

Sistema FIEB



PELO FUTURO DA INOVAÇÃO

CENTRO UNIVERSITÁRIO SENAI CIMATEC
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
COMPUTACIONAL E TECNOLOGIA INDUSTRIAL
Mestrado em Modelagem Computacional e Tecnologia Industrial

Victor Rocha Santos

REDES NEURAIS COM CAMADAS COMPLEXAS PARA
APRIMORAMENTO DE IMAGENS DIGITALIZADAS QUE
APRESENTAM DEFEITOS

Novembro, 2021

Salvador, BA

Victor Rocha Santos

**Redes neurais com camadas complexas para aprimoramento de
imagens digitalizadas que apresentam defeitos**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação Stricto Sensu do Centro Universitário SENAI CIMATEC como requisito parcial para a obtenção do título de Mestre em Modelagem Computacional e Tecnologia Industrial sob orientação do Prof. Dr. Erick Giovani Sperandio Nascimento e da Profa. Dra. Ingrid Winkler.

Centro Universitário SENAI CIMATEC
Programa de Pós-graduação Stricto Sensu MCTI
Mestrado em Modelagem Computacional e Tecnologia Industrial

Novembro, 2021
Salvador, BA

Ficha catalográfica elaborada pela Biblioteca do Centro Universitário SENAI CIMATEC

S237r Santos, Victor Rocha

Redes neurais com camadas complexas para aprimoramento de imagens digitalizadas que apresentam defeitos / Victor Rocha Santos. – Salvador, 2021.

68 f. : il. color.

Orientador: Prof. Dr. Erick Giovani Sperandio Nascimento.

Orientadora: Prof^a. Dr^a. Ingrid Winkler.

Dissertação (Mestrado em Modelagem Computacional e Tecnologia Industrial) – Programa de Pós-Graduação, Centro Universitário SENAI CIMATEC, Salvador, 2021.

Inclui referências.

1. Aprendizagem de máquina profunda. 2. Melhoria da qualidade de imagens. 3. Remoção de ruídos em imagens. 4. Redes adversárias generativas. 5. Números complexos. I. Centro Universitário SENAI CIMATEC. II. Nascimento, Erick Giovani Sperandio. III. Winkler, Ingrid. IV. Título.

CDD 006.32

CENTRO UNIVERSITÁRIO SENAI CIMATEC**Mestrado Acadêmico em Modelagem Computacional e Tecnologia Industrial**

A Banca Examinadora, constituída pelos professores abaixo listados, aprova a Defesa de Mestrado, intitulada “**Redes neurais com camadas complexas para aprimoramento de imagens digitalizadas que apresentam defeitos**” apresentada no dia 01 de dezembro de 2021, como parte dos requisitos necessários para a obtenção do Título de Mestre em Modelagem Computacional e Tecnologia Industrial.

Orientador:

DocuSigned by:



FA8177266C9F451...

Prof. Dr. Erick Giovani Sperandio Nascimento
SENAI CIMATEC

Coorientadora:

DocuSigned by:



EB83F32E864047D...

Prof.^a Dr.^a Ingrid Winkler
SENAI CIMATEC

Membro Interno:

DocuSigned by:



F4FD1E982FE04A...

Prof. Dr. Valter de Senna
SENAI CIMATEC

Membro Externo:

DocuSigned by:



78F6C0D8D8E4343A...

Prof. Dr. Claudio Fabiano Motta Toledo
USP

Membro Externo:

DocuSigned by:



E7184C86D391407...

Prof. Dr. Ewerton Lopes Silva de Oliveira
HP INC

Agradecimentos

Este trabalho não seria possível sem a colaboração de muitas pessoas. Gostaria de agradecer primeiramente a Deus, agradecer a meus pais, por terem feito tudo possível para que eu pudesse sentir de tudo, trilhar caminhos, voar para longe, sabendo que sempre teria suporte.

Gostaria de agradecer também aos meus orientadores, Prof. Dr. Erick Sperandio e a Profa. Dra. Ingrid Winkler pela paciência, disponibilidade e parceria e a banca avaliadora pela disponibilidade e contribuições.

Agradeço ao SENAI CIMATEC e a HP pelo auxílio financeiro da bolsa de formação e desenvolvimento tecnológico concedida. Agradeço aos integrantes do grupo de pesquisa VIALAB, em especial aos participantes do desafio 01, Lucas Ortega e José Vinícius que tiveram grandes contribuições no debate do tema.

Também gostaria de agradecer a todos os meus familiares e amigos que me auxiliaram direta ou indiretamente para a conclusão deste trabalho, em especial ao meu amigo Rosalvo Matos pelas inúmeras vezes que esteve próximo me apoiando em busca da resolução dos problemas.

Por fim, agradeço aos demais estudantes e funcionários do SENAI CIMATEC, desejando sucesso na conclusão das suas pesquisas.

Resumo

Nos últimos anos, muitas formas de digitalização foram propostas com destaque para a digitalização por imagem, que foi popularizada com o advento de smartphones com câmeras de alta qualidade. Entretanto, variados fatores podem acarretar em imagens defeituosas de baixa qualidade. Nesse contexto, este trabalho tem como objetivo propor uma técnica baseada em aprendizagem de máquina profunda, denominada ComPlix2Denoise, para realizar o aprimoramento de imagens digitalizadas com defeitos de sombra, reflexo de luz, papel amassado, baixa luminosidade e mudança de temperatura. Para tal, após uma revisão da literatura sobre métodos de melhoria de imagens com os defeitos em questão, foi proposto um modelo de rede adversária generativa (*Generative Adversarial Network* - *GAN*) que utiliza camadas convolucionais com números complexos no discriminador. O modelo proposto se caracteriza sendo um dos trabalhos pioneiros na utilização de redes com camadas complexas para a remoção de defeitos. Também foi criado um *dataset* com dados reais e artificiais para preencher a lacuna causada pela falta de *datasets* que representem os defeitos estudados. Por fim, algumas técnicas do estado arte para remoção de defeitos foram testadas e os resultados comparados. Os resultados obtidos apontaram que a técnica proposta conseguiu atenuar alguns defeitos com intensidades variadas e, em alguns casos, removê-los, e que o *dataset* desenvolvido para auxiliar na solução do problema também representa grande contribuição na área de remoção de defeitos em imagens, pois, apesar das redes terem sido treinadas no *dataset* artificial, conseguiram bom desempenho quando testadas em imagens com defeitos reais. Conclui-se que estes experimentos iniciais demonstram que a técnica de aprendizagem profunda de máquina é bastante promissora para a resolução de alguns dos defeitos estudados, como baixa luminosidade, mudança de temperatura e papel amassado, e que avanços significativos foram alcançados na melhoria da qualidade das imagens, apesar de algumas limitações encontradas nos casos de imagens que possuem defeitos de reflexo de luz.

Palavras-chaves: Aprendizagem de máquina profunda, Melhoria da qualidade de imagens, Remoção de ruídos em imagens, Redes Adversárias Generativas, Números complexos.

Abstract

Millions of photos are captured and scanned daily in the world, and several factors can influence the quality of an image, causing distortions. Research has shown that there are several ways to remove defects from images. In this context, this work aims to propose a technique based on deep machine learning, to enhance the scanned images with shadow defects, light reflection, crumpled paper, low light and temperature change. To this end, after a review of the literature on methods of improving images with the defects in question, a model of generative adversarial network (GAN) was proposed that uses convolutional layers with complex numbers in the discriminator, being one of the pioneering works in the use of networks with complex-valued layers for the image denoising. In addition to a dataset with real and artificial data to fill the gap caused by the lack of datasets that represent the studied defects. Finally, some techniques from the state of the art for defect removal were tested and the results compared. The results obtained showed that the proposed technique was able to mitigate some defects with varying intensities and, in some cases, remove them, and that the dataset developed to assist in the solution of the problem also represents a major contribution in the the image denoising area, because, despite the networks have been trained in artificial data, they have achieved good performance when tested in real defective images. On this basis, these experiments demonstrate that the proposed deep learning technique is very promising for the resolution of some of the studied defects, such as low luminosity, temperature change and crumpled paper and that significant advances have been achieved in improving the quality of the images, despite some limitations found when images have light reflection defects.

Palavras-chaves: Deep Learning, Image quality enhancement, Image denoising, Generative Adversarial Network, Complex numbers.

Lista de Figuras

Figura 1 – Modelo de um neurônio. Fonte: (HAYKIN et al., 2009, p. 42)	17
Figura 2 – Arquitetura de uma MLP com duas camadas ocultas. Fonte: Próprio autor	18
Figura 3 – Representação de uma convolução 2D com avanço (<i>stride</i>) 1. Fonte: (GOODFELLOW et al., 2016, p. 325)	20
Figura 4 – Esquema de representação do funcionamento de uma GAN. Fonte: Próprio autor	22
Figura 5 – Representação do aprendizado em uma rede profunda. Nas camadas iniciais são aprendidos conceitos mais simples como arestas e contornos, enquanto nas camadas mais profundas são aprendidas representações mais complexas. Fonte: (GOODFELLOW et al., 2016)	23
Figura 6 – Protocolo da revisão do estado da arte.	27
Figura 7 – Distribuição dos defeitos e suas combinações no <i>dataset</i> sintético.	41
Figura 8 – Arquitetura da U-Net utilizada no gerador e discriminador da pix2pix.	42
Figura 9 – Imagens reais capturadas para serem utilizadas como teste.	47
Figura 10 – Imagens reais capturadas para serem utilizadas como teste.	48
Figura 11 – Imagens reais capturadas para serem utilizadas como teste.	48
Figura 12 – Imagens artificiais criadas pelo algoritmo de geração de defeitos para serem utilizadas no treinamento.	49
Figura 13 – Imagens artificiais criadas pelo algoritmo de geração de defeitos para serem utilizadas no treinamento.	49
Figura 14 – Imagens artificiais criadas pelo algoritmo de geração de defeitos para serem utilizadas no treinamento.	50
Figura 15 – Comparação das imagens artificiais previstas pelas redes	53
Figura 16 – Comparação das imagens artificiais previstas pelas redes	54
Figura 17 – Comparação das imagens artificiais previstas pelas redes	55
Figura 18 – Comparação das imagens artificiais previstas pelas redes	56
Figura 19 – Comparação das imagens artificiais previstas pelas redes	57
Figura 20 – Comparação das imagens artificiais previstas pelas redes	58
Figura 21 – Comparação das imagens previstas pelas redes	58
Figura 22 – Comparação das imagens previstas pelas redes	59
Figura 23 – Comparação das imagens previstas pelas redes	59

Figura 24 – Imagem real predita pela ComPlix2Denoise apresentando distorções . . . 60

Lista de Tabelas

Tabela 1 – Resumo dos estudos selecionados na revisão do estado da arte.	31
Tabela 2 – Tabela final comparativa das métricas. Mud: Mudança percentual. Def: Média imagens defeituosas. Pred: Média imagens preditas. Legenda das métricas: (a) WadiQaM, (b) SSIM, (c) MSE, (d) MAE, (e) R^2 , (f) MS-SSIM, (g) PIE. As células marcadas em verde indicam qual técnica obteve a melhor mudança em relação a imagem defeituosa. $*1 \times 10^6$ (valores grandes foram colocados em base 10).	52
Tabela 3 – Comparação do tempo médio em segundos e o desvio padrão para inferência (predição) das imagens reais.	52

Lista de Siglas

ADP	Assistentes Digitalizadores Pessoais
CNN	Convolutional Neural Network
CPU	Central Processing Unit
DA	Data Augmentation
DCNN	Deep Convolutional Neural Network
DL	Deep Learning
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
IA	Inteligencia Artificial
ML	Machine Learning
OCR	Reconhecimento ótico de caracteres
RN	Redes Neurais
SDAE	Stacked Denoising Autoencoder
VC	Visão Computacional

Sumário

1	INTRODUÇÃO	11
1.1	Objetivos	14
1.2	Estrutura do Trabalho	14
2	CONCEITOS E REVISÃO DO ESTADO DA ARTE	16
2.1	Redes neurais artificiais	16
2.1.1	Neurônio artificial	16
2.1.2	Redes multicamadas	18
2.1.3	Redes Convolucionais	20
2.1.4	Aprendizagem profunda de máquina	21
2.1.5	Redes com camadas complexas	24
2.1.6	Métricas de qualidade da imagem	24
2.2	Caracterização do estado da arte	27
2.2.1	Defeitos de baixa luminosidade	32
2.2.2	Defeitos em fotos reais	33
2.2.3	Defeitos a partir de dispositivos móveis	34
2.2.4	Miscelânea de técnicas e defeitos correlatos	35
2.2.5	Datasets semelhantes	37
3	MATERIAIS E MÉTODOS	39
3.1	Criação do dataset	39
3.2	Criação de técnica para aprimoramento (ComPlix2Denoise)	42
3.3	Validação do modelo	45
4	RESULTADOS E DISCUSSÃO	47
4.1	Criação do <i>dataset</i>	47
4.2	Criação da técnica e validação final	50

5	CONSIDERAÇÕES FINAIS	61
	REFERÊNCIAS	62

1 Introdução

Os digitalizadores modernos podem ser considerados sucessores dos aparelhos de fax e telefotografia. O pantelégrafo foi a primeira máquina conhecida desenvolvida para este fim e pode ser considerado o predecessor da máquina de fax. Criado em 1865 por Giovanni Caselli, era capaz de emitir e de receber imagens sobre distâncias longas através das linhas de telégrafo.

