplos objetivos em simultâneo. Esta variável representou assim um avanço significativo no uso de GANs para funções complexas que exigem a integração de múltiplos objetivos num único modelo de aprendizagem (Y. Bai *et al.*, 2018).

2.4 EDITGAN

A maioria dos métodos de edição de imagens baseados em GANs, exigem conjuntos de dados de grande escala com anotações de segmentação semântica para treinar, apenas fornecem controlo de alto nível ou intercalam entre imagens diferentes. O *EditGAN* é um novo método de edição de imagens semânticas de alta qualidade e precisão, que recria imagens e as suas segmentações semânticas. Permite que os utilizadores alterem as imagens modificando as suas características específicas, como por exemplo, se o utilizador quiser alterar um objeto, terá de desenhar uma máscara por cima da área do objeto que quer editar. Particularmente, adiciona-se uma imagem ao espaço latente do GAN e otimiza-se o código latente condicional usando a edição de segmentação, que vai alterar a imagem. Já no espaço latente, os vetores editoriais realizam edições a fim de amortizar a otimização.

A estrutura permite adquirir uma quantidade ilimitada de vetores de edição, podendo ser usados diretamente em outras imagens, a taxas interativas. Experimentalmente o *EditGAN* tem a capacidade de manipular imagens com bastante profundidade e também é capaz de combinar facilmente várias edições para produzir edições plausíveis.

Por referência, esta secção do método em estudo (*EditGAN*), foi baseada na informação do único documento legítimo online com o nome de “*EditGAN: High-Precision Semantic Image Editing*”, produzido por Huan Ling *et al.*, 2021, em parceria com a NVIDIA, Universidade de Toronto, *Vector Institute* e o MIT.

2.4.1 StyleGAN2

O *StyleGAN2*, é o componente de geração de imagens do *EditGAN*. Atualmente é uma das redes adversarial mais avançada para síntese de imagens (Tero Karras *et al*, 2019). Consegue transformar os códigos latentes $z \in Z$ de uma distribuição normal multivariada em imagens realistas. Um código latente z é inicialmente transformado num código intermediário $w \in W$ por meio de uma função de mapeamento não linear. De seguida, os vetores $K + 1$ são transformados em $w^0, \dots, w^K$ por meio de uma transformação semelhante. Os mapas de características detalhadas, são criados por introduzir esses códigos latentes transformados em blocos de síntese. Ao utilizar esta distribuição conjunta $p(x, y)$, o *EditGAN* pode realizar uma edição de imagens semânticas de alta precisão a partir de imagens reais e sintetizadas. O *DatasetGAN*, aplica um classificador

simples de três camadas a mapas de características associadas. Este classificador deteta uma etiqueta de segmentação para cada pixel (Yuxuan Zhang *et al.*, 2021).

2.4.2 Incorporação de Imagens no Espaço Latente

O objetivo de treinar este codificador é aprender a reconstruir perdas L2 e LPIPS perceptuais utilizando dados de treino reais e amostras do próprio GAN. Além disso, regulariza explicitamente o codificador, utilizando os códigos latentes subjacentes conhecidos para as amostras do GAN. Na prática, inicia-se as imagens com o codificador e refina-se iterativamente o código latente w^+ com otimização e objetivos de reconstrução padrão (Rameen Abdal *et al.*, 2019).

Desse modo, incorpora-se as imagens registadas x de um conjunto de dados rotulados com segmentações semânticas no espaço latente e treina-se o ramo de segmentação do gerador com objetivos padrão de supervisão, como a perda de entropia cruzada. Apenas retro propaga-se a perda para o ramo de segmentação e mantêm-se os pesos do gerador de imagens congelados (Elad Richardson *et al.*, 2020).

2.4.3 Encontrar a Semântica no Espaço Latente

O principal objetivo do *EditGAN* é usar a distribuição conjunta $p(x, y)$ de imagens e segmentações semânticas para edição de imagens de alta precisão. É possível editar uma imagem x recém-criada no espaço latente W^+ do *EditGAN*. Alternativamente, podemos fazer o próprio modelo e usá-lo. Como as segmentações e as imagens RGB têm os mesmos códigos latentes w^+ , o ramo de segmentação vai produzir a segmentação y correspondente. Após isso, o utilizador consegue alterar manualmente a segmentação para uma edição desejada utilizando ferramentas simples de pintura digital interactiva ou com um rótulo. A máscara de segmentação editada é indicada por y . Para assim obter um novo $w^+ = w^+ + \delta w^+$ a partir da incorporação w^+ da imagem (x) não editada e da segmentação (y), fazendo uma otimização dentro de W^+ . Isso permite que a saída RGB mude dentro da região de edição (Huan Ling *et al.*, 2021).

2.4.4 Maneiras de Edição com a Interface

1) Edição em tempo real com vetores de edição

Para edições localizadas e separadas, é utilizado apenas vetores de edição aprendidos com escalas variáveis s_{edit} e manipula-se as imagens a taxas interactivas.

2) Edição baseada em vetores com refinamento supervisionamento automático

Inicia-se a edição utilizando os vetores de edição aprendidos para remover artefactos de otimização em tempo de teste, para edições que não estão perfeitamente desempenhadas com outras partes da imagem.

3) Edição com base na otimização

Os vetores de edição não transferem edições específicas e muito grandes de imagens para outras imagens. Assim, realiza-se a otimização a partir do zero para essas operações (Huan Ling *et al.*, 2021).

2.4.5 Código em estado BETA

Em fevereiro de 2022, Huan Ling *et al.* lançaram o código da versão BETA do *EditGAN* no *Github* (https://github.com/nv-tlabs/editGAN_release). O código foi concebido maioritariamente em linguagem Python. A fim de o utilizador testar a versão BETA, foi lançada a seção em classe de carros já treinada pelo *StyleGAN* e escrito um guião de instalação, outro guião de requerimentos e o por fim os componentes necessários para fazer *download*. Também são fornecidos documentos de API (*Application Programming Interface*), explicando o propósito de cada *input*, *output* e *usage*. Para apresentação dos resultados desta fase experimental, a equipa apresenta vídeos de demonstrações dos carros a serem editados, classe de edição que foi inserida na versão experimental. Também apresenta mais exemplos em outras classes, como periquitos, retratos faciais e gatos, não estando estas no BETA.

2.5 Implementação e Resultados

Foram utilizadas divisões detalhadas de componentes de *DatasetGAN* (Yuxuan Zhang *et al.*, 2021). O esquema de anotação face a face é mostrado na Figura 3. Todas as informações

adicionais estão presentes no apêndice. Quando a edição é apenas baseada na otimização ou quando se aprende os vetores de edição, executa-se sempre 100 passos de otimização utilizando o modelo de otimização sofisticada. A fim de demonstrar a funcionalidade da edição, utilizaram-se imagens genuínas do conjunto de teste do *DatasetGAN* que não faziam parte do treino do GAN para Carro, Gato e Rostos (Diederik P *et al.*, 2014).



Figura 3: Esquema de rótulos do rosto de uma pessoa (Huan Ling et al., 2021).

2.5.1 Resultados Qualitativos

2.5.1.1 No domínio

A Figura 4 mostra a estrutura do *EditGAN* aplicando vetores de edição δw_{edit}^+ em imagens novas e refinando com 30 passos de otimização. A operação de edição mantém as imagens de alta qualidade e são separadas para cada classe. É importante observar que muitas das operações de edição produzem imagens manipuladas plausíveis, ao contrário das imagens que aparecem nos dados de treino do GAN. Por exemplo, os dados de treino não incluem gatos com orelhas ou olhos excessivamente grandes.

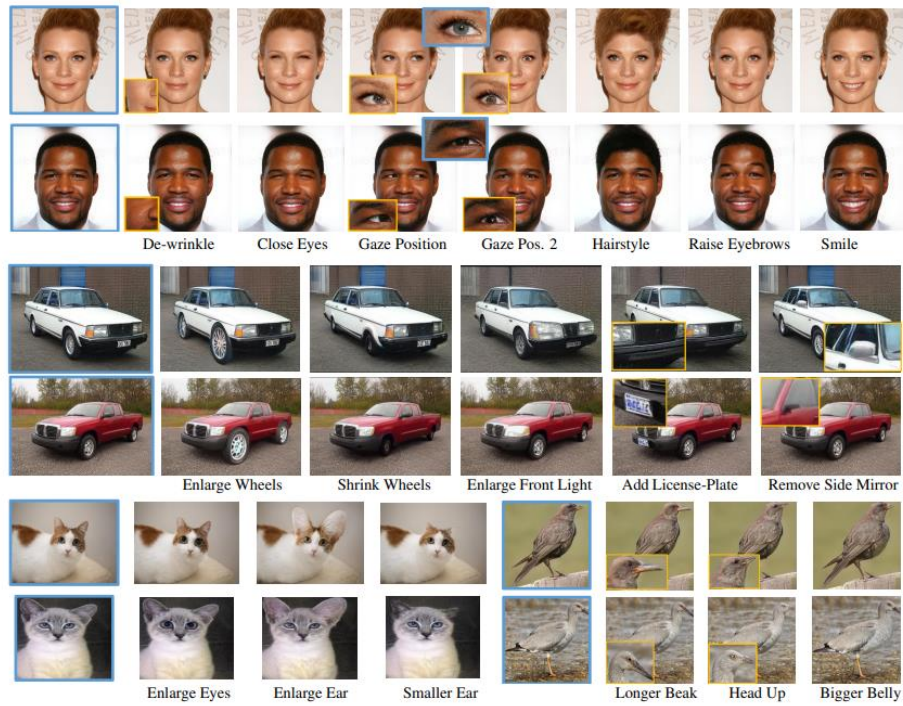


Figura 4: Exemplos de edições orientadas para a segmentação com *EditGAN*

Na Figura 4, os resultados apresentados foram baseados em refinamento com supervisionamento automático de 30 passos e edição com vetores. Caixas azuis indicam representações originais. Caixas laranja indicam a capacidade de ampliar a visão (Huan Ling *et al.*, 2021).

2.5.1.2 Fora do domínio

No conjunto de dados *MetFaces*, desenvolvido para avançar a pesquisa no reconhecimento facial e análise de rosto, oferecendo outra dimensão para o seu treino e avaliação, foi demonstrada a capacidade do *EditGAN* de se generalizar para dados fora do domínio (Tero Karras *et al.*, 2020). Foi utilizado o modelo *EditGAN*, para criar vetores de edição δw_{edit}^+ utilizando caras reais no domínio. Em seguida, foi aplicado os vetores de edição com 30 passos de refinamento supervisionado automático e incorporado pelos retratos *MetFaces*, que têm 100 passos de otimização. Mesmo para exemplos tão distantes do domínio, verifica-se que as operações de edição se traduzem facilmente.

2.5.2 Resultados Quantitativos

O parâmetro de referência *smile edit*, desenvolvido para alterar expressões faciais específicas, que o *MaskGAN* introduziu para medir a quantidade de habilidades de edição de imagem do *EditGAN*. O foco principal do *MaskGAN* é fornecer a geração de imagens com atributos específicos e também recuperar partes de uma imagem que estão em falta, complementando a coerência visual e a qualidade dos pormenores (Cheng-Han Lee *et al.*, 2020). O desempenho é avaliado por meio de três métricas:

- a. **Correção semântica:** Foi utilizado um classificador de atributos de sorriso pré-treinado para avaliar se os rostos mostravam expressões sorridentes após a edição.
- b. **Qualidade da imagem ao nível da distribuição:** as distâncias de compreensão de *Frechet Inception Distance* (FID), métrica utilizada para avaliar a qualidade de imagens geradas via GANs e as suas distâncias em termos de distribuições de características reais e geradas (Min Jin *et al.*, 2020). Já a *Kernel Inception Distance* (KID), é uma métrica com o mesmo propósito, mas opera com uma fórmula diferente (Mikołaj Binkowski *et al.*, 2018).
- c. **Conservação da identidade cultural:** Foi utilizada a rede de extração de características *ArcFace* (Jiankang Deng *et al.*, 2019) pré-treinada para verificar se a identidade dos sujeitos é mantida ao fazer a edição. A *ArcFace* é uma abordagem bastante inovadora para verificar e reconhecer o retrato facial de uma pessoa. Destaca-se por melhorar a precisão do reconhecimento facial com a sua técnica de otimização de perda durante o treinamento das redes neurais profundas. Reportou-se assim a semelhança entre as imagens originais e editadas em particular.

2.6 Comparação de Métodos Generativos

O *EditGAN*, foi comparado com outros quatro métodos concorrentes, que utilizam técnicas bastante semelhantes e avançadas para a manipulação de atributos faciais em imagens geradas:

- 1) *MaskGAN2* utiliza imagens sem sorrisos, além de suas máscaras de segmentação e uma máscara de segmentação de sorrisos alvo. O treino do *MaskGAN2*, requer conjuntos de dados anotados consideráveis (Cheng-Han Lee *et al.*, 2020).
- 2) *LocalEditing3* (técnica avançada para editar imagens locais, realiza modificações precisas e realistas em zonas específicas da imagem), combina características do GAN para realizar uma edição local e usa imagens de referência, neste caso rostos com expressões sorridentes (Edo Collins *et al.*, 2020).
- 3) *InterFaceGAN4* é outra base idêntica ao *EditGAN*, o *InterFaceGAN* procura vetores de edição no espaço latente. Mas este utiliza classificadores de atributos auxiliares e baseia-se em conjuntos de dados consideráveis anotados. Em geral, ele não consegue obter o controle de edição completo do nosso *EditGAN* (Yujun Shen *et al.*, 2020).
- 4) *StyleGAN2 Distillation*, como referido em cima, é um método alternativo que não exige a inclusão de imagens genuínas e usa um modelo de vetor de edição para criar um conjunto de dados de treino (Yuri Viazovetskyi *et al.*, 2020).

Tabela 1: Comparações quantitativas com várias linhas de base de edição de sorrisos baseada na de Huan Ling *et al.*, 2021.

MÉTRICA	ANOTAÇÃO DA MÁSCARA	ANOTAÇÃO DE ATRIBUTO	PRECISÃO DE ATRIBUTO (%)	FID	KID	PONTUAÇÃO DA IDENTIDADE
MaskGAN	30,000	-	77,3	46,84	0,020	0,4611
LocalEditing	-	-	26,0	41,26	0,012	0,5823
LocalEditing - Enconding4Editing	-	-	41,75	48,28	0,016	0,6603
InterfaceGAN	-	30,000	83,5	39,42	0,010	0,7295
EditGAN	16	-	91,5	41,74	0,013	0,7047
EditGAN+30	16	-	85,8	40,83	0,012	0,7452
StyleGAN2 Distillation	-	30,000	98,3	45,09	0,013	0,7823

Nesta comparação, utilizou-se 1,875 vezes menos rótulos de treino para superar o *MaskGAN*, em cada uma das três métricas. De forma semelhante, os resultados são significativamente superiores aos do *LocalEditing*. De acordo com uma avaliação, o *LocalEditing* não funciona bem em inserir imagens reais e fornece um desempenho ligeiramente menor nas pontuações FID e KID. O *EditGAN* supera o *InterFaceGAN* em termos de preservação da identidade e precisão da classificação de atributos, enquanto o *InterFaceGAN* obtém pontuações FID e KID mais altas. Quando a escala do vetor de edição é pequena, a pontuação da identidade é elevada enquanto a pontuação do atributo é baixa, pois não há modificação nas imagens originais. Verificou-se assim, que a edição em tempo real com vectores de edição é comparável ao *InterFaceGAN*. Contudo, *EditGAN* supera o *InterFaceGAN* ao realizar o refinamento supervisionado automático em tempo de teste.

Por último, examinou-se o *StyleGAN2 Distillation*, que funciona bem, mas só permite a edição de atributos de imagem de alto nível, com anotações em grande escala, pois depende de classificadores treinados anteriormente, como o *InterfaceGAN*. Além disso, separa as edições em redes *Pixel2PixelHD* (técnica para a tradução de imagens de alta resolução) distintas, o que significa que cada edição requer treinamento de uma nova rede, e que limita a aplicabilidade abrangente e interativa do utilizador (Yuri Viazovetskyi *et al.*, 2020).

Foi também examinado o tempo de execução da edição utilizando a GPU NVIDIA Tesla V100. A otimização condicional, dada uma máscara de segmentação modificada, leva 11,4 segundos com 30 etapas de otimização. Este procedimento fornece o vetor de edição. Por conseguinte, uma edição interativa complexa em tempo real é possível graças à rapidez quase instantânea da aplicação dos vetores de edição, que dura apenas 0,4 segundos.

3. Metodologia de Investigação

A metodologia utilizada nesta Dissertação é apresentada neste capítulo. A técnica de design está centrada na edição de imagem e a pensar nas interações que o utilizador tem de fazer desde o início até ao final da edição. Os métodos utilizados, como a análise de dados de pesquisas, observação, entrevistas e testes de usabilidade são posteriormente elaborados, exemplificando como são aplicados.

A informação desta investigação foi obtida maioritariamente através do *Scopus* e do *Google Scholar*, após procurar e organizar a informação recolhida de fontes que me transmitiam confiança e devida importância.

Dessa informação recolhida, foi feita uma síntese de texto que era pertinente para esta investigação. Foram também utilizados filtros de pesquisa, assim como: por data de publicação, por número de citações exercidas, autores, títulos e subtítulos. Qualquer informação adicional, foi obtida em sites fidedignos e com reputação. Todo este processo foi estudado e aprendido na cadeira de Metodologias de Investigação, que introduziu o programa *Zotero*, que foi bastante útil para este processo.

A análise de dados recolhidos dos participantes, foi baseada em tabelas ou gráficos exercidos no site *Canva* (<https://www.canva.com>).

Em relação à arquitetura dos *mockups*, foi utilizado o *Figma*, um site de edição online, próprio para fazer *mockups* de aplicações de alta qualidade, proporcionando assim serviços interativos, ícones, ecrãs, modelo, etc., tal como um editor necessita para a realização da arquitetura de uma aplicação.

