Politécnico
Castelo Branco
Escola Superior
de Tecnologia

BordadoBOT

Projeto I

João Rafael de Matos Pires

Nº 20220424

Gabriel Horta Charro

Nº 20220414

Orientadores

Arlindo Ferreira da Silva

Ana Paula Neves Ferreira da Silva

Trabalho de Projeto apresentado à Escola Superior de Tecnologia do Instituto Politécnico de Castelo Branco para cumprimento dos requisitos necessários à obtenção do grau de Licenciatura em Engenharia Informática, realizada sob a orientação científica do Professor Doutor Arlindo Ferreira da Silva e coorientarão científica da Professor Doutora Ana Paula Neves Ferreira da Silva, do Instituto Politécnico de Castelo Branco.

Fevereiro de 2025

Composição do júri

Presidente do júri

Doutor, Pedro Nuno Moreira da Silva

Professor Adjunto, Escola Superior de Tecnologia

Vogais

Doutor, Arlindo Ferreira da Silva

Professor Adjunto, Escola Superior de Tecnologia

Doutor, Fernando Reinaldo da Silva Garcia Ribeiro

Professor Adjunto, Escola Superior de Tecnologia

Resumo

O setor das artes e do design, especialmente no contexto do bordado, possui uma rica tradição que reflete estilos e técnicas específicas, como o bordado de Castelo Branco. No entanto, com o avanço da tecnologia e da Inteligência Artificial (IA), novas formas de expressão artística têm surgido, possibilitando a reinterpretação inovadora de estilos tradicionais.

Este projeto busca resolver o desafio de adaptar imagens contemporâneas ao estilo do bordado de Castelo Branco por meio de técnicas de IA, explorando a interseção entre arte tradicional e tecnologia. Para isso, foram utilizados modelos de Stable Diffusion (SD), que permitem gerar e modificar imagens de forma a refletir as características visuais específicas deste bordado.

Para atingir o nosso objetivo, foi realizada uma revisão do estado da arte para compreender e aprofundar as melhores práticas na aplicação de IA para a transferência de estilos artísticos. Em seguida, foi construído um dataset composto por diversas imagens do bordado de Castelo Branco, servindo de base para o treino do modelo. Com o modelo já treinado, foram conduzidos testes iniciais com um pipeline completo que permite a aplicação do modelo de SD para transformar a imagem fornecida pelo utilizador no estilo do bordado.

O Stable Diffusion é uma técnica de IA capaz de gerar imagens a partir de descrições textuais ou modificar imagens existentes para se alinharem a um estilo específico. Neste projeto, o foco é adaptar qualquer imagem para refletir as características visuais do bordado de Castelo Branco, destacando os seus padrões florais e cores vibrantes.

Em síntese, este projeto não só preserva e celebra a rica herança cultural do bordado de Castelo Branco, como também inova ao integrar inteligência artificial, possibilitando a aplicação desse estilo tradicional a novas imagens e contextos.

Palavras-chave

Inteligência Artificial

Deep Learning

Transferência de Estilo

Modelos de Difusão

Arte Generativa

Abstract

The arts and design sector, especially in the context of embroidery, has a rich tradition that reflects specific styles and techniques, such as the Castelo Branco embroidery. However, with advancements in technology and Artificial Intelligence (AI), new forms of artistic expression have emerged, enabling the innovative reinterpretation of traditional styles.

This project aims to address the challenge of adapting contemporary images to the style of Castelo Branco embroidery through AI techniques, exploring the intersection between traditional art and technology. To achieve this, Stable Diffusion (SD) models were used, allowing the generation and modification of images to reflect the specific visual characteristics of this embroidery.

To reach our objective, a state-of-the-art review was conducted to understand and deepen the best practices in applying AI for artistic style transfer. Subsequently, a dataset comprising various images of Castelo Branco embroidery was built, serving as the foundation for training the model. With the model already trained, initial tests were carried out using a complete pipeline that enables the application of the SD model to transform a user-provided image into the embroidery style.

Stable Diffusion is an AI technique capable of generating images from textual descriptions or modifying existing images to align with a specific style. In this project, the focus is to adapt any image to reflect the visual characteristics of Castelo Branco embroidery, highlighting its floral patterns and vibrant colors.

In summary, this project not only preserves and celebrates the rich cultural heritage of Castelo Branco embroidery but also innovates by integrating artificial intelligence, making it possible to apply this traditional style to new images and contexts.

Keywords

Artificial Intelligence

Deep Learning

Style Transfer

Diffusion models

Generative Art

Índice geral

Índice de figuras

Lista de tabelas

Lista de abreviaturas, siglas e acrónimos

AIGC (Artificial Intelligence Generated Content)

API (Application Programming Interface)

BO (Bayesian Optimization)

CNN (Convolutional Neural Network)

DCGAN (Deep Convolutional Generative Adversarial Network)

DL (Deep Learning)

FID (Fréchet Inception Distance)

GAN (Generative Adversarial Network)

IA (Inteligência Artificial)

IES (Interactive Evolutionary Systems)

KID (Kernel Inception Distance)

LoRA (Low-Rank Adaptation)

LDM (Latent Diffusion Model)

ML (Machine Learning)

NPR (Non-Photorealistic Rendering)

RNN (Recurrent Neural Network)

SEAN (Semantic Image Synthesis with Controllable Layered Feature Decomposition)

SD (Stable Diffusion)

SGDM (Style-Guided Diffusion Model)

StySim (Style Similarity)

T2I (Text-to-Image)

VAE (Variational Autoencoder)

1. Introdução

A transferência de estilos é uma área englobada dentro da visão computacional [1] que consiste em transportar características estilísticas de um domínio para outro mantendo os elementos do segundo [2]. Resulta assim numa fusão entre duas realidades, permitindo resultados onde a estrutura da imagem inicial fica enriquecida com os elementos artísticos do estilo de preferência.

Devido à natureza complexa dos estilos artísticos, esta tarefa é muito desafiadora, uma vez que têm de ser detetados inúmeros detalhes como cor, textura e até possíveis geometrias presentes no estilo. No entanto, o grande número de avanços dentro da área da Inteligência Artificial, incluindo recentemente os modelos de difusão, como o Stable Diffusion, têm alavancado um rápido e positivo avanço na forma como a transferência de estilos é realizada, permitindo ter resultados convincentes [3]

O uso de tecnologias como os modelos de difusão torna possível explorar a realidade da transferência de estilos complexos e densos, como o nosso caso de estudo, o Bordado de Castelo Branco. Para alcançarmos esse objetivo, iremos realizar um estudo do estado da arte deste tipo de aplicações, construir um dataset para o nosso domínio específico e treinar um novo modelo de transferência de estilo a partir dum modelo pré-existente recorrendo ao novo dataset.

1.1 Enquadramento

A utilização de Inteligência Artificial no mundo artístico tem sido um tópico bastante debatido [4]. A geração de conteúdos nunca vistos em segundos, quase indistinguíveis de criações humanas, levanta sérias questões sobre a criatividade, autoria e limites éticos destas tecnologias. O uso de ferramentas que abordam o uso de Stable Diffusion têm demonstrado um grande potencial para a criação de imagens com estilos complexos permitindo ao utilizador diversas personalizações numa imagem à sua escolha ou então a sua criação do zero.

1.2 Objetivos

No âmbito da unidade curricular de Projeto I, pretende-se analisar o sucesso de técnicas que dão uso a modelos de difusão, como é o caso de Stable Diffusion, para fazer uma transferência de estilo que permita dar as características estilísticas presentes no Bordado de Castelo Branco a uma outra imagem. Foram então delineados os principais objetivos para confirmar a viabilidade do uso desses modelos para a dada transferência de estilo:

- Fazer uma revisão sistemática do uso de modelos de difusão para transferência de estilo
- Criação de um dataset de referência que recolha o máximo de imagens possíveis do caso de estudo
- Identificar possíveis tecnologias que possam auxiliar o processo
- Criação de um modelo experimental, que, ao ser treinado com o dataset construído, seja capaz de gerar resultados promissores que lembrem aos elementos fundamentais do Bordado de Castelo Branco

1.3 Planeamento do Projeto

Com o objetivo de chegarmos ao nosso objetivo, este projeto foi dividido em 6 principais tarefas espalhadas pelo 1º semestre do ano letivo de 2024/2025.

- Tarefa A – **Elaboração do relatório:** A redação deste documento tem como objetivo documentar o trabalho e pesquisa desenvolvidos, estendendo-se assim por todo o semestre.
- Tarefa B – **Construção de um dataset:** Um dos principais objetivos deste projeto é a criação de um dataset de referência relacionado ao bordado de Castelo Branco.
- Tarefa C – **Revisão Sistemática:** Investigação da eficácia relacionada com o uso de modelos que utilizam Stable Diffusion para efetuar transferências de estilo.
- Tarefa D – **Procura por ferramentas:** Analisar quais as ferramentas mais comuns dentro da aplicação de modelos de difusão.
- Tarefa E – **Estudo das ferramentas:** Com base na análise anterior fazer um estudo sobre quais serão mais adequadas ao trabalho
- Tarefa F – **Resultados iniciais:** Através do uso do dataset construído e ferramentas mais promissoras, conseguir resultados iniciais de transferência de estilo

Tabela 1 - Cronograma de tarefas do projeto

Tarefa	2024								2025	
	Setembro		Outubro		Novembro		Dezembro		Janeiro	
	1ª Quinzena	2ª Quinzena	1ª Quinzena	2ª Quinzena	1ª Quinzena	2ª Quinzena	1ª Quinzena	2ª Quinzena	1ª Quinzena	2ª Quinzena
A										
B										
C										
D										
E										
F										

1.4 Estrutura do Relatório

Este relatório é constituído por 8 capítulos. Neste primeiro capítulo, o objetivo é identificar o problema que nos dispomos a resolver e assim como a sua possível solução. É também apresentada uma introdução ao trabalho, assim como o enquadramento e as principais tarefas delineadas numa fase inicial.

O segundo capítulo, tem como função apresentar as bases teóricas que foram necessárias para o desenvolvimento da nossa solução. Vão ser principalmente abordados temas relacionados com IA, sendo este o principal tópico do projeto. Dentro da área, vai ser dado ênfase a dois principais tópicos fundamentais para a compreensão dos modelos utilizados, nomeadamente Machine Learning e o uso de modelos de difusão. Estes irão desempenhar um papel chave na nossa abordagem ao problema.

No terceiro capítulo vai ser feita a análise do estado da arte, onde serão exploradas e analisadas as diversas soluções já existentes que dão uso a modelos de Stable Diffusion para a transferência de estilos. Neste capítulo vão ser descritas as fontes utilizadas para a pesquisa dos artigos relevantes, assim como os critérios para a inclusão dos mesmos no nosso estudo do estado da arte. Após a análise dos mesmos, vão ser tiradas as principais conclusões e feito um balanço dos resultados obtidos.

No quarto capítulo, é feita uma análise mais detalhada acerca da transferência de estilo. O seu principal objetivo é esclarecer o que realmente é a transferência de estilos, assim como analisar outras técnicas usadas para alcançar o objetivo de transportar os elementos estilísticos de umas imagens para outras. É também feita uma reflexão sobre a ética deste tipo de métodos, assim como dado um exemplo concreto de transferência de estilo com o uso de modelos de difusão.

No quinto capítulo, vamos fazer uma análise ao nosso caso de estudo, o bordado de Castelo Branco, para dar a conhecer um pouco da sua história e influência na região. Ainda neste capítulo vai ser abordado em mais detalhe o procedimento utilizado para a construção do nosso dataset. Iremos descrever a recolha das imagens e preparação das mesmas para poderem ser utilizadas com treino para o nosso modelo, bem como algumas dificuldades encontradas.

No sexto capítulo vamos listar todas as tecnologias/ferramentas que utilizámos para alcançar o nosso objetivo, apresentando a sua função na nossa solução.

No sétimo capítulo, é onde pretendemos apresentar os nossos resultados através da demonstração de algumas imagens geradas pelo nosso modelo após o treino. Esta demonstração vai incluir imagens geradas por diversas formas: através de image-to-image onde o modelo terá de efetuar a transferência de estilo à imagem fornecida, e também através de text-to-image para ser possível avaliar a capacidade do modelo em criar imagens que tenham os elementos do bordado de Castelo Branco. Após a apresentação das imagens vai ser feito um balanço dos resultados obtidos.

Já no oitavo capítulo, sendo este a conclusão, vamos fazer uma retrospetiva ao trabalho realizado apresentando as principais conclusões retiradas. Este capítulo encerra com possíveis sugestões para Projeto II para fazer melhorias ao modelo utilizado.

2. Enquadramento Teórico

O objetivo deste capítulo é aprofundar os conhecimentos de forma teórica, apresentando as tecnologias e metodologias relevantes para a transferência de estilo do Bordado de Castelo Branco para outra imagem. Além disso procura-se fundamentar as bases relacionadas com Inteligência Artificial, Machine Learning, Modelos de Difusão e o uso de Stable Diffusion que vão ser fundamentais para o entendimento do processo aplicado a cada imagem.

2.1 Inteligência Artificial

A inteligência artificial é uma área que engloba vários campos incluindo Machine Learning, Deep Learning, processamento de linguagem natural, robôs inteligentes entre outros. O seu principal objetivo é através de diversas técnicas e algoritmos, conseguir resolver problemas simples da vida humana como tomadas de decisão e resolução de problemas, baseando-se principalmente em conceitos como lógica, probabilidade e matemática. O problema reside no facto de não podemos fornecer a capacidade de raciocínio às máquinas. Assim, será necessário criar uma forma de simular ou copiar o funcionamento do cérebro humano [5]. Permitindo assim a estas tecnologias otimizar e automatizar tarefas.

Com o avanço das tecnologias, é possível, atualmente, ter acesso a múltiplas ferramentas capazes de fornecer suporte a diversas atividades humanas. O que nos interessa principalmente é capacidade de observação e aprendizagem de um estilo específico. Ou seja, neste contexto, não será apenas utilizada para a deteção dos elementos visuais extraídos de uma imagem, mas também para transportar esses elementos para outras imagens possibilitando a criação de novas realidades artísticas. Com isto, a IA sai do mundo analítico do qual foi inicialmente desenhada para uma nova realidade onde pode ser parte do mundo artístico.

Machine Learning surge quando permitimos à máquina aprender por si [6] através de dados fornecidos, de forma a conseguir realizar previsões e análises. A técnica mais explorada neste trabalho vai ser o uso o de Latent Diffusion Models que através de estudos recentes [7] demonstrou obter melhores resultados quando comparada com abordagens anteriores, este modelo vai utilizar técnicas tais como Deep Learning através de redes neurais profundas para transformar as imagens para num formato que as redes possam entender e interpretar. Estas imagens passarão para o Espaço Latente, onde ficarão extremamente comprimidas, mas mantendo os detalhes necessários para uma transferência de estilo eficaz, o que

diminui a quantidade de recursos computacionais necessários [8] , temas estes serão explorados mais a fundo nos próximos capítulos.

2.1.1 Definição e Conceitos

A inteligência artificial é atualmente uma das áreas mais estudadas no mundo da ciência da computação [9]. Este ramo permite que as máquinas se comportem de forma inteligente, permitindo-lhes lidar autonomamente com problemas cada vez mais diversos. Resumidamente o termo IA refere-se a sistemas ou máquinas que simulam a inteligência humana para realizar tarefas e podem melhorar iterativamente com base na informação que recebem.

Esta área não é apenas um tópico da ciência da computação, mas também filosófico, uma vez que, desde o início, se abordam questões sobre a possibilidade de dotar as máquinas da capacidade de pensar como seres humanos. Desde cedo foram definidos dois principais tipos de IA. O primeiro, mais simples e frequentemente referido como Weak AI (ou IA Fraca) tem como principal objetivo realizar uma função específica. Este tipo de soluções é capaz de resolver problemas específicos dentro de um escopo limitado. Normalmente é utilizado para tarefas simples que, por vezes, podem causar dificuldades aos humanos [10]. É possível encontrar a IA Fraca em tarefas como reconhecimento de padrões, análises médicas, e jogos de estratégia, mas não tem entendimento do que está para além daquele cenário, nem compreende perfeitamente a tarefa que executa, limita-se responder com base na informação com o qual foi treinada [11] .

Por outro lado, a Strong AI (ou IA Forte) consiste na possibilidade das máquinas terem a capacidade de raciocínio e compreensão equivalente à humana, sendo capazes de analisar uma série de parâmetros diferentes e tomar decisões em diversos contextos, sem depender de um conjunto de regras predefinidas ou de um domínio controlado. Esta não é uma área existente, mas sim uma possibilidade que tem sido debatida desde o início do estudo da IA [11]. Com o passar dos anos e devido aos enormes avanços desta área, a IA Fraca tornou-se altamente avançada, sendo possível encontrar ferramentas amplamente utilizadas nas mais diversas áreas do conhecimento, já estando presente em sistemas de reconhecimento de voz, traduções automáticas, carros autónomos e ferramentas de recomendação.

2.1.2 Evolução e Aplicações

A IA é uma das áreas mais promissoras na ciência da computação, tendo como principal foco o desenvolvimento de sistemas que realizam tarefas que normalmente exigem inteligência humana. Este sistema tem como objetivo aprender e melhorar iterativamente utilizando métodos que atualmente denominados de Machine Learning. Estes algoritmos têm como principal objetivo analisar grandes quantidades de dados e aprender com esses dados, através da deteção de padrões e extração de informações.

Alan Turing, um dos pioneiros no campo da IA, é amplamente conhecido pela criação do conceito do Teste de Turing. Neste teste, originado no seu famoso artigo seminal “Computing Machinery and Intelligence” (1950), é levantada uma questão bastante pertinente: “Serão as máquinas capazes de pensar?”. Nesse artigo, ele apresenta uma série de argumentos em que, em vez de se focar no conceito de “pensar”, seria mais eficaz criar um teste prático que respondesse à sua questão original. Toda esta linha de pensamento foi o que deu origem ao que hoje conhecemos como o Teste de Turing, no qual a premissa original é testar a inteligência da máquina através da análise da sua capacidade de enganar um interlocutor humano. Ou seja, se uma máquina conseguisse, de forma convincente, passar-se por humana, aos olhos de Alan Turing faria com que se tornasse inteligente [12] [13]. Apesar de inovador para altura, este teste demonstrou-se ser ineficiente e até injusto devido à natureza humana.

De seguida, já nos anos 80, foram feitos grandes avanços dentro da área do Machine Learning, devido à introdução de novos e mais complexos algoritmos, como o caso da retropropagação. Este novo método permitiu à IA aproximar-se do que é o cérebro humano, uma vez que usa uma arquitetura semelhante à dos nossos neurónios. Através da receção de dados de entrada, estes neurónios artificiais aplicam funções de ativação gerando assim valores de saída. Estas inovações dos anos 80 fazem parte hoje do núcleo do que entendemos por IA [14].

2.2 Machine Learning

A consequente evolução dentro da área da IA abriu espaço para a área de Machine Learing (ML) tornar-se uma das mais impactantes subáreas da IA, estando atualmente presente em diversas áreas de operação, como saúde e finanças, sendo principalmente utilizada pela sua capacidade de análise de dados [15]. O que capacitou este crescimento, foi a sua especialização na deteção de padrões em vez da aplicação de regras pré-estabelecidas, como ocorre em outras áreas

das ciências da computação, onde um agente aplica as regras num conjunto para obter respostas. Já em Machine Learning, esse processo varia um pouco, pois ao fornecer o nosso conjunto de dados e as respostas, o modelo irá nos devolver as regras, promovendo assim a análise dos dados, permitindo tirar conclusões muitas vezes escondidas entre os dados e de difícil análise para nós, seres humanos [16].