As máquinas de fax realizavam inicialmente a transmissão analógica dos dados e somente no final da década 60 do século XX surgiu a primeira máquina que digitalizava e comprimia os dados, possibilitando a transmissão de mais informações. Na década de 80, a tecnologia de digitalização se tornou comum e acessível, especialmente com o advento das impressoras multifuncionais que são capazes de realizar impressão colorida e preto e branco, digitalização e outras funções.

Historicamente os digitalizadores de fotos e documentos têm sido dispositivos grandes, planos, posicionados no local de trabalho seja ele doméstico ou profissional e que ocupam uma superfície maior do que o documento a ser digitalizado. Foram desenvolvidos digitalizadores menores alimentados por folhas nas quais durante a varredura, as folhas passam em frente ao sensor de captura. A maioria desses métodos são conectados a um cabo de alimentação e a um computador, o que geralmente não é um problema, tendo em vista que os locais onde eles estão instalados sempre dispõem de tomadas e energia elétrica.

A internet possibilitou a agilidade no sistema de troca de informações e o envio de documentos entre longas distâncias, eliminando o tempo e o custo da entrega física deles. O surgimento de digitalizadores veio com grande importância para o próximo passo na direção de um sistema de troca de dados e informações mais ágil e com maior portabilidade. Define-se digitalizador como aquele dispositivo que converte informação ou dados analógicos para formato digital (WILLIAMS, 1979).

No início do século XXI, muitas formas de digitalização foram propostas, com destaque para a digitalização por imagem, que é uma alternativa mais prática e reduz os custos de utilização, eliminando a necessidade da compra de um equipamento específico para a digitalização, por exemplo. Neukermans, Downing e Slater (2000) desenvolveram um dos primeiros Assistentes Digitalizadores Pessoais (ADPs) que reduzia o tamanho do digitalizador e realizava o tratamento das informações utilizando técnicas de Reconhecimento Óptico de Caracteres (OCR).

Ainda no início do século XXI, surgiram os primeiros celulares equipados com câmeras e as primeiras tentativas de utilizá-las para a digitalização de arquivos. [Azimi e Michejda \(2006\)](#) foram um dos pioneiros na utilização das câmeras desses aparelhos, mas seu trabalho ainda tinha algumas limitações como a necessidade de conectar o celular a um computador ou ter que tirar várias fotos quando o documento era grande demais. Limitações de *hardware* impediam o funcionamento ideal, mas a ideia já existia.

Atualmente, as câmeras de maior resolução instaladas em alguns *smartphones* podem capturar imagens de documentos com alta qualidade e processá-las com algoritmos de melhoria de imagem contidos em aplicativos disponíveis para a maioria dos sistemas operacionais telefônicos. Muitos desses aplicativos podem processar documentos de várias páginas com sucessivas imagens e exportá-los como um único arquivo. Isso justifica a crescente demanda por soluções dessa natureza, que apresentem melhoria na qualidade da imagem e a supressão dos defeitos resultantes das condições do ambiente onde acontecem as capturas das imagens.

Os avanços alcançados na qualidade do *hardware* e nos algoritmos de tratamento de imagens solucionaram alguns problemas e permitiram a popularização dos métodos de digitalização por imagem, contudo, diversos fatores podem influenciar a qualidade de uma imagem.

Defeitos podem ser criados no momento da captura, do armazenamento, da transmissão e da exibição, entre outros. Além disso, podem ocorrer outros inconvenientes, tais como a falha da câmera, iluminação inadequada de ambiente, inexperiência do fotógrafo e diversas outras situações que podem acarretar em imagens de baixa qualidade ([LI et al., 2019](#)).

Os defeitos causados pela inexperiência dos fotógrafos são muito comuns, especialmente os de sombra, reflexo de luz e papel amassado. Defeitos como mudança de temperatura, que é geralmente ocasionado pela exposição da fotografia a uma fonte de luz amarela e defeito baixa luminosidade também acontecem com frequência, sendo causados em um dos estágios mencionados anteriormente. A imagem resultante de uma digitalização com esses defeitos apresenta baixa qualidade e até perda de informações em alguns casos.

É possível classificar o problema de digitalização como melhoria de documentos e como melhoria de fotos. Esses dois tipos se distinguem principalmente pelas cores presentes em cada categoria. Enquanto que na digitalização de documentos tem-se a cor branca com grande predominância, na digitalização de fotos há uma variedade de cores maior. A comunidade acadêmica e a indústria denominam a atividade de remoção de defeitos

em imagens em termos muito semelhantes: aprimoramento de imagem, restauração de imagem e remoção de defeitos. Subentende-se por esses termos, a melhoria da imagem relativo à remoção dos defeitos trabalhados.

Há várias formas de tentar remover os defeitos e melhorar a qualidade das imagens. A forma mais tradicional utiliza técnicas convencionais de visão computacional, geralmente com o uso da biblioteca OpenCV ou bibliotecas similares. OpenCV é uma biblioteca multiplataforma, totalmente gratuita para o uso acadêmico e comercial, com foco no desenvolvimento de aplicativos voltados à visão computacional (Culjak et al., 2012). Kulkarni et al. (2020) desenvolveram um algoritmo que ajuda a decifrar o texto em documentos decrépitos, cujo objetivo é converter documentos manchados, vincos e desbotados em um formato mais limpo e legível. Entretanto não é capaz de tratar outros problemas que são muito comuns, tais como sombra e reflexo de luz.

Para esses problemas de digitalização, foi observado que alguns defeitos já possuem métodos satisfatórios de remoção quando se trata do defeito isolado, entretanto ainda há espaço para o estudo de métodos para aprimoramento das imagens com defeitos simultâneos e dos defeitos sem soluções consolidadas. Ou seja, embora os estudos tenham obtido avanços significativos, problemas com defeitos de sombra, reflexo de luz, papel amassado, baixa luminosidade e mudança de temperatura em imagens ainda seguem sem uma solução satisfatória devido à sua sensibilidade aos dados de treinamento e à natureza complexa do ruído de imagens reais (ZENG et al., 2019).

Para a melhoria da qualidade da imagem, além dos métodos de visão computacional convencionais, diversas técnicas de aprendizagem de máquina, mais especificamente de aprendizagem profunda, têm obtido avanços importantes (O'MAHONY et al., 2019). Esses avanços com inteligência artificial só foram possíveis graças aos algoritmos e *hardwares* (GPUs) desenvolvidos nos últimos anos, que permitem o processamento dos dados de forma paralela, acelerando o processo de treinamento e predição das redes neurais. Redes totalmente convolucionais, *autoencoders*, redes adversárias generativas, redes com camadas complexas etc, são exemplos de técnicas e arquiteturas mais comumente utilizadas em tarefas de visão computacional. Apesar de ser aplicada nos últimos anos para solução de problemas de visão computacional, redes convolucionais com camadas complexas pouco foram utilizadas para a recuperação de imagens (QUAN et al., 2021).

1.1 Objetivos

Neste cenário, este trabalho tem como objetivo geral propor uma técnica baseada em aprendizagem de máquina profunda, denominada ComPlix2Denoise, para realizar o aprimoramento de imagens digitalizadas com defeitos de sombra, reflexo de luz, papel amassado, baixa luminosidade e mudança de temperatura utilizando camadas complexas.

Para alcançar este objetivo geral, foram propostos como objetivos específicos:

1. Criar um *dataset* de imagens com defeitos artificiais que se assemelhem aos defeitos encontrados em imagens de casos reais;
2. Criar técnica baseada em aprendizagem profunda que realize a melhoria da qualidade de imagens defeituosas;
3. Validar e refinar o modelo para garantir sua qualidade final, funcionalidade e aplicabilidade.

1.2 Estrutura do Trabalho

Este trabalho foi organizado no formato de capítulos para uma melhor apresentação e entendimento do mesmo, possuindo no total cinco capítulos, conforme descrito a seguir:

- Este capítulo **1 - Introdução**: Apresenta os aspectos estruturantes da pesquisa, como apresentação do problema de pesquisa, posicionamento em relação aos estudos atuais sobre o tema, os Objetivos e a Organização do documento apresentado.
- Capítulo **2 - Conceitos e revisão do estado da arte**: São definidos e discutidos termos e conceitos importantes para a fundamentação teórica deste estudo, como redes neurais artificiais, funções de ativação, algoritmo de retropropagação, redes convolucionais, redes adversárias generativas e métodos de avaliação de imagens. Também é apresentada uma revisão do estado da arte no contexto de aprimoramento de imagens e remoção de defeitos.
- Capítulo **3 - Materiais e métodos**: São descritos os procedimentos metodológicos adotados para alcançar os objetivos propostos.
- Capítulo **4 - Resultados e discussão**: São apresentados, discutidos e analisados os resultados alcançados para cada objetivo específico proposto.

- Capítulo 5 - **Considerações finais:** São apresentadas as conclusões do trabalho e sugestões de pesquisas futuras.

2 Conceitos e revisão do estado da arte

Creswell e Clark (2015) defendem que, ao propor uma pesquisa, é importante que o pesquisador identifique e defina os termos que os indivíduos fora do campo do estudo possam não entender e que extrapolam a linguagem comum. Assim, nas próximas subseções são definidos alguns termos importantes para esta pesquisa. Diante disto, neste capítulo serão apresentados conceitos importantes no âmbito de aprendizado de máquina e avaliação de imagens.

2.1 Redes neurais artificiais

O sistema nervoso humano é composto por uma rede de neurônios, cada um com um dendrito e um axônio. Suas sinapses (região que ocorre a comunicação), estão sempre entre o axônio de um neurônio e o dendrito de outro. As redes neurais artificiais surgiram com o intuito de imitar tal funcionamento (MCCULLOCH; PITTS, 1943).

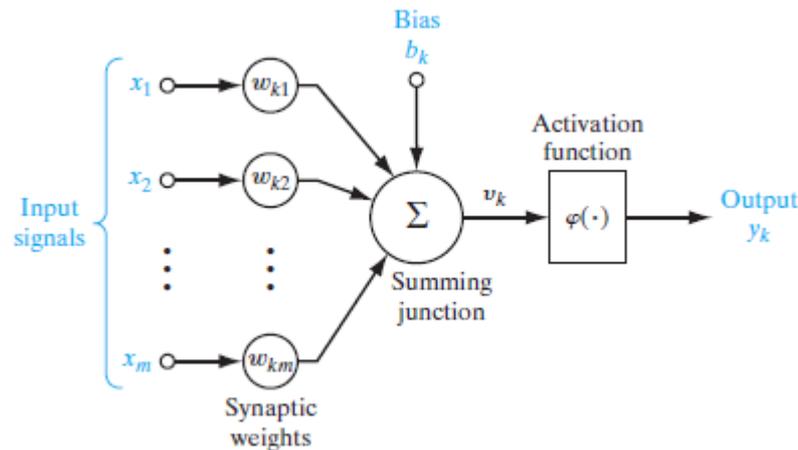
2.1.1 Neurônio artificial

O primeiro modelo de neurônio artificial foi criado por McCulloch e Pitts (1943) em 1943, onde definiram que o neurônio era a soma ponderada de entradas, acrescido de um fator *bias*. Na Equação 2.1 temos y sendo o resultado, n a quantidade de entradas, x_i sendo as entradas, w_i seus respectivos pesos e b sendo o *bias*.

$$y = \sum_{i=1}^n x_i w_i + b \quad (2.1)$$

Para mais, há também uma função de ativação (ou de transferência) que serve para limitar seu potencial de ativação e definir o quão influente esse neurônio é na rede.