3.1 Plano de Pesquisa

Para atingir os objetivos estabelecidos nesta investigação, foi criado um cronograma, que é composto por nove macro tarefas: 1. Revisão da literatura, 2. Organização dos fundamentos do design, 3. Inquérito preliminar ao utilizador, 4. Criação do protótipo, 5. Testes de avaliação, 6.

Análise da eficácia do protótipo, 7. Plano estratégico de mercado, 8. Guia de usabilidade do gerador, 9. Redigir a Dissertação.

Estas tarefas foram organizadas e priorizadas de acordo com a sua importância para facilitar a gestão do projeto, como mostra a Tabela 2.

Tabela 2: Plano de Pesquisa definido para esta Dissertação.

TAREFAS	OUT	NOV	DEZ	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET
Revisão da literatura												
Organização dos fundamentos do design												
Inquérito preliminar ao utilizador												
Criação do protótipo												
Guia de usabilidade												
Testes de avaliação												
Análise da eficácia do protótipo												
Observações finais e ajustes												
Redigir a tese												

3.2 Design centrado no utilizador

As interfaces padrões por vezes podem ser difíceis de utilizar, levando o utilizador a deparar-se com dificuldades na realização de tarefas simples. A fim de criar uma proposta de interface interativa, recorreremos a uma abordagem de design centrado no utilizador.

Esta é uma abordagem que envolve os utilizadores requisitados, no processo de design para influenciar a forma como o design é concessionado (Abrás *et al.*, 2004). Engloba uma variedade de métodos que procuram melhorar a usabilidade e experiência do utilizador em produtos e sistemas. Este grau de envolvimento do utilizador varia entre o seu espectro, passando por uma consulta das suas necessidades até à conceção do processo todo. Pois o principal conceito da UCD

(*User Centered Design*), é que os utilizadores estejam envolvidos durante este processo de conceção do início ao fim.

3.3 Métodos

Os métodos utilizados neste estudo, são descritos nesta secção. O design e interface do protótipo deve ser gerido e planeado de forma cuidada para que no final suceda às expectativas do utilizador. Contudo, este estudo focou-se em dois tipos de dados: A fase inicial envolveu a utilização de dados e estudos das redes adversárias generativas já inseridas em aplicações concessionadas e totalmente funcionais, como tal foi registado os seus principais problemas, as suas principais vertentes, as opiniões da comunidade e os comentários dos investigadores. Estas fontes de dados forneceram informações pertinentes sobre os requisitos e as práticas gerais para uma aplicação de edição de imagem via inteligência artificial, assim como também ajudou na criação de cenários e na *interface*.

Na fase seguinte foram feitas entrevistas, questionários e testes de usabilidade da *interface* com possíveis utilizadores. Estes participantes foram escolhidos aleatoriamente dentro da faculdade de Engenharia da Universidade do Porto, sendo estudantes, docentes e até visitantes. Esta última fase, focou-se na avaliação e aperfeiçoamento dos requisitos da *user interface* estabelecidos na primeira fase. Após serem feitos todos os testes e obtidos diversos *feedbacks* da comunidade e participantes, o objetivo de obter uma visão abrangente das necessidades e preferências do utilizador foi concluído.

Finalizando assim a concessão do protótipo com todas as modificações necessárias, obtivemos uma parte interativa da *user interface* direcionada a obter *feedbacks* finais, resultando também numa amostra funcional do protótipo.

3.4 Tipos de Aplicações de Edição de Imagem com IA

A *DeepArt* foi uma aplicação de edição de imagem automática gratuita, que revolucionou bastante o mercado em 2016. Utiliza redes neurais convolucionais (CNNs) e o seu processo envolve uma imagem de conteúdo fornecida pelo utilizador e a imagem de estilo artístico escolhida na aplicação, resultando numa junção das características das duas imagens para obter

uma nova imagem gerada resultante destas duas. Este processo tenta equilibrar ao máximo tanto os aspetos de conteúdo como os de estilo, de modo a obter uma boa imagem final. (Informações recolhidas do site oficial da *DeepArt* (<https://www.deeparteffects.com/>).

Um dos maiores problemas de aplicações com este método de edição é o facto da qualidade de imagem variar significativamente e não existir precisão alguma na concessão da edição. Sendo também bastante limitada na personalização da imagem. Outras aplicações do mesmo género e que utilizam o mesmo método é a *Prisma* (criada em 2016) e a *DeepDream* (criada em 2015) sendo das aplicações com mais impacto significativo na exploração artística da inteligência artificial (Gatys *et al*, 2015).

Mais tarde em 2017, foi criada a *FaceApp*. Uma aplicação que ficou viral nesta altura simplesmente pelo fato da sua precisão e realismo na edição facial. Aqui o utilizador era capaz de editar o rosto das suas fotografias, tornando-as mais envelhecidas (funcionalidade mais usada) ou até adicionar sorrisos e alterar o género da pessoa. Esta aplicação utiliza na mesma as CNNs, mas o treinamento de dados é muito mais amplo a fim de alterar e identificar características específicas. Tendo melhores resultados que as aplicações anteriores (Jun-Yan Zhu *et al.*, 2017). Por outro lado, a *FaceApp* gerou novos problemas relacionados com a privacidade e segurança do utilizador, pois a empresa por detrás desta aplicação é acusada de utilizar as fotografias dos utilizadores para treinamento do algoritmo das redes neurais convolucionais, mesmo não estando autorizados a fazê-lo. Após estas acusações a empresa teve de alterar a sua política de privacidade. (Informação recolhida do Jornal Público: <https://www.publico.pt/2020/06/19/p3/noticia/faceapp-voltou-riscos-são-1921223>).

Em 2021, foi lançado uma aplicação chamada *DALL-E*, baseada numa versão do GTP-3 e feita para a geração de uma imagem única a partir de uma descrição textual feita pelo utilizador. Este programa de edição produz conteúdo muito criativo e inovador, com uma alta resolução e qualidade de imagem (informações recolhidas do site oficial da *DALL-E*: <https://openai.com/index/dall-e-2/>). Em 2022, obteve uma versão melhorada da aplicação, ficando com o nome de *DALL-E 2*, otimizando a coerência e qualidade de imagens geradas, permitindo também adicionar novos elementos ou filtros e regenerar elementos específicos. Esta aplicação foi bastante importante, pois abriu novas portas na arte e design inseridos pela inteligência artificial como também foram criadas mais aplicações utilizando o mesmo método, assim como o *MidJourney*, *Stable Diffusion*, *Artbreeder* e entre muitas outras (Nichol, Alex *et al.*, 2021).

Mesmo assim, atualmente o *Adobe Photoshop*, é a ferramenta de edição de imagem mais comum aos olhos de um editor, sendo que a empresa desenvolveu uma ferramenta adicional de

edição automática e *machine learning*, chamada de *Adobe Sensei*. Com as principais funcionalidades de remover objetos indesejados de uma determinada imagem, selecionar automaticamente uma área principal da imagem como também selecionar objetos específicos, substituição do céu da imagem, filtros com ajustes criativos e corretivos assim como também mudança de cor, contraste e tom automático. Funcionalidades bastante interativas e práticas, com bastante precisão e eficácia, pois simplificam tarefas complicadas e oferecem bastante criatividade ao utilizador. Os maiores problemas de acordo com os comentários de utilizadores, podem ser encontrados em imagens com muita complexidade e fora do comum, pois dificulta bastante o algoritmo treinado da inteligência artificial. (Informações recolhidas do site oficial da Adobe: <https://business.adobe.com/pt/products/sensei/adobe-sensei.html>.)

De uma forma geral, a melhor aplicação a utilizar foi o *Adobe Photoshop*, simplesmente pelo fato de conseguir obter um resultado como se esperava e poder modificar partes específicas de uma imagem com bastante precisão e qualidade. O mesmo não se pode afirmar da *DeepArt* e *DeepDream*, pois geram automaticamente uma imagem enviada a partir de outra (ou com um filtro), nunca fica um resultado espetável, mas sim uma imagem estranha. A aplicação mais recente e bem conseguida na geração de imagens automáticas a partir de uma descrição textual foi a *DALL-E*, tendo também como principal problema o facto de o resultado não ser de todo como esperado e o serviço ser pago. Por último a *FaceApp* é uma aplicação excelente para quem não quer passar muito tempo na edição e consegue produzir um conteúdo divertido e bonito ao aplicar os filtros, a única menos valia é a precisão da edição que por vezes estraga a imagem original enviada.

Concluindo, uma edição automaticamente com mais interatividade e personalização por parte do utilizador é uma mais-valia para este tipo de aplicações, sendo que o importante é que o utilizador goste do resultado, não importando assim o tempo e a dificuldade que irá necessitar para o obter. Por conseguinte, o *Editable-Gen* relaciona-se mais ao *Adobe Photoshop* do que às outras aplicações apresentadas.

De acordo com o físico teórico Stephen Hawking: “A criação bem-sucedida de inteligência artificial seria o maior evento na história da humanidade. Infelizmente, pode também ser o último, a menos que aprendamos a evitar os riscos”. Esta citação é inspiradora e faz perceber que a inteligência artificial pode e deve ser usada somente como uma ferramenta de ajuda e não como a chave de um produto final.

3.5 Requisitos de Design

Os requisitos de *design* considerados para o protótipo foram:

Os ícones escolhidos são bastante simples (todos retirados do site do Figma, gratuitamente), de forma a existir clareza, consistência visual, e serem adequados para cada botão, bastante genéricos e acessíveis para toda a gente compreender. Sendo que o principal objetivo, é melhorar significativamente a usabilidade e estética do protótipo.

Em termos de funcionalidades, estas consistem em todas as básicas que uma aplicação oferece, assim como definições gerais, ecrãs de *loading*, registo de uma conta social, edição de perfil, comentários, gostos, partilhas e entre outras, realçando a funcionalidade mais importante, o editor de imagem automática.

Já na navegação sobre o protótipo, o utilizador tem de estar confortável quer visualmente quer na interação sobre o protótipo. Por isso, este consiste numa barra de navegação bastante simples e eficaz em baixo do ecrã do dispositivo móvel, permitindo o utilizador de navegar entre todas os ecrãs sem conflito algum e quando quiser. Ilustrando assim a estrutura geral da aplicação.

Em termos de cores, foram unicamente utilizadas as cores da paleta explicadas mais à frente, cores simpáticas a nível visual, com o objetivo de não cansar os olhos dos utilizadores. Cada cor foi definida da melhor forma para cada componente do protótipo, oferecendo um bom *feedback* visual.

A documentação sobre o funcionamento do protótipo, decisões de design, as diversas funcionalidades são explicadas na seção seguinte, algo que é bastante fundamental existir em qualquer aplicação.

A *interface* do utilizador (UI) está completa, contendo assim botões, menus, ícones, texto, guias e entre outros elementos visuais importantes para a estética e *layout* do protótipo

4. Protótipo Interativo

Esta secção do relatório centra-se na conceção do protótipo de interface otimizado para dispositivos móveis com um sistema operacional IOS ou Android. Esta decisão foi tomada pelo simples fato de que a aplicação de demonstração do *EDITGAN* foi desenvolvida apenas para o computador, por isso seria vantajoso existir um protótipo para demonstração da aplicação em dispositivos móveis. Com isto poderá trazer inúmeras vantagens, como o fácil acesso da sociedade, pois é bastante mais popular o uso de aplicações de edição em dispositivos móveis, atualmente, do que o uso do computador, pois apresenta maior conveniência e portabilidade. A maior parte da sociedade tem consigo um telemóvel sempre pronto a usar, o que facilita bastante em edições rápidas como este tipo.

Esta proposta de *interface* foca-se somente em obter *feedbacks* pessoais de participantes escolhidos, e não se destina a partilhar dados de interesse com outras partes sem existir um consentimento pelo investigador. Os protótipos interativos são representações tangíveis de um produto ou sistema que permitem aos utilizadores explorar e avaliar funcionalidades e interações, simulando uma experiência real. Ao contrário dos protótipos estáticos, que apenas mostram visualizações dos ecrãs ou fluxos de trabalho, os protótipos interativos incluem comportamentos dinâmicos e respostas às ações dos utilizadores, proporcionando uma compreensão mais detalhada da usabilidade e da experiência do utilizador. Os protótipos interativos apresentam características tais como a interatividade, a fidelidade, o *feedback* em tempo real e interatividade no desenvolvimento:

- A nível interativo, baseiam-se todas as ações que o utilizador terá de fazer no protótipo, assim como cliques, desenhos, pinturas, arrastar o ecrã para cima e para baixo e muitas outras ações.
- A fidelidade consiste em vários níveis, tendo esta como objetivo uma representação do produto final, tanto em termos de funcionalidade como de aparência, sendo que esta pode ir de simples esboços a protótipos altamente detalhados, como podemos observar no Anexo 6 um esboço de baixa qualidade somente para guia de concessão do protótipo de alta qualidade.

- O *feedback* em tempo real, é uma forma de identificar no protótipo eventuais problemas e otimizações que o mesmo possa necessitar através da utilização do protótipo pelos utilizadores, assim como oportunidades de melhoria na aplicação.
- A interatividade no desenvolvimento, facilita a rápida iteração no *design*, permitindo que *designers*, programadores e *stakeholders* testem e aprimorem ideias de forma ágil. Vantagem esta que está presente no Figma, o que ajudou bastante na concessão do protótipo.

Os protótipos interativos têm bastante relevância no desenvolvimento de *design* destinado ao utilizador (UCD). Considerando algumas possibilidades ao realizar a criação deste tipo de protótipo, são por exemplo a validação de conceitos, que permite testar e validar ideias de *design* com utilizadores reais como forma de teste, poupando tempo e recursos aos desenvolvedores do protótipo para o produto final. Oferece um produto tangível a possíveis *stakeholders* para que estes tenham uma melhor noção do produto em questão assim como forma de comunicar ideias e obter *feedback* dos mesmos (Don Norman, 2008).

A parte social da aplicação consiste na afirmação de Zhaoqing Pan et al., em 2019, realçando que uma partilha social da arte gerada, a fim de o ser humano avaliar a sua qualidade e desempenho é melhor forma métrica de avaliação, face a métricas baseadas em fatores de informação, como por exemplo a FID e KID.

4.1 Figma

Existem diversas técnicas e ferramentas para a criação de protótipos interativos, exemplos destas são o Adobe XD, Sketch, Axure entre outras, a ferramenta utilizada para a produção do protótipo Editable-Gen foi o Figma, sendo esta uma ferramenta de *design* colaborativo que funciona puramente na *web*, maioritariamente utilizada para a criação de *interfaces* para o utilizador (UI – *User Interface*) e para a experiência do utilizador (UX – *User Experience*).

O que diferencia o Figma de algumas das ferramentas de *design*, é a facilidade da criação de protótipos interativos, permitindo simulações de pesquisa, também oferece ferramentas para criação de *layouts*, e proporciona formas de reutilizar e criar componentes derivados á UI, assim como integrar e exportar arquivos, tais como imagens e diversos outros formatos, PNG, JPG e entre vários outros.

O protótipo interativo desenvolvido pode ser interagido em: <https://www.figma.com/proto/eIPTUav1ghjqWhC9bLCne/Editable-Gen?node-id=119-3402&node-type=CANVAS&t=JemiFP2vvJ5k4WTO-1&scaling=min-zoom&content-scaling=fixed&page-id=0%3A1&starting-point-node-id=119%3A3402&show-proto-sidebar=1>

4.2 Criação do Protótipo

O nome introduzido para este protótipo foi de Editable-Gen, que significa *editable generator*, ou seja, gerador editável. Este nome foi escolhido a pensar no seu método em causa *EDITGAN* e numa aplicação de edição de imagem automática, sendo que parte do processo usado é de gerir imagens numa só.

Seguidamente à escolha do nome, foi escolhida a paleta de cores pois é das partes mais importantes no desenvolvimento de um protótipo. As cores são dos primeiros elementos que os utilizadores irão reparar ao usar qualquer outra aplicação, realçando bastante importância na atração visual. É também deveras importante a escolha da cor por motivos como a identidade da marca, o contraste e a legibilidade das cores, sendo que as cores devem garantir que o texto e elementos do protótipo sejam legíveis assim como a própria emoção e psicologia, visto que as cores influenciam o humor e as emoções dos variados utilizadores (Josef Alber, 2013):

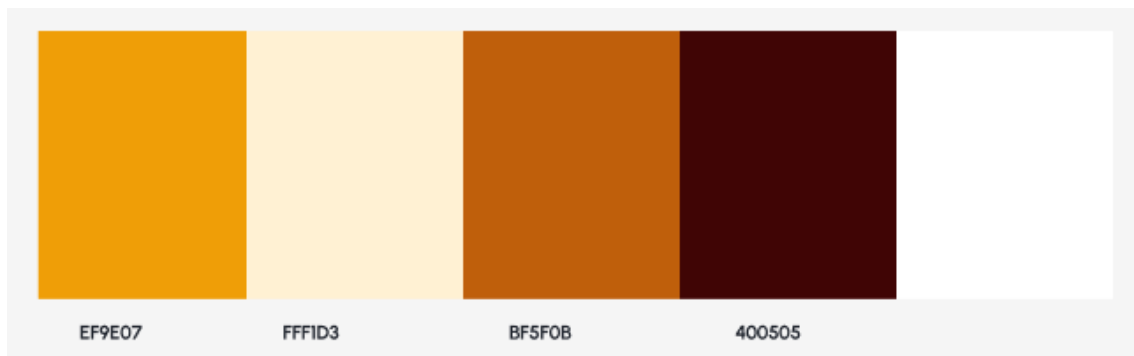


Figura 5: Paleta de cores escolhidas para o protótipo desenvolvido.

1. O laranja vivo (EF9E07) é uma cor com uma tonalidade intensa e energética, é adequada e utilizada para chamar à atenção dos utilizadores da aplicação.

2. O bege claro (FFF1D3) é uma cor neutra e suave, sendo excelente para se usar como fundo permitindo por exemplo que conteúdos ou imagens editadas sobressaíam.

3. O laranja-escuro (BF5F0B) é uma tonalidade agressiva e maioritariamente usada como forma de destaque visual como por exemplo para botões secundários relevantes ou para ícones.

4. O castanho avermelhado escuro (400505) é usado como forma de aplicar um contraste com o resto das cores da paleta de cores, realça-se com a exposição.