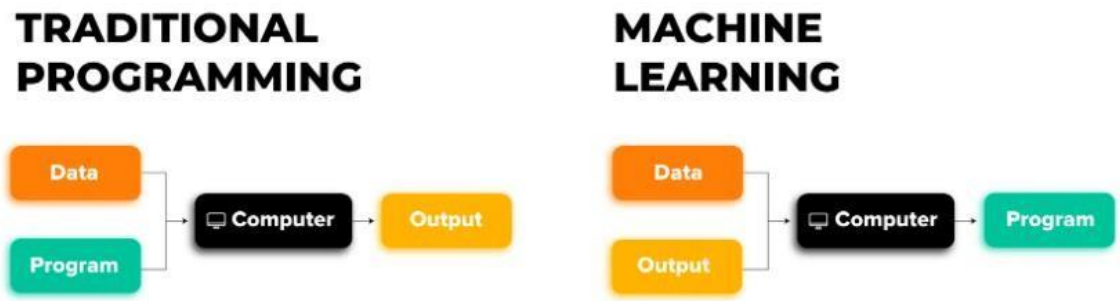

Figura 1 – Comparação entre programação tradicional e Machine Learning (Retirado de: [17])

2.2.1 Tipos de Aprendizagem

Para os modelos de ML conseguirem fazer esta análise dos dados, é fundamental que primeiro sejam treinados. O processo de treino pode ser feito utilizando diferentes estratégias, cada uma com as suas aplicações e pontos fortes específicos. Esses tipos diferentes podem ser classificados como:

2.2.1.1 Aprendizagem Supervisionada

Ao utilizar esta estratégia de treino, estamos a empregar um conjunto de dados que possuem uma característica alvo, onde todos os dados estão rotulados com um valor dessa dada característica. O que faz com que todas as entradas possuem uma saída, permitido, com base nas entradas recebidas, que o modelo faça uma previsão do atributo alvo[18].

2.2.1.2 Aprendizagem Não Supervisionada

Nesta abordagem, os dados já não possuem rótulos nem um atributo que possamos considerar como objetivo. Aqui, o trabalho do modelo é interpretar os dados na totalidade e tentar encontrar padrões ou regras de agrupamento, que permitam separar a totalidade dos dados em diferentes grupos com características semelhantes entre eles [18].

2.2.1.3 Aprendizagem Por Reforço

Na aprendizagem por reforço, o modelo vai aprender com base no feedback recebido das interações com ambiente. Diferentemente de outras abordagens, aqui não existem rótulos; em vez disso, o modelo vai explorar os dados e, conforme os resultados obtidos, vai modificar a sua estratégia e alterar ações de modo a maximizar as recompensas positivas[18].

2.2.2 Redes Neuronais

As redes neuronais (ou ANNs) são uma classe de algoritmos associados ao ML que têm como principal tarefa a deteção de padrões e aprendizagem com base numa grande quantidade de dados. A arquitetura destes modelos, tal como nos primeiros tempos de IA, vem da biologia humana, que, de forma muito simplificada, tenta simular o comportamento de um neurónio humano [19]. É possível verificar esta inspiração na forma como as camadas estão interligadas.

O funcionamento deste mecanismo envolve principalmente a organização de três componentes essenciais: a camada de entrada, uma (ou várias) camadas ocultas, e por fim a camada de saída. A camada de entrada tem como principal função receber os dados e, consequentemente, transmiti-los para a(s) camadas seguintes. Na(s) camada(s) oculta(s) serão realizadas uma série de operações matemática.

Estas operações variam entre somas ponderadas e a aplicação de várias funções de ativação, como ReLU (Rectified Linear Unit), Sigmoid ou Tahn. As funções de ativação são de grande importância, uma vez que o seu principal papel é capturar e interpretar as relações não lineares entre os diversos atributos dos dados, garantindo assim que consiga lidar facilmente com dados onde a variação é maior [20]. Após o processamento nas camadas ocultas, o resultado das operações é passado para a última camada, a camada de saída, que, com base nos valores recebidos, vai gerar um resultado.

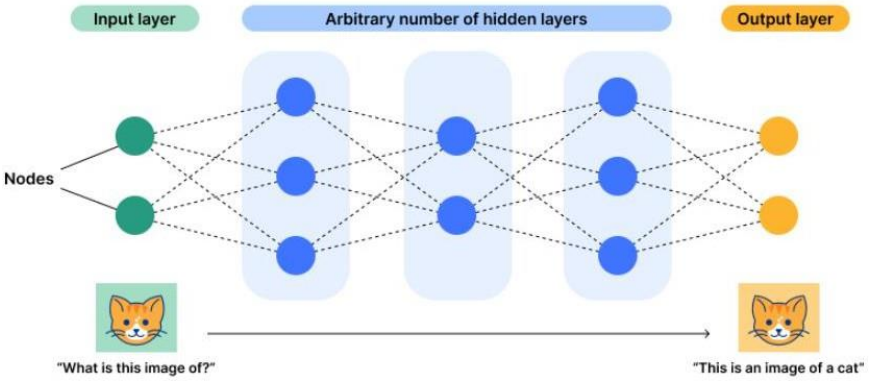

Figura 2 – Exemplo de uma rede neuronal (Retirado de: [21])

2.2.3 Deep Learning

Deep Learning (DL) é uma subárea do Machine Learning que tem como principal foco o uso e melhoramento de redes neuronais artificiais em diversas tarefas como classificação, regressão e reconhecimento de padrões. Nesta subárea o principal foco está centrado em redes neuronais profundas (com 3 ou mais camadas de neurónios) e as suas capacidades de realizar essas tarefas.

Avanços recentes dentro desta área são os principais fatores para existência de modelos generativos tão eficazes como os que temos atualmente, inovações como Transformers e Autoencoders Variacionais têm um papel central nos modelos de processamento de linguagem natural mais usados [22]. Além disso, as evoluções das arquiteturas das redes neuronais, através da criação de redes convulsionais (CNNs) e redes recorrentes (RNNs), permitiram a obtenção de resultados muito positivos. Estas redes tornaram-se hoje os alicerces de modelos baseados em visão computacional, permitindo o seu uso nas mais diversas áreas onde a deteção de objetos é fundamental. Consolidaram-se assim, como ferramentas robustas dentro da medicina, transportes e segurança.

Os progressos feitos também na área do hardware não só proporcionaram uma nova capacidade computacional, mas também trouxeram diversas técnicas de otimização que nos permitiram superar limitações no treino [23].

2.3 Modelos de Difusão

Os modelos de difusão são sistemas probabilísticos que foram propostos por Jascha Sohl-Dickstein em 2015, inspirados pela física termodinâmica fora do equilíbrio. A sua conceção baseia-se principalmente na reversão de uma distribuição complexa de dados e na transformação desta numa nova distribuição mais acessível, onde seria possível o seu tratamento e análise. O problema inicial que levou à ideia deste modelo vem do processo de difusão da física termodinâmica. Nele, ao colocar uma gota de um líquido colorido dentro de outro líquido, o primeiro, após o primeiro contacto com o segundo, acaba por ter uma distribuição completamente aleatória e impossível de prever [24]. Ou seja, por outras palavras, estes modelos foram criados com o objetivo inicial de prever a posição inicial de moléculas após entrarem em contacto com outro líquido pré-existente.

2.3.1 Definição e Conceitos

O núcleo do conceito destes modelos baseia-se na adição consecutiva de ruído nos dados através de um número finito de passos até a amostra inicial já não ser reconhecível, ficando apenas uma imagem de ruído puro. Em seguida, o modelo deve aprender a reverter este processo, ou seja, gerar uma nova imagem a partir de ruído puro e criar assim uma imagem clara. Este processo também é realizado progressivamente através de uma cadeia de Markov. Para alcançar esta capacidade de reversão do processo de difusão, o modelo, durante a sua fase de treino, será exposto a diferentes versões dos dados de treino com ruídos cada vez maiores. O que torna o processo completamente diferente de outros modelos semelhantes, como os GANs, é o fato de o modelo não aprender diretamente com os dados inseridos, mas sim aprender a estimar a melhor forma de remover o ruído e restaurar gradualmente a imagem, em vez de depender de uma competição adversária [24].

Durante o processo de treino, de forma a otimizar a aprendizagem, serão utilizadas métricas, como a divergência de Kullback-Leibler (métrica usada para medir a diferença entre duas distribuições de probabilidade) ou o erro quadrático médio, uma vez que a diminuição dos valores obtidos dessas métricas indicará que o modelo está a conseguir produzir novas imagens realistas com base no ruído fornecido [24].

2.3.2 Como Funcionam

Para descrever o funcionamento destes modelos, é necessário, primeiramente, descrever os seus principais componentes e etapas fundamentais para a criação de novas imagens, incluindo a formulação matemática e a estrutura das redes neuronais envolvidas em todo o processo. Estes modelos fundamentam-se principalmente em duas etapas distintas: o processo de difusão e o processo de difusão reversa [24].

Na primeira etapa, o processo de difusão será feito progressivamente, adicionando ruído Gaussiano ao longo de um número finito de passos T, seguindo a seguinte equação [25]:

$$\boldsymbol{q(x_t \mid x_{t-1}) = N(x_t\,;\, \sqrt{\alpha_t}\, x_{t-1}\,, (1 - \alpha_t) I)}$$

O que esta equação faz é, a cada passo no espaço de tempo T, a amostra x_t vai ser obtida através da versão anterior x_{t-1} aplicando uma transformação linear multiplicativa ($\sqrt{\alpha_t}\ x_{t-1}$), e somando ruído gaussiano, ($(1 - \alpha_t)\ I$). Este processo é fundamental para garantir que, após um número suficientemente grande de passos, a amostra original vai ser substituída por uma constituída de apenas ruído [25].

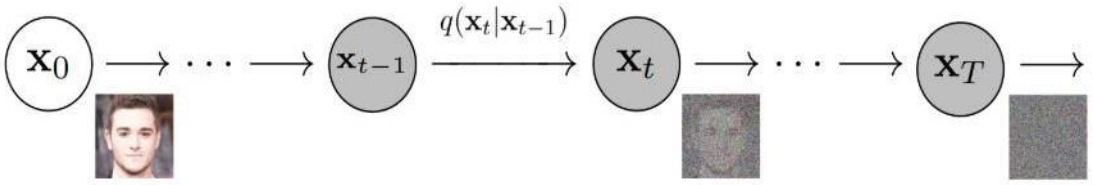

Figura 3 – Processo de difusão (Retirado de: [26])

Na segunda etapa, no processo de difusão reversa, o modelo aprende a reverter os passos feitos no processo de difusão, começando com um dado de ruído puro que será progressivamente refinado até chegar a uma imagem sem ruído. Esta reversão é feita através da seguinte distribuição condicional [25]:

$$p_\theta(x_{t-1} \mid x_t) = N\,(x_{t-1}\,;\, \mu_\theta(x_t\,, t), \Sigma_\theta(x_t, t))$$

A partir desta distribuição, o modelo neuronal vai prever a média $\mu_\theta(x_t\ ,\ t)$ e a covariância $\Sigma_\theta(x_t,\ t)$, permitindo que a amostra seja refinada gradualmente até atingir a distribuição dos dados reais, ou seja, a imagem sem ruído. A intuição por de trás desta equação é que o modelo aprende a prever como era a versão anterior, com menos ruido, utilizando um processo de aproximação estatística. O objetivo do modelo é, após a previsão, minimizar a função que quantifica a diferença entre o ruído previsto e ruído real através da seguinte função [25]:

$$\mathbb{E}_{x_0,t,\epsilon}\left[\,||\epsilon - \epsilon_\theta(x_t,t)||^2\,\right]$$

Esta função vai calcular a diferença entre o ruído gerado e a estimativa proveniente do processo de difusão reversa gerado pela rede neuronal profunda. Através da diminuição dos valores gerados, isso permite que o modelo consiga desfazer a difusão e assim gerar novos dados de qualidade através de ruído gaussiano puro.

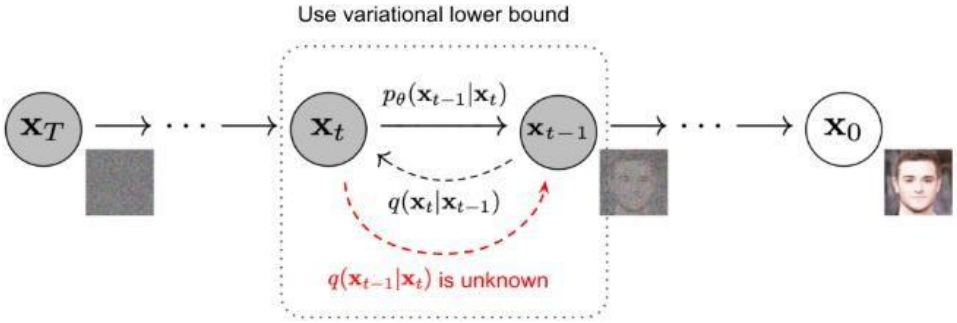

Figura 4 – Processo de difusão reverso (Retirado de: [26])

Na figura abaixo é possível visualizar de forma ilustrativa o funcionamento destes modelos, com o processo de difusão a amarelo, e o processo reverso em azul.

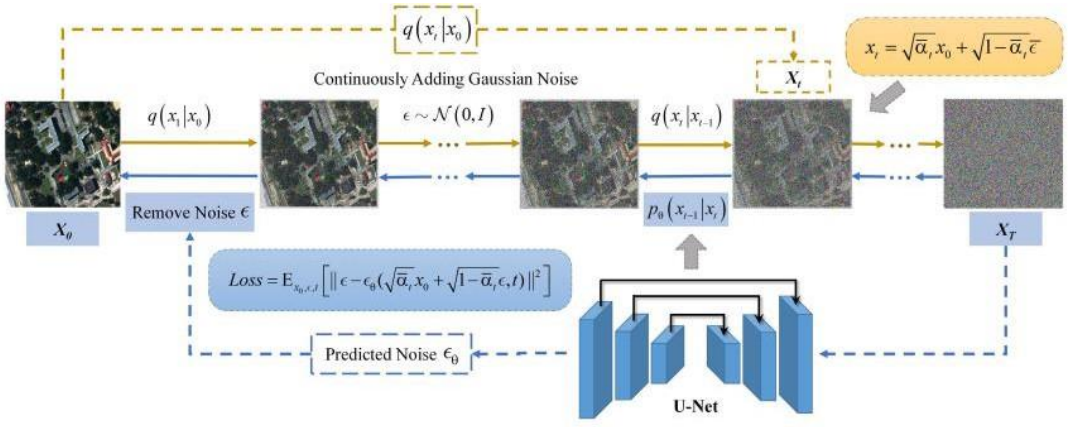

Figura 5 – Demonstração visual de ambos processos (Retirado de: [27])

Em suma, estes tipos de modelos focam-se na aplicação e remoção de ruído Gaussiano, característica essa que os difere de outros. Nestes modelos, o ruído é progressivamente adicionado e removido através de um processo probabilístico, permitindo a reconstrução de dados a partir de ruído. Enquanto isso, as GANs geram dados por meio da interação de um discriminador e um gerador, sendo a principal função do gerador criar dados enquanto o discriminador avalia-os [28] . O principal objetivo das VAEs é converter os dados para o espaço latente, onde são armazenadas as principais características, e reutilizando-as para a gerar novas imagens [29]. Os modelos baseados em fluxo seguem uma abordagem semelhante à dos modelos de difusão, mas, em vez de corromper os dados com ruído, simplificam as imagens durante a fase de treino, com o objetivo de transformar as distribuições complexas das imagens originais, em versões mais simples [30].

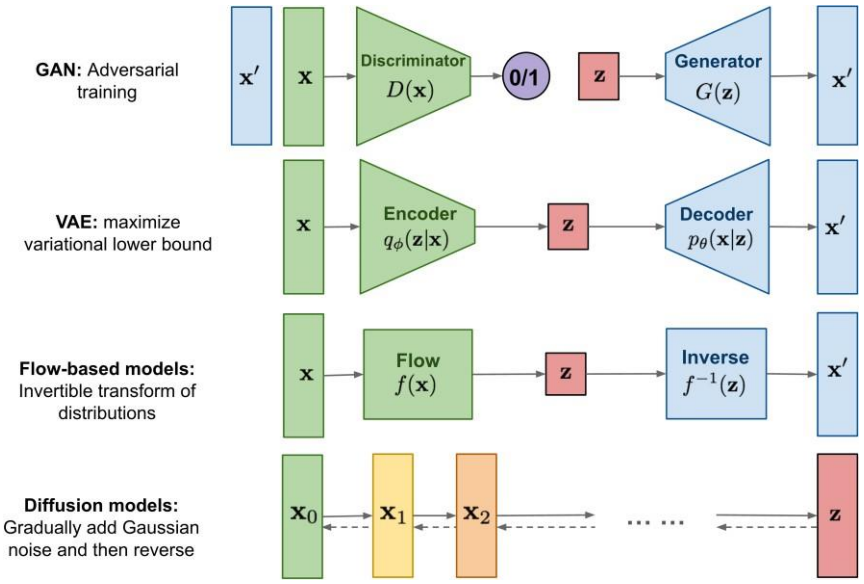

Figura 6 – Comparação visual entre diferentes modelos

2.4 Stable Diffusion

Stable Diffusion é uma abordagem mais recente e otimizada em relação aos tradicionais modelos de difusão. As principais diferenças estão na forma com os dados, ou neste caso as imagens, são transformadas em dados computacionais. Introduz o uso do espaço latente que simplifica o processamento das imagens. Estas deixam de ser transformadas com base nos valores de R, G e B dos pixels e são transformadas em vetores latentes que capturam de maneira mais generalizada as características principais da imagem, como textura, cor e padrões visuais [31].

2.4.1 Arquitetura

A arquitetura dos modelos de Stable Diffusion é composta por vários componentes que, em conjunto, permitem a criação de imagens de alta qualidade com grande eficiência computacional. Isto possibilita que os processos de difusão operem no espaço latente, ao invés de tratar cada pixel individualmente. Os seus principais componentes são:

2.4.2.1 VAE (Variational AutoEncoder)

O VAE, um tipo de rede neuronal, é adotado em modelos de Stable Diffusion pela sua capacidade de simplificação dos dados. Este componente é responsável por transformar os dados do mundo dos pixéis para o espaço latente, sendo um dos principais fatores para a otimização computacional destes modelos [29].

Para serem representadas computacionalmente, as imagens precisam de ser transformadas em um formato interpretável. Normalmente, esta é feita e mantida no universo dos pixéis. Ou seja, por exemplo, numa imagem 50x50, são precisas três matrizes de 50x50 onde cada uma representa o valor de R (Red), G (Green) e B (Blue) de cada pixel. Embora seja forma comum e viável de representar imagens, não é computacionalmente eficiente para o processo de difusão. Portanto, o VAE é o componente responsável por transformar uma imagem para o espaço latente. Neste novo espaço, as imagens são representadas por vetores numéricos que codificam as informações semânticas como imagem, textura, cor e padrões [29] . Tornando assim o tamanho da estrutura dados muito mais pequeno.