Figura 1 – Modelo de um neurônio. Fonte: (HAYKIN et al., 2009, p. 42)



A função de ativação, denotada por $\varphi(v)$, define a saída de um neurônio em termos do campo local induzido v . Haykin et al. (2009) afirmam que existem dois tipos básicos de função de ativação: a função de limite (*threshold function*) (Equação 2.2), onde a saída de um neurônio assume o valor 1 se o campo local induzido desse neurônio for não negativo e caso contrário, a saída é zero.

$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{if } v < 0 \end{cases} \quad (2.2)$$

A segunda função de ativação básica é a função sigmoide, que é uma das mais utilizadas. A função pode ser definida pela Equação 2.3, onde a é o parâmetro de inclinação. Com o parâmetro de inclinação tendendo ao infinito, a função sigmoide se torna apenas uma função de limite, mas tendo a grande vantagem de ser diferenciável, enquanto a função de limite não é (HAYKIN et al., 2009).

$$\varphi(v) = \frac{1}{1 + \exp(-av)} \quad (2.3)$$

Existem outras funções de ativação que podem ser usadas em diferentes momentos durante o treinamento de uma rede neural, entre elas a função de ativação linear, tangente hiperbólica (\tanh), *Rectified Linear Unit* (reLU), *Leaky Rectified Linear Unit* (Leaky reLU), *SoftMax*, entre outras (SHARMA, 2017).

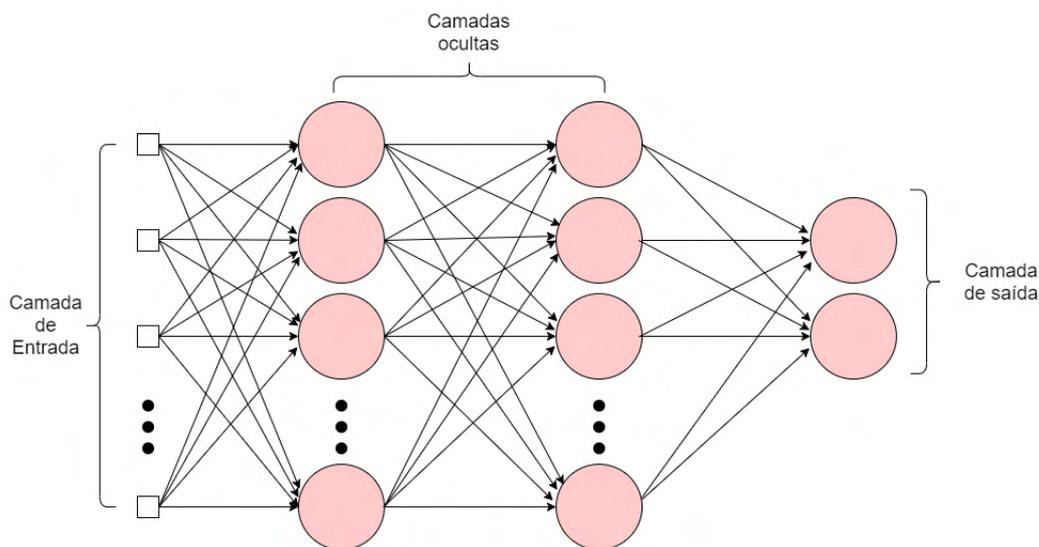
Para um melhor desempenho de uma rede, muitos aspectos devem ser considerados, como o número de camadas, quantidade de neurônios, qualidade do *dataset* utilizado no

treinamento, métodos de treinamento, validação e teste, ajuste de hiperparâmetros etc. A escolha da função de ativação e dos hiperparâmetros ideais para qualquer tarefa pode ser um processo entediante e exigir muitas pesquisas e experimentos.

2.1.2 Redes multicamadas

Um conjunto de neurônios artificiais em uma rede neural de uma única camada está limitada à classificação de problemas linearmente separáveis. Para superar tal limitação, existem as redes neurais de multicamadas, conhecidas como redes *Multilayer perceptrons* (MLP)(Figura 2). Segundo Haykin et al. (2009), essas redes possuem três características básicas: o modelo de cada neurônio possui uma função de ativação não-linear que é diferenciável, as redes possuem uma ou mais camadas ocultas além das camadas de entrada e de saída, e a rede mostra um alto grau de conectividade cuja extensão é determinada pelos pesos da rede. Entretanto, estas características também trazem problemas no entendimento do funcionamento da rede. A própria presença de não-linearidade e alta conectividade tornam a análise teórica de uma rede MLP muito difícil de ser realizada. O processo de aprendizagem também é dificultado porque a solução deve ser conduzida em um espaço muito maior de funções possíveis a partir das representações alternativas do padrão de entrada.

Figura 2 – Arquitetura de uma MLP com duas camadas ocultas. Fonte: Próprio autor



A maneira mais popular para se calcular o erro nas redes multicamadas é através do algoritmo de retropropagação de erro (*back-propagation*), cujo processo é dividido em

duas etapas: na primeira, na fase de avanço (*forward phase*), os pesos da rede estão fixos e o sinal de entrada é propagado através da rede, camada a camada, passando por suas respectivas funções de ativação, até encontrar a camada de saída. Neste momento o fluxo dos dados deve ser sempre “para frente” buscando evitar que a rede entre em um ciclo repetitivo.

Na segunda etapa, na fase para trás (*backward phase*), o resultado obtido é comparado com o resultado esperado. O gradiente da função de perda é calculado na camada final e os pesos são atualizados de trás para frente, recursivamente, através da derivada da regra da cadeia, fazendo com o que a fase para trás tenha maior custo computacional do que a fase de avanço. Logo, para aplicar o algoritmo de retropropagação de erro, as funções de ativação devem ser diferenciáveis, possibilitando que o erro parcial de cada neurônio seja calculado. O método de retropropagação de erro representou um grande avanço na área de aprendizagem de máquina e possibilitou um método computacionalmente eficiente para o treinamento de MLPs (HAYKIN et al., 2009).

Um dos objetivos das redes multicamadas é encontrar modelos que representem a distribuição probabilística dos diferentes tipos de dados encontrados nos diversos problemas do mundo real. Com o aprimoramento dos algoritmos de retropropagação do erro, de regularização, de generalização e de computação paralela, aliados ao desenvolvimento acelerado de hardware cada vez mais especializado no treinamento de redes neurais, foi possível o advento e a aplicação de modelos maiores e mais profundos.

Como explicado anteriormente, as redes multicamadas possuem um relacionamento muito complexo entre seus neurônios para ter modelos expressivos, sendo um constante desafio para quem desenvolve e constrói esses modelos garantir que tenham a qualidade esperada e consigam generalizar bem para diferentes conjuntos de dados. Quando um modelo se ajusta muito a um determinado conjunto de dados no qual foi treinado e não consegue ter desempenho similar em outro conjunto de dados que foram retirados da mesma distribuição, se diz que esse modelo está sofrendo de sobreajuste ou *overfitting*, ou seja, o modelo decorou os dados ao invés de extrair as características principais. Da mesma forma, quando um modelo não consegue ter bom desempenho e extrair as características de um determinado conjunto de dados, chama-se subajuste ou *underfitting* (HAYKIN et al., 2009).

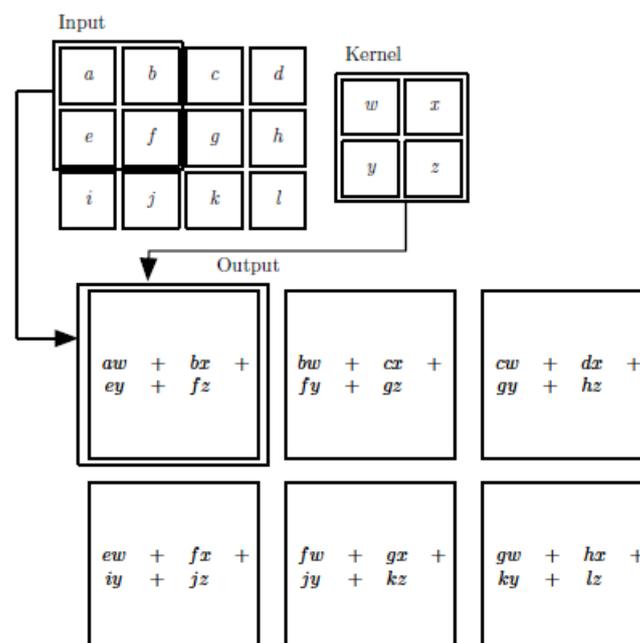
Com o intuito de reduzir o *overfitting* alguns métodos foram desenvolvidos, sendo o *dropout* um dos mais utilizados. A técnica consiste em desativar alguns neurônios de uma rede neural, que significa remover temporariamente um neurônio da rede junto com todas

as suas ligações de entrada e saída. A escolha é feita de forma aleatória e a quantidade pode ser definida pelo usuário (SRIVASTAVA et al., 2014).

2.1.3 Redes Convolucionais

Uma das variações das redes multicamadas são as redes convolucionais (CNN), que recebem esse nome por empregar uma operação matemática chamada convolução. Convolução é uma operação linear que mede a soma dos produtos de duas funções ao longo de determinado espaço. As redes convolucionais utilizam esta operação no lugar da multiplicação geral da matriz em pelo menos uma das camadas da rede. No lugar do neurônio tradicional explicado anteriormente, existe um filtro com valores que funcionam como pesos (que armazenam o aprendizado) e possuem tamanho variado a ser definido pelo usuário. A retropropagação do erro acontece nesses pesos da mesma forma como ocorre nos neurônios de uma MLP tradicional. A Figura 3 exemplifica as operações de convolução em uma matriz, que pode ser encarada analogamente como uma imagem. Esse tipo de rede tem sido utilizada principalmente no tratamento de imagens fazendo uso de convolução 2D, mas apresenta ótimos resultados em outras aplicações com dimensões diferentes de 2 (GOODFELLOW et al., 2016)(HAYKIN et al., 2009).

Figura 3 – Representação de uma convolução 2D com avanço (*stride*) 1. Fonte: (GOODFELLOW et al., 2016, p. 325)



As camadas convolucionais geralmente são usadas concomitantemente com camadas

de *pooling*. A função de *pooling* é utilizada para ressaltar os valores e diminuir o tamanho da matriz de entrada. Existem três operações principais: *MaxPooling*, *MinPooling* e *AveragePooling*. Nesses casos, são considerados os valores máximos, mínimos ou médios de determinada parte da matriz de entrada da respectiva camada. A utilização correta do *pooling* pode melhorar muito a eficiência da rede.