5. O branco (FFFFFF) é a cor mais versátil, usada maioritariamente para fundos ou para criar fortes contrastes com as tonalidades mais escuras, assim como a cor (BF5F0B) ou a (400505), o branco ajuda também a manter uma experiência visual clara e organizada da aplicação ao utilizador. É utilizado mais no editor automático, para desenhos e pinturas.

Após a escolha de cores foi concebido o logótipo no Photoshop (programa de edição pormenorizada), uma simples imagem que representa o protótipo num todo. Foram desenhados vários tipos de logótipos numa tentativa erro, e discutido verbalmente com estudantes de multimédia, qual o que ficaria melhor numa aplicação deste género. Após isso foi feita a escolha, como podemos ver na Figura 7, é parecido com uma tomada elétrica, mas com um raio de eletricidade no meio e a sair da tomada, significando assim um gerador com energia. Este resultado foi bastante positivo pelo *feedback* obtido de outros estudantes de multimédia. Este logótipo teve como medidas 200 pixels de altura e 180 pixels de largura.

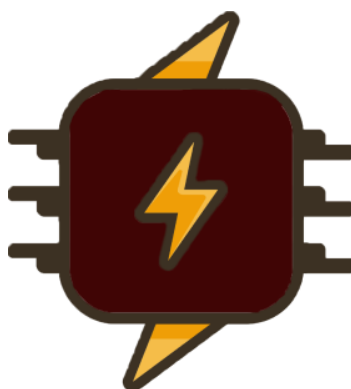


Figura 6: Logótipo escolhido para representação do protótipo desenvolvido.

Para a criação do protótipo, foi necessário criar uma divisória entre dois termos como forma de organização do protótipo, sendo estes dois conceitos: Módulo de Autenticação e Gestão de Utilizador e o Módulo de Interação e Conteúdo (Steve Krug, 2009).

4.3 Autenticação e Gestão do Utilizador

Este módulo representa todo o conjunto de funcionalidades e ferramentas utilizadas para as necessidades básicas do protótipo, tais como para o *login*, senha, registo, os termos e serviços, recuperação da palavra-passe e entre outras componentes e funcionalidades que garantem a administração e a segurança dos utilizadores. Na Figura 8, podemos ver os diferentes ecrãs deste módulo.

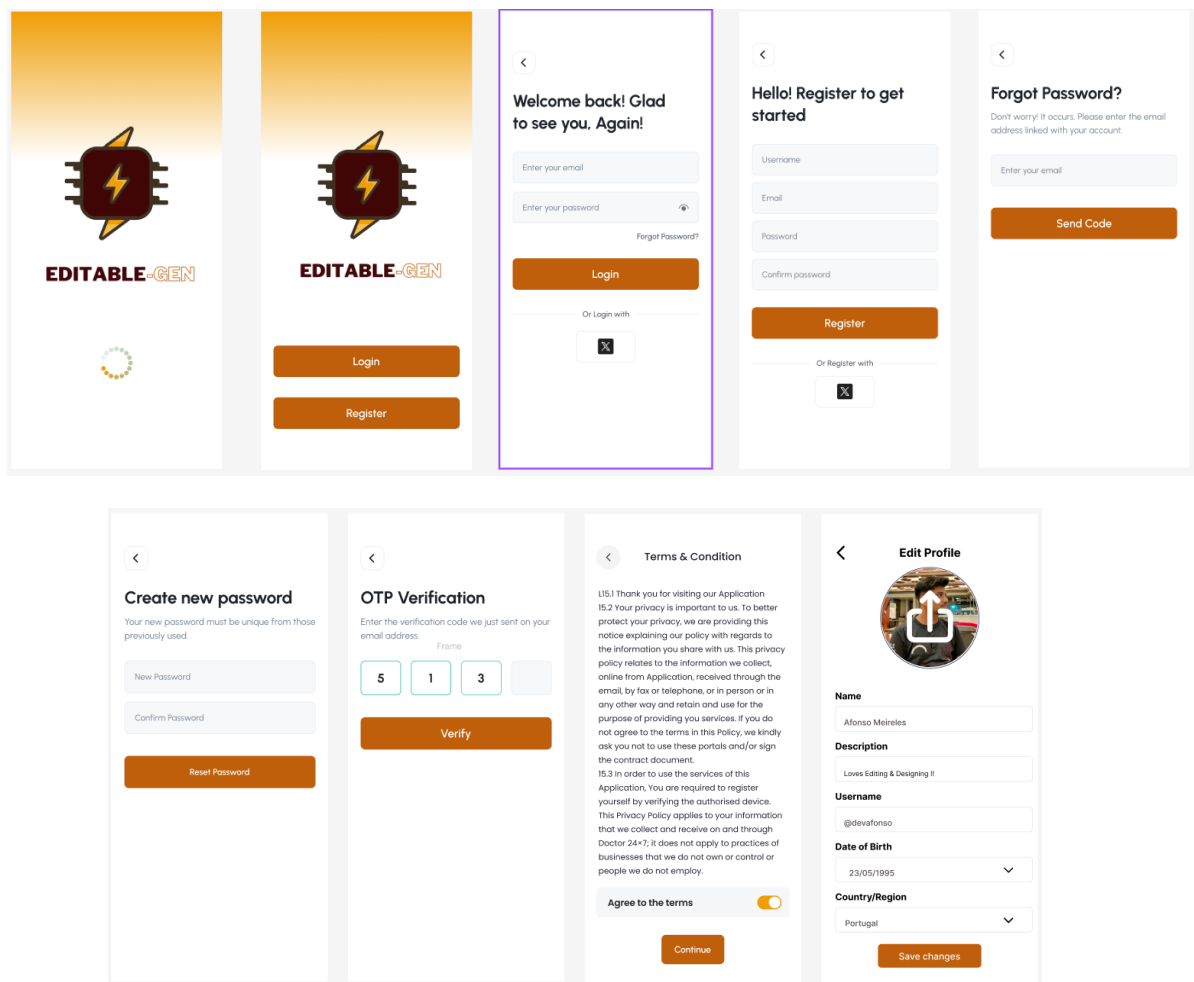


Figura 7: Representação dos ecrãs do Módulo de Autenticação e Gestão do Utilizador, início do protótipo desenvolvido.

O ecrã de *login* inicial apresenta uma interface intuitiva e simples, com o logótipo da aplicação centralizado como forma de reforçar a identidade da aplicação. Como forma de realçar o logótipo é usada a cor laranja forte na parte superior do ecrã sendo que esta se vai desvanecendo para o branco ao chegar ao logótipo e aos botões de *Login* e de Registrar como forma de realçar e enaltecer tanto o logótipo como ambos os botões que apresentam a funcionalidade, foi aplicado aqui um gradiente. Ambos os botões de “Login” e de “Register” apresentam uma cor mais viva (laranja) que fornecem um fácil contraste com o branco de forma a facilitar a visão do utilizador

O ecrã de registo da aplicação foi desenvolvido com um *design* funcional e direto em mente, com o objetivo de simplificar o processo de criação de uma nova conta. No topo do ecrã, está o campo do “Username” designado pelo nome do utilizador, onde estes inserem o seu nome de pretendido. Abaixo, encontram-se dois campos dedicados à palavra-passe: “Password” e “Confirm Password”, garantindo que a palavra-passe escolhida é repetida corretamente para evitar erros.

Na parte inferior do ecrã de registo, destacam-se dois botões principais. O botão de “Register” que finaliza o registo, assim como também existe um botão para registo através do Twitter, sinalizado com o logótipo X, permitindo aos utilizadores de efetuar o registo de forma mais rápida utilizando a sua conta da rede social. Rede esta escolhida com o propósito de os utilizadores partilharem lá as suas edições, visto que é uma rede bastante popular e mais usada para a arte gerada e todo o tipo de informações.

O ecrã de *Login* na aplicação após o registo estar feito, apresenta uma mensagem de boas-vindas ao utilizador agradecendo ao utilizador por estar de volta na aplicação, incentivando-o ao uso da aplicação. É possível visualizar também uma caixa de texto onde é colocado o “Email” e a “Password” do utilizador. É importante salientar o uso na caixa de texto da *password* de um emoji na parte final da caixa de texto que ao ser usado permite ao utilizador de visualizar a sua palavra-passe para verificar se a mesma se encontra correta.

A verificação OTP (*One Time Password*) é uma camada extra de segurança que ajuda a proteger o acesso a contas dos utilizadores no protótipo assim como é usada para confirmar transações de forma mais segura (componentes *premium*). A verificação baseia-se num código de 4 dígitos que é gerado de forma aleatória e enviado ou para o telemóvel ou para o email de determinado utilizador. O código só pode ser usado exclusivamente uma vez e existe um condicionante temporal de forma a não ser possível forçar a descoberta do código por parte de terceiros. O processo é rápido e simples e exige pouco esforço por parte dos utilizadores, tornando a experiência mais segura.

O ecrã de “Termos e Condições” é deveras importante em qualquer aplicação ou protótipo, ele introduz aos utilizadores a definição das normas que regulam a relação entre o utilizador e quem está a fornecer o serviço. É um documento essencial para que ambos os lados compreendam e concordem com os direitos, privacidade e obrigações associados ao uso da aplicação.

A nível do programador do protótipo, o ecrã serve como função protetora, ao indicar aos utilizadores claramente as condições de uso da aplicação assim como as regras que devem ser seguidas para a utilização correta do serviço. Algumas aplicações informam ainda de forma legal, possíveis consequências a nível de taxas que podem ser aplicadas caso exista uma violação dos termos estabelecidos.

A edição do perfil numa aplicação ou num protótipo é importante, sendo que esta é uma ferramenta essencial para os utilizadores que pretendem personalizar e gerir as suas informações pessoais. Através deste ecrã é possível alterar a foto de perfil, modificar o nome, criar ou ajustar a descrição pessoal do perfil, criar ou alterar o *username*, a data de nascimento e o país de residência dos utilizadores.

Cada uma destas funcionalidades permite que o utilizador controle a forma como se apresenta no protótipo, seja para refletir mudanças pessoais ou simplesmente para manter o perfil atualizado. A troca da foto de perfil, por exemplo, permite ao utilizador escolher uma imagem que o represente melhor num dado momento. Já a opção de alterar o nome e o *username* proporciona flexibilidade na forma como o utilizador quer ser identificado pelos outros.

A nível dos ícones é importante realçar que para se realizar o *upload* da foto de perfil encontra-se um ícone onde está a foto de perfil, sendo este um ícone universal para *upload*. O botão “Save changes” destaca-se com uma cor laranja como forma de guardar toda a informação introduzida.

A página de perfil dos utilizadores do protótipo é deveras importante visto que vai ser nesta página que os utilizadores se vão conseguir expressar e demonstrar os vários aspetos de edição de fotografias, *design*, arte e entre outros.

Acentuam-se no ecrã do perfil do utilizador as interações importantes que os mesmos podem ter no ecrã do perfil, sendo estas os botões de edição de perfil, denominado por “Edit profile”, assim como o botão de partilha, denominado por “Share” e o botão de “+” que leva o utilizador para o editor automático, que se encontra colocado no centro do ecrã na parte inferior junto dos vários ícones que permitem ao utilizador navegar pelo protótipo, como podemos observar na Figura 9.

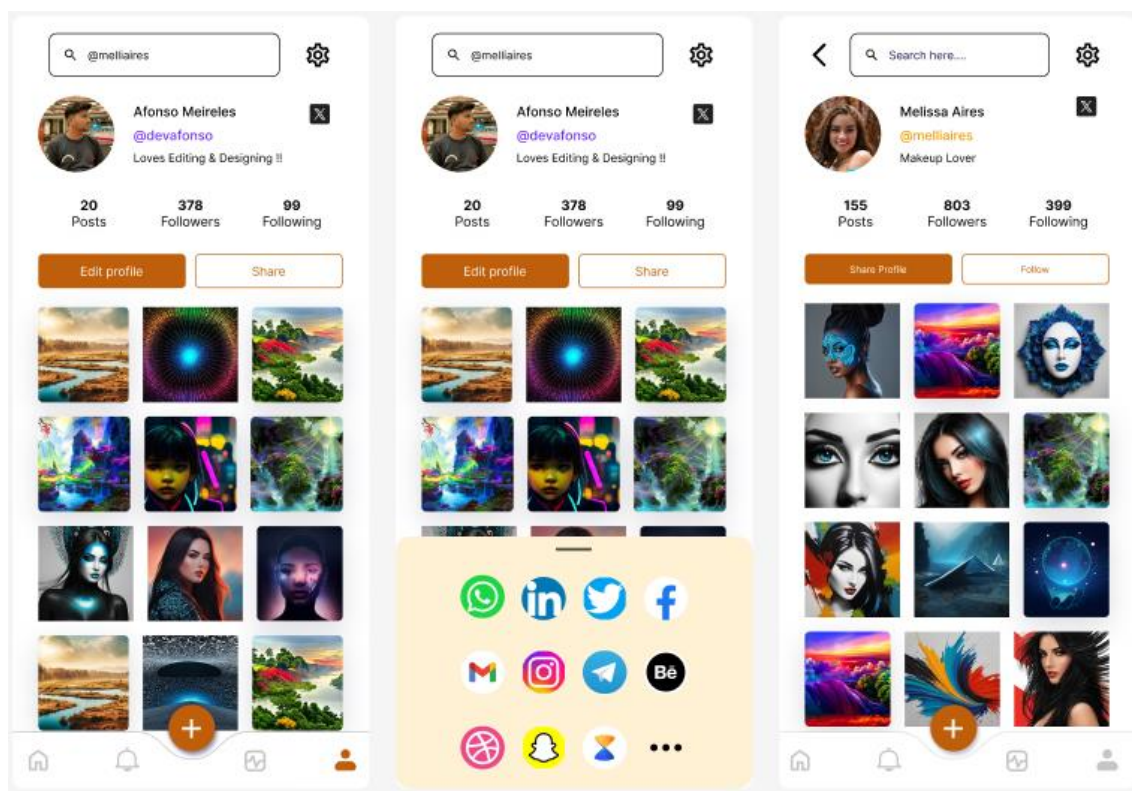


Figura 8: Ecrãs do perfil do utilizador e a sua pesquisa do protótipo desenvolvido.

Os ícones principais do protótipo inseridos na barra inferior do ecrã, permitem ao utilizador navegar pela aplicação, designados por: *Home*, Notificações, Galeria e a do Perfil, designadamente. Estes ícones encontram-se no protótipo em todos os ecrãs a partir do momento em que os utilizadores se encontrem no Módulo e Interação de Conteúdo do protótipo relatado mais á frente. Cada ícone tem um significado diferente, assim como:

- O ícone “Home”, é designado por um ícone de uma casa permite ao utilizador de ir para o ecrã principal do protótipo onde este poderá ver várias publicações dos diversos utilizadores que este possa estar a seguir, assim como também ver as suas próprias publicações. É o botão que se encontra em primeiro na ordem, a contar da esquerda, dos ícones com o objetivo de facilitar a experiência aos usuários.
- O ícone das notificações, é designado por um sino, é um ícone universal em várias aplicações e protótipos e este permite aos utilizadores de aceder a todas as notificações relacionadas com a aplicação.
- O ícone da galeria designado por quadrado com bordas arredondadas com uma linha de pulsação no centro do quadrado. O quadrado apresenta bordas arredondadas que adicionam

suavidades e modernidade ao design do mesmo enquanto as linhas de pulsação representam um símbolo universal de vida, movimento e energia, sendo este ícone excelente para representar toda a atividade que irá ocorrer na galeria, desde vídeos, fotos ou edições de fotografias por parte do editor.

- Para finalizar, o último ícone, designado pela forma de um ser humano, remete os utilizadores até à sua própria página de perfil, onde este poderá visualizar as suas fotografias, vídeos, ver o seu número de publicações, seguidores, pessoas a seguir, editar o perfil, partilhar novas publicações, aceder às configurações, e inclusivamente ter acesso ao X (antigo Twitter).

Os utilizadores ao carregarem no botão de partilha, denominado por “Share” deparam-se na parte inferior central do ecrã com várias aplicações diferentes, como por exemplo o WhatsApp, Facebook, Instagram, Telegram e entre outras com as quais eles podem partilhar.

4.4 Módulo de Interação e Conteúdo

Este módulo representa todas as funcionalidades e elementos do protótipo após os utilizadores terem acesso à aplicação depois de terminarem o processo do Módulo de Autenticação e Gestão do Utilizador. Ele define-se através de funcionalidades tais como a criação do perfil, a publicação de conteúdo social. Juntamente com interações sociais tais como gostos, comentários e partilhas. Este módulo representa ainda o sistema de notificações, pesquisa, descoberta e as configurações e preferências da aplicação, como podemos observar na Figura 10, e a representação do conteúdo publicado e as suas interações.