Para entender melhor o efeito deste componente, podemos analisar o seu funcionamento com o seguinte exemplo: uma imagem de 1024x1024, com um fator de redução de 4, vai ser transformada numa matriz de 256 por 256 valores, resultando num total de 262.144 valores presentes. Enquanto no universo dos pixéis, precisaríamos de representar com três matrizes, cada uma de 1024 por 1024, dando um total de 3.145.728, oferecendo assim uma diminuição de cerca de 12 vezes o número total de dados.

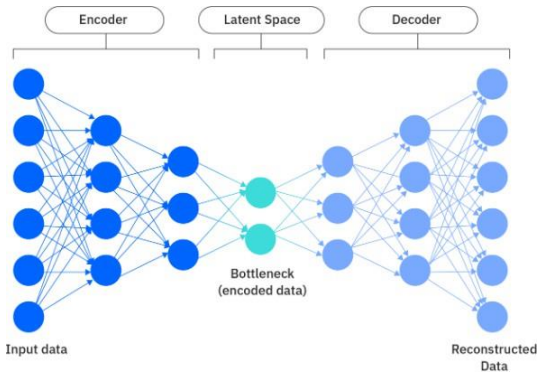

Figura 7 – Arquitetura de um VAE (Retirado de: [32])

2.4.2.2 LDM (Latent Diffusion Model)

O LDM será o nosso modelo de difusão, uma versão aprimorada dos modelos tradicionais de difusão. Nele, os dados são tratados no espaço latente, originados pelo processo do VAE. Este componente segue o mesmo princípio dos modelos de difusão, no qual o ruído é progressivamente adicionado aos dados até que se transformem em uma distribuição de ruído pur. Após esse processo, o modelo aprende a gerar, com base numa distribuição aleatória de ruído, novos dados limpos e de alta qualidade [7] .

2.4.2.3 Rede U-Net

A rede U-Net é uma arquitetura baseada em redes neuronais profundas originalmente criada para a segmentação de imagens médicas, apesar de se demonstrado extremamente eficaz para o processamento das imagens, incluindo modelos generativos como é o caso de Stable Diffusion [33]. A sua estrutura é composta por dois principais elementos, um codificador e um decodificador. O codificador vai progressivamente diminuir a dimensão da imagem enquanto ao mesmo tempo captura as principais características. Já o descodificador vai reconstruir a imagem preservando as características obtidas pelo codificador até ela ter o mesmo tamanho da original. Devido à natureza dos modelos de Stable Diffusion, esta rede vai operar por cima do espaço latente, ou seja, o redimensionamento que ocorre dentro de uma U-Net vai ser feito simultaneamente que o processo de difusão reversa, quando estes terminarem, o output vai ser colocado numa VAE e então é retornada uma imagem final visível [34] .

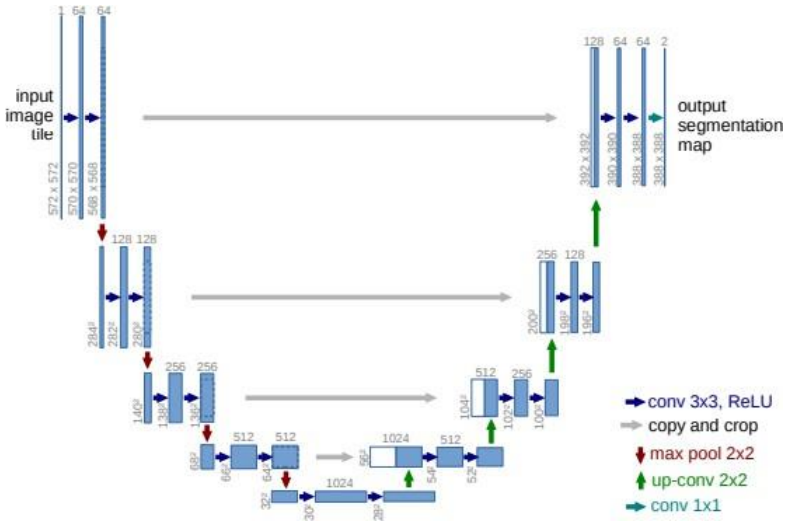

Figura 8 – Arquitetura de uma rede U-Net (Retirado de: [34])

2.4.2.4 CLIP (Contrastive Language-Image Pretraining)

O CLIP é um modelo capaz de processar vários tipos de dados como texto, imagens, vídeos e áudio permitindo as entender relações entre os diferentes tipos de inputs. Num contexto dos modelos de Stable Diffusion, o CLIP tem como principal função compreender o conteúdo dos prompts textuais. Após a interpretação do texto, o modelo irá mapear os prompts para vetores latentes, para depois serem processados pelo VAE [35].

Este componente utiliza duas redes neuronais: uma para processar o conteúdo visual e outra para as descrições textuais. Dessa forma, o modelo consegue criar um alinhamento eficaz entre ambos, permitindo estabelecer correspondências entre o texto fornecido e as imagens geradas [35].

2.4.3 Processo de Treino

O processo de treino de modelos Stable Diffusion ocorre em vários passos. Primeiramente, é fornecida uma imagem de treino com uma quantidade aleatória de ruído ao modelo. A rede neuronal tenta então separar o conteúdo original da imagem do ruído. Com base na previsão do ruído, é possível subtraí-lo da imagem para obter uma estimativa de como seria a imagem sem ele. Em seguida, parte do ruído calculado na etapa anterior, por exemplo, 75% do ruído, é adicionado novamente à imagem prevista como limpa no passo anterior. Isso resulta em uma versão ligeiramente menos ruidosa da imagem original, que será utilizada como input na próxima etapa. O processo é repetido diversas vezes até que o número de passos de treino seja completado. Por exemplo, se for usado como parâmetro de treino 100 passos, o ciclo será repetido 100 vezes por cada imagem. A figura abaixo

representa de forma ilustrativa como uma etapa do processo de treino funciona [36] [25].

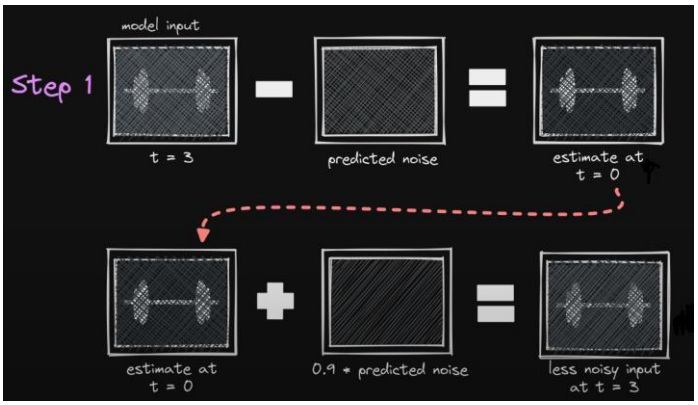

Figura 9 – Ilustração do processo de treino de um modelo de Stable Diffusion (Retirado de: [36])

3. Estudo do Estado de Arte

Nesta secção será feita uma revisão sistemática da literatura sobre o uso de modelos de Stable Diffusion na transferência de estilo. Esta revisão permite aos autores conhecer e compreender os trabalhos e investigações já feitas tornando assim possível explorar novas soluções, adaptar soluções já existentes, e/ou confirmar se a investigação já feita é replicável.

3.1 Metodologia e Processo de Pesquisa

Para a elaboração de uma revisão sistemática da literatura sobre o uso de modelos de Stable Diffusion vamos nos focar nos seguintes aspetos:

- Propósito e questões da investigação;
- Fonte utilizadas;
- Estratégia de pesquisa;
- Critérios de exclusão;
- Análise dos artigos encontrados;
- Conclusão e discussão dos resultados.

3.1.1 Objetivo da Pesquisa

Nesta revisão sistemática vamos investigar o uso de estratégias para treino de modelos, metodologias e arquiteturas de redes neurais empregadas em tecnologias Stable Diffusion, assim como fazer uma comparação direta com outras metodologias e tecnologias. Mantendo a ênfase na área da geração de imagens para compreender como diferentes abordagens podem impactar a coerência visual e os resultados obtidos.

Uma das motivações para a realização desta revisão é o rápido avanço das tecnologias de geração de conteúdos através de Inteligência Artificial, como é possível ver na figura 10, e também entender como esta emergente abordagem diverge das restantes. Na figura 11 é possível visualizar uma comparação entre o uso de IA regular com IA Generativa nas empresas, mais concretamente o número de organizações que adotaram IA para pelo menos uma função dentro da empresa.

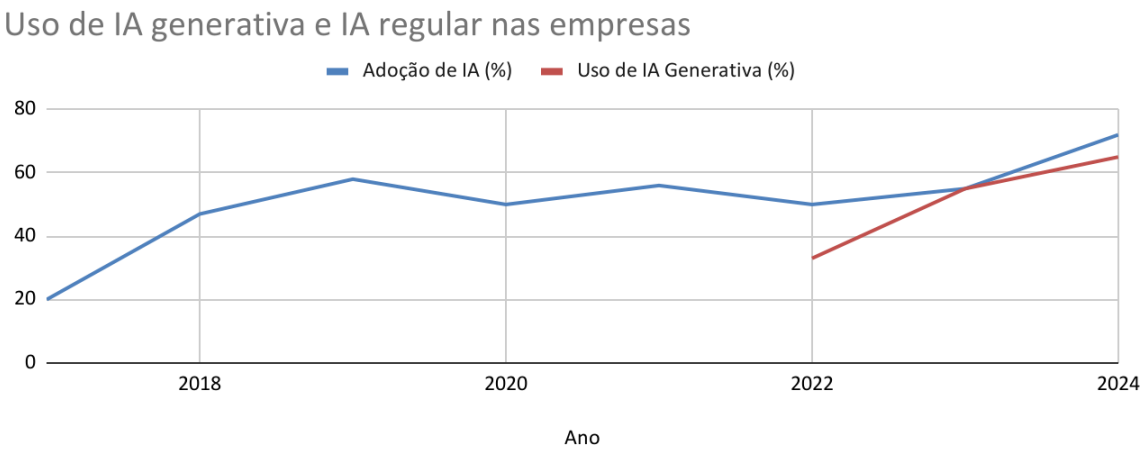

Figura 10 – Uso de IA generativa e IA regular nas empresas (Adaptado de:[37])

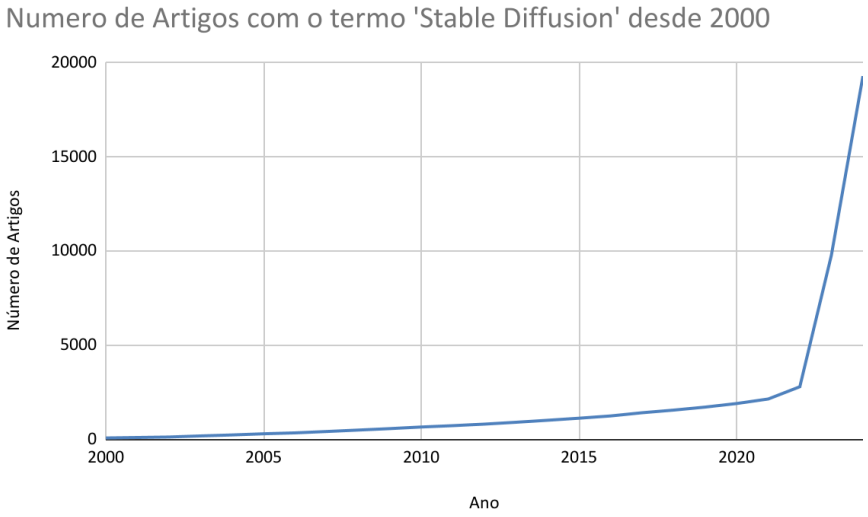

Figura 11 – Número de Artigos com o termo ‘Stable Diffusion’

3.1.2 Fontes Utilizadas

Para a pesquisa de trabalhos e estudos relevantes ao nosso tema, foram utilizadas duas principais plataformas web que disponibilizam acesso a documentos científicos, revistas, artigos, e-books e relatórios dos mais diversos temas dentro da área das ciências da computação.

A IEEE Xplore [38] é uma biblioteca digital disponibilizada pelo Institute of Electrical and Electronics Engineers, que fornece acesso a literatura técnica de alta qualidade relacionada sobre engenharia elétrica, ciência da computação, eletrónica

e outras disciplinas relacionadas. Nesta plataforma é possível ter acesso a mais de 6 milhões de documentos técnicos.

A B-On [39] ou também conhecida como a Biblioteca do Conhecimento Online, foi criada pela FCCN (Fundação para a Computação Científica Nacional) em 2004, com a função de fornecer acesso permanente a diversos textos completos de milhares de publicações científicas. O seu principal objetivo é centralizar e ampliar o acesso à informação científica internacional.

Ambas as plataformas têm mecanismos para refinar a pesquisa, sendo possível a inclusão de operadores lógicos como AND, OR e NOT que se demonstraram altamente úteis para procura de artigos que não só faziam referência às tecnologias em questão, como também nos permitiu filtrar para os casos de uso mais relevantes. Tornando a análise em específico daqueles que tinham como foco a geração de imagens mais fácil, tópico relevante para o nosso objetivo.

3.1.3 Estratégia de Pesquisa

Com o objetivo de alcançar todos os possíveis artigos que se pudessem demonstrar relevantes ao estudo do estado da arte, é importante fazer pesquisas com base nas palavras-chaves que melhor identificam o nosso tema. No nosso caso, procuramos especificamente por artigos que combinam o uso de técnicas que usam Stable Diffusion para a geração de imagens e transferência de estilos. A string de pesquisa utilizada foi:

AB “Style Transfer” AND "Stable Diffusion"

Nesta pesquisa, selecionamos todos os artigos que mencionam 'transferência de estilo' e 'Stable Diffusion' no resumo. Optámos por essa abordagem porque, em muitos casos, essas tecnologias eram citadas, mas não representavam o foco principal do trabalho desenvolvido. Como resultado, identificamos 21 artigos na plataforma B-On.

3.1.4 Critérios de Exclusão

O primeiro critério aplicado foi a remoção dos artigos que não foram escritos em inglês, este critério levou à remoção de um artigo que estava escrito em mandarim. O segundo critério aplicado foi o acesso ao texto integral, que eliminou um trabalho

ficando com 19 artigos no total. Com o objetivo de alcançar apenas os artigos em que foco principal era a transferência artística de estilos, removemos todos os artigos onde abordavam o uso destas técnicas para fins diferentes do desejado, isto inclui os artigos onde o foco é a medicina, biologia e a edição/alteração de vídeos através de modelos de difusão, sobrando assim 10 artigos que seguiriam à fase de análise.

3.2 Análise dos Artigos Encontrados

No artigo [40] os autores realizaram um teste baseado na geração de avatares animados utilizando tecnologias GAN. Eles propuseram um mecanismo de transferência de estilo capaz de intercalar dois estilos de avatares até alcançar uma fusão entre ambos. Para isso, utilizaram as frameworks StyleRig (uma técnica que combina StyleGAN com edição baseada em rigging) e SEAN (Semantic Image Synthesis with Controllable Layered Feature Decomposition). O StyleRig aprende um espaço de estilo controlável, ajustando múltiplas dimensões de parâmetros até atingir um estilo intuitivo e manipulável. Já o SEAN explora os estilos dos avatares por meio de aprendizagem não supervisionada.

Os autores destacam a importância da geração de avatares animados não supervisionados, um método que reduz a necessidade de um conjunto de dados para treino. Isso significa que essa tecnologia pode gerar avatares sem depender de amostras reais, dispensando um pré-treino, e realizando a geração de forma aleatória.

Além disso, exploram a criação de avatares animados multimodais, que podem ser gerados a partir de diferentes tipos de entrada, como voz e texto.

Para o treino, utilizaram a abordagem baseada na DCGAN (Deep Convolutional Generative Adversarial Network), uma arquitetura de redes neurais que conta com dois componentes principais: o gerador, cria as imagens a partir dos vetores de ruído, e o discriminador que avalia os dados criados pelo gerador e tenta diferenciar as amostras reais daquelas que foram criadas artificialmente.

O objetivo do estudo foi explorar o potencial da DCGAN na geração automática de avatares animados de alta qualidade. Os autores reconhecem, no entanto, que ainda há espaço para avanços, abrindo caminho para futuras pesquisas que aprimorem a transformação e fusão de estilos nos avatares animados.

Figura 12 – Comparação dos resultados obtidos pelos autores e outras abordagens (Retirado de: [40])

Os autores do artigo [41] procuraram criar imagens incorporadas em códigos QR de forma a manter o seu normal funcionamento. Para isso, desenvolveram o Text2QR, um método que equilibra a estética com a funcionalidade deste tipo de tecnologias. Como parte da solução, introduziram o módulo QAB (QR Aesthetic Blueprint), que gera uma “blueprint” que faz a união da imagem desejada com o código QR. Após ter a união de ambos os elementos usam o SELR (Enhancing Latent Refinement) irá melhorar a combinação entre a imagem e código QR, mas mantendo o foco na funcionalidade dos códigos.

A framework utilizada pelos autores baseia-se em 3 etapas, primeiro os utilizadores geram as suas imagens utilizando o modelo de Stable Diffusion, enquanto simultaneamente escolhem uma mensagem para inserir via código QR, de seguida vai ser usado o módulo QR Aesthetic Blueprint que tem o conteúdo da imagem gerada refletindo a mensagem, e por fim é construída uma equação para quantificar o conteúdo e a mensagem que for gerada, resultando assim num código QR legível e esteticamente apelativo.

Em suma o Text2QR usa Stable Diffusion para transformar imagens em códigos QR absorvendo o estilo de uma imagem fornecida, uma maneira inovadora de estilizar uma tecnologia muito utilizada mantendo a capacidade de leitura intacta. Apesar de não fazerem referência ao treino, é um grande indicador da capacidade de tecnologias de Stable Diffusion em estilizar imagens.

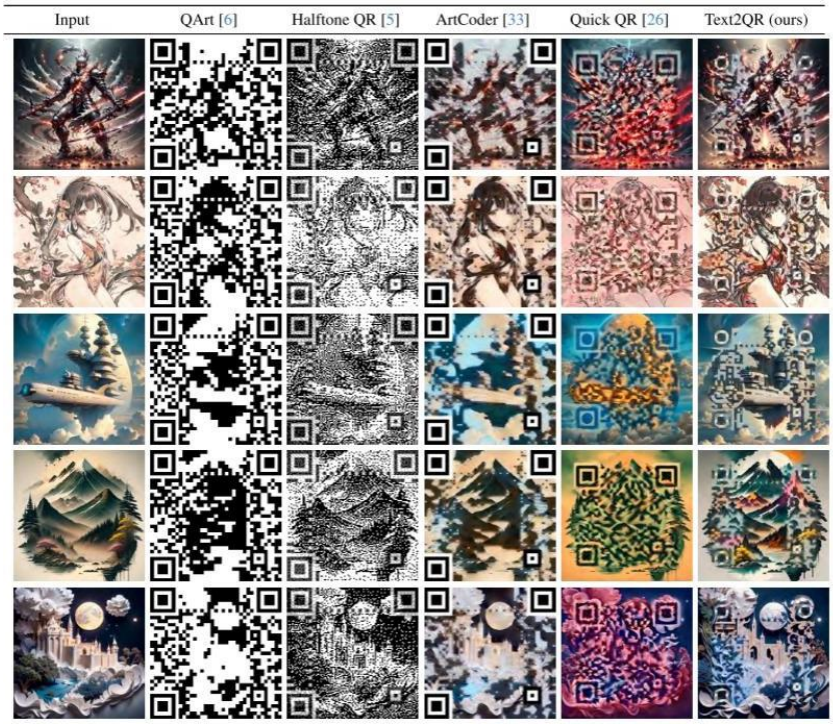

Figura 13 – Comparação de imagens geradas com outras alternativas (Retirado de: [41])

No artigo [42] os autores propõem-se a apresentar um modelo de difusão capaz de aprender e aplicar diversos estilos de forma eficiente e sem a necessidade de muitos dados para treino. Esta abordagem consiste principalmente em duas fases, na primeira é adaptar distribuição do ruído de modo a refletir melhor as características do estilo alvo. E de seguida um refinamento de um componente U-Net do Stable Diffusion vai utilizar esta nova distribuição mais eficaz, para o efeito pretendido.