2.1.4 Aprendizagem profunda de máquina

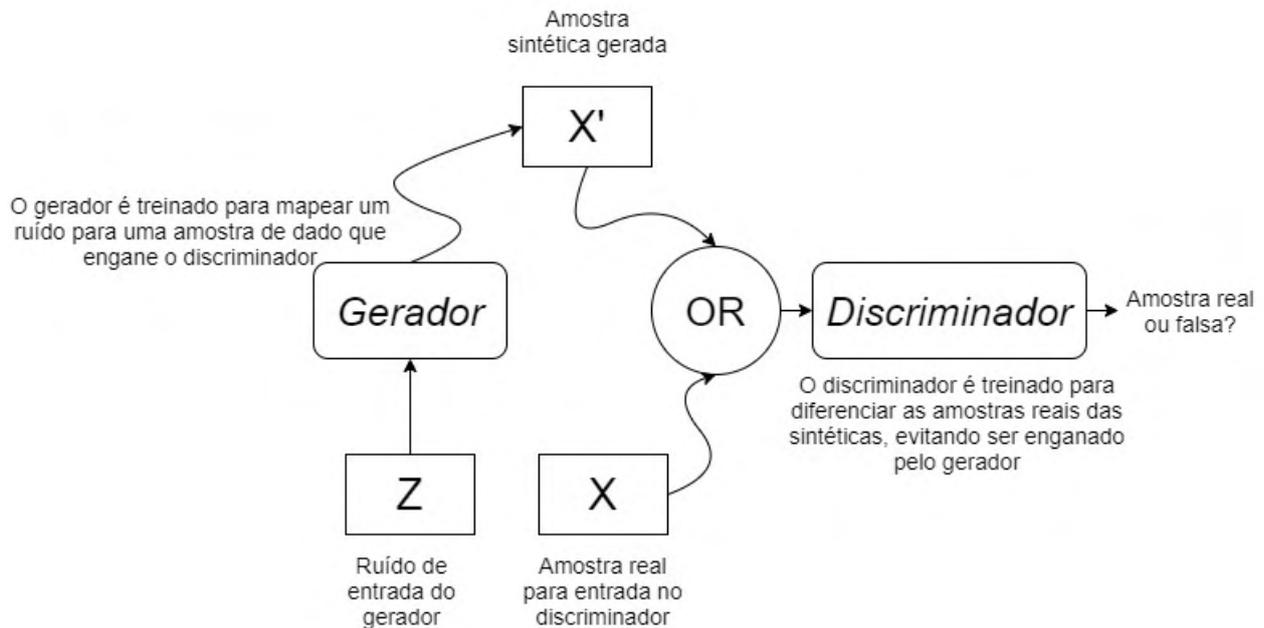
Aprendizagem de máquina permite que os computadores aprendam conceitos do mundo real e seus relacionamentos através das experiências (dados). A partir desses dados são extraídos os fatores de variação, que são as características que influenciam o resultado final. Valores que nem sempre podem ser facilmente observados nos dados, mas que são influenciadores dos resultados. Ao analisar o rosto de uma pessoa, os fatores de variação incluem a idade, sexo, tom de pele, cor dos cabelos. Ao analisar a previsão do tempo em um determinado local, os fatores de variação são umidade, sazonalidade, temperatura, entre outros.

Um exemplo clássico de arquitetura para representação do aprendizado em aprendizagem de máquina é o codificador-decodificador (Autoencoder). Como o nome sugere, esse tipo de arquitetura é dividido em duas partes: Codificador e Decodificador. O Codificador é uma função que converte os dados de entrada em uma representação diferente, com uma dimensão menor do que a da entrada e o decodificador é uma função que converte a representação reduzida de volta ao formato original (GOODFELLOW et al., 2016).

Outro exemplo de arquitetura para representação do conhecimento em aprendizagem de máquina são as redes adversárias generativas (GANs) propostas por Goodfellow et al. (2014). Elas representam um divisor de águas no campo de aprendizagem profunda e sua arquitetura consiste basicamente em duas redes que são treinadas simultaneamente e concorrem entre si. A primeira rede é chamada de gerador e tem o objetivo de capturar a distribuição dos dados e gerar amostras semelhantes às amostras que recebe de entrada, enquanto a segunda rede, chamada de discriminador, tem o objetivo de estimar uma probabilidade de um dado recebido ser real ou falso. Conforme ilustrado na Figura 4, o processo de treinamento pode ser resumido como a rede geradora tentando enganar a rede discriminadora.

Esta arquitetura têm sido bastante utilizada para os mais variados fins, como geração de objetos 3D, criação de medicamentos, utilizada em estudos sobre pandemias, processamento de imagens, detecção de faces, tradução de imagem para imagem, trans-

Figura 4 – Esquema de representação do funcionamento de uma GAN. Fonte: Próprio autor

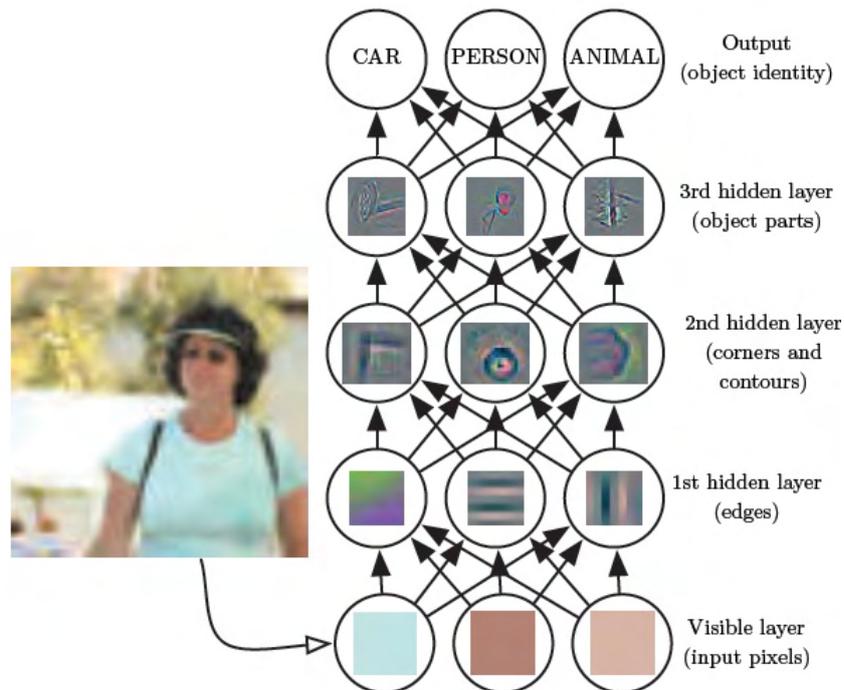


ferência de textura, controle de tráfego, remoção de ruídos de imagens, entre outros (AGGARWAL; MITTAL; BATTINENI, 2021).

Malgrado as técnicas citadas e outras existentes, inúmeras dificuldades permanecem para aplicações de inteligência artificial. No mundo real, muitos dos fatores de variação influenciam todos os dados que podemos observar. Os *pixels* individuais em uma imagem da cor de pele de uma pessoa podem sofrer influência da iluminação. Grande parte das aplicações nos obrigam a selecionar os melhores fatores de variação para cada problema, o que nem sempre é uma tarefa trivial. Muitos desses fatores são difíceis de serem identificados sem uma compreensão de dados em um nível muito abstrato, quase humano, e aparenta que a solução de um problema tão complicado não será possível através dessas técnicas. A aprendizagem profunda permite que o computador resolva este problema com conceitos complexos através de uma representação mais simples como cantos, contornos, bordas e arestas (Figura 5).

Segundo Goodfellow et al. (2016), há principalmente duas formas de se calcular a profundidade de uma rede. A primeira é contar a quantidade de instruções sequenciais que são executadas até o fim da arquitetura, como se fosse um fluxograma, percorrendo todos os estados até o final. O problema é que cada pessoa pode dar um grau de detalhamento

Figura 5 – Representação do aprendizado em uma rede profunda. Nas camadas iniciais são aprendidos conceitos mais simples como arestas e contornos, enquanto nas camadas mais profundas são aprendidas representações mais complexas. Fonte: (GOODFELLOW et al., 2016)



maior ou menor a seu fluxograma, deixando a contagem injusta por permitir representações diferentes para a mesma arquitetura. A segunda abordagem é feita a partir do grafo dos conceitos relacionados a um problema, entretanto, a profundidade do grafo para calcular o resultado pode ser muito maior do que a profundidade do grafo que explica os conceitos.

Logo, não fica claro qual das formas é mais adequada para se medir a profundidade de uma rede e nem o porquê das pessoas escolherem maneiras diferentes para representá-las. Da mesma forma, se for solicitado para duas pessoas desenvolverem um algoritmo simples, cada uma criará um algoritmo diferente e com uma quantidade de comandos (o que seria a profundidade) diferente. Portanto, não existe uma profundidade específica para que um modelo seja considerado profundo, todavia, a aprendizagem profunda pode ser definida como o estudo de modelos que relacionam uma quantidade maior de conceitos e funções do que o aprendizado de máquina tradicional. Então conclui-se que aprendizagem profunda é um tipo de aprendizagem de máquina com grande poder para representar os conceitos complexos do mundo real em conceitos mais simples e abstratos.

2.1.5 Redes com camadas complexas

Números complexos são frequentemente usados em robótica, bioinformática, processamento de imagem, sonar, radar e reconhecimento de voz. Nos últimos anos, redes neurais com números complexos começaram a receber atenção por sua melhor performance de generalização, maior capacidade representativa etc (QUAN et al., 2021). No entanto, a maioria das implementações atuais de Redes Neurais Artificiais (RNAs) e ferramentas de aprendizagem de máquina usam números reais em vez de números complexos.

A camada convolucional de valor complexo é construída simplesmente substituindo o *kernel* de valor real pelo *kernel* de valor complexo no processo de convolução. Como muitas ferramentas ainda não suportam convoluções com números complexos, pode-se implementar a operação com 4 camadas reais. Os algoritmos de retropropagação de erro e de *batch normalization* são implementados de forma similar às redes com valores reais. E como função de ativação, pode ser utilizada a *CReLU*, que é a função ReLu aplicada separadamente nas camadas imaginárias e nas camadas reais.

Quan et al. (2021) reivindicaram o título de primeiro trabalho a investigar o potencial de camadas convolucionais complexas para tarefas de visão computacional de baixo nível e reafirmaram a importância dessas abordagens para problemas de *image denoise*.

2.1.6 Métricas de qualidade da imagem

A avaliação de imagem visa distinguir computacionalmente fotos de alta qualidade das de baixa qualidade com base em regras pré-estabelecidas. Normalmente o estado da arte em *Image Quality Assessment* (IQA) utiliza métricas que quantificam a relação técnica ou subjetiva que as imagens possuem com as imagens de referência ou estimam uma nota para a qualidade da imagem. Várias abordagens têm sido propostas na literatura para tentar resolver este problema desafiador, podendo ser classificadas como uma avaliação *full-reference* (FR), quando existe uma imagem de referência ou *no-reference* (NR), quando não há imagem de referência.

O MAE (*Mean Absolute Error* ou erro médio absoluto) mede o valor médio dos erros num conjunto de previsões, sem considerar sua direção. É a média das diferenças absolutas entre previsão e observação real em que todas as diferenças individuais possuem peso igual. O valor sempre será não negativo e quanto mais próximo de zero, implica numa melhor qualidade da imagem. Na Equação 2.4 temos M e N sendo a dimensão (altura x

largura) das imagens, \hat{g} e g são as imagens a serem comparadas e i e j são as coordenadas dos pixels.

$$MAE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |\hat{g}(i,j) - g(i,j)| \quad (2.4)$$

O MSE (*Mean Squared Error* ou Erro quadrático médio) calcula a média dos quadrados dos erros, ou seja, o MSE mede o quão próximo a linha de regressão está de um conjunto de dados. Isto é feito com base na distância dos pontos para a linha de regressão e elevando-os ao quadrado, assim removendo os valores negativos e atribuindo maior peso para os valores mais distantes. O valor sempre será não negativo e quanto mais próximo de zero, implica numa melhor qualidade da imagem. Assim como no MAE, na Equação 2.5 temos M e N sendo a dimensão das imagens, \hat{g} e g são as imagens a serem comparadas e i e j são as coordenadas dos pixels.