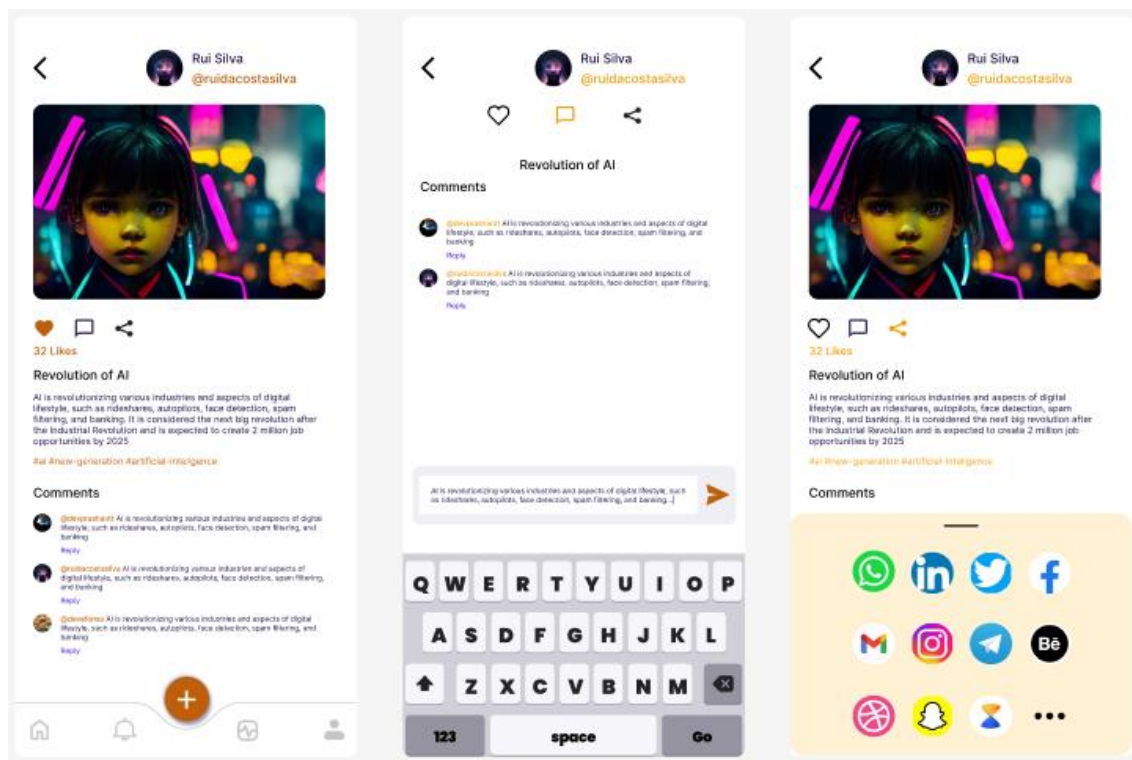


Figura 9: Representação das publicações no protótipo desenvolvido.

Através da visualização deste ecrã, é importante de realçar a estrutura assim como as cores dos vários aspetos que o definem. Primeiramente, na parte superior do ecrã encontra-se o nome do utilizador, assim como o seu *username*, sendo o *username* denominado a laranja. Em seguida encontra-se a publicação do utilizador, no centro do ecrã da Figura 10, podemos observar três ícones como o número de gostos, o título da publicação, a descrição da publicação e a arroba na qual representa o endereço de cada utilizador. É ainda possível realizar comentários na publicação pelos vários utilizadores do protótipo.

Os três ícones usados nestes ecrãs (Figura 10) são os seguintes:

O primeiro ícone, com a forma de um coração, indica o número de gostos (*Likes*) da publicação. O segundo ícone, com a forma de uma caixa de texto, permite aos utilizadores comentar na publicação. O terceiro ícone, com uma representação de uma “*network*”, uma ligação de vários pontos através de uma rede, permite aos utilizadores partilharem a publicação.

É importante também salientar que neste ecrã do protótipo, a presença da cor laranja, é a cor que define o protótipo da Editable-Gen, no número de *likes* da publicação, dos arrobas embutidos na publicação e nos nomes dos vários perfis que comentaram, com o objetivo de realçar aos

utilizadores estas questões de relevância. O botão que permite a funcionalidade de responder a um determinado comentário da publicação é com a cor azul. De forma a enaltecer esta outra funcionalidade do protótipo aos utilizadores, como forma de incentivo à comunicação entre os vários utilizadores.

A “Home Page” do protótipo da Editable-Gen é o local de grande relevância, assim como é o sítio onde os utilizadores passam a sua maior parte do tempo enquanto estiverem a usar o protótipo da aplicação para fins sociais. Esta página foi criada com atenção ao detalhe e ao pormenor, de uma forma moderna e limpa, oferecendo uma boa experiência de navegação ao utilizador, como podemos observar na Figura 11.

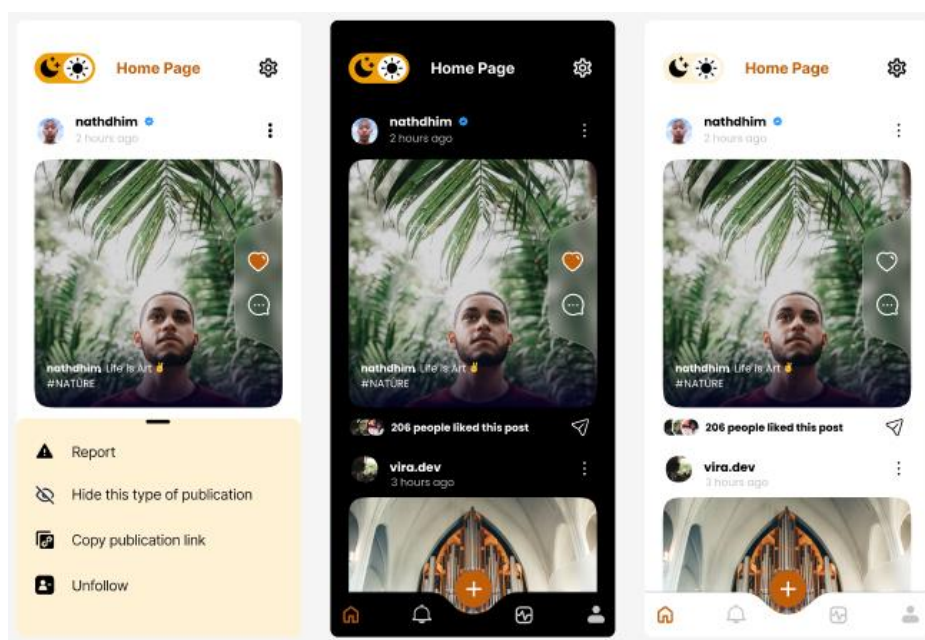


Figura 10: Ecrãs da página principal do protótipo desenvolvido.

Primeiramente, na parte superior do ecrã (Figura 11) é possível identificar 3 ícones que permitem ao utilizador várias funcionalidades do protótipo, sendo estes o de alterar entre o modo noturno (lua) ou diurno (sol) do *background* e o ícone que permite entrar nas configurações, designado por uma ferramenta. É importante também salientar a presença do texto “Home Page” bem centrado na página com uma cor de laranja viva, que permite ao utilizador de uma forma rápida, saber em que parte do protótipo se encontra.

O primeiro ícone, com a forma de uma lua, é o ícone que permite ao utilizador de tornar o protótipo escuro, transformando o que se encontra da cor branca na página principal em preta (Figura 11). O segundo ícone, com a forma de um sol, é o modo predefinido do protótipo, mantendo as cores brancas do fundo da página principal do protótipo. O último ícone representado por uma ferramenta, é o que permite ao utilizador, rapidamente, ir as definições em caso de necessidade.

O foco principal do ecrã principal “Home Page” são as variadas publicações dos utilizadores no protótipo. As publicações apresentam-se ordenadamente de forma temporal sendo possível deslizar (*scroll down/up*) para visualizar as diversas publicações diferentes no protótipo. Estas encontram-se com uma estrutura de forma a facilitar o uso de acesso aos utilizadores e a enaltecer o uso das várias funcionalidades que são possíveis de realizar com as publicações (Figura 11).

As estruturas das publicações encontram-se em retângulos com bordas arredondadas de forma a manter o design moderno e limpo do protótipo. Encontram-se dois botões de forma também subtil e moderna na parte central direita das publicações que permitem aos utilizadores a funcionalidade tanto de gostar (*like*) ou de comentar na publicação, a posição destes dois ícones relembra também aos utilizadores destas funcionalidades e gera de forma inconsistente um maior compromisso (Figura 11).

O nome do utilizador que partilhou a publicação encontra-se em antecedente da publicação de forma a ser possível identificar quem fez a publicação, assim como anteriormente referido, uma caixa de texto que indica há quanto tempo a publicação foi feita. Se o utilizador que fez a publicação for um utilizador “verificado” pela aplicação, este terá também um ícone que o indicará através de um ícone que se define por uma nuvem azul com um certo branco por dentro da nuvem (Figura 11).

Para finalizar, é visível no retângulo da publicação na parte inferior esquerda o nome de utilizador com o texto a negrito de forma a realçar o mesmo, seguido do título da publicação e a arroba no qual o utilizador da publicação é identificado. Já na parte inferior da publicação, fora do quadrado que define a publicação, o número de utilizadores que usou a funcionalidade de dar gosto na publicação, assim como um ícone que permite partilhar a publicação (Figura 11).

As configurações são divididas em três secções, sendo estas, as de configurações de conta (*Account*), as de suporte e sobre (*Support & About*) e as de ações (*Actions*), tendo estas três todas objetivos e funcionalidades diferentes para os utilizadores.

As configurações associadas à conta (*Account*) são as configurações do perfil do utilizador, a segurança (*Security*), as notificações (*Notifications*) e as configurações da privacidade (*Privacy*).

O suporte e sobre (*Support & About*) aglomera as configurações das subscrições (*My Subscription*), ajuda e suporte (*Help & Support*), e os termos e políticas (*Terms and Policies*).

Para finalizar, as configurações das ações (*Actions*), sendo possível através destas reportar um problema aos criadores e programadores do protótipo (*Report a Problem*), adicionar uma conta ao protótipo (*Add account*) e a opção de terminar a sessão (*Log Out*), como podemos observar na Figura 12.

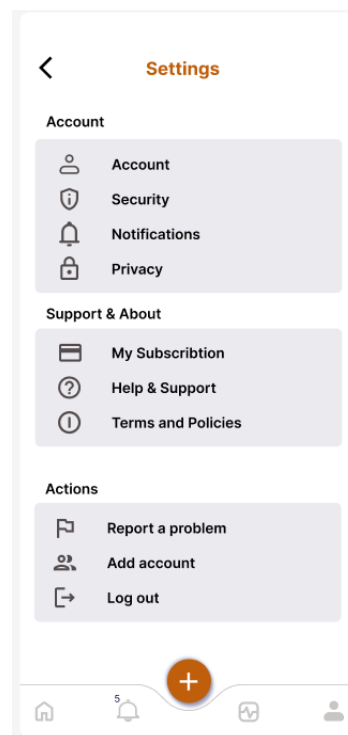


Figura 11: Ecrã das definições do protótipo desenvolvido.

O menu das notificações (*Notifications*), permite aos utilizadores verificar todas as notificações referentes a atividades que ocorrem ou ocorreram no protótipo da Editable-Gen. As notificações avisam um utilizador caso aconteçam determinadas ações tais como gostos em publicações, seguidores novos e comentários de pessoas que os utilizadores seguem.

A caixa de texto das notificações, dependendo da ação que tenha ocorrido, apresenta uma cor laranja viva, como forma de salientar a importância do evento. É importante indicar também

que notificações que tenham ocorrido recentemente e que o utilizador não tenha verificado, apresentam um pequeno círculo azul ao lado da foto de perfil como forma de manter o utilizador informado temporalmente sobre os eventos mais recentes.

O ícone de notificações apresentado na parte inferior esquerda do ecrã, no menu de navegação da aplicação, indica possíveis notificações que o utilizador ainda não tenha verificado, variando conforme o número destas, como podemos ver na Figura 13.

Existe também um ícone com três pontos no canto superior em cima do lado direito do ecrã, para o utilizador ordenar as notificações recebidas, como as lidas, não lidas e todas. Pode também clicar no botão “*Mark all as read*” a fim de marcar todas as notificações como lidas.

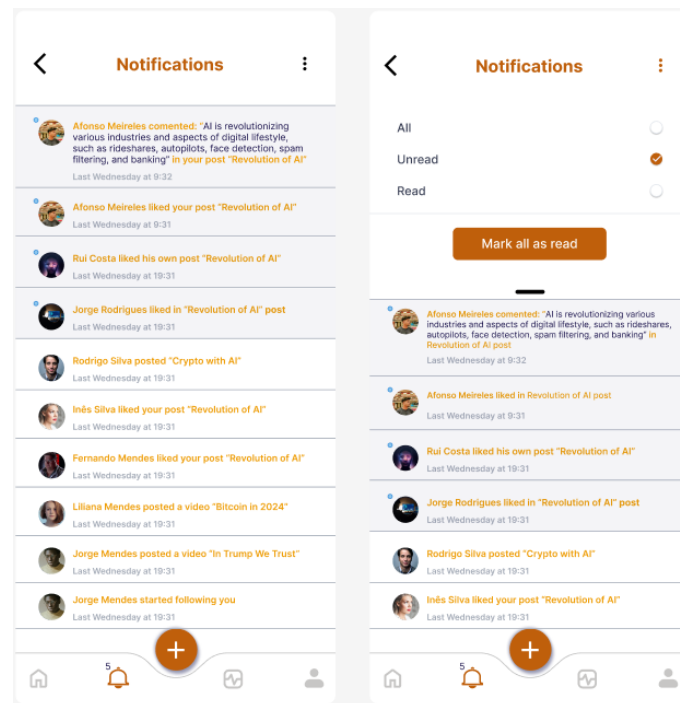


Figura 12: Ecrã das notificações do protótipo desenvolvido.

A galeria do protótipo (*Project Gallery*), permite aos utilizadores uma série de funcionalidades. A funcionalidade principal é a gestão e visualização das fotos e vídeos que o utilizador tenha guardado na galeria após utilizar o editor ou feito o *upload* do dispositivo móvel para a aplicação.

É possível no protótipo, visualizar e ordenar o conteúdo da galeria de três formas diferentes, sendo estas as “recents” (recentes), as de “cloud” (nuvem *online* para guardar conteúdo) e as “favorites” (as favoritas), como podemos ver na Figura 14.

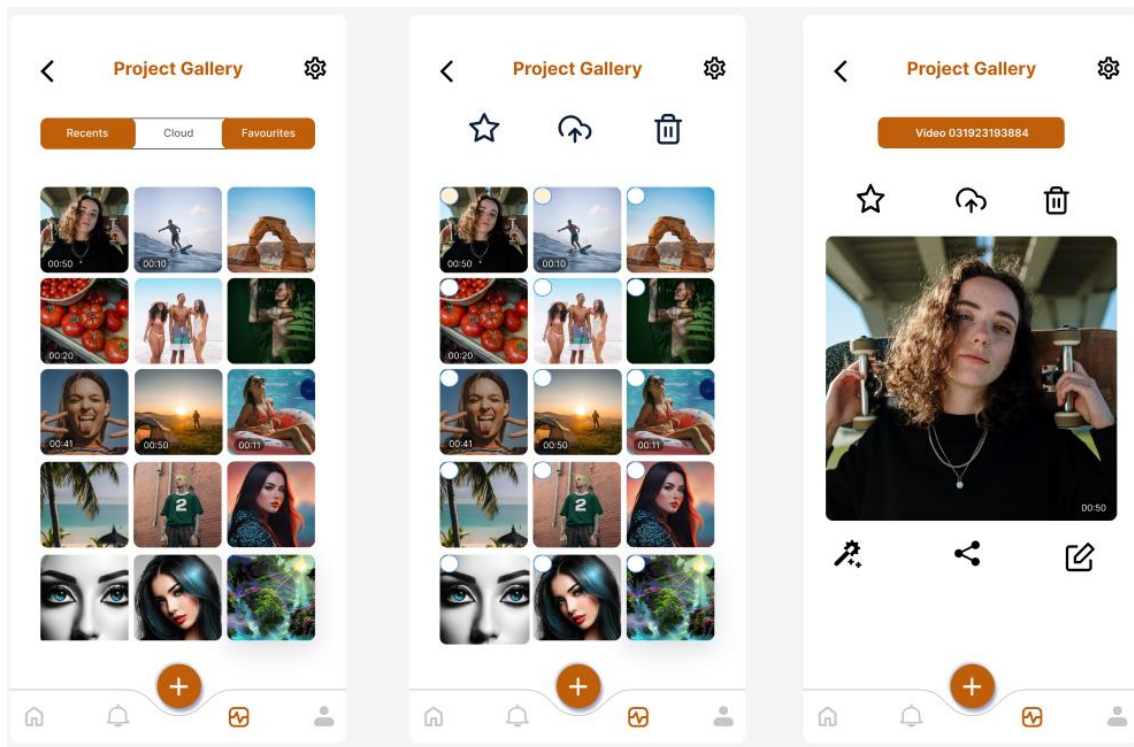


Figura 13: Ecrãs da galeria do protótipo desenvolvido.

Através destas funcionalidades, que quando premidas apresentam uma bolinha de fundo branca (seleccionar - passa a círculo com um visto azul se escolhido), sendo possível o utilizador ordenar o conteúdo que lhe é apresentado de várias formas diferentes, ou até apagar o conteúdo seleccionado no ícone representativo do lixo. Se o utilizador optar por clicar somente numa imagem em específico, irá abrir o último ecrã na Figura 14, com as opções de meter o conteúdo como favorito, ou dar *upload* para a nuvem caso ainda não estiver lá, ou apagar permanentemente, ou melhorar o conteúdo automaticamente em termos de cores, contraste, iluminação e nitidez com o ícone da varinha mágica, pode também partilhar em outras redes sociais, ou no ícone mais à direita em baixo da fotografia, abrir a mesma no editor da aplicação.

4.5 Editor do protótipo

Passando á parte mais importante deste protótipo, o editor de imagens automático (representativo com um “+” no centro em baixo) com o método *EditGAN* inserido. Contém

também todas as funcionalidades que estão apresentadas no site da NVIDIA (<https://research.nvidia.com/labs/toronto-ai/editGAN/>) na versão experimental.

A primeira tela, é destinada ao utilizador escolher a imagem que gostaria de editar. Na parte limite superior do ecrã encontram-se de forma clara o nome do protótipo, designadamente Editable-Gen no lado esquerdo, e o ícone que representa o protótipo do lado direito, pois são nestas telas que o nome e o logótipo mais se identificam.

O ecrã de “Upload” de conteúdo digital apresenta ao utilizador a funcionalidade de introduzir no protótipo fotografias ou vídeos em determinados formatos, sendo estes especificados num texto da parte superior central do ecrã. As imagens têm como formatos JPG ou PNG.

No centro do ecrã encontra-se a funcionalidade principal deste ecrã em específico, sendo esta a de “Upload” ao conteúdo, representada em tamanho grande e com uma cor viva laranja de forma a ser imediatamente visível ao utilizador onde ela se encontra, como podemos observar na Figura 15.