Os autores demonstraram resultados significativos e muito acima das técnicas anteriormente utilizadas. Este tipo de tecnologia juntamente com um eficaz Prompt Engineering demonstram ter uma precisão muito alta na transferência de um estilo, dados os exemplos usados de esboços de anime, personagens do universo Pokémon e pinturas de Salvador Dalí.

Por fim, fazem referência a alguns pontos fracos como a herança de defeitos com base no modelo em que foi efetuado o treino, juntamente com a necessidade

de ajustes manuais para uma otimização da qualidade dos resultados. Mesmo assim, apesar de alguns impasses, continua a representar um avanço significativo devido à sua surpreendente velocidade de treino, e capacidade de diversas aplicações artísticas demonstrando assim, ter um grande potencial para a expandir o mundo a geração e personalização de imagens através do uso de IA.

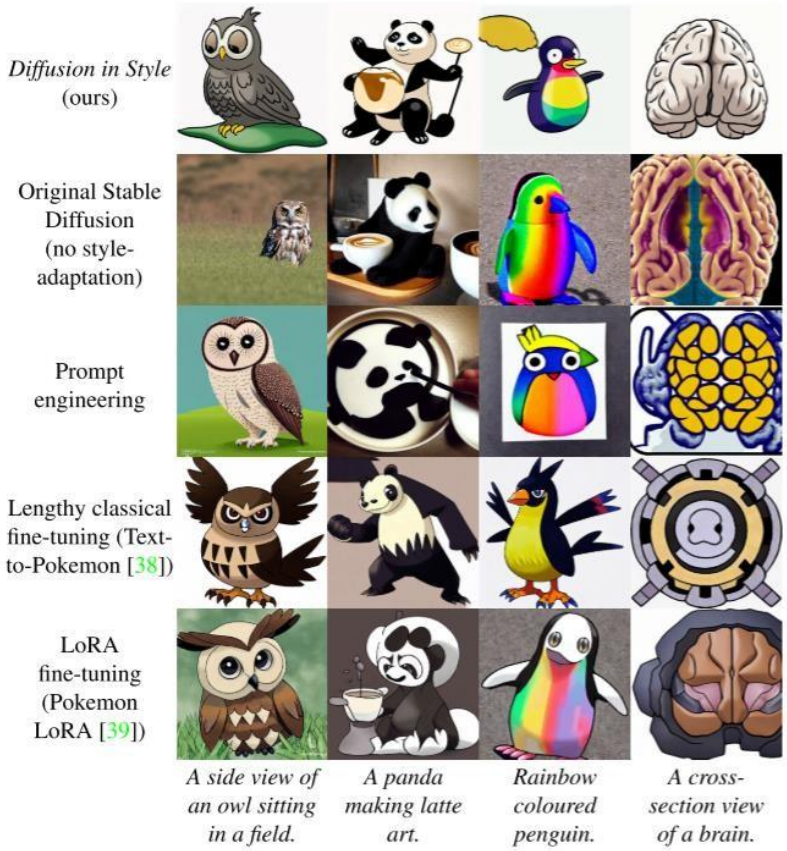

Figura 14 – Comparação de imagens geradas com outras abordagens (Retirado de: [20])

Os autores em [43] propuseram a junção de IES (Interactive Evolutionary Systems) e BO (Bayesian Optimization) para avaliar interactivamente um lote de imagens fornecidas, e com base nos parâmetros dados, o modelo de difusão iria gerar novas imagens adaptando as fornecidas com as características estilísticas pretendidas.

Dessa forma, é possível criar múltiplos designs alternativos através do IES, baseados em um conjunto reduzido de parâmetros. Os autores destacaram um exemplo de uma aplicação do modelo onde era gerado uma série de designs que evoluem a partir de um número mínimo de parâmetros, seguindo um processo de mutação e seleção, similar à evolução biológica.

Essa abordagem expandiu-se para diversas aplicações que utilizam técnicas de interação com o utilizador, permitindo que participe ativamente da seleção ou avaliação das soluções geradas. Em essência, o sistema incorpora as preferências do utilizador, refinando os resultados com base em suas escolhas ao longo do processo evolutivo.

Neste estudo, o sistema gera um conjunto de animações a partir do input do utilizador, onde os parâmetros são selecionados a partir de pequenas opções e tratados de forma independente. A BO é um método de otimização que utiliza uma função de aquisição para equilibrar média e variância, sendo especialmente útil em situações onde a função principal é complexa e difícil de avaliar.

Figura 15 – À esquerda a imagem utilizada como treino para o modelo apresentado, e à direita uma recriação da imagem original feita pelo modelo (Retirado de: [43])

Em [44] os autores abordam técnicas de personalização de estilos com o uso de uma abordagem de nome SGDM (Style-Guidede Diffusion Model) que, através da rede neuronal VGG19, em conjunto com matrizes de Gram, conseguem extrair e aplicar estilos artísticos na geração de imagens. Apesar deste trabalho não referir em particular a transferência de estilos demonstra avanços na capacidade de personalizar imagens fornecidas sem haver a necessidade de treino para cada novo estilo.

O principal objetivo deste artigo foi superar limitações observadas em modelos de geração de imagens. Os autores propõem um método inovador que elimina a

necessidade de treino específico para cada novo estilo, além de aprimorar o alinhamento entre os prompts textuais e o estilo desejado. Para isso, apresentam o SGDM, um modelo que extrai automaticamente o estilo desejado a partir de um conjunto de imagens de referência. Com base nessas imagens, o sistema gera uma máscara de ruído, que orienta a difusão do modelo para criar imagens alinhadas ao estilo fornecido.

Esta abordagem mostrou ter melhores resultados do que as técnicas usadas anteriormente, como ControlNet e T2I-Adapter, nos seguintes aspetos: Personalização de estilo uma vez que através de métricas como o StySim, foi possível chegar à conclusão de que o modelo apresenta um melhor desempenho na personalização adaptativa de estilos. Além disso, as métricas FID e KID demonstraram que o modelo conseguia gerar imagens de alta qualidade. Por fim, a última grande conclusão retirada foi o aumento da consistência na geração text-to-image. Embora esse não fosse o foco do trabalho desenvolvido, houve um melhor alinhamento entre o estilo desejado e o prompt textual fornecido. Os autores referem ainda que o desempenho da solução aplicada sofreu alterações devido ao dataset utilizado, uma vez que este apresentava algumas inconsistências artísticas.

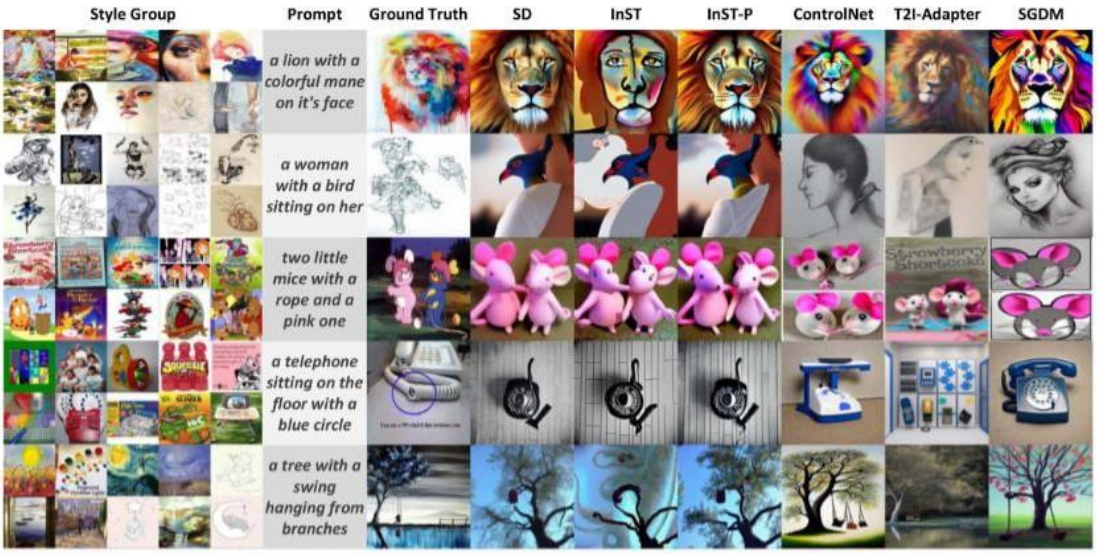

Figura 16 – Exemplos de imagens geradas pelos diversos elementos utilizados no modelo apresentado, juntamente das imagens utilizadas para o estilo e prompt textual (Retirado de: [44])

No artigo [45], os autores propõem um modelo baseado em Stable Diffusion de nome DEPM (Diffusion-Enhanced PatchMatch) que tem como objetivo realizar uma transferência de estilo sem a necessidade de ajustes ou pré-treino.

O DEPM, consegue capturar as características estilísticas e ao mesmo tempo preservar detalhes finos nas imagens permitindo assim uma transferência de estilos arbitrária, flexível e eficiente, destacando-se pela sua não necessidade de ajustes finos. Para alcançar este o objetivo os autores utilizaram métodos inovadores tais como o PM (PatchMatch) e o WCT (Whitening and Colour Transform).

O PostMatch é essencialmente uma técnica que ajuda na transferência de estilo através da aplicação de patches nas imagens, ou seja, a imagem é dividida em pequenos blocos (ou patches) para depois ser feita uma comparação de cada bloco obtido com o estilo alvo a que se pretende aplicar a transferência, assim, para cada patch, vai ser encontrado o pedaço mais próximo com base na textura ou padrões encontrados. Após este processo ser aplicado a todos os patches, começa a fase de substituição, onde cada um vai ser trocado pelo seu pedaço correspondente da imagem de estilo. Por fim os pedaços substituídos são combinados novamente, de forma a originar uma nova versão da imagem original ajustada à imagem do estilo escolhido. Esta técnica demonstrou ser muito eficaz a transportar detalhes de estilo desejado, mas, uma vez que altera a ordem dos pixéis, pode trazer problemas futuros a alguns modelos de difusão.

Já o WCT (Whitening and Coloring Transform) tenta resolver o problema que surge da utilização do PostMatch. Para o fazer, divide o processo em duas etapas: a primeira Branqueamento (ou Whitening) é responsável por remover as características estilísticas da imagem original, esta etapa é importante pois permite criar uma base limpa para aplicação dos efeitos estilísticos da imagem de referência. A segunda etapa denominada de Colorização (ou Coloring) vai reintroduzir os detalhes do estilo da imagem de referência através da matriz de correlação que vai ser criada através da análise da imagem do estilo no espaço latente.

Este trabalho evidenciou resultados experimentais com desempenhos muito positivos incluindo uma superior transformação das cores, melhor preservação de conteúdo e uma alta qualidade na transferência de estilo. O sistema apresentado demonstra uma grande flexibilidade para diferentes aplicações criativas permitindo um controle sobre o equilíbrio entre o conteúdo e estilo. Um dos aspetos mais notáveis no trabalho é a sua capacidade de efetuar a transferência de estilo sem a necessidade de prompts textuais, mas mantendo a sua opção para a modificação das imagens. Esta nova abordagem representa um avanço significativo na área, oferecendo uma solução mais eficiente e flexível.

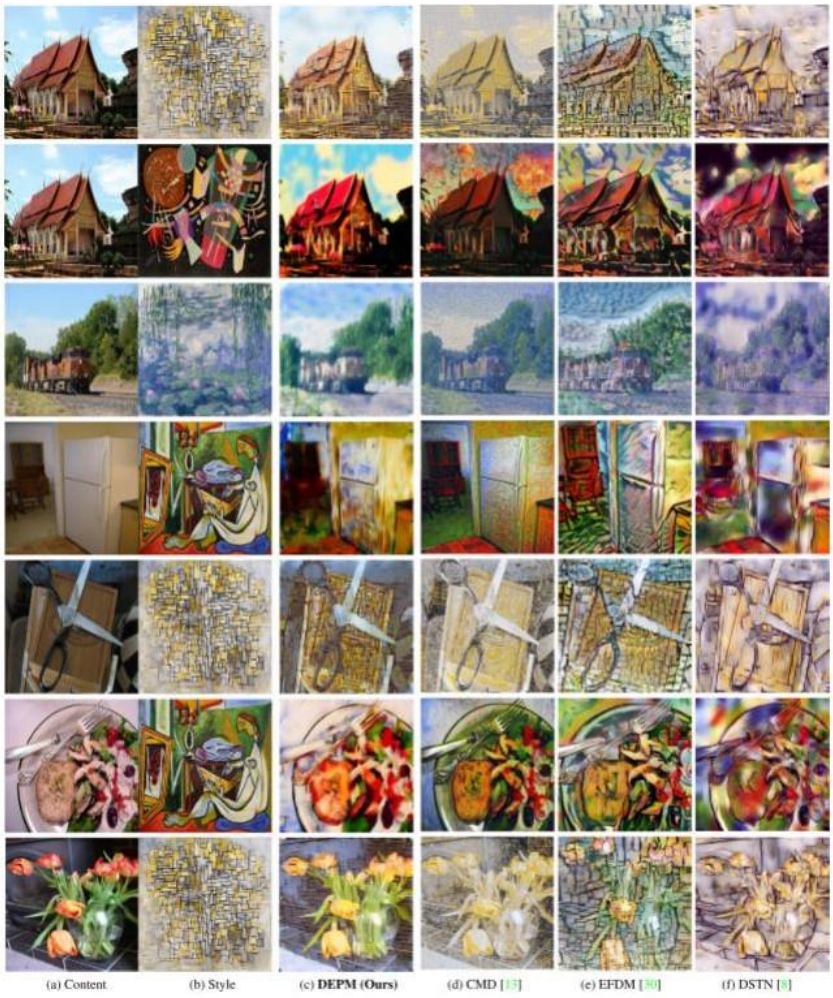

Figura 17 – Comparação da aplicação de estilos entre diversos modelos (Retirado de: [45])

Em [46] os autores fazem referência à dificuldade da escrita chinesa para pessoas que não estão familiarizadas com esta cultura e os seus símbolos ideográficos. Assim, basearam-se na tecnologia Stable Diffusion com adaptações LoRa (Low-Rank Adaptation) para que seja possível redesenhar caracteres chineses com base nos seus significados. Ou seja, os utilizadores poderão compreender e aprender a semântica de cada caracter de uma forma mais rápida e intuitiva.

O modelo de difusão utilizado nesta solução foi adaptado para a tarefa de redesenhar os caracteres chineses que tem como entradas sequências de escrita introduzidas pelos utilizadores. Para o bom funcionamento desta abordagem foram utilizados ficheiros SVG para permitir uma fragmentação precisa dos caracteres introduzidos, permitindo assim ao modelo interpretar todas as linhas de forma individual.

Após a seleção das partes pretendidas, é inserido um prompt que orienta o modelo a conseguir transformar o traço do caractere chinês, em uma figura que se assemelha ao significado do símbolo com o uso de técnicas de difusão latente e adaptações LoRa para o ajuste dos parâmetros. Estando este trabalho numa fase inicial, o foco dos autores é principalmente garantir a qualidade do redesenho dos traços, preservando assim a estrutura do caractere e a sua legibilidade

Este artigo demonstra que os modelos de difusão, mais propriamente Stable Diffusion, podem ter um grande sucesso a adaptar figuras permitindo que pessoas não nativas possam aprender chinês com mais facilidade. Os autores após testarem o modelo desenvolvido, comprovaram um aumento de cerca de 12% na precisão do reconhecimento em caracteres redesenhados.

Além disso a integração de SVG e possibilidade de utilização de prompts permitiu um controlo refinado dos resultados de reconstrução dos caracteres mantendo a sua estrutura e significado. Fazem referência a algumas principais dificuldades como a necessidade de prompts muito específicos e dificuldade em abranger todas as possibilidades de caracteres da língua. Apesar de tudo, continua sendo uma solução inovadora que junta tecnologias que dão uso a modelos de difusão com a área educacional e preservação de cultura para o transporte de estilos alternativos à escrita chinesa.

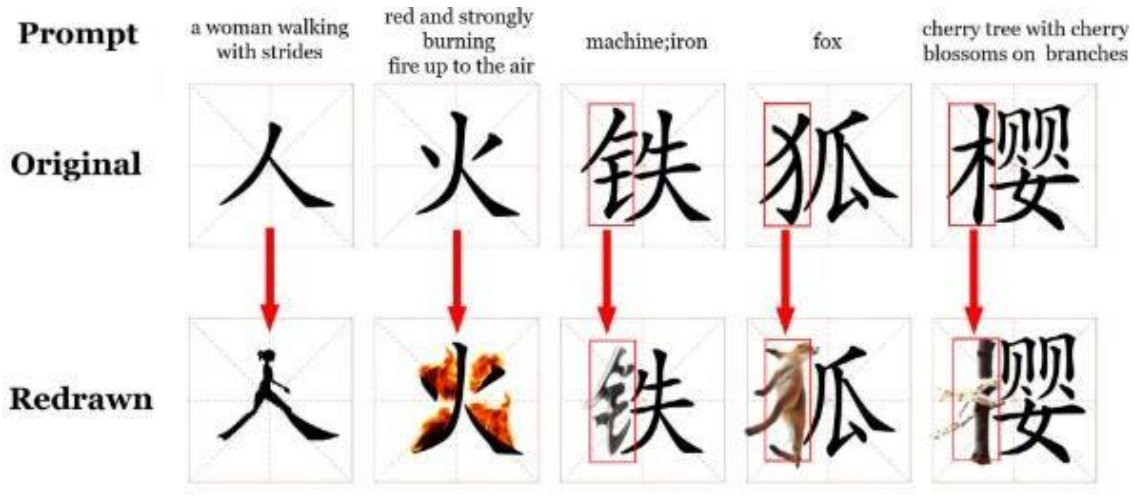

Figura 18 – Exemplo da aplicação do modelo apresentado (Retirado de: [46])

Em [47] os autores fizeram uma pesquisa para introduzir uma nova metodologia de transferência de estilo, através do uso do modelo base Stable Diffusion juntamente com incorporações LoRa (Low-Rank Adaptation) e ControlNet. O principal objetivo deste estudo é o teste da geração de imagens de plantas de diferentes espécies e em diferentes fases do seu crescimento para facilitar a

criação de datasets para outros modelos de IA que tenham o objetivo de fazer análises de fetotipagem (processo de observação das características físicas de uma planta).