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [\hat{g}(i,j) - g(i,j)]^2 \quad (2.5)$$

O R^2 ou coeficiente de determinação, mede a porcentagem de variação de y explicado pelas variáveis de x . Na Equação 2.6 temos o R^2 , que é calculado dividindo a soma dos quadrados dos resíduos do modelo de regressão (dado por SQ_{res}) pela soma total dos quadrados dos erros do modelo médio (dado por SQ_{tot}) e então subtraindo-o de 1. A melhor pontuação possível é 1 e pode ser negativa se o modelo for arbitrariamente pior, onde y_i são os valores a serem preditos e \hat{y}_i e \bar{y}_i são os valores preditos de y_i .

$$R^2 = 1 - \frac{SQ_{Res}}{SQ_{Tot}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \quad (2.6)$$

Imagens naturais são bem estruturadas, seus *pixels* demonstram forte dependência entre eles, sobretudo quando estão próximos e carregam muita informação sobre a imagem. A visão humana é adaptada para extrair informações estruturais de uma imagem, então o *Structural Similarity Index Measure* (SSIM) busca avaliar as imagens baseado nas suas distorções estruturais. O valor varia de zero a um e quanto mais próximo de um, implica numa melhor qualidade da imagem (WANG et al., 2004). Temos a Equação 2.7, onde μ_A , μ_B , σ_A e σ_B são os valores da média e desvio padrão de A e B, σ_{AB} é a covariância entre A e B, $c_1 = (k_1 L)^2$ e $c_2 = (k_2 L)^2$, sendo L o valor máximo de cada pixel ($2^{bitsporpixel} - 1$) e

$k_1 = 0.01$ e $k_2 = 0.03$ são constantes.

$$\text{SSIM}(A, B) = \frac{(2\mu_A\mu_B + c_1)(2\sigma_{AB} + c_2)}{(\mu_A^2 + \mu_B^2 + c_1)(\sigma_A^2 + \sigma_B^2 + c_2)} \quad (2.7)$$

Uma forma mais avançada que o SSIM, o *Multi Scale Structural Similarity Index Measure* (MS-SSIM) fornece mais flexibilidade pois incorpora variações do ponto de vista. O método de multicamadas é uma forma conveniente de incorporar imagens com mais de um canal de cor. Assim como no SSIM, o valor varia de 0 a 1 e quanto mais próximo de um, implica numa melhor qualidade da imagem (WANG; SIMONCELLI; BOVIK, 2003). Utiliza os módulos do SSIM que calcula a luminância, contraste e a estrutura. Temos a Equação 2.8, onde a comparação de contraste $c_j(A, B)$ e estruturas $s_j(A, B)$ são calculadas até a j -ésima iteração, já o cálculo da luminância $l_M(A, B)$ é feita apenas na iteração de número M , ou seja, M define o tamanho da escala.

$$\text{MS-SSIM}(A, B) = [l_M(A, B)]^{\alpha M} \cdot \prod_{j=1}^M [c_j(A, B)]^{\beta j} [s_j(A, B)]^{\gamma j} \quad (2.8)$$

As abordagens de aprimoramento de imagens existentes usam predominantemente medidas de erro ao nível de pixel, por exemplo, distâncias L1 e L2 ou uma combinação de ambas (TIMOFTE et al., 2018). Uma vez que essas medidas encapsulam apenas informações locais em nível de pixel, as imagens resultantes nem sempre fornecem resultados perceptivamente bons. Como apresentado em (BULAT; YANG; TZIMIROPOULOS, 2018), algumas avaliações com PSNR e SSIM altos oferecem imagens excessivamente suaves com baixa qualidade perceptiva. Para endereçar estes problemas, várias métricas de avaliação foram propostas. Os métodos mais recentes são ensinados a modelar a percepção humana de imagens, como o PIE (PRASHNANI et al., 2018) e o WaDIQaM (BOSSE et al., 2017). Cada uma dessas métricas tem seus próprios casos de falha, logo, não existe uma métrica perceptiva universal que funcione e quantifique perfeitamente em todas as condições a qualidade da imagem (ANWAR; KHAN; BARNES, 2020).

Criada por Bosse et al. (2017), a *Weighted AverageDeep Image Quality Measure* (WaDIQaM-FR) surgiu para direcionar o problema FR *Image Quality Assessment* adotando o conceito de redes neurais siamesas, conhecidas por resolver problemas de classificação, realizando uma fusão de características para permitir a regressão articulada das mesmas. É uma rede neural voltada para medir a qualidade de uma imagem (*distorted patch*) baseada em uma imagem de referência (*reference patch*). Esse estudo faz o uso de redes convolucionais profundas, e não depende de conhecimento prévio sobre o sistema visual

humano ou as estatísticas da imagem. O valor varia de 0 a 100 e quanto mais próximo do mínimo, implica numa melhor qualidade da imagem.

O *Perceptual Image Error* (PIE) é uma rede neural que apresenta um método para prever o erro perceptivo da imagem como os humanos fazem. Ele utiliza um *framework* de aprendizagem em pares, com o intuito de estimar a preferência de uma imagem distorcida em relação a uma outra. A avaliação gerada pela rede, chamada de *PieAPP*, é altamente relacionada com a opinião humana (PRASHNANI et al., 2018).

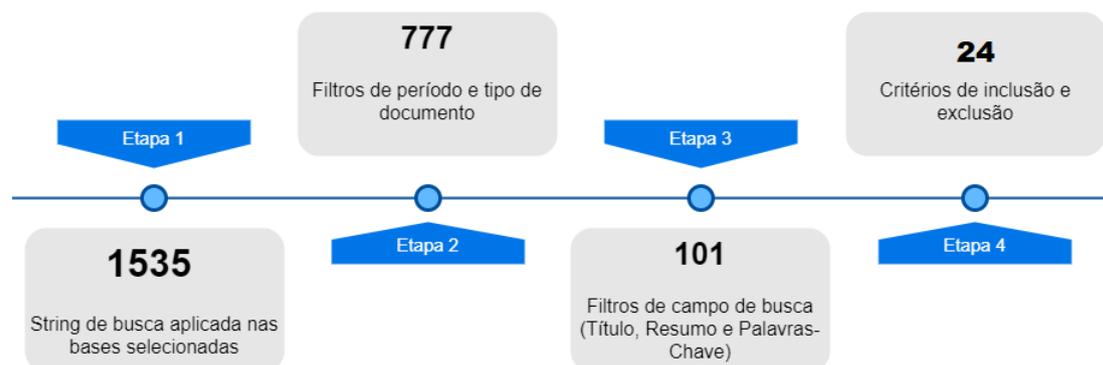
2.2 Caracterização do estado da arte

É possível analisar a fronteira do conhecimento de *image denoise* (remoção de defeitos em imagens) e *enhancement* (aprimoramento de imagens) de várias maneiras. Nesta subseção serão apresentados estudos relevantes para o problema de pesquisa investigado, seja em relação aos tipos de defeitos, forma de aquisição das imagens ou tipo de imagens.

Foi realizado o levantamento do estado da arte utilizando as bases de dados IEEEExplore, Science Direct e Google acadêmico no período 2016-2021 com os descritores:

("image" AND ("enhancement" OR "denoise")) AND (("machine learning") OR ("deep learning")) AND NOT ("segmentation" OR "super resolution" OR "classification")

Figura 6 – Protocolo da revisão do estado da arte.



Seguindo o protocolo da revisão (Figura 6) foram utilizados como critérios de exclusão:

- Estudos relacionados a medicina: Muitos dos artigos encontrados realizavam a remoção de defeitos no âmbito de exame de imagens relacionados a área de saúde, sendo assim, não representam os defeitos de digitalização de imagens.

- Estudos com defeitos diferentes: Defeitos como distorção em onda e mudança de perspectiva estão fora do escopo da pesquisa.
- Estudos com aplicações fora do escopo: Aplicações com o objetivo de segmentação, classificação e *super resolution* também foram excluídos.

Com o objetivo de resumir toda a revisão do estado da arte, foi criada a Tabela 1 que reúne os métodos, objetivos e resultados.

Referência	Método	Objetivo	Resultados
Ke et al. (2020)	Autoencoders para extrair informações do contexto	Remover defeitos de baixa luminosidade	Extrair apenas informações úteis do contexto
Wang et al. (2018)	Autoencoder + CNN para estimar a luminosidade ideal	Remover defeitos de baixa luminosidade	Melhora desempenho de outras tarefas de visão computacional como detecção e reconhecimento de objetos
Wang et al. (2020)	Redes neurais com métodos de transferência de domínio	Remover defeitos de baixa luminosidade	Remoção dos defeitos de baixa luminosidade de maneira isolada
Yang et al. (2020)	Aprendizagem semi supervisionada com duas fases de restauração	Remover defeitos de baixa luminosidade	Correção de iluminação das imagens
Hua e Xia (2018)	Utiliza GAN no aprimoramento do defeito e CNN na avaliação da imagem gerada	Remover defeitos de baixa luminosidade e imagens desfocadas simultaneamente	Correção do desfoque da luminosidade na imagem

Wang et al. (2020)	Compara os métodos de transformação de cinza, equalização de histograma, Retinex, domínio da frequência, fusão de imagem, <i>de-fogging model</i> e aprendizado de máquina	Realizar revisão sistemática do estado da arte sobre remoção do defeito de baixa luminosidade	Revisão do estado da arte sobre remoção do defeito de baixa luminosidade
Wan et al. (2020a)	Variational Autoencoders e GANs	Remover defeitos de imagens antigas	Restauração de imagens antigas
Bannigidad e Gudada (2017)	Redes neurais	Aprimorar manuscritos antigos	Remoção de defeitos causados pela degradação do tempo em documentos
Rani, Jain e Kiran (2018)	Redes neurais	Remover defeitos de imagens antigas	Restauração de imagens antigas
Guo et al. (2019)	Autoencoder + GAN	Remover ruídos gaussianos (imagem chuviscada) em fotografias reais	Remoção dos ruídos gaussianos com defeitos pouco presentes na digitalização de imagens
Chen et al. (2018b)	Autoencoder + GAN	Remover ruídos gaussianos (imagem chuviscada) em fotografias reais	Remoção dos ruídos gaussianos com leve perda de nitidez e com defeitos pouco presentes na digitalização de imagens
Zeng et al. (2019)	Regularização Laplaciana em redes neurais	Remover ruídos gaussianos (imagem chuviscada) em fotografias reais	Remoção dos ruídos gaussianos com defeitos pouco presentes na digitalização de imagens

Yang et al. (2019)	Utiliza um método matemático, a transformada discreta de cosseno	Remover reflexos em imagens	Remoção dos reflexos em vidro que se assemelha bastante com os reflexos causados pelo por <i>flashes</i> de dispositivos móveis
Lin, Chen e Chuang (2020)	Redes neurais que separam o <i>background</i> do <i>non-background</i> para remover apenas a sombra	Remover sombra em documentos	Bom resultados na remoção da sombra em imagens onde a maior parte da área é branca (documentos)
Ignatov et al. (2017)	<i>Region Based Convolutional Neural Networks</i> (RCNN)	Transformar imagens de baixa qualidade em imagens de alta qualidade	Melhora nas cores e nitidez das imagens
Ignatov et al. (2018)	Utiliza GANs	Transformar imagens de baixa qualidade em imagens de alta qualidade	Melhora nas cores e nitidez das imagens com possibilidade de um dataset mais versátil em termos de categoria de imagens (não necessita de imagens semelhantes)
Li et al. (2019)	Utiliza CNN	Remover diferentes tipos de distorções geométricas	Remove defeitos referentes à perspectiva geométrica
Vu et al. (2018)	Utiliza CNN	Remover defeitos de marcas d'água em imagens e aumenta a resolução das imagens	Remove as interferências e as marcas d'água e aumenta significativamente a legibilidade dos documentos.

Yu, Park e Jeong (2019)	Utiliza Autoencoder	Remover ruídos gaussianos (imagem chuviscada)	Técnica destaque no desafio NTIRE 2019 de remoção de ruídos
Sun et al. (2019)	Utiliza Autoencoder	Trocar a iluminação da imagem para uma iluminação alvo	Possibilita a troca de iluminação em imagens reais
Zhu et al. (2017)	Utiliza GAN	Trocar imagens entre os domínios de uma para o outro domínio	Permite a troca entre os domínios das imagens com aplicações variadas
Isola et al. (2017)	Utiliza GAN	Trocar imagens entre os domínios de uma para o outro domínio	Permite a troca entre os domínios das imagens com aplicações variadas
Tian, Xu e Zuo (2020)	Utiliza CNN com convoluções dilatadas	Remover defeitos de imagens diminuindo o custo computacional	Modelo eficiente, robusto e leve para a execução em dispositivos móveis

Tabela 1 – Resumo dos estudos selecionados na revisão do estado da arte.