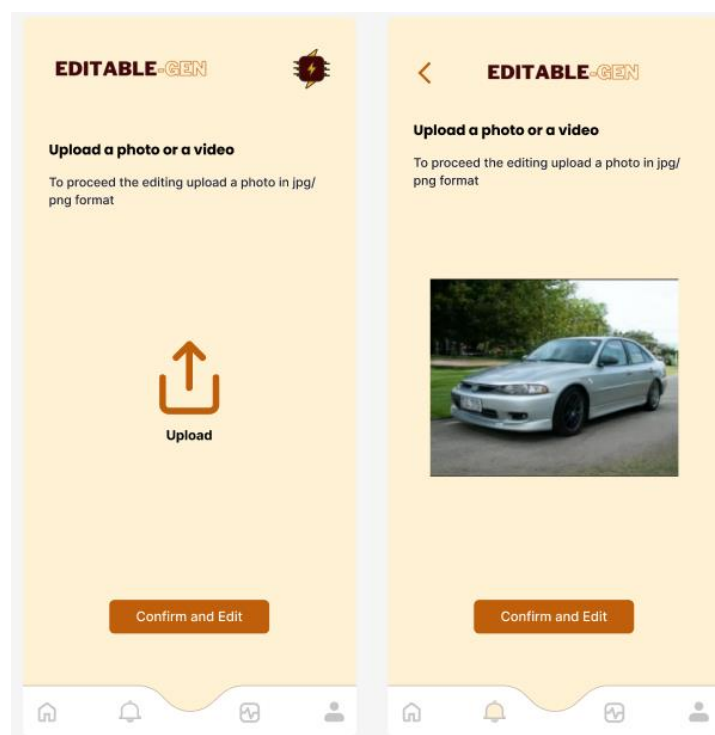


Figura 14: Ecrãs do início do editor Editable-Gen desenvolvido.

Finalizando, na parte inferior central do ecrã, encontra-se a funcionalidade de confirmar e editar o conteúdo, representando uma caixa do texto branco rodeado por uma cor laranja viva de forma a facilitar o uso do protótipo.

Após isso a imagem é enviada para o painel de edição, com diversas funcionalidades de edição da Editable-Gen, onde é o ponto fulcral do protótipo, sendo esta baseada num programa de edição de conteúdo da NVIDIA.

A funcionalidade de edição do protótipo permite aos utilizadores de realizar edição do seu conteúdo através da cor (“Color”), do tamanho (“Size”), da sua forma (“Shape”), e através de Vetores de edição (“Edit Vektor”). Estas opções encontram-se disponíveis na parte superior do ecrã, estando ordenadas da Cor até ao Vektor de Edição, como podemos observar na Figura 16.

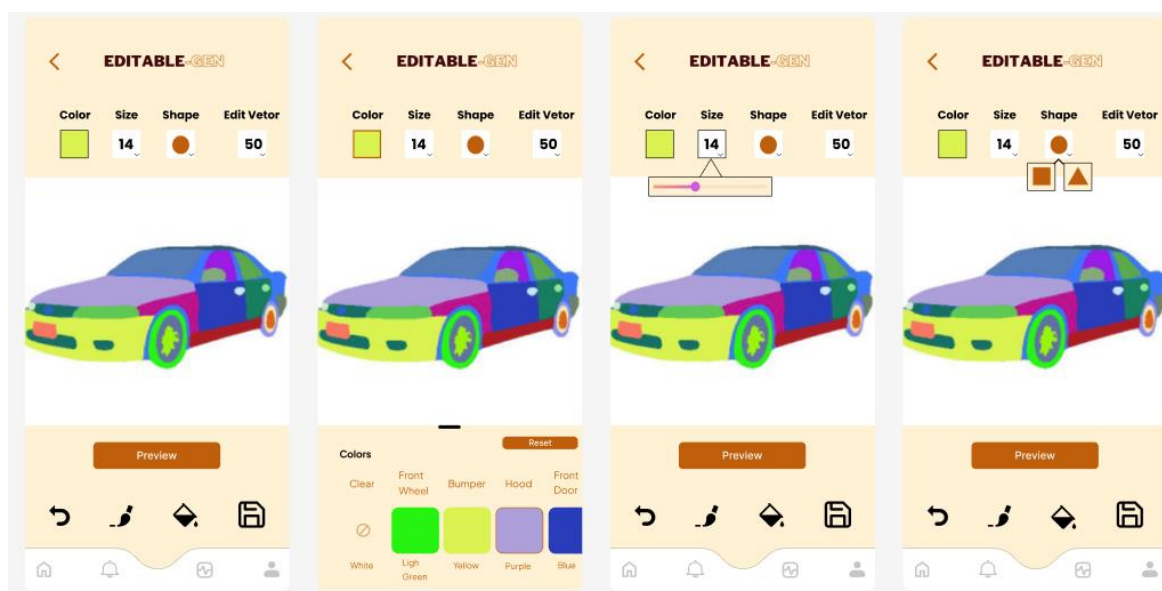


Figura 15: Funcionalidades de edição do editor Editable-Gen desenvolvido.

A cor (“Color”) é utilizada como forma de definir as diferentes componentes de determinado conteúdo para realizar a edição, sendo cada cor atribuída a uma parte diferente do conteúdo, como por exemplo num carro, as várias cores representariam as partes diferentes do carro. Também é possível visualizar através do menu da Cor, na parte superior direita, a funcionalidade de dar “Reset”, que permite aos utilizadores de recomeçarem o processo de edição.

O tamanho (“Size”) permite aos utilizadores definir o tamanho de determinada componente do conteúdo que está a ser editado. Por exemplo, num carro, escolher uma roda e aumentar ao gosto do utilizador o tamanho da roda. Ao utilizarem esta funcionalidade, deparam-se com uma

“barra horizontal”, que permite aos utilizadores “arrastar” a barra como forma de escolher aumentar ou diminuir o tamanho de determinada componente, sendo que estas componentes, como anteriormente foi definido na variável da Cor (“Color”), são definidas por cores.

A forma (“Shape”) permite aos utilizadores através de três formas diferentes, sendo estas um círculo, quadrado ou triângulo, de colorir partes do conteúdo que os utilizadores tenham como objetivo de editar. Existem três formas diferentes de forma a facilitar o uso da ferramenta de edição consoante as necessidades do utilizador, por exemplo, na roda de um veículo será mais útil utilizar a forma (“Shape”) de um círculo do que a de um quadrado ou triângulo, devido à forma da roda, enquanto para o capô de um carro faria mais sentido a forma quadrada.

O vetor de edição (“Edit Vetor”) permite aos utilizadores, através do uso de Vetores de edição, definir a quantidade de IA (Inteligência Artificial) que será usada em determinada componente do conteúdo que o utilizador tenha ou esteja a alterar, sendo que, quanto maior o valor do vetor de edição (“Edit Vetor”), maior quantidade de IA será usada por consequência. Esta funcionalidade é utilizada, assim como o tamanho (“Size”), através de uma “barra” que permite aos utilizadores de escolherem a quantidade pretendida, como mostra a Figura 17.

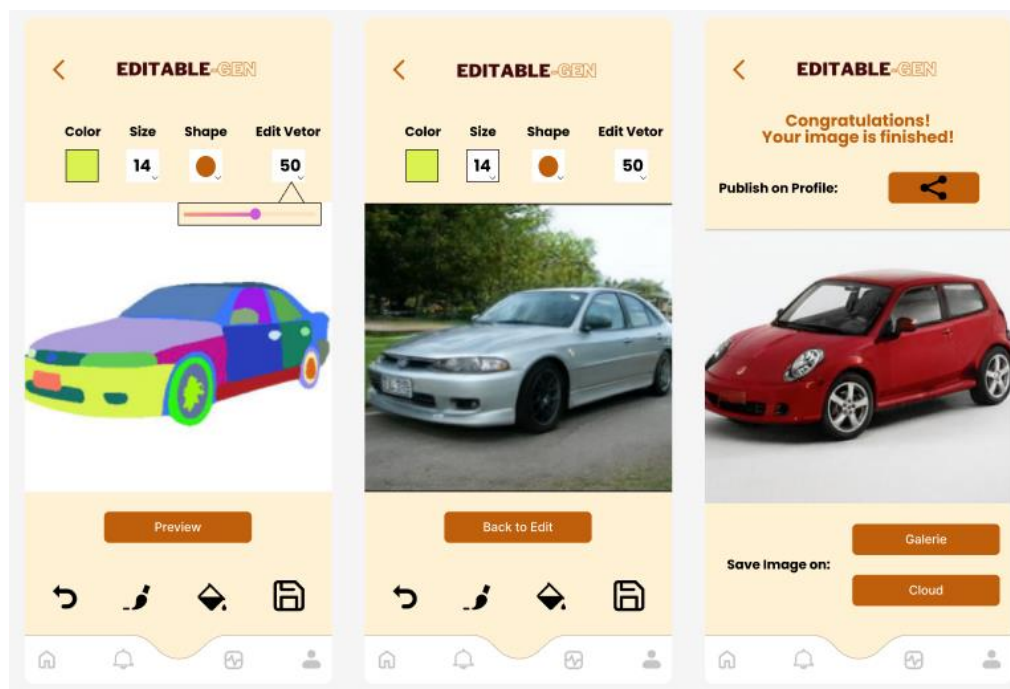


Figura 16: Ecrãs finais do editor Editable-Gen desenvolvido.

O ecrã de edição apresenta inclusivamente na parte inferior do ecrã, vários ícones que permitem aos utilizadores de utilizarem mais funcionalidades a nível editorial. Estas funcionalidades permitem aos utilizadores realizar uma visualização de como o produto está a ficar, através da funcionalidade “Preview”, a funcionalidade de retroceder, sendo que esta remete os utilizadores uma ação para trás, de forma a estes não perderem o trabalho de edição por completo caso tenha ocorrido um erro. Existem ainda as funcionalidades: o pincel, que permite aos utilizadores de escolher o pincel quando quiserem realizar uma pintura simples, também existe o balde de tinta que permite aos utilizadores colorir uma grande parte de um determinado espaço fechado com uma determinada cor (em relação à versão do computador foi modificado pela pipeta, que não fazia grande sentido no dispositivo móvel). Finalizando a funcionalidade de guardar, que permite aos utilizadores de guardar o seu conteúdo editado.

O ecrã da finalização da edição ocorre após os utilizadores terem dado como terminado o processo de edição do conteúdo e este encontra-se de forma estruturada e cuidada de forma aos utilizadores saberem o produto final ao colocar uma imagem no centro do ecrã das edições realizadas. É apresentada uma breve mensagem aos utilizadores a explicar que o conteúdo foi terminado na parte superior do ecrã: “Congratulations! Your image is finished!” seguido de um ícone que ao ser utilizado permite aos utilizadores partilhar o resultado através do protótipo da Editable-Gen ou de outras redes sociais, denominado por: “Publish on Profile”.

Finalizando, são fornecidas mais duas opções aos utilizadores na parte inferior do ecrã, sendo estas as de guardar o conteúdo ou na galeria ou na *cloud*, sendo esta opção denominada por “Save Image on: Galerie; Cloud”.

5. Testes de Usabilidade e Inquéritos

5.1 Questionários

Foram desenvolvidos 2 questionários completamente anónimos através do *Google Forms*, com o objetivo de obter diferentes perspetivas, experiências e opiniões de participantes com diversas idades, género e especialização de trabalho.

O primeiro questionário teve como principal objetivo conhecer o utilizador de forma a estruturar a arquitetura da aplicação com base da necessidade do mesmo. Por conseguinte apresentou questões sobre a idade, género, o grau de ensino, a especialização, o seu trabalho, se utiliza as redes sociais, se faz *upload* de imagens para as redes sociais, se faz *upload* de imagens somente para a *cloud* ou um espaço restrito, se seria interessante existir uma aplicação ou um plugin para esconder algum objeto de uma imagem, que redes sociais utiliza, qual o dispositivo que mais usa para a gravação de uma imagem/vídeo, se gostava de obstruir algum objeto de uma imagem, cenários em que o utilizador acharia útil uma aplicação deste tipo e em que dispositivo se via a utilizar esta aplicação.

No total foram obtidas 21 respostas, no qual mais de 70% eram de estudantes de diversos cursos. Resultados estes podem ser encontrados neste link de *Excel online* (anexo 1):

<https://docs.google.com/spreadsheets/d/1º0ykLtq6CGTPghAWrP6S4oaJUEUrtiAQRuqDbZbO7og/edit?usp=sharing>

Neste primeiro questionário foram obtidas respostas de 10 pessoas com o grau de ensino de licenciatura, e idades entre os 22 e os 27 anos, 4 indivíduos com um mestrado, de idades entre 23 e 53 anos e 2 pessoas com um curso superior técnico, com idades entre os 24 e 25 anos. As especializações destes candidatos são bastante diversas, tendo como área de especialização mais comum, a de informática e engenharias. Relativamente ao trabalho de cada candidato é muito variado o que torna este questionário bom a fim de obter diferentes perspetivas a nível de empregabilidade, sendo que 12 pessoas estão empregadas e 9 ainda a completar os estudos. Todos eles utilizam as redes sociais, 4 não costumam fazer *upload* de conteúdo para as redes sociais, e 3 também não fazem upload para um espaço mais restrito como a *cloud*. Na seguinte questão

houve somente 4 respostas que não achariam interessante uma aplicação ou *plugin* para esconder alguma coisa/objeto de uma imagem ou vídeo. Foi também questionado que redes sociais utilizavam, sendo que todos eles utilizam o *instagram*, muitos também usam o *Twitter*, quem está empregado utiliza o *LinkedIn*, já o *Facebook* tem pouco uso. Todos utilizam o telemóvel para gravação de vídeo ou tirar fotografias e alguns uma câmara profissional (observação que levou à concessão de um protótipo para dispositivo móvel). Já na edição de conteúdo, 8 utilizam o telemóvel, mas as restantes utilizam o computador. Foi questionado aos candidatos, se sentiam a necessidade de obstruir algo do conteúdo que gravam, e a maior parte das respostas foi dada com um “sim”. Por conseguinte, gostavam de obstruir mais conteúdos pessoais e sensíveis (pessoas, crianças), depois conteúdo explícito, de seguida conteúdo de objetos (matrículas de carros e entre outros). Na questão “Indique cenários em que esta aplicação seria útil”, foram obtidas diferentes respostas, destacando-se entre elas: a censura de caras, objetos que identifiquem a localização do utilizador de forma a obter privacidade, publicidade, desfoque de molduras com fotografias pessoais em casa, esconder certos objetos sem recorrer a um *software* de edição. Na seguinte questão “Em que aparelho consideraria utilizar mais esta aplicação?” foi respondido maioritariamente o telemóvel. Finalizando alguns dos candidatos deram sugestões importantes, como ser uma aplicação *user friendly* e permitir que a aplicação seja ativada e desativada através de apenas um clique. Foram também obtidos 7 *emails* de diferentes candidatos com o objetivo de ajudar numa fase mais avançada.

Após a primeira fase do protótipo estar concluída, foi exercido um questionário final mais completo e dinâmico que o primeiro, explicando tudo sobre o modelo *EDITGAN* e a ideia para a concessão de uma aplicação utilizando o mesmo. Pois a participação da comunidade é fundamental para garantir a otimização da realização, implementação e avaliação das iniciativas resultantes (Vargas *et al.*, 2022), por isso, é essencial entender quem usa este quotidiano, tornando-os parte da criação. O âmbito deste questionário foi entender erros ou problemas do protótipo, perspetivas de como seria melhor o estilo de design e as suas cores, se seria interessante ter uma parte de partilha da edição de imagem e também uma parte social, se a aplicação funcionava para pessoas com alguma incapacidade, comentários acerca dos problemas gerais de um editor de imagem automático, questões de modo a obter respostas para solucionar parte destes problemas, se seria utilizavam uma editor de imagem com o modelo *EDITGAN*, se um *cloud* para guardar os projetos e não sobrecarregar o dispositivo era vantajoso e por fim sugestões finais. Sendo este o questionário mais importante para este estudo, foi enviado como e-mail dinâmico para todos os estudantes e docentes da Universidade do Porto, obtendo assim 44 respostas

maioritariamente de estudantes e alguns docentes. Resultados estes podem ser encontrados neste link de *Excel online* (anexo 2):

<https://docs.google.com/spreadsheets/d/1U7YWzls9Eeyw774YglOW7wQ4BYXRwEuzRvMUTcMPQm0/edit?usp=sharing>

Este segundo questionário teve candidatos de dois gêneros, 51 % masculino e feminino 48% com idades entre os 19 e os 74 anos, sendo a maioria jovens estudantes (88%) e os restantes trabalhadores (10%). As suas especializações e graduações são de diversas áreas distintas, contudo 97% utiliza redes sociais. Maior parte dos candidatos, só edita imagens às vezes, 30% não faz edições e 23% fazem edições constantes. Metade edita imagens no telemóvel, 36% dos candidatos edita no computador e os restantes 12% fazem edições nos dois dispositivos. Sendo que mais de metade deles fazem *upload* das edições para redes sociais. Foi questionado que tipo de *design* teria melhor impacto numa aplicação deste tipo, e 68% escolheu a opção “*Design* Simples e eficaz, agradável aos olhos do utilizador sem o sobrecarregar de informação” (observação que levou a concessão de um protótipo com este estilo de *design*). Na questão “Seria vantajoso existir um perfil social para cada utilizador partilhar a sua arte gerada?”, maior parte respondeu talvez e 34% responderam que sim (observação que resultou em deixar ficar uma parte de partilha social no protótipo). Foi questionado também se existia a possibilidade deste tipo de aplicação resultar para educar crianças e jovens e pessoas com algumas incapacidades, a resposta mais frequente foi o talvez para crianças e jovens e o sim para as pessoas incapacitadas. Uma questão bastante importante foi “Quais os maiores problemas de um editor de imagem por inteligência artificial?”, as respostas foram dadas como escolha múltipla, destacando-se entre elas, o resultado da imagem não ser o esperado, falta de controlo na edição feita pelo utilizador e *deepfakes*. Contudo foi consentido que se o utilizador tivesse mais controlo na edição de imagem poderia solucionar estes problemas. Também foi consentido existir uma *cloud* para guardar as imagens editadas e não sobrecarregar o dispositivo (93%). Salientando a pergunta “Considerando o método EDIT-GAN, um gerador de imagem editado ao pormenor, acha que iria ter grande impacto em relação aos outros editores de imagem por inteligência artificial já existentes?”, 63% responderam como sim. Finalizando, foi questionado se davam uso a este editor de imagem e se sim em que contexto, obtivemos mais de metade das respostas como sim no contexto pessoal e para entretenimento.