A proposta apresentada pelos autores tem a seguinte arquitetura: Uso de LoRa para ajudar a aprendizagem do domínio alvo, neste caso plantas, em que as suas estruturas têm rosetas. Dois módulos ControlNet que através da sua operação em paralelo conseguem identificar com maior precisão as bordas do elemento desejado com o uso de um algoritmo de nome Canny, com esta estrutura é garantida a preservação detalhada da geometria original da planta, para depois, começar o processo de transferência de estilo e gerar assim um conjunto de novas imagens realistas.

Deste trabalho é possível retirar alguns destaques, como a sua capacidade de trabalhar com quantidades reduzidas de dados, os autores dizem que cerca de 10-30 imagens são suficientes para um treino eficaz. Esta abordagem apresentou também resultados superiores aos de tecnologias anteriores, superando assim algumas dificuldades dos métodos convencionais que têm como base o uso de GANs e CycleGANs, tendo assim uma maior consistência e capacidade de adaptação em diversos cenários de utilização.

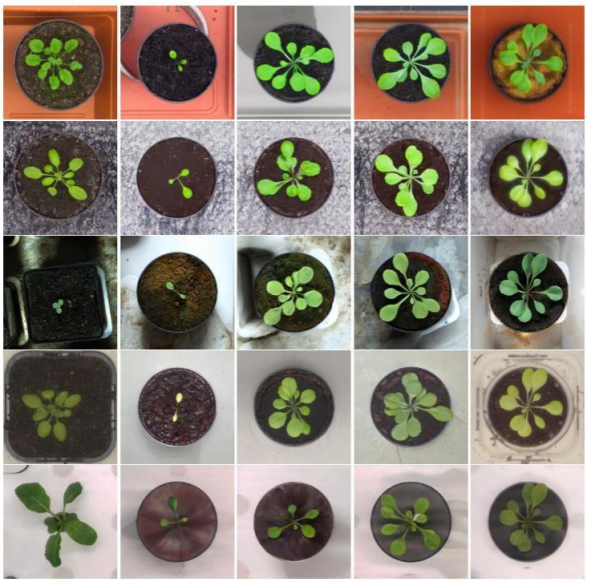

Figura 19 – Exemplo de imagens geradas pelo modelo apresentado, a primeira coluna representa a imagem utilizada como treino, enquanto o resto são diferentes outputs do modelo (Retirado de: [47])

Com o artigo [48] os autores apresentam o MRStyle, um framework que inova na área transferência de estilo focada na cor. Através de dois tipos diferentes de prompt, consegue alterar as cores de uma imagem fornecida. O primeiro prompt

representa a imagem original que queremos editar, enquanto no segundo prompt vai uma descrição textual do estilo desejado. Para a parte da imagem, os autores utilizaram um modelo de nome IRStyle, que, através de tabelas de consulta 3D otimizadas, faz uma rede de mapeamento dupla e um pipeline com aprendizagem híbrida, conseguindo assim alcançar uma eficiência computacional muito elevada, mantendo a consistência e qualidade das imagens, mesmo em cenários onde existe muita variação de cor. Já nos prompts textuais, é utilizado o TRStyle, que em conjunto de um modelo de Stable Diffusion, alinha as características do texto apresentado à imagem fornecida, permitindo assim a customização das cores, minimizando artefactos desnecessários causados pelo uso destes modelos.

O principal desafio deste trabalho foi unificar as informações relativas aos estilos presentes nas imagens com a descrição textual fornecida. Para ultrapassar esta dificuldade, os autores focaram-se em treinar o modelo que efetua a transferência de estilo. De seguida, é feito um ajuste na descrição textual. Segundo os autores esta é a primeira abordagem que mistura a transferência de estilo via imagem e via texto em simultâneo, demonstrando-se assim um trabalho muito positivo por inovar dentro do campo da transferência de estilo com o uso de tecnologias que usam como base modelos de difusão.

Além disso o MRStyle destaca-se também pela sua elevada capacidade lidar com transferências estilísticas em cenários abertos, permitindo transformações eficientes. O uso de um pipeline híbrido combina supervisão pareada e não pareada, foi, segundo os autores, essencial para melhorar a qualidade e eficiência do framework desenvolvido, garantindo que os resultados respeitem as características originais da imagem, assim como, corretamente, alterar a cor de forma que o resultado seja uma fusão entre ambos os inputs recebidos pelo utilizador. Quando comparado a métodos, o MRStyle apresenta vantagens significativas como: maior precisão na transferência de estilo, uma presença reduzida de artefactos visuais produzidos, e melhorias na eficiência do uso de memória e tempos de execução reduzidos mesmo sendo uma abordagem que está preparada para trabalhar com imagens de alta resolução e descrições textuais complexas.

Figura 20 — Comparação entre outros modelos, acompanhada pela imagem de input e descrição textual do estilo (Retirado de [48])

Com [49] é apresentado uma nova abordagem para realizar transferência de estilo com uso de modelos de difusão pré-treinados, esta estratégia destacou-se de outras metodologias principalmente pela sua elevada capacidade de gerar imagens estilizadas, altamente realistas, com foco na estrutura original da imagem, este equilíbrio era algo muito difícil de alcançar em outras técnicas analisadas pelos autores.

A principal inovação que separa esta abordagem das outras, está no sistema de controlo criado para separar o processo de difusão em duas dimensões. O modelo apresentado divide a rede U-Net em três camadas distintas, a grosseira (coarse), a moderada (moderate) e a fina (fine) para conseguir capturar os detalhes estilísticos com maior precisão, através destas divisões, conseguem minimizar a probabilidade de erros na deteção do estilo desejado. O processo feito pelos modelos de difusão também foi alterado, e, segundo os autores, alterar a quantidade de etapas na remoção de ruído trouxe vantagens, nesta abordagem, foram usados 1000 passos divididos em dez etapas distintas que permitem ao modelo ganhar controlo granular sobre a evolução da imagem. Durante ambas estas fases, o sistema apresentado aplica diferentes controlos em cada fase. Na fase inicial, o prompt textual controla a parte grosseira da rede U-Net que preserva as características principais da imagem, já na fase intermédia o foco fica com os detalhes estruturais da imagem através da camada moderada, e, por fim, são feitos os retoques finos e aplicações minuciosas do estilo desejado na última e mais detalhada camada que irá tornar a transição entre a imagem original e o estilo mais suave e natural.

A principal limitação desta alternativa foi o tempo de geração de cada imagem quando comparada a outros métodos passados que utilizam GANs, uma vez que, a divisão por etapas e a criação de vários novos parâmetros atrasou o processo, apesar de resultar em melhorias positivas na imagem final. Os híper parâmetros criados também se demonstraram uma fonte alguns problemas, uma vez que este modelo necessita de um ajuste cuidadoso dos mesmos para maximizar a qualidade dos resultados.

Esta solução representa uma superação significativa nas dificuldades apresentadas por outras soluções que tentam fazer semelhante, pois maior parte dependia de prompts globais ou divisões em etapas mais simples que as apresentadas pelos autores. Este novo sistema que divide o prompt, permite não só apenas ao modelo aprender melhor o estilo desejado, mas também tornar o seu armazenamento mais otimizado.

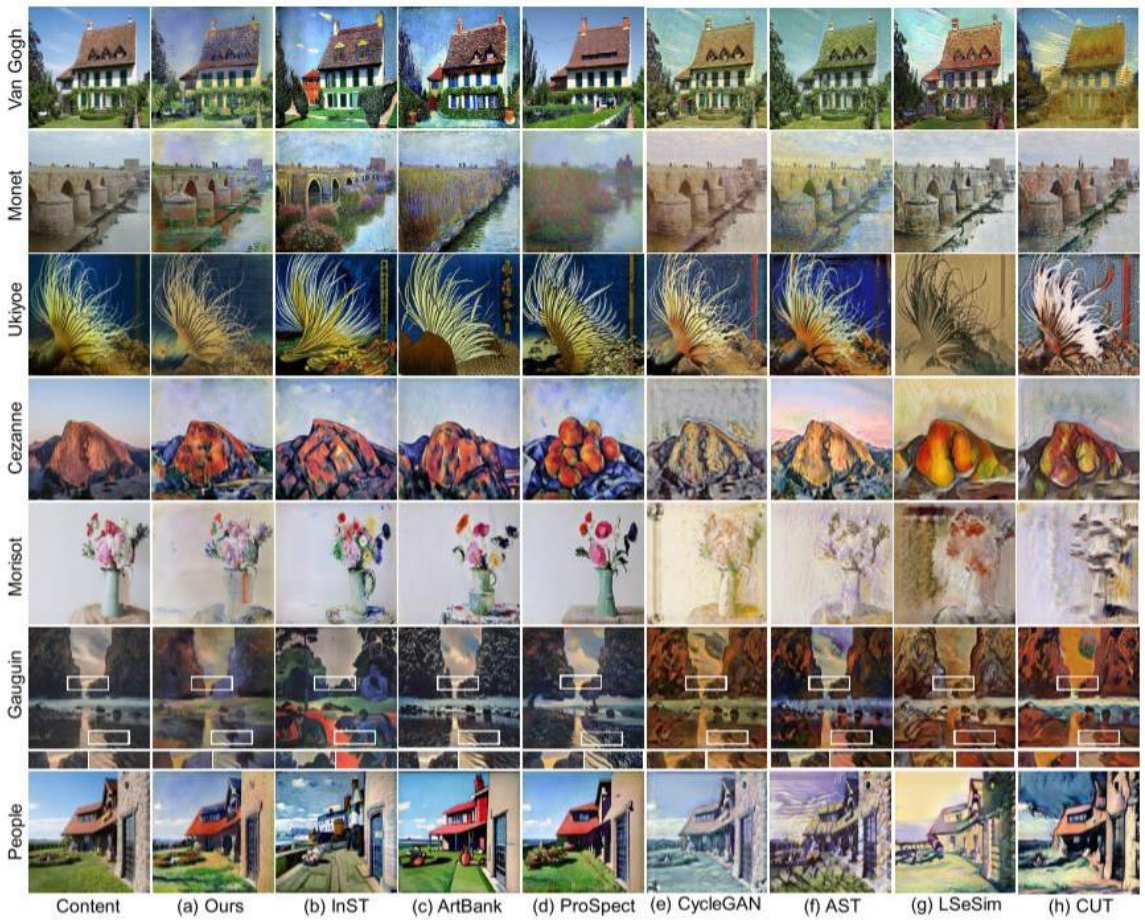

Figura 21 — Comparação dos resultados obtidos com outros modelos, as colunas B-D representam os modelos de Stable Diffusion estudados pelos autores, e as colunas E-H representam outras técnicas como o uso de GANs (Retirado de [49])

Os autores de [50] apresentam o DiffStyler, uma ferramenta criada com o objetivo de realizar transferências de estilos de maneira eficaz. A abordagem apresentada baseia-se na utilização de modelos de difusão, mais especificamente Stable Diffusion que, com o uso de LoRa (Low-Rank Adaptation), permite de uma maneira mais eficaz capturar o estilo desejado.

Para alcançar este objetivo, os autores exploraram algumas técnicas presentes nos modelos text-to-image como: inversão DDIM, mecanismo responsável por reconstruir as imagens fornecidas do espaço dos pixéis para o espaço latente permitindo modificações sem perder a semântica original, a injeção e características e mecanismos de atenção LoRa, que irá permitir a combinação dos estilos desejados, ou seja, manter os elementos estruturais da imagem original enquanto é aplicada detalhes do estilo desejado. O treino LoRa, é um método muito eficiente que permitiu treinar um modelo pré-treinado mediante matrizes de baixa complexidade, isto, essencialmente, permitiu capturar os atributos do estilo desejado sem haver a necessidade de alterações grandes dos pesos originais do modelo utilizado, ao invés disso, são feitas pequenas alterações de forma que seja possível aprender o estilo apresentado. Por fim a remoção ruído por meio de máscaras obtidas pelo modelo FastSAM, ajudaram a preservar certas áreas da imagem original, de modo que haja um equilíbrio entre o novo estilo inserido na imagem e os seus elementos originais.

Após os testes realizados, esta abordagem demonstrou uma consistência notável nas imagens obtidas, ajudando a fundamentar a viabilidade do uso de modelos de Stable Diffusion juntamente com LoRa. Concluindo, o DiffStyler representa uma significativa contribuição na área da transferência de estilos com modelos de difusão, permitindo a aprendizagem de estilos via uma única referência e integrando o estilo eficazmente numa imagem fornecida.

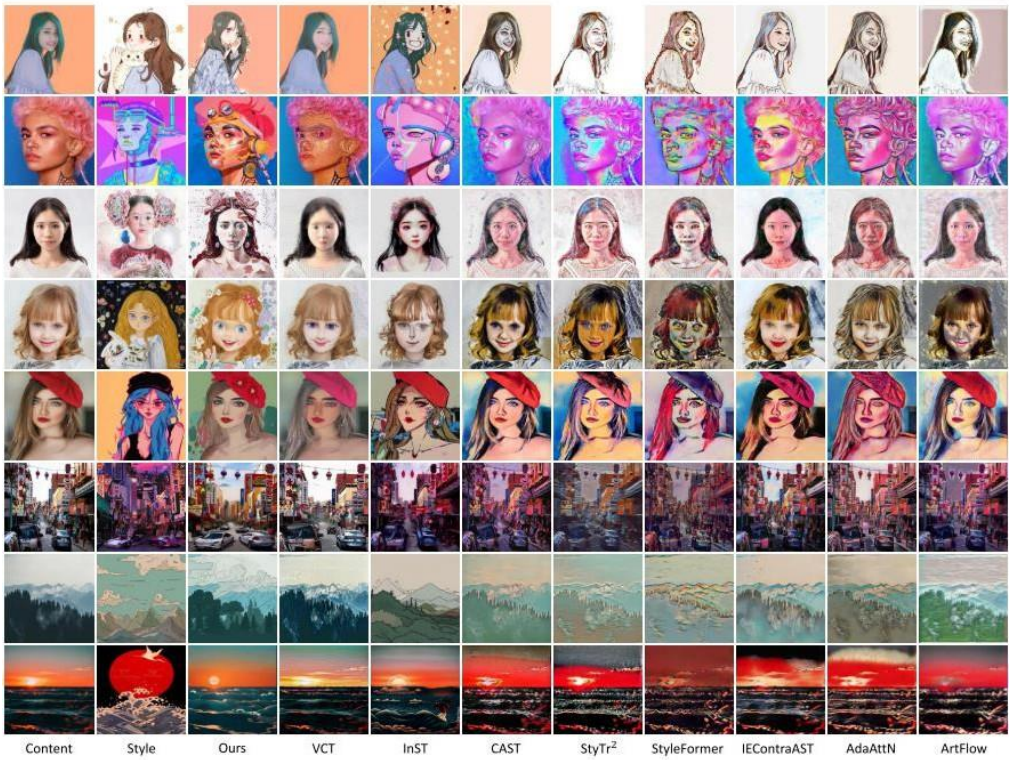

Figura 22 — Exemplos de imagens geradas por outros modelos e pelo modelo apresentado pelos autores, com base num input (Content) e num estilo (Style) submetidos (Retirado de: [50])

Para sintetizar a análise realizada, é possível visualizar na Tabela 2 as principais características analisadas em cada artigo

Tabela 2 - Descrição dos Atributos

Atributo	Descrição
Referência	Número de referência do artigo
Tema	Nome ou tema principal do artigo
Ano	Ano em que estudo/artigo foi realizado
Técnicas Utilizadas	Métodos, frameworks, modelos e algoritmos aplicados pelos autores para realizar os objetivos do artigo.
Resultados	Contribuições significativas do artigo, como, os principais resultados obtidos e as suas dificuldades.

Na tabela 3 é possível visualizar os principais tópicos da cada um dos artigos analisados consoante os parâmetros da tabela 2.

Tabela 3 - Resumo das caraterísticas analisadas em cada artigo

Ref	Tema	Ano	Técnicas Utilizadas	Resultados
[40]	Geração de Avatares Animados com GAN	2023	GAN (DCGAN), StyleRig, SEAN	Alta flexibilidade
[41]	Text2QR: Geração de QR Codes Estéticos	2023	Stable Diffusion, QR Aesthetic Blueprint (QAB), SELR	Equilíbrio entre arte e legibilidade dos códigos QR
[42]	Modelo de Difusão com Alta Eficiência de Estilo	2024	Stable Diffusion, refinamento de U-Net, ajuste de ruído	Alta precisão na transferência de estilos em pinturas famosas
[43]	Combinação de IES e BO para Estilo Personalizado	2023	IES, BO, biomorphs	Alta personalização com geração interativa de designs
[44]	SGDM: Personalização Sem Treino Prévio	2024	VGG19, matrizes de Gram	Alta personalização de estilo em text-to-image
[45]	DEPM: Transferência de Estilo com PatchMatch	2023	PostPatch, WCT, Stable Diffusion	Preservação de detalhes finos e flexibilidade
[46]	AIGC para Redesenho de Caracteres Chineses	2024	Stable Diffusion, LoRA, SVG	Aumento de 12% no reconhecimento dos caracteres redesenhados
[47]	Geração de Plantas para Fenotipagem	2024	Stable Diffusion, LoRA, ControlNet,	Treino eficaz, sem a necessidade de

			algoritmo Canny	um grande dataset
[48]	MRStyle: Transferência de Estilo via Texto e Imagem	2024	IRStyle, TRStyle, Stable Diffusion	Ganho na precisão da transferência de estilo e redução de artefactos
[49]	Nova Abordagem de Estilo com Difusão Pré-Treinada	2023	Divisão U-Net em coarse, moderate, fine; 1000 passos de difusão	Imagens realistas com preservação estrutural
[50]	DiffStyler: Transferência Eficaz de Estilo	2024	Stable Diffusion, LoRA, DDIM, FastSAM	Aprendizagem eficiente de estilos com preservação estrutural

3.3 Discussão dos Resultados

Os avanços analisados nos modelos de difusão, mais concretamente, o Stable Diffusion, têm transformado o cenário dentro da área da IA generativa. Estes modelos têm se demonstrado extremamente versáteis e capazes de gerar imagens que correspondem às mais diversas realidades. Através deste subcapítulo pretendemos explorar os principais atributos que contribuíram para estes resultados a partir das conclusões a retirar de cada um dos artigos.

O Stable Diffusion tem ganho bastante destaque nos últimos anos no que toca a transferência de estilos, demonstrando melhores resultados que estratégias até então como as GANs (redes generativas adversárias) tornando-se assim a escolha acertada para o transporte do estilo do bordado de Castelo Branco. Abordagens analisadas como o caso do DiffStyler [50], mostram como a utilização de modelos de difusão com outras técnicas auxiliares como LoRas, pode trazer resultados muito positivos, permitindo uma aprendizagem rápida e eficiente de um estilo ao mesmo tempo que mantendo a estrutura e semântica da imagem original fornecida, oferecendo um alto nível de consistência e realismo.

Os trabalhos realizados em [41] e [46] demonstram também uma muito positiva prova da capacidade de juntar à transferência de estilos uma utilidade fora do universo artístico. Em [47] também é possível vermos um cenário onde o foco principal não está na criação de imagens para fins artísticos, mas sim utilizando a ideia da transferência de estilos para criar imagens realistas o suficiente para serem utilizadas como parte de um dataset, onde, caso não utilizassem este tipo de técnicas, iram ter dificuldades para atingir um número satisfatório de imagens.