Os vinte e quatro artigos restantes foram separados em quatro grupos: estudos referentes a defeitos de baixa luminosidade que estão em grande quantidade na literatura atual, estudos com fotos reais, estudos que as fotos foram adquiridas por dispositivos móveis e o último grupo reúne uma miscelânea de estudos com técnicas e defeitos correlatos.

Nos últimos anos as pesquisas para o tratamento dos defeitos de baixa luminosidade nas imagens têm crescido bastante, embora alguns pontos ainda possuam pouca investigação. Alguns estudos foram encontrados na revisão do estado da arte em relação aos defeitos de baixa luminosidade.

2.2.1 Defeitos de baixa luminosidade

Ke et al. (2020) afirmam que os métodos de aprimoramento de imagem para defeitos de baixa luminosidade foram muito desenvolvidos nos últimos anos, entretanto, os métodos geralmente têm desempenhos instáveis quando o ambiente é muito diferente. A partir disso, propõe um *autoencoder* para a extração de mapa de características para melhorar a utilização de informações de contexto e extrair as informações mais úteis. Já Wang et al. (2018) abordam o problema de baixa luminosidade através de uma rede batizada GLADNet que utiliza um *autoencoder* para realizar a estimativa inicial da iluminação global e em seguida, uma rede convolucional reconstrói os detalhes utilizando a previsão de iluminação e a entrada original. Utiliza *datasets* pequenos para o treinamento (até centenas de imagens) e sua principal aplicação é ajudar a melhorar o desempenho de outras tarefas de visão computacional, como detecção e reconhecimento de objetos. Ainda com o intuito de resolver os problemas de baixa luminosidade, Wang et al. (2020) propuseram utilizar técnicas de transferência de domínio (ou estilo) entre imagens com iluminação baixa e normal. Isso foi possível graças ao ciclo de aprendizado por ele desenvolvido, onde as imagens passam do domínio da luminosidade baixa para a luminosidade normal e depois de volta para o estado inicial.

Uma alternativa para as técnicas anteriores foi proposta por Yang et al. (2020), um novo método de aprendizagem semi supervisionado, utilizando o conhecimento de imagens sintéticas emparelhadas com luz baixa/normal e dados de alta qualidade desemparelhados para aprimoramento de imagens com luz baixa. Para este fim, criaram uma rede de duas fases que na primeira restaura o sinal com base na fidelidade da imagem de referência e, em seguida, aprimora os resultados parciais para melhorar a qualidade visual.

O defeito de baixa luminosidade é um dos mais estudados quando se trata de *image denoise* e *enhancement*, entretanto, na maioria dos casos ele é tratado de forma individual como nos estudos citados acima ou com defeitos que não são comuns na digitalização de imagem. Hua e Xia (2018) afirmam que a maioria dos métodos tradicionais de remoção de defeito e aprimoramento de imagem são menos robustos por não levar em consideração o desfoque e o ruído existente no mundo real. Com o objetivo de aprimorar a qualidade de imagem, eles propuseram um novo método baseado em redes adversariais generativas (GAN) e avaliação da qualidade de imagem (IQA). Fazem o uso de GAN para atacar o defeito de baixa luminosidade, e o método IQA usa CNN para avaliar cada qualidade de imagem aprimorada sem perder a nitidez. O trabalho traz avanços importantes e defeitos compostos (baixa luminosidade e desfoque), mas tais defeitos compostos são diferentes

dos estudados neste trabalho.

Para finalizar a revisão referente a baixa luminosidade, Wang et al. (2020) apresentam seu trabalho de revisão sistemática que aborda as principais soluções apresentadas nas últimas décadas comparando os métodos de transformação de cinza, equalização de histograma, Retinex, domínio da frequência, fusão de imagem, *defogging model* e aprendizado de máquina. Um estudo que implementa, testa e compara dezessete técnicas de modelos consideradas o estado da arte dentro das sete áreas citadas e discute os pontos fortes e fracos de cada abordagem. No geral eles afirmam que os modelos de aprendizagem de máquina têm excelentes resultados, mas que dependem de *datasets* robustos e possuem uma alta complexidade computacional. Sugere que os pesquisadores devem melhorar a robustez e a capacidade adaptativa dos métodos existentes para obter melhores resultados em aplicações práticas.

2.2.2 Defeitos em fotos reais

Para corrigir defeitos de imagens antigas (como aspecto granulado de filme, desbotamento das cores, arranhões, marcas de poeira etc.), Wan et al. (2020a) formulam o problema como um *triplet domain translation*, aproveitando dados de três domínios diferentes dentro do espaço latente: fotos antigas reais, imagens antigas sintéticas e as imagens de referência correspondentes. Uma das premissas assumidas no estudo é a possibilidade de aproximar a interseção do espaço latente das imagens sintéticas com o espaço latente das imagens reais para obter uma boa generalização para ambos os casos. O trabalho se comunica diretamente com o estudo de Bannigidad e Gudada (2017) e Rani, Jain e Kiran (2018) que desenvolveram formas de restaurar imagens de documentos antigos degradados. Tais estudos podem ser considerados como soluções na área de *denoise*, *enhancement* e *restoration* (Restauração), mas têm uma aplicação muito restrita as imagens antigas por terem características peculiares.

Com o objetivo de aprimoramento em fotografias reais, Guo et al. (2019) e Chen et al. (2018b) desenvolveram métodos semelhantes que utilizam *autoencoders* e GANs para a remoção de defeitos reais, sobretudo ruídos gaussianos (imagem chuviscada). Apesar de abordagens diferentes, ambos conseguem eliminar os ruídos com pouca perda de nitidez, resultando na melhoria da qualidade das imagens. Zeng et al. (2019) tentam uma nova abordagem ao integrar a regularização Laplaciana de grafos como um módulo treinável em uma estrutura de aprendizado profundo reduzindo a chance de ocorrer *overfitting*. Os três estudos conseguem bons resultados comprovados com as métricas de avaliação

e na avaliação subjetiva comparada com técnicas anteriores. Os estudos demonstram a necessidade de pesquisas na área de aprimoramento de fotografias reais, mas não abrangem alguns defeitos mais complexos e defeitos existentes na digitalização por imagem.

Para a remoção de reflexo em imagens, [Yang et al. \(2019\)](#) desenvolveram um método em que o resultado não depende de algoritmos iterativos e é obtido através da resolução de uma equação diferencial parcial, que pode ser feita de forma eficiente usando a Transformada Discreta de Cosseno. Um ponto forte é que essa abordagem dispensa *dataset* ou treinamento como os métodos de aprendizagem de máquina. O trabalho foca principalmente na remoção de reflexo de vidro, por exemplo, quando a foto é capturada atrás de uma janela de vidro, mas a aplicação se assemelha bastante com o defeito de reflexo de luz causado por *flashes* de dispositivos móveis ao fotografar um objeto que está próximo como no caso da digitalização de fotos por imagens.

Por fim, para imagens com defeitos de sombra, que se diferem do defeito de baixa luminosidade por ocuparem só uma parte da imagem, [Lin, Chen e Chuang \(2020\)](#) trouxeram a *Background Estimation Document Shadow Removal Network* (BEDSR-Net), desenvolvida especialmente para a remoção de sombras em documentos. Foi desenvolvido um módulo que separa o *background* do *non-background*, e extrai informações espaciais dos *pixels* e utiliza estas informações para remover apenas a sombra. Essa é a principal diferença entre digitalizar documentos onde a maior parte da área é branca, sendo possível realizar a separação do fundo e conteúdo para a digitalização de imagens coloridas e fotografias.

2.2.3 Defeitos a partir de dispositivos móveis

[Ignatov et al. \(2017\)](#) apresenta uma nova abordagem para transformar imagens capturadas por câmeras de baixa qualidade, como smartphones, em imagens de alta qualidade, capturadas por câmeras DSLR. DSLR é uma sigla para *Digital Single Lens Reflex* e se refere a um modelo de câmeras profissionais que usam um conjunto de espelho e prisma para refletir a imagem que você está prestes a capturar para o visor. A abordagem proposta utiliza *Region Based Convolutional Neural Networks* (RCNN) e é capaz de melhorar a reprodução de cores e a nitidez das imagens. Nesta abordagem são utilizadas loss functions para os parâmetros textura, conteúdo e cor. A partir da combinação entre estas funções, a qualidade da imagem é melhorada e defeitos como desfoque são resolvidos. Ao final os autores realizam comparações qualitativas e confirmam que as imagens de baixa qualidade estão semelhantes às de alta qualidade.

Ignatov et al. (2018) propõe uma nova arquitetura, chamada em *Weakly Supervised Photo Enhancer* (WESPE), baseada em *DSLR Photo Enhancement*, capaz de transformar automaticamente imagens capturadas por câmeras de capacidade limitada em imagens com qualidade superior, equivalentes a capturadas por câmeras robustas.

A arquitetura proposta, é baseada em GANs, tendo como principal característica realizar um alinhamento de *pixels* na imagem de referência, utilizando as imagens com melhor qualidade como parâmetro e padronizando as imagens com baixa qualidade. Durante este processo, as métricas *Point Signal-to-Noise Ratio* (PSNR) e *Structural Similarity Index Measure* (SSIM) são aplicadas.

Os autores concluem apresentando que diferente das demais arquiteturas baseadas em WESPE, sua arquitetura não necessita de muita supervisão, entretanto, para aplicá-la, é necessário que o *dataset* seja composto por além das imagens com qualidade inferior, por imagens de boa qualidade também. Informam também que outro diferencial, é que nesta proposta, os *datasets* não precisam ser especificamente de categorias semelhantes, permitindo que as imagens de alta qualidade sejam diferentes das de baixa qualidade.

Entre as duas pesquisas apresentadas nesta categoria, é perceptível o avanço da tecnologia no curto prazo de um ano. A maior diferença é a necessidade do uso de *datasets* correlatos, além disso a maior necessidade de interação com os parâmetros para treino na proposta de Ignatov et al. (2017).

2.2.4 Miscelânea de técnicas e defeitos correlatos

Li et al. (2019) desenvolveram um *framework* que utiliza redes neurais convolucionais (*Convolutional Neural Network - CNN*), treinado num *dataset* sintético, corrige automaticamente diferentes tipos de distorções geométricas em uma única imagem. Esta abordagem, porém, trata apenas os defeitos de perspectiva geométrica, que no processo de digitalização de imagem constituem uma parte menor dos casos.

Já Vu et al. (2018) propuseram uma forma de melhorar a qualidade das imagens com um menor custo computacional e que pode ser treinado para os problemas de *super resolution* e *image enhancement*. Chiang et al. (2017) desenvolveram um método para melhorar imagens de documentos manuscritos, que não só remove as interferências e as marcas d'água como também aumenta significativamente a legibilidade dos documentos.

Criada por Yu, Park e Jeong (2019), a DIDN tem arquitetura baseada na U-net e na DBPN, e se propõe a remover camadas de ruído gaussiano com diferentes intensidades,

sem que exista a entrada das informações dos defeitos na rede. Sendo um *autoencoder*, teve suas camadas de codificação e decodificação modificadas para a tarefa de remoção de defeitos em imagens. Apresentou excelentes resultados no NTIRE 2019, desafio criado exclusivamente para técnicas de remoção de ruídos.

A denominada *Portrait-Relighting Network* (PR-Net) é uma rede neural que tem como objetivo trocar a iluminação de uma imagem capturada por um celular comum para uma iluminação alvo. Foi treinada em um *dataset* pequeno de imagens com uma iluminação direcional. Também apresenta uma arquitetura de *autoencoder*, recebe como entrada a imagem inicial e a iluminação desejada (iluminação alvo) no *bottleneck* (ponto de menor dimensão em um *autoencoder*), enquanto o decodificador produz a imagem com a nova iluminação. A técnica alcança bons resultados, chamando a atenção para seu método diferente de trabalhar com as imagens (SUN et al., 2019).