5.2 Testes de Usabilidade

Para serem realizados testes de usabilidade, é bastante importante a observação sistemática durante todo o processo que o participante começa até ao final das tarefas propostas na *interface* baseada em cenários representativos da aplicação, Turner *et al.*, 2006. De tal forma que os participantes foram gravados em vídeo ao realizarem este teste, para também conseguir analisar com calma o processo todo mais tarde, e entender onde estes teriam mais dificuldades e problemas na *interface*. Por conseguinte foi apresentado aos participantes um consentimento para a participação e gravação em vídeo do teste de usabilidade, a fim de visualizar os vídeos mais tarde e entender onde existiu mais concentração, rapidez, percepção e entretenimento, como mostra no Anexo 3.

De seguida, foi desenvolvido um guia de tarefas do teste de usabilidade para o participante perceber que tarefas teria de solucionar a fim de conhecer todo o protótipo e as suas funcionalidades mais relevantes (desde o *login* e registo de uma conta, até á tela final do editor automático com o modelo *EDITGAN*), como mostra no Anexo 4.

O processo do teste de usabilidade consistiu em duas fases, a primeira foi a interação dos participantes com o protótipo, e a segunda foi o preenchimento de um questionário de satisfação para medir a experiência do participante no protótipo testado. Por isso, após o teste de usabilidade, foi pedido aos participantes que preenchessem esse questionário de satisfação (UEQ), como mostra um exemplo no Anexo 5. Questionário retirado do site: <https://www.ueq-online.org/>.

Os resultados obtidos foram bastante positivos, realçando os pontos mais fortes de ser um teste de usabilidade compreensível, agradável, organizado, eficiente, de fácil aprendizagem e seguro. Estes resultados vão ser analisados mais á frente, pois têm bastante relevância para este estudo.

No final, obtivemos 6 respostas de participantes com idades entre os 18 e os 30 anos, de diferentes áreas de trabalho ou estudo. Este teste foi realizado na Faculdade de Engenharia da Universidade do Porto. A escolha dos participantes foi de forma aleatória a estudantes e dois docentes. No total foram questionadas 11 pessoas, sendo que houve mais indivíduos que tinham interesse na participação do teste de usabilidade (6), os outros queriam continuar com o seus estudos e trabalhos na sala de estudo.

Durante as sessões de teste de usabilidade, foi sugerido aos participantes que utilizassem o protocolo *think-aloud* (Charters & Elizabeth, 2003). Esta abordagem exigia que os participantes verbalizassem os seus pensamentos e ações enquanto realizavam as tarefas em causa. Sendo que todos concordaram em fazê-lo, obtivemos assim informações qualitativas sobre os seus processos

cognitivos e deu para compreender melhor a tomada de decisão do participante durante todo o processo de realização da tarefa.

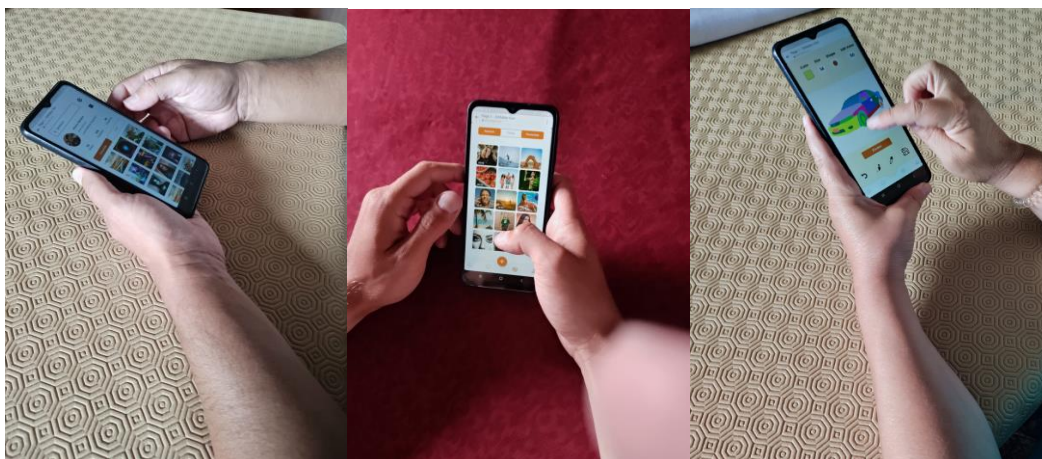


Figura 17: Imagens dos testes de usabilidade da sessão de gravação.

5.3 Análise de Dados

A fim de analisar os dados de investigação anteriores relativamente aos testes de usabilidade, foi utilizada uma tabela com o *feedback* dos participantes, o que facilitou a identificação de tendências, problemas e melhoramentos a fazer a nível de *design* e *interface*.

Realçando mais uma vez o protocolo *think-aloud*, obtivemos assim uma análise pessoal de cada candidato, por mera discussão verbal. O candidato número 1 (identificado na tabela 3), teve bastante rapidez em executar cada tarefa proposta, mas ficou um bocado aborrecido no ecrã de edição de imagem pois comentou que o seguinte editor não teria espaço suficiente para desenhar num ecrã de telemóvel e aconselhou o editor estar no formato horizontal a fim de existir mais espaço para a edição. O candidato número 2, comenta que o início do protótipo está bastante idêntico às redes sociais mais comuns como o *Instagram* e o *Twitter*, contudo finaliza as tarefas todas e sugere alterações na parte do editor de imagem, pois acha que deveria ter tantas funcionalidades como o programa da *Adobe Lightroom*, como por exemplo a distorção automática, ajustes de iluminação automática, máscaras de background automáticas e entre outras. O candidato número 3, gostou bastante da *interface*, das cores escolhidas, do *design* e do nome escolhido para a aplicação, sendo que criticou a tipografia, o tamanho de certos botões da galeria e o editor ser parecido com o apresentado no site da NVIDIA para computador. Já o candidato número 4, teve mais ou menos uma crítica idêntica, sugerindo uma alteração total no

editor de imagem. O candidato número 5, gostou bastante do protótipo no geral, apenas sugeriu baixar o tamanho dos títulos, a fim de obter mais espaço para o utilizador interagir. Finalizando, o candidato número 6, adorou o facto de ser um protótipo simples e eficaz, comentando também o facto de ser muito semelhante a outras redes sociais, e criticando a parte da galeria da *cloud*, pois não tinha muita interação. Sugere também existir mais espaço para o utilizador fazer o desenho da edição nos ecrãs finais.

Tabela 3: Feedback dos participantes no teste de usabilidade

Docente, 55 anos, Informática 1 • Protótipo simples e eficaz • Bastante parecido com o instagram na parte social • No editor de imagem, optar por meter o telemóvel na horizontal, aumentando o espaço para fazer o desenho da edição, botões mais pequenos, ocupar menos espaço nas funcionalidades para o desenho • Tema agradável	Estudante, 21 anos, Mecânica 2 • Parte do Login e Registo bastante simples, já no perfil podia ser mais criativo, pois parece muito com o instagram e twitter combinados. • O principal da aplicação percebe-se que é o editor de imagem, coisa que representa bem no botão do meio da tela. • Incluir mais funcionalidades na parte da edição era bem visto.	Estudante, 23 anos, Multimédia 3 • Boa escolha da paleta de cores da aplicação, nome bastante criativo e o ícone bem pensado para o que se trata. • Na galeria, os botões estão muito grandes para a tela • Na tipografia, podia ser utilizada uma mais futurista • A parte do editor, está semelhante ao editor no computador
Estudante, 24 anos, Multimédia 4 • Gostei da escolha de cores, bastante apelativo e agradável de ver • O design está muito simples, o ícone está bom e atractivo • Botões podiam ter outro formato, são muito simples e às vezes demasiado grandes. • Podia ser feito algumas alterações no editor de imagens, acho que não está destinado a ser usado no telemóvel.	Estudante, 21 anos, Bioengenharia 5 • Fácil de utilização, design apelativo • Títulos com letras grandes • A tela da Home Page, foi o que mais gostei de ver, bastante simples e eficaz. • Pouca carga de funcionalidades, o que torna a app agradável aos olhos • No geral está bastante bom	Estudante, 26 anos, Biomédica 6 • Aplicação muito agradável aos olhos, mas com pouca criatividade, muito semelhante a outras já feitas, a simplicidade é o seu forte, tive maior constrangimento a mexer com a tela da galeria, aqui devia ter mais funcionalidades. • O editor de imagem relacionado com a versão BETA no computador, está bem pensado, deixar mais espaço para desenhar não é mau pensado.

Os dados recolhidos do questionário UEQ forneceram informações significativas sobre a experiência do utilizador. Cada escala do questionário, oferece uma perspetiva distinta sobre a avaliação da aplicação.

A Tabela 4 apresentada em baixo, mostra alguns dos dados mais significativos do teste UEQ (criatividade, facilidade, segurança, motivação, organização, interesse, prático e compreensível) para o protótipo testado. Cada participante está inserido com o seu número da tabela anterior e designa-se por “P” (participante). Foi utilizada a escala como no questionário UEQ, de 1 a 7 (sendo 1 o ponto de avaliação mais baixo, 4 o ponto de avaliação intermédio e 7 o ponto de avaliação mais alto).

Tabela 4: Dados mais significativos do teste de satisfação do participante.

Participantes (P)	CRIATIVIDADE	FACILIDADE	SEGURANÇA	MOTIVANTE	ORGANIZAÇÃO	INTERESSANTE	PRÁTICO	COMPREENSÍVEL
P ₁	3	7	5	5	6	6	7	7
P ₂	5	7	5	5	5	5	6	6
P ₃	5	7	6	6	6	7	7	7
P ₄	5	6	4	5	5	6	6	7
P ₅	6	7	5	5	6	6	6	3
P ₆	4	7	5	4	6	5	6	7

Ao analisarmos os dados da Tabela 4, observamos que os participantes avaliaram positivamente a aplicação em termos de criatividade, facilidade, fiabilidade, motivante, organizado, interessante, prático e compreensível com valores médios de 6 pontos, numa escala de 1 a 7. Estes resultados indicam que a aplicação foi percebida como visualmente apelativa, clara, eficiente e fiável pelos participantes. Os valores mais baixos, indicam que existe alguma falta de criatividade, isto devido à aplicação ter semelhanças com outras. Já os valores mais altos, estão mais presentes na facilidade de utilização da aplicação e em ser compreensível (o quinto participante como não era fluente em inglês avaliou com uma nota mais baixa), também avaliaram a aplicação como bastante prática.

Finalizando este estudo e análise, procedemos ao melhoramento da qualidade da *interface*, assim como *wireframes*, *design*, botões, telas e entre outros pormenores, complementando também com *feedbacks* dos participantes.

6. Resultados

Este Capítulo apresenta os resultados que correspondem às questões de investigação colocadas anteriormente:

RQ1- Considerando o método *EDITGAN*, um editor de imagem ao pormenor, irá ter grande impacto em relação aos outros editores de imagem por inteligência artificial já existentes?

RQ2- Que estratégias podem ser utilizadas para diminuir problemas relacionados com o colapso de modos em GANs, e como é que essas estratégias têm influência na diversidade das amostras geradas?

Após a análise dos métodos e do trabalho de campo efetuado, que conduziu ao desenvolvimento de uma interface e avaliação da mesma, conseguimos dar resposta a estas questões de investigação.

Respondendo á pergunta RQ1, o *EDITGAN* diferencia-se bastante de qualquer outro método de edição de imagem automática, pois utiliza um dos métodos mais avançados para a síntese de imagem, chamado de *StyleGAN2*. Opera com imagens reais e tem uma grande precisão e otimização na imagem final. É uma ferramenta de fácil utilização e com bastante interatividade por parte do utilizador, pois tem a capacidade de modificar detalhes de uma imagem com qualidade e facilidade, coisa que maior parte dos outros editores não conseguem obter. Como analisado anteriormente, muitos editores de inteligência artificial focam-se somente em tarefas mais específicas, como remover o *background* ou substituir cores de uma imagem e o resultado nunca é o esperado. O *EDITGAN* ainda está em versão BETA, poderá ter alguns erros e não ser 100% funcional, mas a empresa da NVIDIA e os seus colaboradores afirmam que futuramente irá ser um editor com destaque. Temos de ter em consideração que este editor não funciona com todo o tipo de imagens, como, por exemplo, imagens muito complexas com vários objetos em questão não resulta, simplesmente destina-se a imagens simples, mas é bastante bom nisso, por

exemplo na edição facial, e em imagens com um só objeto em questão. Deste modo não se pode comparar a todos os editores de imagem automáticos. Existem também algumas limitações que nem todos os utilizadores estão confortáveis de trabalhar, como por exemplo compreender as variáveis do espaço latente e saber modificá-las até obter um resultado esperado. Finalizando, o *EDITGAN*, não substitui todos os editores de imagem automática, mas oferece uma boa ferramenta para quem está mais apático com a edição detalhada e deseja algo mais realista no seu resultado.

Em resposta à pergunta RQ2, os problemas relacionados com o colapso nas redes adversárias generativas, ocorre quando o gerador começa a produzir uma variedade de amostras limitadas e ignora as regiões do espaço de dados. Com isso resulta numa perda de informação e amostras, comprometendo assim a qualidade do modelo. Mesmo assim existem algumas formas de controlar esta perda. Como referido anteriormente, existe a função chamada de WGAN (Wasserstein GAN), que melhora a estabilidade de treino e reduz o colapso de alguns modos, pois a WGAN promove uma geração de amostras mais vasta, melhorando a sua diversidade. Também existe a opção das GANs condicionais, demonstradas anteriormente, que introduzem uma condição explícita no processo de geração, e com isso obriga o gerador a representar diferentes condições para diferentes partes do espaço de dados. Como existem diferentes condições o modelo é levado a gerar diferentes tipos de amostras. Outra técnica bastante comum, é impor limitações nos valores dos espectros da camada do discriminador, o que o leva a controlar a sua capacidade de representar funções mais complexas e reduz o risco de levar o discriminador ao colapso por excesso de dados. Por último, podem ser introduzidas técnicas de aumento de dados ao conjunto de treino, de modo a existir mais variabilidade na aprendizagem automática, basicamente fornece ao gerador mais exemplos, estimulando-o a introduzir uma gama mais elevada do espaço de dados. Temos de ter em consideração, que cada uma destas técnicas terá um impacto grande na promoção da diversidade de amostras geradas e no alívio do seu colapso, melhorando assim a utilidade e qualidade das redes adversárias generativas em aplicações práticas.

7. Conclusão e Oportunidades Futuras

Esta Dissertação centrou-se no desenvolvimento de um protótipo concebido para toda a gente que gosta de editar imagens de forma rápida e precisa, contendo o modelo *EDITGAN*, ainda em estado experimental. Ao examinar o estado da arte, da investigação junto dos utilizadores e da avaliação do protótipo proposto, obtivemos informações preciosas sobre as considerações e os requisitos de conceção desta *interface*. Nesta secção, aprofundamos as conclusões e propomos dicas para trabalhos futuros no domínio do desenvolvimento de uma *interface* proporcionada ao modelo *EDITGAN*.

7.1 Conclusões Finais

A síntese dos resultados da investigação sobre os utilizadores com base na observação, nas entrevistas e nas sessões de testes, combinada com os conhecimentos analisados de investigações anteriores, proporcionou uma base para a definição dos requisitos de conceção para uma *interface* com este modelo *EDITGAN*, assegurando assim que estava alinhado com as necessidades específicas da versão experimental. Tendo na mesma muitos pontos por melhorar, quando sair a fase final deste modelo (atualizações da *interface*, *design* e botões).

Wei Chen *et al.*, em 2020, reconhece que, o campo da visão computacional tem crescido nos últimos anos devido ao impulso das GAN, e várias aplicações surgiram como resultado. A maioria dessas aplicações trabalha com imagens. A pesquisa sobre processamento de vídeo, incluindo geração de vídeo, colorização de vídeo, pintura de vídeo, transferência de movimento e síntese de animação facial, foi limitada. Embora as GANs tenham sido utilizadas para a criação e síntese de modelos em 3D, como colorização em 3D, reconstrução de faces em 3D, animação em 3D e geração de objetos em três dimensões texturizados, os resultados não são totalmente satisfatórios.

As GANs ainda se baseiam em grandes dados de treino. No futuro, a redução da utilização de dados é uma tendência inevitável.

O *EDITGAN* está limitado a imagens que podem ser modeladas pelo GAN, como todos os métodos de edição de imagens baseados em GAN. Por exemplo, é difícil usar a aplicação do *EditGAN* para editar fotografias de cenas urbanas intensas. Embora a maioria das edições de alta precisão possa ser facilmente transferida para outras imagens utilizando vetores de edição aprendidos, também existem edições difíceis que exigem otimização iterativa em cada exemplo. Como resultado, a pesquisa futura visa acelerar a otimização dessas edições e desenvolver modelos generativos melhores com espaços latentes mais desempenhados, demonstrando uma ampla variedade de operações de edição em diferentes tipos de imagens e alcançando assim, um nível sem precedentes de flexibilidade e liberdade em termos de edição. Conservando ao mesmo tempo uma boa qualidade de imagem. (Huan Ling *et al.*, 2021).

A inclusão de uma parte social ao protótipo, é uma coisa adicional que os participantes dos questionários acharam interessante incluir para o lançamento deste tipo de aplicação, isto para chamar mais atenção da comunidade a ver grandes resultados a acontecer nas imagens editadas pelo *EDITGAN* e existir uma avaliação humana e não somente uma avaliação baseada em fatores de informação. Sendo também uma parte bastante interativa para os utilizadores conhecerem-se, divertirem-se e partilharem as suas edições.