Apesar disso, pelos trabalhos analisados, é possível analisar que o forte deste tipo de modelos está nas adaptações artísticas, como é possível ver pelos artigos [42], [43], [44], [45], [48], [49] e [50] sendo nestes trabalhos explorado a vertente artística destes modelos, desde a transferência arbitrária de estilos, até à transferência de estilos das mais conhecidos como estilo de pinturas de artistas famosos para a outras imagens fornecidas como input. Uma análise a estes artigos, mostrou-nos que todas as abordagens analisadas, demonstraram-se ser superiores aos métodos utilizados até então em diversas métricas, como, tempo de execução, flexibilidade, redução quantidade de imagens para treino e capacidade de manter os elementos principais da imagem fornecida.

3.4 Conclusões

O estudo da análise dos artigos permitiu-nos verificar e perceber os avanços realizados na área da transferência de estilos e personalização de imagens via IA generativa, com o uso de modelos de difusão. Estes avanços tendem sempre em focar na criação de soluções mais eficientes e flexíveis, enquanto mantém um nível bastante positivo de realismo e fidelidade às imagens fornecidas. Um dos principais pontos em comum com os estudos analisados é o uso de pequenas adaptações e inovações nos modelos de difusão, tais como Stable Diffusion, o uso de ControlNet e LoRa que quando utilizados, elevam os resultados esperados oferecendo uma maior precisão na transferência estilo. Além disso, a utilização destes avanços torna os modelos muito mais flexíveis e preparados para diferentes estilos, tornando assim estas soluções extremamente adaptáveis a outros meios assim como preparadas para receber não só uma imagem para efetuar a transferência de estilo, mas também prompts textuais que permitem ao utilizador refinar os resultados ou alterar as características mais pequenas.

Apesar de tudo, foi ainda possível verificar algumas das mais frequentes dificuldades destas técnicas, a necessidade de ajustes finos mediante prompts textuais foi uma limitação em vários trabalhos, o tempo de geração de imagens e os desafios na otimização de alguns parâmetros dificultaram a geração de resultados consistentes.

4. Transferência de Estilo com Stable Diffusion

Neste capítulo, pretendemos entrar mais a fundo na transferência de estilo, através da introdução de alguns conceitos fundamentais para o entendimento desta técnica. Vai também ser feita uma referência a alguns problemas e questões a serem levantadas com a aplicação de técnicas como esta, assim como uma breve evolução da transferência de estilos.

4.1 O que é Transferência de Estilo

A transferência de estilo é uma técnica desenvolvida há perto de duas décadas, baseia-se na manipulação de um conjunto de dados, normalmente imagens ou vídeos. O objetivo deste processo é criar amostras de dados utilizando como base outros dados, combinando essencialmente as características estilísticas de uma imagem com o conteúdo de outra. Uma das suas principais dificuldades, é ter ao seu dispor uma quantidade suficientemente ampla de dados para conseguir capturar todas as nuances e detalhes de um dado estilo, já que a qualidade dos resultados obtidos vai depender diretamente com a qualidade e flexibilidade do conjunto de treino. Outra dificuldade muito comum, é a dificuldade na distinção entre o estilo e o conteúdo da imagem, o que pode levar a uma incorreta seleção das características do estilo, criando assim inconsistências e resultados menos positivos[51].

4.2 Implicações Éticas e Legais

A transferência de estilo, quando junta de ferramentas poderosas de inteligência artificial, pode levantar sérias questões éticas e legais. Como, por exemplo, ao treinarmos um modelo para aprender o estilo de um dado artista, podemos obter resultados tão convincentes que é natural surgir a dúvida de se a obra gerada foi feita pelo autor em questão, ou não, criando assim problemas sérios, e situações onde ocorre uma violação da lei da propriedade. Além disso, há o risco de plágio, pois obras geradas com transferência de estilos, embora muitas vezes tragam consigo fusões novas de estilos e criação de arte nunca vista, acabam, por natureza, carregar características identificáveis das imagens originais.

Outro grande problema muito discutido na atualidade está relacionado com o uso de tecnologias como esta para a manipulação de imagens. Vivendo nós na era das notícias falsas onde existe sempre a necessidade de verificar os factos, estas tecnologias abrem espaço para a criação de imagens com os seus conteúdos

visuais alterados ou então até mesmo a imitação de documentos oficiais com mensagens alteradas.

Em suma, apesar de ser uma tecnologia revolucionária dentro dos campos criativos e artísticos, é importante haver uma responsabilização pelo conteúdo criado e atenção cuidadosa aos seus utilizadores e intenções.

4.3 Técnicas Tradicionais de Transferência de Estilo

As primeiras técnicas de transferência de estilo baseadas em renderização conhecidas como NPR (Non-Photorealistic Rendering) foram criadas em 2001 por Hertzmann, sendo desenvolvidas visando a geração imagens estilizadas que imitavam os estilos mais tradicionais de arte como desenhos e pinturas, ao invés de ir atrás da foto realismo. Estas abordagens faziam comparações pixel a pixel para a deteção de contornos e aplicações de filtros simples que tinham como principal objetivo, remover algumas imperfeições criadas. Ao aplicar estas técnicas, o objetivo é alterar a aparência de uma dada figura e computacionalmente aplicar efeitos estilísticos que remetessem a formas de arte tradicionais.

Estes métodos clássicos seguem frequentemente um modelo de “um para um”, no qual todo o processo de transferência ocorre tendo apenas como base duas imagens, uma onde o estilo vai ser extraído e outra onde vai ser aplicado. Estas abordagens embora simples, tinham grandes dificuldades quando o estilo desejado era rico em detalhes uma vez que a simples análise dos contornos não iria permitir capturar todos os detalhes do estilo[52].

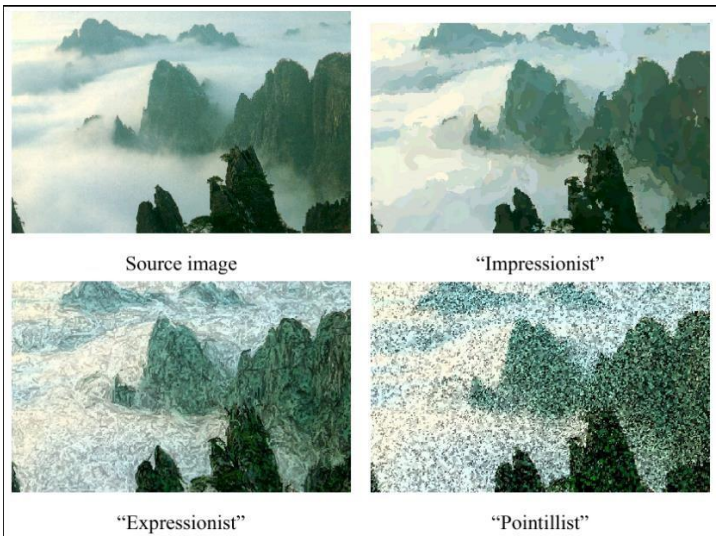

Figura 23 — Exemplo do uso do NPR para três estilos distintos de pinturas (Retirado de: [53])

4.4. Aplicação de Transferência de Estilo com Stable Diffusion

Modelos que utilizam Stable Diffusion tem capacitado cada vez mais a aplicação de transferência de estilo no dia a dia das pessoas por meio de aplicações que incorporam este tipo de modelos. Permitindo assim, o transporte de estilos para novas realidades, como, por exemplo, aplicar o estilo de famoso pintor numa foto nossa [45].

A principal vantagem ao utilizar este tipo de modelos difusão é a sua capacidade em gerar resultados muito positivos sem a necessidade de grandes conjuntos de dados para treino, a sua eficiência, flexibilidade e capacidade de fazer ajustes finos e alta customização do conteúdo de cada foto.

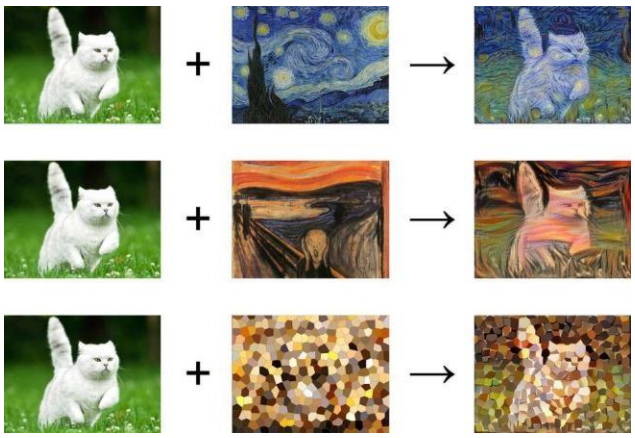

Figura 24 — Exemplos de transferências de estilo com Stable Diffusion (Retirado de: [54])

5. Caso de Estudo: Bordado de Castelo Branco

Neste capítulo, temos como objetivo explorar com mais detalhe o caso de estudo a ser analisado, assim como documentar a criação do dataset, passando pelos pontos mais pertinentes como a recolha, tratamento e dificuldades deste processo.

5.1 História

O bordado de Castelo Branco foi e continua a ser um dos símbolos mais característicos da região de Castelo Branco, é feito a partir de colchas de linho bordadas com fio de seda, com desenhos influenciados por estilos e técnicas tradicionais, tornando-se assim conhecido a partir do meio do século XVI. Este tipo de arte ficou fixa na região de Castelo Branco devido à sua cultura do linho que vinha do fato de amoreira se desenvolver com grande facilidade nesta região [55] .

A maioria dos elementos das representações bordadas têm significados, como, por exemplo, os cravos e rosas representarem o homem e a mulher, corações o amor, gavinhas para a amizade entre outros. Uma das variadas características do tema do bordado de Castelo Branco, é que ele se propaga pela cidade quer nas ruas, edifícios e calçadas assumindo um dos maiores símbolos da cidade [56].

O bordado de Castelo Branco torna-se diferente de todos os outros, pois a simbologia embutida nele é própria e original, atualmente ainda se produz algumas peças, apesar do processo de criação das mesmas ser, por vezes, bastante difícil.

Algumas dessas peças estão expostas localmente em museus, como por exemplo, o Museu Francisco Tavares Proença Júnior ou então também visíveis em feiras ou eventos como a Feira Internacional de Turismo de Madrid, uma vez que o município de Castelo Branco marcou a sua presença fundamentalmente por causa do bordado [56].

Figura 25 – Exemplo do Bordado de Castelo Branco

5.2 Criação do Dataset

Com o objetivo de ensinar o estilo do bordado de Castelo Branco ao nosso modelo, era de grande importância conseguir criar um dataset que represente com a máxima clareza e diversidade possível os elementos do bordado. O dataset utilizado foi feito com base nos conteúdos digitais disponíveis pela internet, que retratavam parcialmente ou completamente os elementos estilísticos do bordado.

5.2.1 Recolha das Imagens

No desenvolvimento desta fase do projeto, encontrámos algumas dificuldades técnicas que fizeram com que os resultados obtidos não fossem os mais satisfatórios.

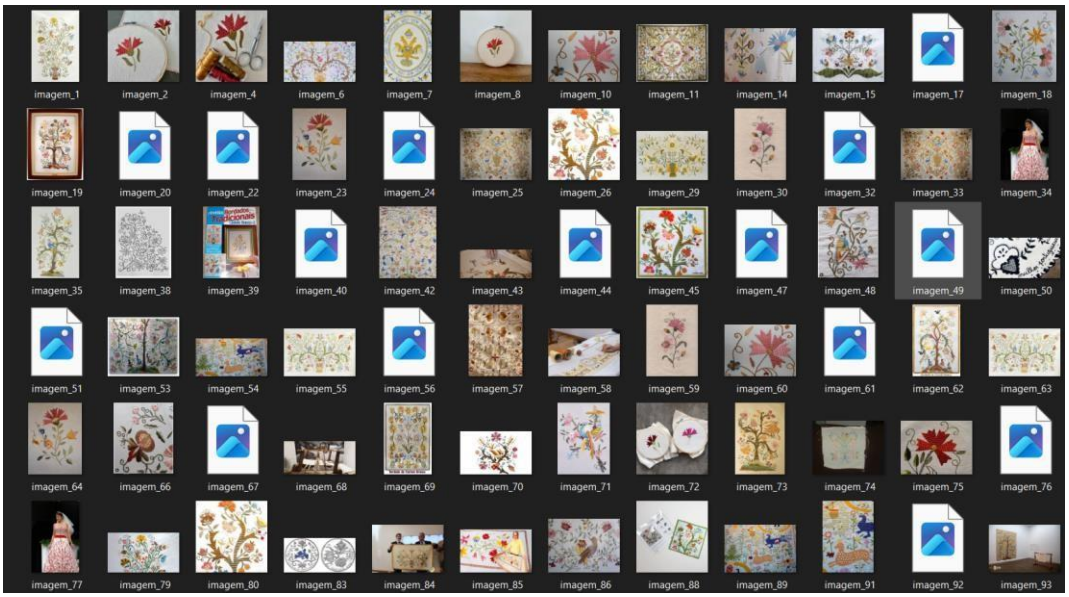

Figura 26 – Seleção das imagens para o dataset

Foram recolhidas 200 (duzentas) fotografias, todas com origem na pesquisa via internet com o termo ‘Bordado de Castelo Branco’. Para a obtenção destas imagens, o optámos pela utilização da API do Google Custom Search JSON API [57] que tem como principal objetivo, criar, com base numa query, um ficheiro JSON com os URLs de todas as imagens obtidas através da pesquisa, e através da biblioteca requests disponível nativamente com a linguagem Python, conseguimos transferir cada uma das imagens retornadas pela pesquisa feita.

5.2.2 Desafios

Após termos as imagens obtidas, enfrentámos alguns desafios, principalmente no que toca na estrutura das fotos, uma vez que nem todas têm a mesma luminosidade, ângulo, e até mesmo a estrutura em volta do bordado serem diferentes, o facto de algumas das fotos terem sido duplicadas ou então não estarem no formato correto tornou o seu acesso impossível. Na figura 27 é possível visualizar um de muitos exemplos de fotos não tão satisfatórias obtidas pelo processo de recolha anteriormente descrito.

Figura 27 – Exemplo de uma foto obtida

5.2.3 Preparação dos Dados

De forma a minimizar os desafios que surgiram pela falta da qualidade das imagens foi feita uma escolha, dentro de todos os resultados obtidos, de forma a excluir as situações mais extremas, onde seria muito difícil de extrair apenas os elementos presentes no bordado. Para todas as outras, com o objetivo de aumentar significativamente o tamanho do nosso dataset, começamos por recortar as imagens de forma a ficar apenas com as partes mais relevantes. Após os recortes, optámos por efetuar diversas rotações e espelhamentos de forma a multiplicar as imagens obtidas e maximizar assim a informação que o modelo poderia aprender com cada uma das imagens obtidas.

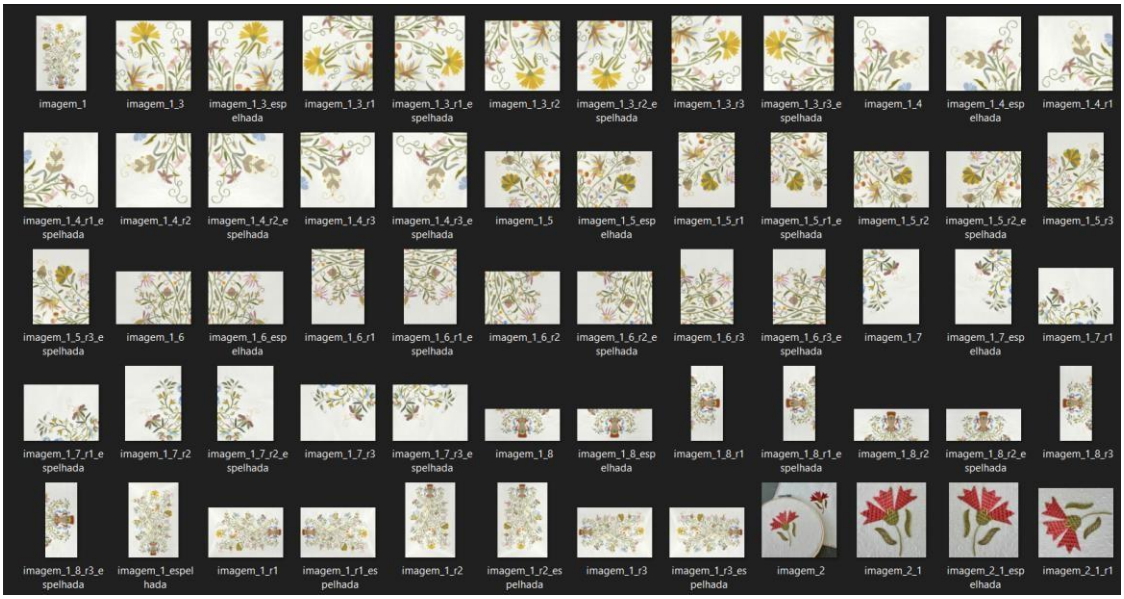

Figura 28 – Seleção das imagens para o dataset após recortes e rotações

Com este processo conseguimos transformar as 200 (duzentas) imagens obtidas para 1785 (mil setecentos e oitenta e cinco), classificadas como Bordado de Castelo Branco.

Por fim, com o objetivo de criar um standard de tamanhos, convertemos todas as imagens para um tamanho igual, aceite de forma generalizada, com diversos modelos de Stable Diffusion, tornado assim todas as imagens no tamanho 512x512.

6. Ferramentas Utilizadas

Com este capítulo pretendemos abordar as mais predominantes ferramentas utilizadas para o desenvolvimento deste projeto.

6.1 Python

Python [58] é uma linguagem de programação de alto nível que devido à sua fácil sintaxe permite aos desenvolvedores escreverem o seu código de forma clara e concisa. É, também, uma das ferramentas mais populares dentro do campo da IA devido ao seu amplo suporte e número extenso de módulos e bibliotecas que dão suporte a esta área.

6.2 Google Colab

O Google Colab [59] é uma plataforma baseada em nuvem que permite a execução de código remotamente, diretamente do navegador. Esta ferramenta foi fundamental para o desenvolvimento do nosso projeto, especialmente na secção de treino do modelo utilizado, uma vez que, permitiu realizá-lo rapidamente devido à alta capacidade de computação das máquinas disponibilizadas, assim como ter um ambiente controlado onde o controlo de versões de cada um dos módulos necessários não é um problema.

6.3 Hugging Face

Hugging Face [60] é uma plataforma com foco na comunidade de ML e ciência de dados, muito semelhante com o GitHub, onde o foco principal está na disponibilização de ferramentas, modelos e datasets open source. Neste projeto o seu uso está relacionado com a interligação com as plataformas Google e o seu fácil uso junto do Google Colab.

6.4 Birme

Birme [61] é uma aplicação web que permite o redimensionamento e recorte em massa de imagens. É uma plataforma gratuita, que facilita o processo de ajuste do tamanho de um grande número de imagens, tornando-se assim importante para este projeto, uma vez que a normalização do tamanho das imagens tornou a aprendizagem mais fácil.

6.5 Automatic 1111

O Automatic 1111 [62] é uma plataforma open source disponível no GitHub, que oferece uma interface gráfica simples e intuitiva para o uso de modelos SD. Esta permite o uso de vários tipos de prompts, assim como a alteração minuciosa de diversos parâmetros. Além disso, suporta uma ampla diversidade de extensões abrindo o caminho para o uso de ferramentas como ControlNET e LoRAs.