A CycleGAN (ZHU et al., 2017) tem sido testada em diferentes aplicações nos últimos tempos, por ter grande capacidade de generalização com seus modelos atuando de forma adversária. Consiste em dois pares compostos por discriminador e gerador: D_x com F , e D_y com G . O gerador G trabalha para traduzir o domínio X ao domínio Y enquanto tenta convencer o discriminador D_y que sua saída é autêntica, o mesmo de F para D_x . Esse método é conhecido como *cycle-consistency loss* e visa garantir a consistência da tradução entre os domínios. Se baseia no pressuposto de que o resultado de uma tradução pode ser revertido para voltar à imagem original. Ele compara a imagem original à imagem reconstruída para forçar que o mapeamento entre os domínios seja bijetor e que no decorrer do treino esse emparelhamento não se distancie muito de uma representação ideal.

Já a Pix2Pix (ISOLA et al., 2017), é uma GAN que originalmente trata do problema de tradução de imagem para imagem com excelentes resultados na mudança de domínios. A arquitetura do discriminador e gerador do modelo proposto consiste numa *deep convolutional GAN* (DcGAN) cujo módulos seguem o formato: Convolução + BatchNormalization + ReLU como função de ativação. O gerador utiliza um *autoencoder* com *skip connections* para impedir a perda de informações durante a compactação da entrada. Já para o discriminador, foi utilizada uma arquitetura denominada PatchGAN que verifica a autenticidade estrutural de janelas (*patches*) da imagem mediante convolução. Assim como a CycleGAN, a rede tem sido testada com bons resultados em diversas aplicações, incluindo remoção de defeitos em imagens.

Para corrigir defeitos de imagens antigas (como aspecto granulado de filme, desbotamento das cores, arranhões, marcas de poeira, etc), (WAN et al., 2020b) formula o

problema como um *triplet domain translation*, aproveitando dados de três domínios diferentes dentro do espaço latente: fotos antigas reais, imagens antigas sintéticas e as imagens *ground-truth* correspondentes. Uma das premissas assumidas no estudo é a possibilidade de aproximar a interseção do espaço latente das imagens sintéticas com o espaço latente das imagens reais para obter uma boa generalização para ambos os casos.

Proposta em 2019, a BRDNet (TIAN; XU; ZUO, 2020) tem o objetivo de remover defeitos das imagens resolvendo dois grandes problemas: o grande custo computacional para treinar modelos muito profundos e a dificuldade de evitar a saturação de performance (*vanishing* e *exploding gradient*). A arquitetura da BRDNet é composta por duas redes CNNs concatenadas para abranger uma maior captura de *features*. A rede superior consiste em camadas Conv + BRN + ReLU enquanto a rede inferior, embora similar a rede superior, contém *dilated convolutions* em meio a camadas. *Dilated convolutions* são convoluções que apresentam filtros esparsos, tal técnica permite o aprendizado de conteúdos que não estão espacialmente próximos.

O defeito de baixa luminosidade é um exemplo de defeito que já foi bastante estudado na literatura, mas possui poucos experimentos estando simultaneamente com outros defeitos. Logo, o presente estudo também avança nesse aspecto.

2.2.5 Datasets semelhantes

Também foram identificados na revisão do estado da arte alguns *datasets* que têm como objetivo simular esses problemas na qualidade da imagem (defeitos em imagens). Em (GUO; DAI; HOIEM, 2011) é criado um *dataset* com 108 trios contendo a imagem sombreada, imagem de referência e a máscara da sombra para avaliar os algoritmos de detecção e remoção de sombras propostos no artigo. Vicente et al. (2016) deram um passo adiante fornecendo 5000 imagens sombreadas e anotadas. O SID proposto em (CHEN et al., 2018a) contém 5094 imagens de exposição curta do obturador, pareados com as imagens de exposição longa, simulando fotografias em condições de baixa iluminação para fins de treinamento e *benchmarking*. Já em (LOH; CHAN, 2019), o ExDark apresenta 7363 imagens, sombreadas ou com baixa visibilidade, possuindo defeitos verdadeiros e artificiais, possibilitando treinamento de modelos de correção de imagem ou identificação de objetos, mas somente com um defeito. Buscando contribuir no estudo de flash em fotografias, Aksoy et al. (2018) propuseram 2775 pares de imagens iluminadas somente com flash, e a mesma somente com luz ambiente. Por último, como demonstrado em (LI et al., 2020), foi criado um *dataset* com 220 pares de imagens, dividido em imagens de referência e imagens

contendo reflexo ou brilho adicionado artificialmente à imagem original.

Porém, ainda que estes estudos tenham investigado o problema, os *datasets* supracitados não possuem escalabilidade. Concluiu-se que dentre os encontrados, poucos são otimizados para a resolução de múltiplos defeitos de imagem, a maioria se limitava a reproduzir um único defeito de maneira sintética ou coletar imagens reais e usá-las para o treinamento de modelos de melhoria de imagens. No primeiro caso, muitas das imagens reproduzidas não representam os defeitos da vida real. Já no segundo caso, o conjunto de dados tende a ser muito pequeno.

3 Materiais e métodos

Neste capítulo serão descritos os procedimentos metodológicos adotados para alcançar cada objetivo específico proposto para esta pesquisa. Os procedimentos metodológicos foram divididos em três etapas.

3.1 Criação do dataset

No âmbito da digitalização de imagens por foto, há uma predominância de alguns defeitos, entretanto, poucos *datasets* se propõem a corrigi-los, e menos ainda possuem a imagem de referência para cada imagem defeituosa.

Portanto, a primeira etapa consistiu na criação de algoritmos para a simulação das seguintes classes de defeitos: sombra, reflexo de luz, papel amassado, luminosidade e mudança de temperatura. Optou-se também em criar uma classe com imagens sem defeito sintético, totalizando seis classes individuais.

Os cinco tipos de defeitos foram sintetizados sob os seguintes parâmetros:

1. Baixa luminosidade: Para simular a baixa iluminação foi aplicada a operação *gamma* na imagem com valores aleatórios (entre 0,6 e 0,9); fazendo com o que a imagem ficasse entre 10% e 40% mais escura, para dessa forma, reproduzir imagens com diferentes graus de luminosidade.
2. Sombra: Na simulação de sombras, foi aplicada na imagem uma elipse com dimensões, intensidade e posição aleatória. A elipse foi escolhida por ter um efeito semelhante com a sombra nos objetos. A intensidade da elipse sobre a imagem foi definida empiricamente e encontra-se entre 30% e 80% para evitar sombras muito claras ou escuras.
3. Papel amassado: Foram pesquisadas e coletadas 167 imagens de papéis amassados, que servem como máscaras para serem inseridas aleatoriamente em imagens do *dataset*. A intensidade foi definida empiricamente. Assim, são adicionadas máscaras de papéis amassados em imagens reais com o objetivo de simular a captura de fotografias que apresentam amassados.
4. Reflexo de luz: Foi usada uma máscara circular com gradiente do centro para as bordas com grau de rotação aleatória, que pode variar até 45 graus. A intensidade do efeito é limitada entre 55% e 85% para não ter reflexos muito claros ou pouco

intensos. Os reflexos sobrepõem parte da imagem, observando o limite de tamanho de 30% a 80% em relação à imagem original. Esses parâmetros também foram definidos empiricamente e podem ser ajustados no algoritmo. Com isso, simula efeitos de luminosidade localizados em formato elipsoidal com regiões, intensidades e tamanhos aleatórios, tendo o objetivo de generalizar os reflexos de luz naturais.

5. Mudança de temperatura: Para reproduzir o defeito de desgaste nas cores das imagens, foram usadas 19 intensidades de temperatura que são escolhidas de forma aleatória para atualizar os valores RGB da imagem.

O pseudocódigo a seguir demonstra de forma resumida o funcionamento do algoritmo de inserção de defeitos nas imagens para a criação do *dataset* artificial.

Entrada: N imagens de referência e máscaras utilizadas

Saída: $N * 3$ imagens defeituosas

```

Esses valores são ajustáveis                                     ▷ comentário
qtd_imagens_entrada ← 9900
qtd_imagens_saida ← 3

para  $i = 1$  até  $\leq$  qtd_imagens_entrada faça                       ▷ repetição
    Carrega a imagem  $i$ 
    para  $j = 1$  até  $\leq$  qtd_imagens_saida faça                       ▷ repetição
        img_temp ← Cria uma cópia da imagem de entrada  $i$ 
        defeitos ← Define os defeitos a serem adicionados na imagem  $i$ 

        se não possui defeito a ser adicionado então             ▷ condição
            Salva img_temp apenas trocando o nome para o padrão das defeituosas
        senão
            para cada defeito em defeitos faça                     ▷ repetição
                Chama o método específico de criação do defeito e salva a imagem
            fim para
        fim se
    fim para
fim para
fim para
Salva todas as informações do dataset em um arquivo JSON         ▷ final

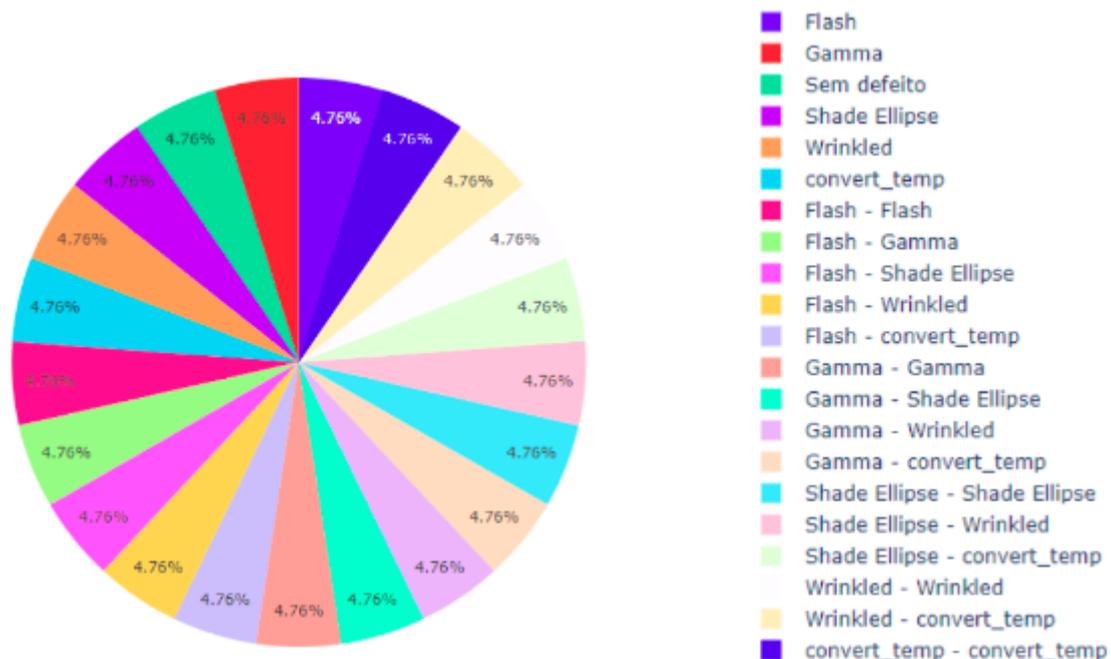
```

Durante a execução do algoritmo, cada imagem de entrada gerou outras três, nas quais cada uma recebeu de zero a dois defeitos. Foi selecionado o COCODataset (LIN et al., 2014) para a adição dos defeitos, devido ao seu extenso número de imagens públicas. Foram utilizadas 9.900 imagens como referência, que serviram de entrada para os algoritmos de simulação de defeitos, resultando em 29.700 imagens defeituosas.