7.2 Oportunidades Futuras

Num futuro próximo, é essencial avançar com o desenvolvimento do protótipo, melhorando o que aprendemos com os resultados dos testes do protótipo. A investigação futura deve explorar mais a fase final do modelo para testar a usabilidade, com diferentes participantes, para compreender melhor o impacto da *interface* desenvolvida. Com a implementação do protótipo numa solução funcional, estes novos testes devem ser executados com o protocolo *think-aloud* para avaliar totalmente o candidato. Ter em consideração também outras métricas de avaliação, como por exemplo a eficácia, que não foi avaliada nesta investigação.

É crucial considerar também a inclusão e a acessibilidade em trabalhos futuros, para as pessoas com incapacidades na futura conceção e desenvolvimento deste tipo de aplicação e garantir que estes indivíduos possam também beneficiar desta tecnologia, que pode ser fundamental para promover uma sociedade equitativa e justa.

Finalizando, é bastante importante atualizar a *interface* com base nos lançamentos da NVIDA e os seus colaboradores, de forma a não existir erros e falhas no algoritmo da edição.

8. Referências

Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. Generative Adversarial Networks An overview 2018, publish on IEEE Signal Processing Magazine.

<https://bit.ly/3uq881A>

Albers, Josef. "Interaction of Color." Kyung Dang (2013).

<https://aaeportal.com/publications/-22763/interaction-of-color--the-complete-digital-edition>

Abras, C., Maloney-Krichmar, D., Preece, J. (2004) User-Centered Design. In Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications.

<https://shorturl.at/52xo4>

Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks, 2015. <https://arxiv.org/abs/1511.06434>

Charters, Elizabeth. "The use of think-aloud methods in qualitative research an introduction to think-aloud methods." *Brock Education Journal* 12.2 (2003).

<https://journals.library.brocku.ca/brocked/index.php/home/article/view/38>

Collins, Edo, et al. "Editing in style: Uncovering the local semantics of gans." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

https://openaccess.thecvf.com/content_CVPR_2020/papers/Collins_Editing_in_Style_Uncovering_the_Local_Semantics_of_GANs_CVPR_2020_paper.pdf

Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. Maskgan: Towards diverse and interactive facial image manipulation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

<https://bit.ly/49oTiKV>

Chong, Min Jin, and David Forsyth. "Effectively unbiased fid and inception score and where to find them." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

https://openaccess.thecvf.com/content_CVPR_2020/papers/Chong_Effectively_Unbiased_FID_and_Inception_Score_and_Where_to_Find_CVPR_2020_paper.pdf

Daiqing Li, Junlin Yang, Karsten Kreis, Antonio Torralba, and Sanja Fidler. Semantic segmentation with generative models: Semi-supervised learning and strong out-of-domain generalization, 2021.

<https://bit.ly/3Uxf76P>

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

<https://arxiv.org/pdf/1412.6980>

Denton, Emily L. "Unsupervised learning of disentangled representations from video." Advances in neural information processing systems 30 (2017).

<https://proceedings.neurips.cc/paper/2017/file/2d2ca7eedf739ef4c3800713ec482e1a-Paper.pdf>

Deng, Jiankang, et al. "Arcface: Additive angular margin loss for deep face recognition." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

https://openaccess.thecvf.com/content_CVPR_2019/papers/Deng_ArcFace_Additive_Angular_Margin_Loss_for_Deep_Face_Recognition_CVPR_2019_paper.pdf

Emily L. Denton, vighnesh Birodkar. Unsupervised Learning of Disentangled Representations from Video, 2017.

<https://bit.ly/3w9551J>

Ehsani, Kiana, Roozbeh Mottaghi, and Ali Farhadi. "Segan: Segmenting and generating the invisible." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

https://openaccess.thecvf.com/content_cvpr_2018/papers/Ehsani_SeGAN_Segmenting_and_CVPR_2018_paper.pdf

Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel CohenOr. Encoding in style: a stylegan encoder for image-to-image translation, 2020.

<https://bit.ly/3HRBFYI>

Gatys, L.A., Ecker, A.S., & Bethge, M. (2015). "A Neural Algorithm of Artistic Style."

<https://arxiv.org/pdf/1508.06576>

Huan Ling, Karsten Kreis, Daiqing Li, Seung Wook Kim, Antonio Torralba, Sanja Fidler. EditGAN: High-Precision Semantic Image Editing 2021.

<https://bit.ly/48Ax6xc>

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron. Courville, and Yoshua Bengio. Generative adversarial networks. In Advances in neural information processing systems, 2014.

<https://bit.ly/48bvx8h>

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks, 2020. Communications of the ACM.

<https://dl.acm.org/doi/abs/10.1145/3422622>

Jie Gui, Senior Member, IEEE, Zhenan Sun, Senior Member, IEEE, Yonggang Wen, Fellow, IEEE, Dacheng Tao, Fellow, IEEE, and Jieping Ye, Fellow. A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications 2021.

<https://bit.ly/3SQFjbB>

Jun-Yan Zhu et al., "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," 2017 IEEE International Conference on Computer Vision (ICCV).

<https://arxiv.org/pdf/1703.10593>

Krug, Steve. Rocket surgery made easy: The do-it-yourself guide to finding and fixing usability problems. New Riders, 2009.

https://books.google.pt/books?hl=pt-PT&lr=&id=9Q3OQVyX_-QC&oi=fnd&pg=PR4&dq=Steve+Krug&ots=6JMR9NcXIV&sig=qV70fIwrtJJW0G-mP7OJYZuLGig&redir_esc=y#v=onepage&q=Steve%20Krug&f=false

Lei Wang; Wei Chen; Wenjia Yang; Fangming Bi; Fei Richard Yu. A State-of-the-Art Review on Image Synthesis with Generative Adversarial Networks 2020.

<https://bit.ly/41VX8Je>

Li, Muyang, et al. "Gan compression: Efficient architectures for interactive conditional gans." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

https://openaccess.thecvf.com/content_CVPR_2020/papers/Li_GAN_Compression_Efficient_Architectures_for_Interactive_Conditional_GANs_CVPR_2020_paper.pdf

Martin Arjovsky, Soumith Chintala, Léon Bottou Proceedings of the 34th International Conference on Machine Learning, 2017. Wasserstein Generative Adversarial Networks.
<https://proceedings.mlr.press/v70/arjovsky17a.html>

Mikołaj Binkowski, Danica J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. In International Conference on Learning Representations, 2018.
<https://par.nsf.gov/servlets/purl/10146725>

Nichol, Alex et al., “GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models.” International Conference on Machine Learning (2021).
<https://www.semanticscholar.org/reader/7002ae048e4b8c9133a55428441e8066070995cb>

Norman, Donald A. Design emocional: por que adoramos (ou detestamos) os objetos do dia-a-dia. Rocco, 2008.
<https://www.bds.unb.br/handle/123456789/761>

Park, Taesung, et al. "Gaugan: semantic image synthesis with spatially adaptive normalization." ACM SIGGRAPH 2019 Real-Time Live! 2019. 1-1.
<https://dl.acm.org/doi/abs/10.1145/3306305.3332370>

Pan, Zhaoqing, et al. "Recent progress on generative adversarial networks (GANs): A survey." IEEE access 7 (2019): 36322-36333.
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8667290>

Rameen Abdal, Yipeng Qin, and Peter Wonka. Image2stylegan: How to embed images into the stylegan latent space? In Proceedings of the IEEE International Conference on Computer Vision, 2019.
<https://bit.ly/4bxCfZ1>

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

<https://bit.ly/4bI2xrV>

Turner, Carl W., James R. Lewis, and Jakob Nielsen. "Determining usability test sample size." International encyclopedia of ergonomics and human factors 3.2 (2006): 3084-3088.

https://www.researchgate.net/profile/James-Lewis-8/publication/242156700_Determining_Usability_Test_Sample_Size/links/00b495355e7630c972000000/Determining-Usability-Test-Sample-Size.pdf

Tero Karras, Samuli Laine, Timo Aila; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019. A Style-Based Generator Architecture for Generative Adversarial Networks.

<https://bit.ly/496yZII>

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.

https://openaccess.thecvf.com/content_CVPR_2020/papers/Karras_Analyzing_and_Improving_the_Image_Quality_of_StyleGAN_CVPR_2020_paper.pdf

Vargas, Carmen, et al. "Co-creation, co-design and co-production for public health: a perspective on definitions and distinctions." Public Health Research & Practice 32.2 (2022).

<https://www.phrp.com.au/wp-content/uploads/2022/06/PHRP3222211.pdf>

Wang, Limin, et al. "Knowledge guided disambiguation for large-scale scene classification with multi-resolution CNNs." IEEE Transactions on Image Processing 26.4 (2017): 2055-2068.

<https://arxiv.org/pdf/1610.01119>

Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "SOD-MTGAN: Small object detection via multi-task generative adversarial network," in Proc. Eur. Conf. Comput. Vis., 2018.
https://openaccess.thecvf.com/content_ECCV_2018/papers/Yongqiang_Zhang_SOD-MTGAN_Small_Object_ECCV_2018_paper.pdf

Yuri Viazovetskyi, Vladimir Ivashkin, and Evgeny Kashin. Stylegan2 distillation for feed-forward image manipulation. In European Conference on Computer Vision, pages 170–186. Springer, 2020.
https://link.springer.com/chapter/10.1007/978-3-030-58542-6_11

Yujun Shen, Ceyuan Yang, Xiaoou Tang, and Bolei Zhou. Interfacegan: Interpreting the disentangled face representation learned by gans. TPAMI, 2020.
<https://bit.ly/3HP7tx4>

Yuxuan Zhang, Huan Ling, Jun Gao, Kangxue Yin, Jean-Francois Lafleche, Adela Barriuso, Antonio Torralba, and Sanja Fidler. Datasetgan: Efficient labeled data factory with minimal human effort, 2021.
<https://bit.ly/3Ux2zwz>

Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan and Y. Zheng, "Recent Progress on Generative Adversarial Networks (GANs): A Survey," in IEEE Access, 2019.
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8667290>

9. Anexos

Anexo 1: Questionário Inicial, 1ª ideia em estudo, obstrução de um objeto da imagem automático: “App de edição de imagem automática com IA”

Perguntas Respostas 20 Definições

App de edição de imagem automática com IA

B I U ↻ ✖

Para a seguinte proposta de dissertação de mestrado em multimédia na FEUP, optámos por desenvolver uma aplicação de câmara com a funcionalidade de editar um vídeo ou imagem, automaticamente após a sua gravação. A edição é feita por um processo de obstrução de objetos/cenários que o utilizador não quer que apareçam. Adaptando assim o vídeo com emojis ou desfoque na área designada.

Não existe nenhuma pós-produção feita pelo utilizador, este só vai ter que escolher de uma lista o quer obstruir no vídeo. Se for uma coisa pessoal ou que a câmara não identifique, o utilizador terá de tirar uma fotografia ao objeto de modo que a Inteligência Artificial faça um modelo novo para aquela coisa em específico e que depois possa ser adaptada no vídeo. Outra funcionalidade, é aplicar o mesmo processo a vídeos ou fotografias já desenvolvidos. Ou seja, o utilizador dá upload a um vídeo e escolhe da lista as coisas designadas que quer censurar. A principal vertente de existir AI nesta aplicação, é melhorar os modelos já treinados da lista de escolha e até adicionar outros novos. Evoluindo assim consoante o tempo e o número de utilizadores a participarem.

Exemplo 1: Estou a gravar um vídeo com o meu telemóvel e não quero que apareça a cara de um bebé. A aplicação vai gravar o vídeo e substituir a cara do bebé por este emoji 🍼.

Exemplo 2: Estou a tirar fotografias para anunciar a venda do meu carro e não quero a matrícula a aparecer. A aplicação vai desfocar a matrícula.

Exemplo 3: Estou a fazer um vídeo para o youtube e não quero que apareça a marca da minha bebida do starbucks. A aplicação vai desfocar só a marca.

Exemplos bastante distintos mas que demonstram a sua utilidade. **Nota: O questionário irá ser completamente anónimo, afim de fazer uma análise com base nas respostas dadas e no público-alvo.**

Que idade tem?

Texto de resposta curta

Qual o seu gênero?

- ☐ Masculino
- ☐ Feminino
- ☐ Outro

Qual o seu grau de ensino?

- ☐ Secundário
- ☐ Curso técnico superior e profissional - CTeSP
- ☐ Pós-graduação
- ☐ Licenciatura
- ☐ Mestrado
- ☐ Doutoramento

Qual a sua especialização?

Texto de resposta curta

Qual o seu trabalho?

Texto de resposta curta

Utiliza as redes sociais?

- ☐ Sim
- ☐ Não



Costuma fazer upload de vídeos/imagens para as redes sociais?

- ☐ Sim
- ☐ Não

Costuma fazer upload de vídeos/imagens para a cloud ou um espaço mais restrito?

- ☐ Sim
- ☐ Não

Seria interessante existir uma aplicação ou um plugin para esconder alguma coisa/objeto da fotografia ou vídeo que faz upload?

- ☐ Sim
- ☐ Não

Que redes sociais utiliza?

- ☐ Instagram
- ☐ Twitter (X)
- ☐ Tiktok
- ☐ LinkedIn
- ☐ Facebook
- ☐ Outra opção...

Para a gravação de um vídeo/imagem, qual o aparelho que utiliza?

- ☐ Telemóvel
- ☐ Computador
- ☐ Camara Profissional
- ☐ Outra opção...



Quando grava um vídeo ou tira uma fotografia pelo dispositivo, alguma vez sentiu a necessidade de obstruir alguma coisa/objeto em específico?

- ☐ Sim
- ☐ Não
- ☐ Talvez

Se sim, o quê?

- ☐ Conteúdo pessoal/sensíveis
- ☐ Crianças
- ☐ Pessoas
- ☐ Conteúdo Explícito
- ☐ Objetos
- ☐ Marcas
- ☐ Matrículas
- ☐ Outra opção...

...

Indique cenários em que esta aplicação seria útil

Texto de resposta longa

Em que aparelho consideraria utilizar mais esta aplicação?

- ☐ Computador
- ☐ Telemóvel
- ☐ Outra opção...

Tem alguma sugestão ou dica para a aplicação?

Texto de resposta longa



Anexo 2: Questionário Final, após finalização do protótipo – “A evolução da edição de imagem com a Inteligência Artificial, protótipo de uma aplicação com o modelo EDIT-GAN”.

Caro/a membro da comunidade académica,

No âmbito da Dissertação de Mestrado em Multimédia da Faculdade de Engenharia da Universidade do Porto, venho por este meio solicitar a sua participação no presente questionário: <https://forms.gle/NxsiEiENWQ2wMcDg5>



Inquérito: A evolução da edição de imagem com a Inteligência Artificial, protótipo de uma aplicação com o modelo EDIT-GAN

Este estudo insere-se no âmbito da dissertação de mestrado em multimédia na Faculdade de Engenharia da Universidade do Porto. Pretende-se desenvolver um protótipo de uma aplicação de edição de imagem com a funcionalidade da inteligência artificial, utilizando o método EDIT-GAN, ainda em estado beta e que foi lançado pela NVIDIA em 2021. O objetivo é considerar diferentes abordagens

forms.gle

Pretende-se com este trabalho fazer uma análise da evolução da edição de imagem automática e desenvolver a prototipagem de uma aplicação de edição de imagem automática, com o método lançado pela NVIDIA em 2021 chamado **EDIT-GAN** (ainda em estado BETA). O objetivo é considerar diferentes abordagens no protótipo para melhor se adequar à comunidade em geral e antecipar as linhas de guia para uma aplicação no futuro, a fim de trazer às pessoas uma edição simples e fácil de ser concessionada.

Qualquer pessoa pode responder a este questionário.

O questionário tem uma duração de cerca de 5 minutos e todos os dados recolhidos são totalmente confidenciais e utilizados apenas para fins de análise estatística.

Caso tenha alguma questão adicional, pode contactar o responsável pelo estudo, através do e-mail institucional: up202204421@edu.fe.up.pt

Muito obrigado desde já, pela sua disponibilidade!

Perguntas Respostas 44 Definições



+ Open Eyes + Close Mouth + Look Right

Inquérito: A evolução da edição de imagem com a Inteligência Artificial, protótipo de uma aplicação com o modelo EDIT-GAN

B I U ↺ ↻

Este estudo insere-se no âmbito da dissertação de mestrado em multimédia na Faculdade de Engenharia da Universidade do Porto.

Pretende-se desenvolver um protótipo de uma aplicação de edição de imagem com a funcionalidade da inteligência artificial, utilizando o método EDIT-GAN, ainda em estado beta e que foi lançado pela NVIDIA em 2021.

O objetivo é considerar diferentes abordagens no protótipo para melhor se adequar à comunidade em geral e antecipar as linhas de guia para uma possível aplicação no futuro, a fim de trazer às pessoas uma edição simples e fácil de ser concessionada.

Desta forma, peço a vossa colaboração no preenchimento deste questionário. O questionário é totalmente anónimo, tem uma duração inferior a 5 minutos e a sua participação é totalmente voluntária, podendo desistir a qualquer momento. Adicionalmente, as informações fornecidas serão confidenciais, utilizadas exclusivamente para este trabalho e conservadas até à defesa da dissertação de mestrado.

Caso tenha alguma questão adicional poderá contactar o responsável pelo estudo, através do seu email institucional: up202204421@up.pt

Qual o seu gênero?

- ☐ Feminino
- ☐ Masculino
- ☐ Outra opção...



Que idade tem?

Texto de resposta curta

Qual a sua situação profissional?

- ☐ Desempregado
- ☐ Estudante
- ☐ Trabalhador
- ☐ Trabalhador a tempo parcial
- ☐ Reformado

Qual a sua especialização e graduação?

Texto de resposta curta

Utiliza as redes sociais?

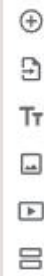
- ☐ Sim
- ☐ Não

Costuma editar fotografias/imagens?

- ☐ Sim
- ☐ Não
- ☐ Às vezes

Edita imagens mais vezes em que dispositivo?

- ☐ PC
- ☐ Telemóvel
- ☐ Os dois em igual
- ☐ Outra opção...



Costuma fazer upload das imagens editadas para as redes sociais?