6.6 Dreambooth

Dreambooth [63] é técnica para o treino para os modelos de difusão que permite a inserção de um novo conceito ao modelo através de um treino com foco na geração text-to-image onde, através das imagens fornecidas, com a devida descrição dos seus conteúdos, torna possível ensinar qualquer novo estilo ou termo desejado. No desenvolvimento deste projeto, juntamente com o Google Colab foi permitido fazer o treino do estilo do bordado de Castelo Branco.

6.7 GitHub

O GitHub [64] é um sistema de alojamento de código e controlo de versões, que, através do Git, permite guardar e armazenar alterações feitas no código e, posteriormente, o seu alojamento online. A sua capacidade em criar diferentes ramificações e registo de bugs torna esta uma plataforma versátil para os desenvolvedores que procuram suporte às suas criações.

7. Resultados Obtidos

Neste capítulo pretendemos documentar todo o processo do trabalho experimental realizado, desde o treino do modelo, até às escolhas de parâmetros nas gerações de imagens, assim como uma pequena demonstração de resultados, através das duas principais formas de utilização destes modelos, a geração de imagens através de text-to-image e com a transferência de estilo a ser aplicada via image-to-image.

7.1 Treino Realizado

Para a fase de treino optámos por utilizar um notebook, disponibilizado pela Dreambooth, na plataforma Google Colab. Apesar de ser um notebook já antigo, e não suportado, demonstrou-se ser das únicas opções que conseguia de maneira eficaz treinar um modelo de Stable Diffusion pré treinado, um novo elemento estilístico do qual passamos a demonstrar a figura abaixo.

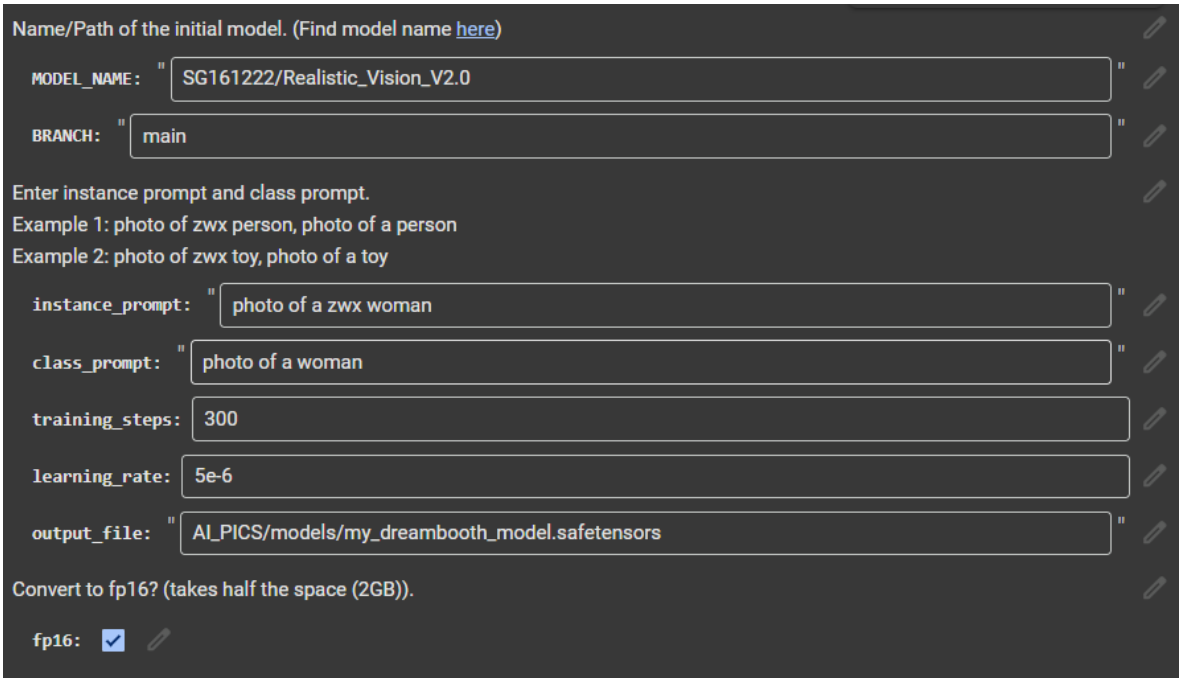

Figura 29 – Parâmetros do notebook de treino da Dreambooth

Estes parâmetros permitem configurar diversas definições do processo de treino do modelo tais como: Escolher o modelo base de Stable Diffusion que queremos utilizar através de MODEL_NAME, este campo irá receber o nome do modelo com base na plataforma HuggingFace, permitindo assim escolher um qualquer modelo disponível dentro da dada plataforma. Em BRANCH é nos permitida escolher uma versão do modelo dentro da plataforma, em quase todos os casos, o uso do branch main é o recomendado, por se tratar da última versão estável.

Com instance_prompt é onde nos permite explicar ao modelo o conceito que vai aprender através de uma pequena descrição, é aqui que palavras-chave como “Bordado de Castelo Branco” vão ser utilizadas, permitindo assim ao modelo ganhar a capacidade de identificar e gerar imagens do bordado. Já em class_promt, permite associar a instance_prompt uma classe mais geral, que ajuda na preservação das características gerais enquanto aprende um conceito mais específico, no nosso caso, utilizámos como instance_prompt = “Bordado de Castelo Branco” e class_prompt = “Bordado”. Na figura a baixo é possível verificar os parâmetros que utilizámos.

Figura 30 — Parâmetros de treino

Após estabelecer conexão com a máquina virtual encarregue de realizar o nosso treino, é nos permitido escolher os ficheiros das imagens com as quais iremos fazer o treino.

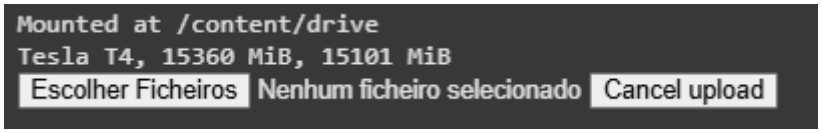

Figura 31 — Escolha das imagens de treino

Após escolher as imagens de treino começa o processo de upload às fotos.

Figura 32 — Upload das imagens de treino

Com este notebook, serão utilizadas técnicas do DreamBooth para realizar o fine-tunning de um modelo pré-treinado através da associação de um token textual único. Este, irá representar o novo tema a ser treinado, neste caso, o bordado de

Castelo Branco. Esta técnica permite ao modelo aprender a gerar imagens coerentes com base no novo conceito, mas sem comprometer a sua capacidade de geração de imagens prévia. Esta abordagem é especialmente importante, pois evita a catástrofe do esquecimento, um problema recorrente em redes neuronais profundas, onde a modificação dos pesos pode levar à perda da capacidade de gerar imagens de outros tópicos. Desta forma, ao contrário de outros métodos de fine-tunning, o overfitting é evitado, perimindo assim ao modelo manter a sua flexibilidade criativa e generalização [65].

Após o processo de upload e treino for concluído o ficheiro que origina do notebook vai ser enviado para o Google Drive, em forma de .safetensors, possibilitando assim o seu uso em aplicações como o A1111, ferramenta escolhida como demonstra a figura a baixo.

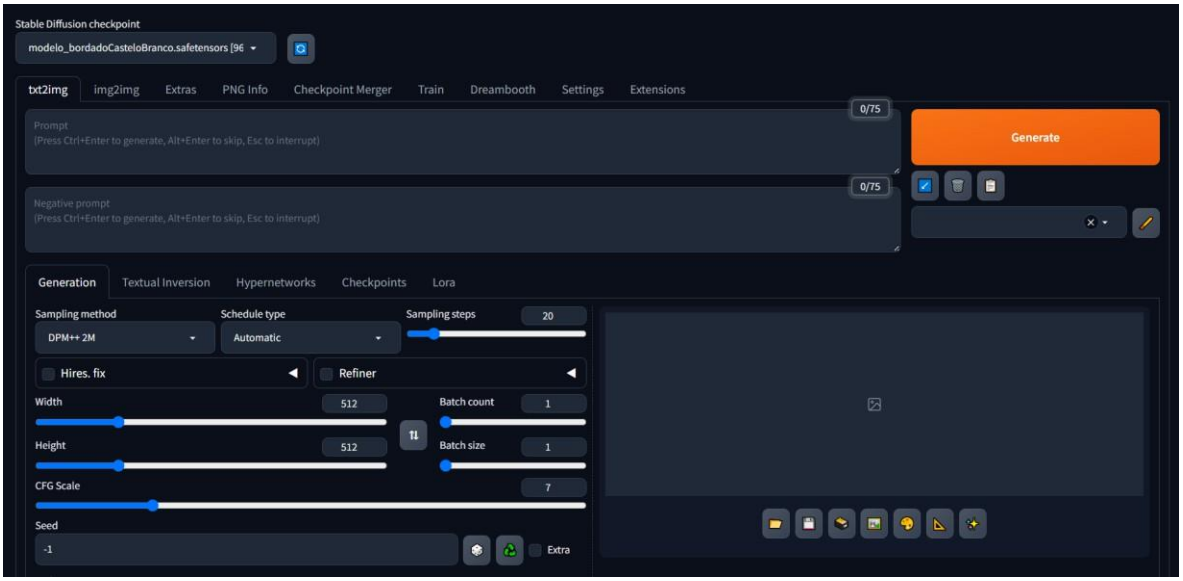

Figura 33 — Interface gráfica do A1111

Vai ser a partir desta interface gráfica que a geração das imagens vai ser feita, dentro desta interface temos a possibilidade de modificar alguns parâmetros do quais passamos a explorar:

Parâmetros Básicos – Estes são os parâmetros mais simples, que não alteram de maneira significativa o resultado e apenas têm como função alterar as propriedades básicas das imagens geradas. São estes: Width (Largura), Height (Altura), Batch Count (número de batches de imagens que vão ser gerados) e Batch Size (número de imagens geradas por cada batch).

Parâmetros Avançados — Estes são os parâmetros que vão ser utilizados pelo modelo, influenciando assim diretamente o resultado. Sendo eles: O modelo em si (é possível selecionar diferentes modelos no canto superior esquerdo da figura

anterior), Prompts Negativos e Positivos (permitem modificação dos resultados a partir de prompts textuais, sendo os positivos as características a gerar, e os negativos características a evitar). Sampling Method (este parâmetro vai definir qual o algoritmo que vai ser utilizado para o processo de remoção de ruído sendo o DPM ++ 2 Karras o recomendado [66]). Já Schedule Type e Sample Steps controlam o processo oposto, permitindo ao utilizador alterar a forma com que a adição de ruído é feita [67]. Hires e Refiner são extensões adicionais do A1111 que permitem o upscaling das imagens e possibilitar a criação das mesmas com múltiplos modelos em simultâneo, respetivamente. CFG Scale representa o peso dos prompts textuais, onde valores altos, como 30 tornam o modelo mais restrito às características dos prompts, e valores mais baixos como 5 dão mais liberdade e criatividade ao modelo. Por fim, seed, permite manter a consistência dos resultados entre execuções, caso seja escolhida uma seed fixa os outputs serão sempre os mesmos enquanto uma seed com um valor de -1 cria resultados novos com cada execução.

7.2 Escolha dos prompts

Nesta secção, pretendemos justificar a escolha dos prompts utilizados. Esta é uma decisão fundamental para a qualidade dos resultados, pois os prompts têm um grande impacto na forma como o modelo gera as imagens. Assim, uma seleção adequada é essencial para a obtenção de imagens de qualidade.

Os prompts positivos escolhidos foram: “Foto do Bordado de Castelo Branco”. Este é um prompt expectável, uma vez que o treino foi realizado com base no termo “Bordado de Castelo Branco”, tornando-o necessário. Os prompts “têxtil”, “feito à mão” e “alto detalhe” têm como objetivo melhorar a qualidade da imagem e transmitir a minuciosidade do bordado nas imagens geradas. Já “simétrico” e “fundo branco” referem-se a características específicas do Bordado de Castelo Branco, garantindo que estas estejam presentes nas imagens.

Quanto aos prompts negativos, foram escolhidos “desfigurado”, “feio”, “mau”, “imaturo” e “anime” como uma série de descritores genéricos para evitar resultados indesejados. Os termos “3D”, “preto e branco” e “fundo negro” foram incluídos para corrigir algumas limitações do treino, garantindo que os resultados sejam coloridos e num fundo branco. Adicionámos ainda “arquitetura” e “castelos” para mitigar um erro do modelo, que, devido à presença do termo “Castelo Branco” em “Bordado de Castelo Branco”, tendia a inserir um castelo na cor branca em várias imagens.

7.3 Exemplos de Imagens geradas através de text-to-image

Neste subcapítulo, pretendemos demonstrar as capacidades do modelo treinado num registo text-to-image, onde serão enviados dois prompts textuais, e o modelo gerará uma imagem com base nestes. Serão apresentadas dez imagens geradas com os mesmos prompts e parâmetros.

Figura 34 – Primeira imagem gerada

Figura 35 – Segunda imagem gerada

Figura 36 – Terceira Imagem gerada

Figura 37 – Quarta Imagem gerada

Figura 38 – Quinta Imagem gerada

Figura 39 – Sexta Imagem gerada

Figura 40 – Sétima Imagem gerada

Figura 41 – Oitava Imagem gerada

Figura 42 – Nona Imagem gerada

Figura 43 – Décima Imagem gerada

```
Foto do bordado de Castelo Branco, têxtil, feito à mão, alto detalhe, simétrico, fundo branco
Negative prompt: Desfigurado, feio, mau, imaturo, anime, 3D, preto e branco, fundo negro, arquitetura, castelos
Steps: 20, Sampler: DPM++ 2M, Schedule type: Karras, CFG scale: 7, Seed: 1317325874, Size: 512x512, Model hash: 963c5aa9aa, Model:
modelo_bordadoCasteloBranco, Version: v1.10.1
```

Figura 44 – Parâmetros para as imagens geradas anteriormente

7.4 Exemplos de Imagens geradas através de image-to-image

Neste subcapítulo pretendemos demonstrar as capacidades do modelo treinado, agora num registo image-to-image, com a demonstração de dez exemplos onde vão ser enviados dois prompts textuais (iguais aos anteriores) juntamente com uma imagem a transformar e o modelo vai gerar uma imagem com base nestes. Nestes 10 exemplos, 6 deles são de imagens genéricas para teste, com elementos mais simples e com menor complexidade, enquanto as últimas 4 são de fotos tiradas por um dos autores com o objetivo de testar uma utilização comum com fotos reais.

Figura 45 — Transferência de estilo com uma flor

Figura 46 — Transferência de estilo com uma árvore

Figura 47 — Transferência de estilo com uma paisagem

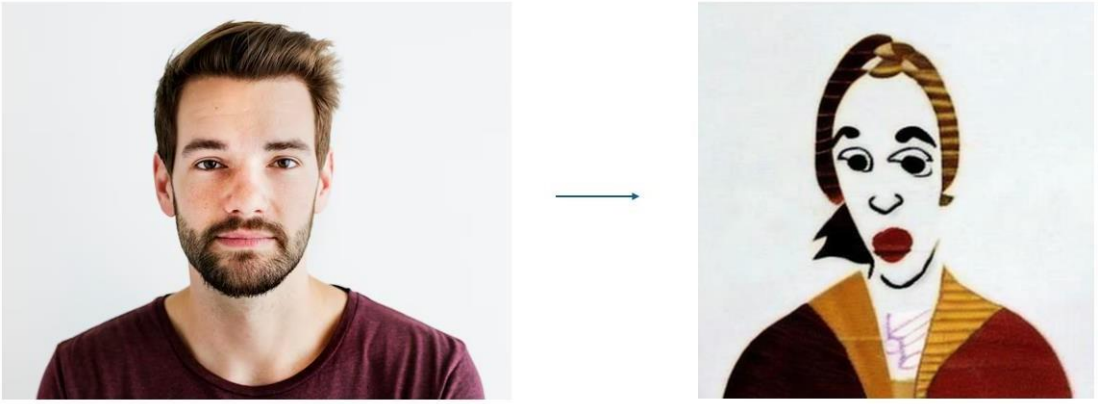

Figura 48 — Transferência de estilo com um rosto

Figura 49 — Transferência de estilo com uma casa

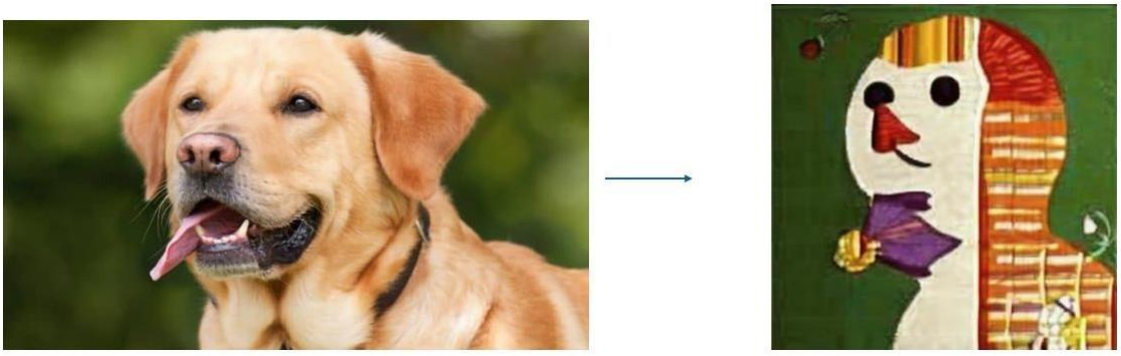

Figura 50 — Transferência de estilo com um cão

Figura 51 — Transferência de estilo com um camelo

Figura 52 — Transferência de estilo com uma águia

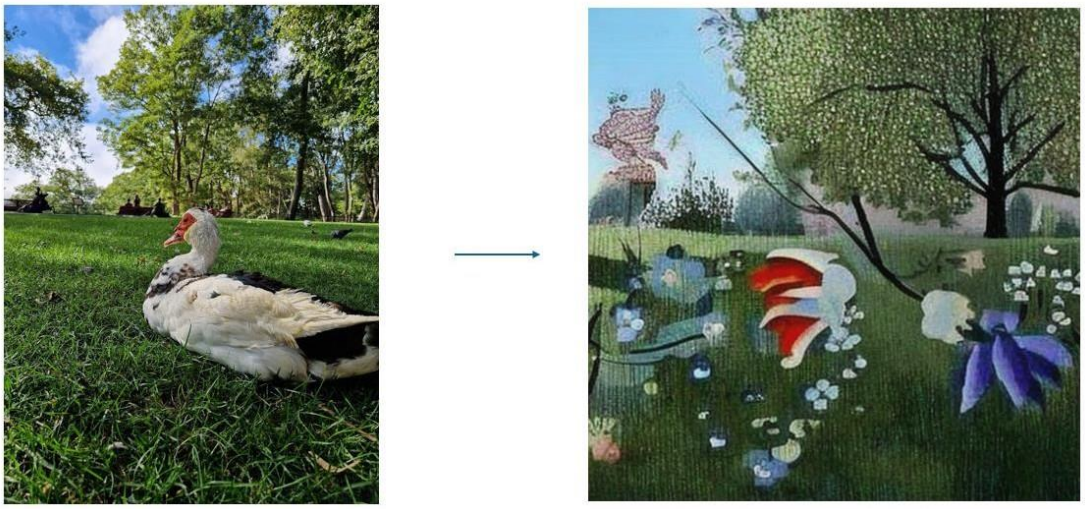

Figura 53 — Transferência de estilo com um ganso

Figura 54 — Transferência de estilo com uma chita

7.5 Avaliação dos Resultados Obtidos

Com base nas imagens obtidas no subcapítulo anterior, podemos concluir que, embora os resultados sejam interessantes, não coincidem totalmente com a natureza do Bordado de Castelo Branco. Este problema é mais evidente no registo img-to-img, onde o modelo apresenta dificuldades em distinguir o conteúdo original da imagem dos elementos estilísticos do bordado. Apesar disso, é possível identificar características que remetem à sua estrutura, como os cravos e flores, apesar de, muitas vezes, desfigurados.