- ☐ Sim
- ☐ Não

Figura 1, exemplo do método EDIT-GAN.

Este método de edição automática, aparenta ser bastante eficaz na edição de alta precisão e semântica de alta qualidade. Funcionando apenas em imagens simples com 1 objecto em questão, o utilizador tem a liberdade de editar cada detalhe da imagem, pois o EDIT-GAN trabalha com dados rotulados a fim de fornecer uma máscara de edição simples ao utilizador.

O utilizador edita a imagem original como se fosse um desenho, este remove ou adiciona as cores destinadas a cada detalhe. À medida que é feita uma edição nas cores, o utilizador pode ver a imagem final a ser alterada ao mesmo tempo.

Mais informações em:

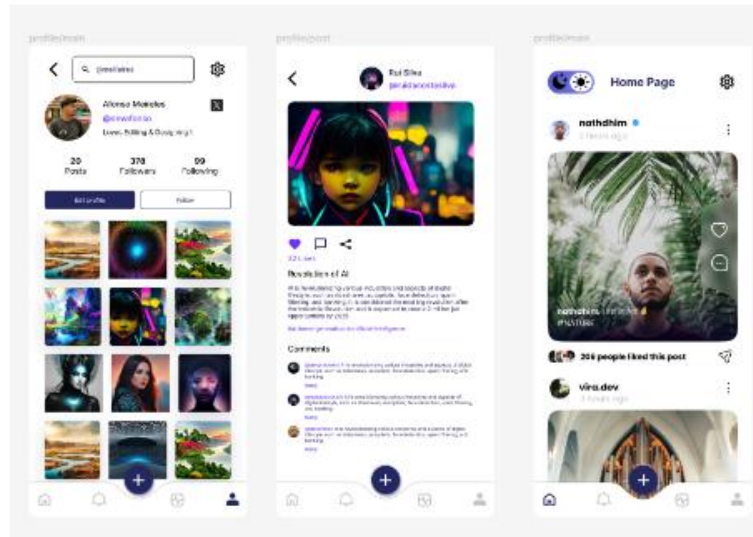
<https://research.nvidia.com/labs/toronto-ai/editGAN/>



Qual o design de interface que gostaria de ver numa aplicação para smartphone, utilizando este método EDIT-GAN?

- ☐ Design Simples e eficaz, agradável aos olhos do utilizador sem o sobrecarregar de informação

Seria vantajoso existir um perfil social para cada utilizador partilhar a sua arte gerada? (estilo twitter mas sem a parte social das mensagens)



- ☐ Sim
- ☐ Não
- ☐ Talvez
- ☐ Outra opção...

Uma aplicação com este tipo de edição, podia resultar bem para educar crianças/jovens?

- ☐ Sim
- ☐ Não
- ☐ Talvez

Uma aplicação com este tipo de edição, podia resultar bem para pessoas com alguma incapacidade?

- ☐ Sim
- ☐ Não

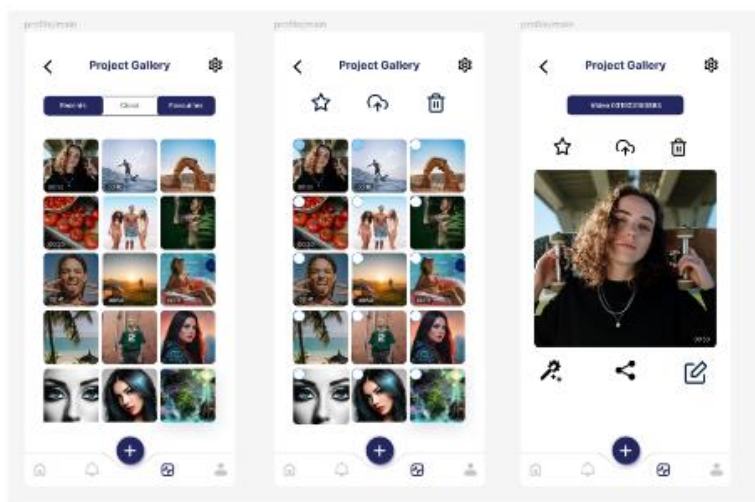
Quais os maiores problemas de um editor de imagem por inteligência artificial?

- ☐ Resultado final da imagem, não ser o espectável
- ☐ "DeepFakes" (distorção da realidade, com o objectivo de enganar o visualizador)
- ☐ Má qualidade e eficácia da imagem
- ☐ Falta de controlo na edição feita pelo utilizador
- ☐ Baixa velocidade e eficiência de imagens geradas
- ☐ Falta de complexidade nas tarefas de edição (remover objetos, adicionar objetos...)
- ☐ Outra opção...

Se o utilizador tivesse mais controlo na edição da imagem, podia solucionar estes problemas?

- ☐ Sim
- ☐ Alguns, mas não todos
- ☐ Não

Seria vantajoso existir uma "cloud online" para guardar as imagens geradas e não sobrecarregar o dispositivo?



Considerando o método EDIT-GAN, um gerador de imagem editado ao pormenor, acha que iria ter grande impacto em relação aos outros editores de imagem por inteligência artificial já existentes?

- ☐ Sim
- ☐ Não
- ☐ Outra opção...

Utilizava este editor de imagem?

- ☐ Sim
- ☐ Não

Se sim, em que contexto?

- ☐ Pessoal
- ☐ Para entretenimento
- ☐ Uso diário
- ☐ Outra opção...

Alguma sugestão?

Texto de resposta longa

Obrigado pelo tempo disponibilizado e continuação de um bom dia!



Anexo 3: Teste de usabilidade – Guia do Participante

Teste de Usabilidade

Editable-Gen 2024

O objetivo deste teste é avaliar a usabilidade da *interface* que foi desenvolvida de forma a garantir que seja intuitiva e de fácil utilização. Ao participar, está a ajudar-nos a identificar possíveis problemas e aperfeiçoar a experiência do usuário. Durante o teste, será solicitado a realizar uma série de tarefas na *interface*. Essas tarefas são projetadas para concluir uma edição de imagem de forma automática e partilhar o seu resultado. Vamos visualizar como irá interagir com a *interface*, compreender onde existe mais dificuldade e onde existe mais facilidade de compreensão pelo utilizador. Esteja à vontade para refletir em alta voz e partilhar as suas opiniões durante este teste. As suas respostas vão nos ajudar a entender o que está a funcionar direito e o que precisa ser alterado. Lembre-se: Este teste não avalia as suas habilidades, mas sim o design da interface. Não há respostas certas ou erradas. Queremos apenas saber como a interface se comporta e como podemos torná-la melhor.

1ª Tarefa: Login e Registo de uma conta

1.1: Faça login na aplicação ou registe-se na aplicação

1.2: Aceite os termos e condições

1.3: Guarde as informações pessoais do seu perfil

2ª Tarefa: Teste a aplicação na totalidade, menos a parte do gerador de imagem automático designado pelo botão principal do meio “+”.

2.1: Interaja com a tela do perfil, aceda as imagens, comentários e o botão de partilhar

2.2: Interaja com a tela seguinte da página inicial, designada pelo botão de uma casinha. Teste a alteração de plano de fundo para escuro e as definições de uma publicação.

2.3: Teste o botão das definições gerais da aplicação, inserido no canto superior em cima à direita. Partilhe a opinião sobre as definições gerais escolhidas.

2.4: Teste o botão das notificações designado por um “sino”, inserido no painel em baixo da barra de tarefas à esquerda. Após isso, teste o botão em cima canto superior, a fim de organizar as notificações.

2.5: Por último, interaja com o botão da galeria, inserido na barra de tarefas em baixo a direita e designado por uma “imagem”. Após isso clique na primeira fotografia e fale-nos da sua opinião, quanto as opções mostradas para a fotografia. Finalmente veja a organização das fotografias na seção “recentes”.

3ª Tarefa: Interaja com o botão principal da aplicação, designado por um “+”. Aqui irá observar um gerador de imagem automática, em que no início terá de fazer upload a uma imagem. Após isso tem diversas ferramentas de edição e desenho da imagem já escolhida a cores. Estas cores são atribuídas a diferentes pormenores do carro (roda, chassis, teto, jantes...), sendo que o objetivo é o utilizador desenhar/alterar diferentes cores de modo ao editor automático alterar na imagem original. Após isso grave a imagem no botão designado por uma “cassete” e diga-nos a sua opinião quanto a este gerador de imagem automático.

Anexo 4: Teste de usabilidade – Consentimento da gravação de vídeo

CONSENTIMENTO PARA PARTICIPAÇÃO NA INVESTIGAÇÃO “EDITABLE-GEN” SESSÃO DE TESTE DE USABILIDADE

O projeto “EDITABLE-GEN” pretende realizar um estudo de investigação para o melhoramento de um protótipo em desenvolvimento de uma interface para uma aplicação com o gerador de imagem automática e uma funcionalidade social e manejo de parâmetros relacionados com a facilidade e percepção do utilizador ao executar certas tarefas. Neste projeto participa a seguinte entidade: Faculdade de Engenharia da Universidade do Porto. No âmbito deste estudo, pretendemos realizar testes de usabilidade do protótipo proposto.

Esta atividade tem o objetivo de perceber:

- Melhorar a qualidade da interface identificando falhas e compreendendo o que funciona bem no design da interface;
- Entender a intuição nas tarefas propostas à realização na interface
- Compreender a satisfação e a experiência do usuário. As sessões serão individuais e privadas.

Materiais

As sessões serão realizadas na Faculdade de Engenharia da Universidade do Porto. Irá ser utilizado um telemóvel para gravar vídeo da realização das tarefas do teste de usabilidade, apenas apanhando as mãos dos participantes (pois é um teste anónimo) para depois ser analisado mais tarde pelo investigador Diogo Pereira com mais atenção. Esta atividade deverá decorrer num ambiente/espço privado, calmo e silencioso, de forma a garantirmos uma comunicação efetiva e foco absoluto.

Procedimentos

A sessão terá a duração estimada de 10 minutos, com uma fase de interação com o protótipo e outra fase de preenchimento de um questionário de avaliação.

Recolha e Análise de Dados

Os seus dados pessoais serão analisados pelo investigador e eliminados dois anos após o final do projeto. Os dados recolhidos são confidenciais e anónimos. Vão ser exercidas todas as medidas necessárias para a salvaguarda e proteção dos dados recolhidos, de forma a evitar que venham a ser acedidos por terceiros não autorizados.

A participação não envolve qualquer prejuízo ou dano material, nem qualquer benefício. Todo o material necessário para este estudo será fornecido pelo investigador. A sua participação não envolve qualquer tipo de pagamento, não terá consequências no seu trabalho ou avaliação, nem terá custos para si ou para a empresa para quem trabalha. A participação é voluntária, podendo em qualquer altura desistir sem nenhum tipo de consequência. Também poderá pedir a retificação ou destruição da informação recolhida a qualquer momento. Para isso, basta que nos contacte através do e-mail fornecido abaixo. Agradecemos muito o seu contributo, fundamental para a nossa investigação.

O participante:

Declaro ter compreendido este documento, bem como as informações verbais fornecidas e aceito participar nesta investigação. Permito a utilização dos dados que forneço de forma voluntária, confiando que apenas serão utilizados para investigação e com as garantias de confidencialidade e anonimato que me são dadas pelo investigador. Autorizo a comunicação de dados de forma anónima a outras entidades parceiras do estudo para fins académicos e de investigação científica.

Nome do participante:

Assinatura do participante:

Data: ____/____/____

Nome do investigador responsável pelo projeto: Diogo Filipe Pereira

Telefone: 913217691

E-mail: up202204421@edu.fe.up.pt

Assinatura do investigador responsável:

Anexo 5: Exemplo de um questionário feito relativo á experiência do utilizador no teste de usabilidade

Por favor dê-nos a sua opinião.

A fim de avaliar o produto, por favor preencha o seguinte questionário. É constituído por pares de opostos relativos às propriedades que o produto possa ter. As graduações entre os opostos são representadas por círculos. Ao marcar um dos círculos, você pode expressar sua opinião sobre um conceito.

Exemplo:

Atraente	☐	☒	☐	☐	☐	☐	☐	Feio
----------	-----------------------	----------------------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	------

Esta resposta significa que avalia o produto mais **atraente** do que **feio**.

Marque a sua resposta da forma mais espontânea possível. É importante que não pense demasiado na resposta porque a sua avaliação imediata é que é importante.

Por favor, assinale sempre uma resposta, mesmo que não tenha certezas sobre um par de termos ou que os termos não se enquadrem com o produto.

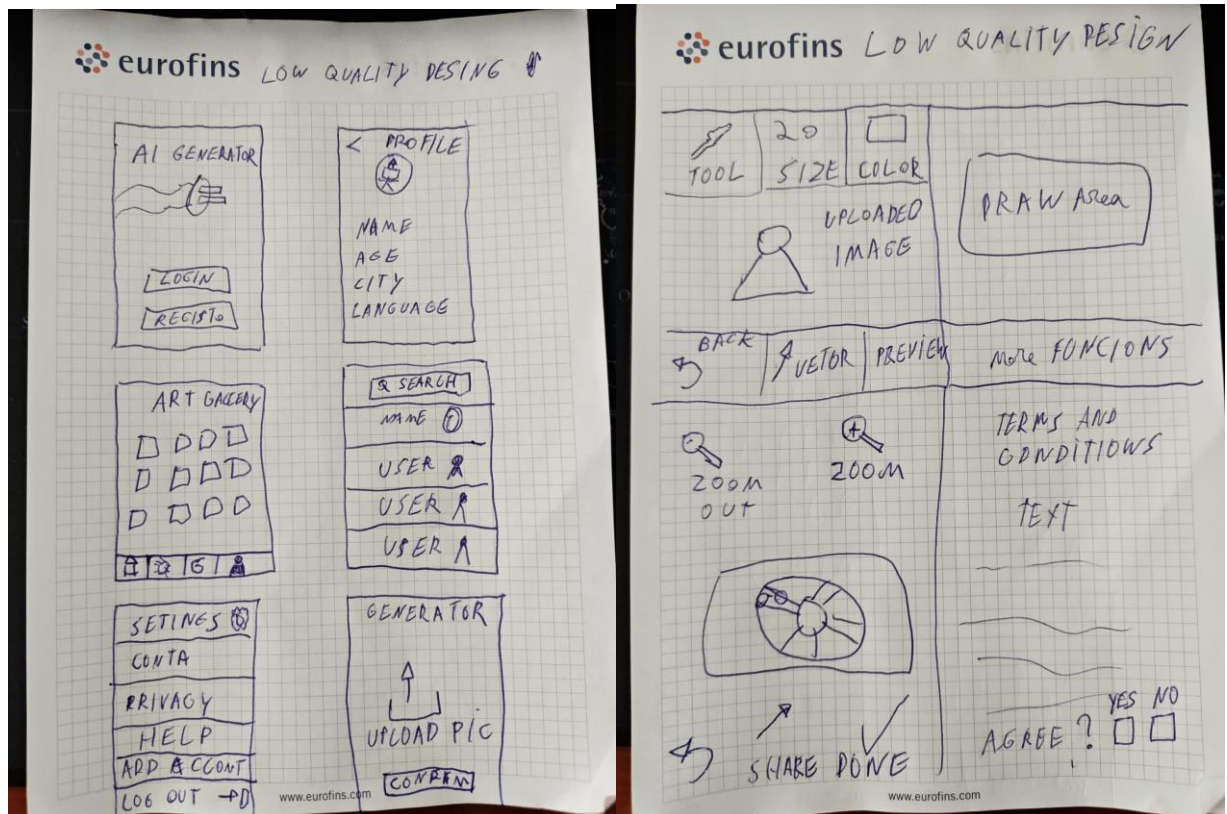
Não há respostas "certas" ou respostas "erradas". A sua opinião pessoal é que conta!

Por favor, dê-nos a sua avaliação atual do produto em causa.

Por favor, marque apenas um círculo por linha.

	1	2	3	4	5	6	7	
Desagradável	☐	☐	☐	☐	☐	☒	☐	Agradável
Incompreensível	☐	☐	☐	☐	☐	☒	☐	Compreensível
Criativo	☐	☐	☒	☐	☐	☐	☐	Sem criatividade
De Fácil aprendizagem	☐	☒	☐	☐	☐	☐	☐	De difícil aprendizagem
Valioso	☐	☐	☒	☐	☐	☐	☐	Sem valor
Aborrecido	☐	☐	☐	☒	☐	☐	☐	Excitante
Desinteressante	☐	☐	☐	☐	☒	☐	☐	Interessante
Imprevisível	☐	☐	☐	☐	☐	☒	☐	Previsível
Rápido	☐	☐	☒	☐	☐	☐	☐	Lento
Original	☐	☐	☒	☐	☐	☐	☐	Convencional
Obstrutivo	☐	☐	☐	☒	☐	☐	☐	Condutor
Bom	☐	☒	☐	☐	☐	☐	☐	Mau
Complicado	☐	☐	☐	☐	☒	☐	☐	Fácil
Desinteressante	☐	☐	☐	☐	☒	☐	☐	Atrativo
Comum	☐	☐	☐	☐	☐	☐	☐	Vanguardista
Incómodo	☐	☐	☐	☐	☐	☒	☐	Cómodo
Seguro	☒	☐	☐	☐	☐	☐	☐	Inseguro
Motivante	☐	☐	☒	☐	☐	☐	☐	Desmotivante
Atende as expectativas	☐	☒	☐	☐	☐	☐	☐	Não atende as expectativas
Ineficiente	☐	☐	☐	☐	☐	☒	☐	Eficiente
Evidente	☐	☒	☐	☐	☐	☐	☐	Confuso
Impraticável	☐	☐	☐	☐	☐	☒	☐	Prático
Organizado	☐	☒	☐	☐	☐	☐	☐	Desorganizado
Atraente	☐	☐	☐	☒	☐	☐	☐	Feio
Simpático	☐	☐	☐	☒	☐	☐	☐	Antipático
Conservador	☐	☐	☐	☐	☒	☐	☐	Inovador

Anexo 6: Esboço do protótipo de baixa qualidade



Anexo 7: Mockups alta qualidade do Figma, design e interface