No registo text-to-image, os resultados foram consistentes, apesar de algumas falhas na clareza dos elementos gerados. Destacam-se a Figura 37 e a Figura 42, que apresentaram melhor alinhamento com o estilo esperado. No entanto, foi notada a inserção de tons de azul que não pertencem à estrutura tradicional do Bordado de Castelo Branco.

Entre os exemplos analisados num registo img-to-img, destacam-se a árvore (Figura 46) e o rosto de um homem (Figura 48), que apresentaram os melhores resultados na distinção entre o estilo e o conteúdo original. Estas imagens demonstram um equilíbrio eficaz entre os elementos estilísticos e a preservação da forma original, resultando em imagens visualmente mais apelativas.

No entanto, ao utilizar fotografias complexas, observou-se um aumento significativo na dificuldade do modelo em realizar a transferência de estilo. Nos quatro últimos exemplos, que incluem fotos reais, o modelo teve dificuldades em reconhecer a silhueta dos objetos, o que levou a resultados distantes das imagens originais. Este problema sugere que um refinamento no dataset ou no ajuste dos parâmetros pode ser necessário para melhorar os resultados futuros.

8. Conclusão

Com o trabalho aqui documentado, foi possível explorar a aplicação de técnicas de inteligência artificial, nomeadamente, modelos de difusão para efetuar transferências de estilo, tendo como base o bordado de Castelo Branco. Esta abordagem permitiu investigar a interseção entre a tecnologia e a arte tradicional, utilizando como base as tecnologias generativas mais recentes.

Numa primeira abordagem, foram analisados e estudados os fundamentos teóricos da inteligência artificial aplicada à geração de imagens, através da análise do funcionamento da IA generativa e das suas potencialidades para a transferência de estilo. Paralelamente, realizou-se uma revisão do estado da arte relacionada com o uso de modelos de difusão para a estilização de imagens. Esta análise evidenciou que os modelos baseados em Stable Diffusion são superiores às abordagens utilizadas até à data, apresentando resultados convincentes em múltiplas métricas e, sobretudo, uma elevada capacidade para aprender estilos artísticos com pequenas quantidades de dados para treino.

O trabalho prosseguiu com o treino de um modelo pré-treinado, no qual foi possível ensinar um novo termo através de técnicas como o DreamBooth, permitindo dar ao modelo a capacidade reconhecer e reproduzir o estilo do bordado de Castelo Branco a partir de um conjunto de dados criado com múltiplas imagens e recortes do bordado. Nesta etapa de criação do dataset, enfrentámos algumas dificuldades, nomeadamente pelo escasso número de imagens de qualidade, o que nos obrigou a recorrer a espelhamentos e recortes nas melhores imagens, podendo isto ter prejudicado o desempenho do modelo. Assim, após o treino, o objetivo consistiu em fazer com que o modelo absorvesse os elementos característicos do bordado, como as suas simetrias e padrões florais, embora tenham surgido novamente algumas dificuldades, especialmente na separação entre o conteúdo das imagens e os efeitos estilísticos nas imagens geradas em cenários de transferência de estilo.

Com o modelo já treinado, passámos à análise dos resultados. Verificámos que as imagens criadas através do método text-to-img apresentavam maior consistência e uma semântica mais próxima da original do bordado, embora existissem problemas com a clareza entre os símbolos e o uso incorreto de algumas cores que não coincidiam com a estrutura do bordado de Castelo Branco. Numa abordagem diferente, através do método img-to-img, o modelo apresentou dificuldades, sobretudo devido à baixa qualidade na deteção das silhuetas nas

imagens fornecidas, o que, por vezes, resultou na geração de imagens com pouca ou nenhuma semelhança com o input.

Apesar das dificuldades enfrentadas, os resultados permanecem interessantes, por sua vez abrindo espaço para melhorias em etapas futuras. A principal conclusão a retirar deste estudo é que, embora o modelo consiga aprender e aplicar o estilo do bordado de Castelo Branco, há ainda grandes espaços para melhorias, especialmente no que se refere à qualidade e precisão na geração das imagens. Melhorias no dataset utilizado para treino e ajustes nos parâmetros do modelo e treino serão os próximos passos para alcançar uma melhor transferência de estilos.

8.1 Trabalho Futuro

Como referido anteriormente, melhorar a qualidade do conjunto de dados de treino e ajustar os parâmetros do modelo constituem os próximos passos a implementar. Nesta fase, é fundamental avaliar a possibilidade de recorrer a modelos pré-treinados mais avançados, bem como de incorporar técnicas de treino mais atualizadas, de modo a melhorar a eficácia na transferência de estilos.

Referências

[1] Y. Zuo *et al.*, «Towards Multi-View Consistent Style Transfer with One-Step Diffusion via Vision Conditioning», Nov. 2024, [Em linha]. Disponível em: http://arxiv.org/abs/2411.10130

[2] L. A. Gatys, M. Bethge, A. Hertzmann, e E. Shechtman, «Preserving Color in Neural Artistic Style Transfer», Jun. 2016, [Em linha]. Disponível em: http://arxiv.org/abs/1606.05897

[3] J. Huang, M. Yan, Y. Liu, e S. Chen, «Color-SD: Stable Diffusion Model Already has a Color Style Noisy Latent Space», em *Proceedings - IEEE International Conference on Multimedia and Expo*, IEEE Computer Society, 2024. doi: 10.1109/ICME57554.2024.10687653.

[4] A. T. Carceller, «The ARTificial Revolution: Challenges for redefining Art Education in the paradigm of generative artificial intelligence», *Digital Education Review*, n. 45, pp. 84–90, Jul. 2024, doi: 10.1344/der.2024.45.84-90.

[5] M. Kaur, A. K. Shukla, e S. Kaur, «An Introduction to Machine Learning in a Nutshell», em *Proceedings of the 2021 10th International Conference on System Modeling and Advancement in Research Trends, SMART 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 17–22. doi: 10.1109/SMART52563.2021.9676315.

[6] Z. Qiu, H. Zhao, e S. Wang, «Applications and Challenges of Artificial Intelligence in Aerospace Engineering», em *2023 6th International Conference on Artificial Intelligence and Big Data, ICAIBD 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 970–974. doi: 10.1109/ICAIBD57115.2023.10206205.

[7] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, e B. Ommer, «High-Resolution Image Synthesis with Latent Diffusion Models», em *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, 2022, pp. 10674–10685. doi: 10.1109/CVPR52688.2022.01042.

[8] A. Horvath, «Stable Diffusion with Continuous-time Neural Networks», Institute of Electrical and Electronics Engineers (IEEE), Set. 2024, pp. 1–4. doi: 10.1109/cnna60945.2023.10652658.

[9] «Artificial Intelligence's Use and Rapid Growth Highlight Its Possibilities and Perils». Acedido: 7 de Dezembro de 2024. [Em linha]. Disponível em: https://www.gao.gov/blog/artificial-intelligences-use-and-rapid-growth-highlight-its-possibilities-and-perils

[10] B. Liu, «"Weak AI" is Likely to Never Become "Strong AI", So What is its Greatest Value for us?», Mar. 2021, [Em linha]. Disponível em: http://arxiv.org/abs/2103.15294

[11] P. Bory, S. Natale, e C. Katzenbach, «Strong and weak AI narratives: an analytical framework», *AI Soc*, 2024, doi: 10.1007/s00146-024-02087-8.

[12] B. J. Copeland, «Early AI in Britain: Turing et al.», *IEEE Annals of the History of Computing*, vol. 45, n. 3, pp. 19–31, Jul. 2023, doi: 10.1109/MAHC.2023.3300660.

[13] L. Li, N. N. Zheng, e F. Y. Wang, «On the Crossroad of Artificial Intelligence: A Revisit to Alan Turing and Norbert Wiener», *IEEE Trans Cybern*, vol. 49, n. 10, pp. 3618–3626, Out. 2019, doi: 10.1109/TCYB.2018.2884315.

[14] L. Cao, «A New Age of AI: Features and Futures», *IEEE Intell Syst*, vol. 37, n. 1, pp. 25–37, 2022, doi: 10.1109/MIS.2022.3150944.

[15] S. K. Puli e P. Usha, «Transforming Healthcare: Advancements, Applications, and Future Directions of Machine Learning», em *2024 10th International Conference on Smart Computing and Communication, ICSCC 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 502–506. doi: 10.1109/ICSCC62041.2024.10690530.

[16] . IEEE Staff, *2009 International Conference on Information Engineering and Computer Science*. I E E E, 2009.

[17] «Machine Learning vs Traditional Programming». Acedido: 22 de Janeiro de 2025. [Em linha]. Disponível em: https://www.avenga.com/magazine/machine-learning-programming/

[18] «Types of Machine Learning». Acedido: 22 de Janeiro de 2025. [Em linha]. Disponível em: https://www.javatpoint.com/types-of-machine-learning

[19] C. Enyinna Nwankpa, W. Ijomah, A. Gachagan, e S. Marshall, «Activation Functions: Comparison of Trends in Practice and Research for Deep Learning».

[20] D. P. Kingma e J. Ba, «Adam: A Method for Stochastic Optimization», Dez. 2014, [Em linha]. Disponível em: http://arxiv.org/abs/1412.6980

[21] «What is a neural network?» Acedido: 25 de Janeiro de 2025. [Em linha]. Disponível em: https://www.cloudflare.com/learning/ai/what-is-neural-network/

[22] D. H. Hagos, R. Battle, e D. B. Rawat, «Recent Advances in Generative AI and Large Language Models: Current Status, Challenges, and Perspectives», *IEEE Transactions on Artificial Intelligence*, 2024, doi: 10.1109/TAI.2024.3444742.

[23] T. Young, D. Hazarika, S. Poria, e E. Cambria, «Recent trends in deep learning based natural language processing [Review Article]», 1 de Agosto de 2018, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/MCI.2018.2840738.

[24] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, e S. Ganguli, «Deep Unsupervised Learning using Nonequilibrium Thermodynamics», Mar. 2015, [Em linha]. Disponível em: http://arxiv.org/abs/1503.03585

[25] «Review: Denoising Diffusion Probabilistic Models (DDPM)». Acedido: 30 de Janeiro de 2025. [Em linha]. Disponível em: https://andycheungyatming.github.io/2022/12/21/Review-DDPM/

[26] «An Introduction to Diffusion Models and Stable Diffusion». Acedido: 30 de Janeiro de 2025. [Em linha]. Disponível em: https://blog.marvik.ai/2023/11/28/an-introduction-to-diffusion-models-and-stable-diffusion/

[27] Y. Liu, J. Yue, S. Xia, P. Ghamisi, W. Xie, e L. Fang, «Diffusion Models Meet Remote Sensing: Principles, Methods, and Perspectives», Abr. 2024, doi: 10.1109/TGRS.2024.3464685.

[28] I. J. Goodfellow *et al.*, «Generative Adversarial Networks», Jun. 2014, [Em linha]. Disponível em: http://arxiv.org/abs/1406.2661

[29] D. P. Kingma e M. Welling, «Auto-Encoding Variational Bayes», Dez. 2013, [Em linha]. Disponível em: http://arxiv.org/abs/1312.6114

[30] L. Dinh, D. Krueger, e Y. Bengio, «NICE: Non-linear Independent Components Estimation», Out. 2014, [Em linha]. Disponível em: http://arxiv.org/abs/1410.8516

[31] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, e B. Ommer, «High-Resolution Image Synthesis with Latent Diffusion Models». [Em linha]. Disponível em: https://github.com/CompVis/latent-diffusion

[32] «What is a variational autoencoder?» Acedido: 30 de Janeiro de 2025. [Em linha]. Disponível em: https://www.ibm.com/think/topics/variational-autoencoder

[33] R. Azad *et al.*, «Medical Image Segmentation Review: The Success of U-Net», *IEEE Trans Pattern Anal Mach Intell*, 2024, doi: 10.1109/TPAMI.2024.3435571.

[34] O. Ronneberger, P. Fischer, e T. Brox, «U-Net: Convolutional Networks for Biomedical Image Segmentation», Mai. 2015, [Em linha]. Disponível em: http://arxiv.org/abs/1505.04597

[35] A. Radford *et al.*, «Learning Transferable Visual Models From Natural Language Supervision», Fev. 2021, [Em linha]. Disponível em: http://arxiv.org/abs/2103.00020

[36] «Training A Diffusion Model - Stable Diffusion Masterclass».

[37] «The state of AI in early 2024: Gen AI adoption spikes and starts to generate value», 2024.

[38] «IEEE Xplore». Acedido: 7 de Dezembro de 2024. [Em linha]. Disponível em: https://ieeexplore.ieee.org/Xplore/home.jsp

[39] «B-On». Acedido: 7 de Dezembro de 2024. [Em linha]. Disponível em: https://www.b-on.pt/

[40] H. Wu, S. Lim, e B. Xiao, «Animated Avatar Generation Technology Research Based on Deep Convolutional Generative Adversarial Network integrated with Self-attention and Spectral Normalization», *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3482989.

[41] G. Wu, X. Liu, J. Jia, X. Cui, e G. Zhai, «Text2QR: Harmonizing Aesthetic Customization and Scanning Robustness for Text-Guided QR Code Generation», em *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2024, pp. 8456–8465. doi: 10.1109/CVPR52733.2024.00808.

[42] M. N. Everaert, M. Bocchio, S. Arpa, S. Susstrunk, e R. Achanta, «Diffusion in Style», em *Proceedings of the IEEE International Conference on Computer Vision*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 2251–2261. doi: 10.1109/ICCV51070.2023.00214.

[43] Y. D. Rueda-Arango, D. Rojas-Velazquez, A. V. Gorelova, J. Garssen, A. Tonda, e A. Lopez-Rincon, «Image Generation with Interactive Evolutionary System using Bayesian Optimization», em *International Conference on Human System Interaction, HSI*, IEEE Computer Society, 2024. doi: 10.1109/HSI61632.2024.10613596.

[44] Y. Xu, X. Xu, H. Gao, e F. Xiao, «SGDM: An Adaptive Style-Guided Diffusion Model for Personalized Text to Image Generation», *IEEE Trans Multimedia*, 2024, doi: 10.1109/TMM.2024.3399075.

[45] M. Hamazaspyan e S. Navasardyan, «Diffusion-Enhanced PatchMatch: A Framework for Arbitrary Style Transfer with Diffusion Models», em *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, IEEE Computer Society, 2023, pp. 797–805. doi: 10.1109/CVPRW59228.2023.00087.

[46] L. Wang *et al.*, «GlyphGenius: Unleashing the potential of AIGC in Chinese character learning», *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3464562.

[47] Z. K. J. Hartley, R. J. Lind, M. P. Pound, e A. P. French, «Domain Targeted Synthetic Plant Style Transfer using Stable Diffusion, LoRA and ControlNet», em *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, Jun. 2024, pp. 5375–5383. doi: 10.1109/CVPRW63382.2024.00546.

[48] J. Huang, Y. Gao, Z. Jie, Y. Zhong, X. Han, e L. Ma, «MRStyle: A Unified Framework for Color Style Transfer with Multi-Modality Reference», Set. 2024, [Em linha]. Disponível em: http://arxiv.org/abs/2409.05250

[49] Z. Zhang *et al.*, «Towards Highly Realistic Artistic Style Transfer via Stable Diffusion with Step-aware and Layer-aware Prompt», Abr. 2024, [Em linha]. Disponível em: http://arxiv.org/abs/2404.11474

[50] S. Li, «DiffStyler: Diffusion-based Localized Image Style Transfer», Mar. 2024, [Em linha]. Disponível em: http://arxiv.org/abs/2403.18461

[51] J. Liao, «A Study on Neural Style Transfer Methods for Images», em *Proceedings - 2022 2nd International Conference on Big Data, Artificial Intelligence and Risk Management, ICBAR 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 60–64. doi: 10.1109/ICBAR58199.2022.00019.

[52] M. Bogacz e M. Iwanowski, «Towards many-to-one neural style transfer method», em *International Conference on Human System Interaction, HSI*, IEEE Computer Society, Jul. 2021. doi: 10.1109/HSI52170.2021.9538780.

[53] «Non-Photorealistic Rendering Pen-and-ink Illustrations Painterly Rendering Cartoon Shading Technical Illustrations».

[54] «How to style transfer your own images». Acedido: 28 de Janeiro de 2025. [Em linha]. Disponível em: https://xebia.com/blog/how-to-style-transfer-your-own-images/

[55] «História do bordado de Castelo Branco». Acedido: 7 de Janeiro de 2025. [Em linha]. Disponível em: https://pt.wikipedia.org/wiki/Bordado_de_Castelo_Branco

[56] «Bordado de Castelo Branco». Acedido: 7 de Janeiro de 2025. [Em linha]. Disponível em: https://www.cm-castelobranco.pt/municipe/noticias/detalhe-noticia/?id=18993

[57] «Google Custom Search JSON API». Acedido: 20 de Janeiro de 2025. [Em linha]. Disponível em: https://developers.google.com/custom-search/v1/overview?hl=pt-br

[58] «Python».

[59] «Google Colab». Acedido: 29 de Janeiro de 2025. [Em linha]. Disponível em: https://colab.research.google.com/

[60] «Hugging Face». Acedido: 29 de Janeiro de 2025. [Em linha]. Disponível em: https://huggingface.co/

[61] «Brime». Acedido: 29 de Janeiro de 2025. [Em linha]. Disponível em: https://www.birme.net/

[62] «A1111». Acedido: 29 de Janeiro de 2025. [Em linha]. Disponível em: https://github.com/AUTOMATIC1111/stable-diffusion-webui

[63] «Dreambooth». Acedido: 29 de Janeiro de 2025. [Em linha]. Disponível em: https://dreambooth.github.io/

[64] «GitHub». Acedido: 30 de Janeiro de 2025. [Em linha]. Disponível em: https://github.com/

[65] «DreamBooth». Acedido: 1 de Fevereiro de 2025. [Em linha]. Disponível em: https://huggingface.co/docs/diffusers/training/dreambooth

[66] «Know these Important Parameters for stunning AI images». Acedido: 26 de Janeiro de 2025. [Em linha]. Disponível em: https://stable-diffusion-art.com/know-these-important-parameters-for-stunning-ai-images/#Sampling_methods

[67] «Basic usage of Stable Diffusion web UI (v1.9.0) Text-to-image section». Acedido: 26 de Janeiro de 2025. [Em linha]. Disponível em: https://www.digitalcreativeai.net/en/post/how-to-use-stable-diffusion-web-ui-text-to-image