Um ponto a ser observado é a quantidade de defeitos possíveis, resultante da combinação das seis classes, tomadas duas a duas, produzindo 15 grupos, somando tal combinação com as outras seis classes individuais têm-se 21 grupos, que posteriormente são produzidos e distribuídos igualmente sobre o *dataset*, como pode ser visto na Figura 7. Os algoritmos foram implementados na linguagem de programação Python utilizando as bibliotecas Opencv e Pillow para o processamento das imagens.

Ainda durante a segunda etapa foi criado um *dataset* real, sendo utilizadas páginas de revistas contendo imagens que cobriam a maior parte da folha. As imagens defeituosas foram obtidas por meio do uso de câmeras de celular para fotografá-las em condições adversas, reproduzindo os defeitos estudados. Ao final do processo, foram adquiridas 294 imagens com defeitos reais que foram utilizadas para o teste final de validação do modelo. A pesquisa tem autorização para o uso destas imagens.

Figura 7 – Distribuição dos defeitos e suas combinações no *dataset* sintético.



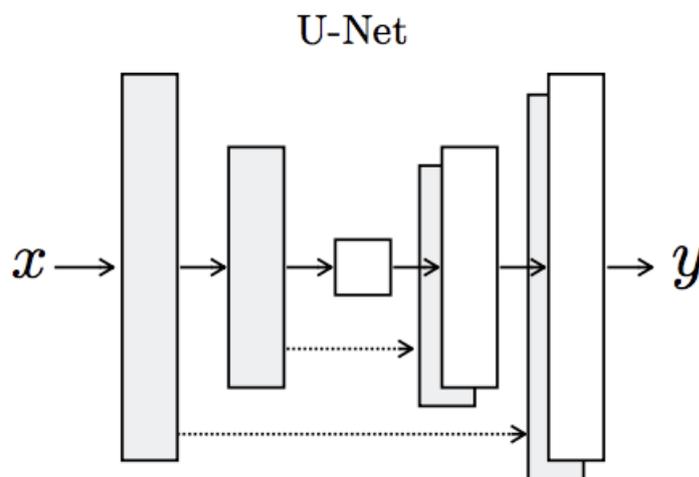
3.2 Criação de técnica para aprimoramento (ComPlix2Denoise)

Para o problema específico de aprimoramento de fotos digitalizadas, foi proposto uma rede generativa adversária que utiliza camadas com valores complexos ao invés de valores reais, denominada: ComPlix2Denoise. A rede proposta varia da arquitetura da pix2pix (ISOLA et al., 2017), já mencionada anteriormente. Por ser uma GAN, possui duas redes competindo entre si: gerador e discriminador.

Ambos os modelos, o gerador e discriminador, utilizam blocos *Convolution-Batch Normalization-ReLU* em suas camadas, como a maioria das redes convolucionais. O gerador recebe como entrada uma imagem do domínio defeituoso e tem como saída uma imagem do domínio aprimorado (sem defeito).

Como mencionado anteriormente, a arquitetura codificador-decodificador se resume a diminuir o tamanho da imagem de entrada para um tamanho mínimo, conhecido como gargalo e depois aumentar o tamanho da imagem novamente até a imagem final ter o tamanho desejado. O modelo utilizado na rede é inspirado na U-Net, que utiliza o modelo codificador-decodificador e insere algumas *skip-connections* entre as camadas do codificador e do decodificador para que a informação não seja perdida durante o processo. A Figura 8 é um exemplo das arquiteturas presentes no gerador e no discriminador da rede proposta (inspirado na U-Net), com as *skip connections* entre as camadas para evitar a perda de informação devido a profundidade da rede.

Figura 8 – Arquitetura da U-Net utilizada no gerador e discriminador da pix2pix.



Enquanto isso, o modelo do discriminador recebe uma imagem do domínio defeituoso e uma imagem sem defeitos (originalmente sem defeito ou aprimorada) e tem como saída a probabilidade de a imagem ser real (sem defeito) ou aprimorada (gerada pelo gerador). No processo de classificação das imagens, é utilizado uma rede convolucional profunda que classifica fragmentos de uma imagem como real ou falso. Denominado PatchGAN, o modelo penaliza as imagens a nível desses fragmentos e em seguida tenta classificar a imagem em real ou falsa a partir da média de todos os resultados. Diante dos experimentos de [Isola et al. \(2017\)](#), foi encontrado que um fragmento de 70x70 é suficiente para a maioria das tarefas de tradução de imagem.

Para o trabalho atual, as camadas do discriminador foram alteradas com o intuito de criar o plano imaginário e permitir a utilização de números complexos, como feito no recente trabalho de [Quan et al. \(2021\)](#), sendo que cada camada passou a ter dois filtros ao invés de um. As camadas que representam a parte imaginária têm seus valores iniciais como uma matriz de zeros.

A convolução é uma das partes mais importantes na construção de uma CNN para aprimoramento de imagens. Um filtro de valor complexo numa rede convolucional tem uma estrutura que se difere da parte de valores reais. Os filtros com seletividade de orientação são geralmente preferidos no processamento de imagens, pois as bordas da imagem local são orientadas em direções diferentes. Seletividade de orientação é a resposta das células do córtex visual a um grau de forma orientado específico apresentado no campo visual humano.

Esses filtros geralmente são inseparáveis, de modo que não podem ser expressos como o produto tensorial de dois filtros 1D de valor real, exceto aqueles ortogonais entre si. Diferentemente, o produto tensorial de dois filtros 1D de valor complexo, denotado por $\mathbf{a}_1 + i\mathbf{b}_1$ e $\mathbf{a}_2 + i\mathbf{b}_2$, não é separável em relação à sua parte real e parte imaginária. Ou seja, o produto tensorial de dois filtros 1D com valores complexos irá resultar em dois filtros 2D com valores reais, um a partir da parte real e outra a partir da parte imaginária (Equação 3.1). Utilizar números complexos permite uma representação mais compacta para a operação de convolução 2D, evitando *overfitting*, justamente por precisar de apenas dois filtros 1D, resultando em $4N$ ao invés de dois filtros 2D resultando em $2N^2$ de complexidade.

$$(\mathbf{a}_1 + i\mathbf{b}_1)(\mathbf{a}_2 + i\mathbf{b}_2)^\top = (\mathbf{a}_1\mathbf{a}_2^\top - \mathbf{b}_1\mathbf{b}_2^\top) + i \cdot (\mathbf{a}_1\mathbf{b}_2^\top - \mathbf{b}_1\mathbf{a}_2^\top) \quad (3.1)$$

A camada convolucional de valor complexo é formada meramente representando

o filtro de valor real pelo filtro de valor complexo na passagem de convolução. Diversos *frameworks* atuais não oferecem suporte a convoluções de valor complexo, por isso foi implementada a convolução de valor complexo usando as convoluções de valor real disponíveis nos *frameworks* existentes. Como expõe-se na Equação 3.2, a convolução de valor complexo pode ser criada por quatro convoluções de valor real. Essas operações podem ser implementadas em uma rede neural de valor real com conexões adicionais o que caracterizaria uma rede CNN de largura dupla (*double-width*), mas é justamente a utilização dessas operações que configura as motivações de convoluções de valor complexo.

$$\mathbf{A} * \mathbf{K} = (\Re(\mathbf{A}) * \Re(\mathbf{K}) - \Im(\mathbf{A}) * \Im(\mathbf{K})) + (\Re(\mathbf{A}) * \Im(\mathbf{K}) + \Im(\mathbf{A}) * \Re(\mathbf{K}))i \quad (3.2)$$

Para o algoritmo de retropropagação, temos \mathbf{K} , \mathbf{A} que representam um filtro complexo e seu respectivo *feature map* e $\mathbf{B} = \mathbf{A} * \mathbf{K}$ e $f(\mathbf{B})$ sendo a função escalar em \mathbf{B} . Ao tratar os filtros como partes separadas, podemos realizar o cálculo dos gradientes encontrados no treinamento de redes convolucionais, através da regra da cadeia, na Equação 3.3. Nesse momento não há mais diferença entre a retropropagação de redes complexas e redes neurais convencionais.

$$\frac{\partial f(\mathbf{B})}{\partial \mathbf{K}} = \frac{\partial f(\mathbf{B})}{\partial \mathbf{B}} \frac{\partial \mathbf{B}}{\partial \mathbf{K}} = \frac{\partial f(\mathbf{B})}{\partial \mathbf{B}} * \mathbf{A} \quad (3.3)$$

A função de ativação utilizada nas redes complexas é muito semelhante a utilizada nas redes com valores reais. Chamada de $\mathbb{C}ReLU$, ela simplesmente aplica a função ReLU na parte real e na imaginária das camadas. Em relação ao *Batch Normalization*, que como explicado anteriormente, serve para melhorar a performance de generalização e de convergência do treinamento, também é aplicado separadamente nas partes reais e imaginárias dos números complexos, como exemplificado na Equação 3.4.

$$BN(\mathbf{z}) = ReBN(\Re(\mathbf{z})) + i \cdot ReBN(\Im(\mathbf{z})) \quad (3.4)$$

Por fim, a última parte que se diferencia das redes com valores reais é que se faz necessário uma camada de mescla dos valores. A Equação 3.5 mostra que os valores das camadas complexas são transformados em valores reais obtendo-se a amplitude do sinal complexo.

$$\mathbf{X} = \sqrt{\Re(\mathbf{B})^2 + \Im(\mathbf{B})^2} \quad (3.5)$$

Depois de explicado as peculiaridades das redes com valores complexos, restam os detalhes da GAN implementada no presente trabalho. O treinamento do discriminador é muito mais rápido se comparado ao do gerador, por isso o erro do discriminador é cortado pela metade para atrasar esse processo e permitir o treinamento simultâneo do gerador (Equação 3.6).

$$DiscriminatorLoss = 0.5 * DiscriminatorLoss \quad (3.6)$$

O treinamento do gerador é feito utilizando tanto o erro adversarial (erro do discriminador) quanto o erro médio absoluto ($L1 loss$) entre os pixels da imagem de entrada e a imagem de saída. Esses dois erros combinados compõem o erro utilizado no gerador. A Equação 3.7 mostra como combinação destes dois erros é controlada por um novo hiperparâmetro Lambda, que visa aumentar a importância do erro médio absoluto durante o processo de treinamento do gerador.

$$GeneratorLoss = AdversarialLoss + Lambda * L1Loss \quad (3.7)$$

Com a utilização de camadas com números complexos, espera-se que o discriminador consiga ter um desempenho melhor na classificação de imagens como reais ou falsas e por consequência, melhorar o desempenho do gerador já que o discriminador estará mais rigoroso e as redes competem entre si.

O modelo foi treinado por 200 épocas, com auxílio do *framework* PyTorch e executado numa placa de vídeo NVIDIA RTX 6000 com 24GB de memória de vídeo com duração aproximada de 72 horas para 12 mil imagens no *dataset* de treino e validação e utilizando Linux Ubuntu 16.4 como sistema operacional.

3.3 Validação do modelo

Na terceira etapa, as técnicas DIDN (YU; PARK; JEONG, 2019), PR-Net (SUN et al., 2019), Pix2Pix (ISOLA et al., 2017), BRDNet (TIAN; XU; ZUO, 2020), CycleGAN (ZHU et al., 2017), MS VAE (WAN et al., 2020a) foram selecionadas e treinadas no *dataset* artificial para efeito comparativo com a técnica desenvolvida no presente trabalho. Foi realizada uma predição com ambos *datasets* (real e artificial) buscando validar dois pontos fundamentais: A semelhança entre os defeitos artificiais e reais e a qualidade da técnica desenvolvida.

Para a análise quantitativa das imagens, foram selecionadas algumas métricas utilizadas para avaliação de imagens, foram elas: WaDIQaM, SSIM, MSE, MAE, R2, MS-SSIM e PIE. Após a escolha das métricas, 100 imagens sintéticas de teste pareadas, nunca vistas pelas redes foram utilizadas para calcular os resultados quantitativos. Também foi destacada a mudança percentual entre a métrica das imagens preditas e defeituosas. Por fim, foi realizado um teste para verificar o tempo de inferência (predição) de cada rede para um conjunto de dados composto por 294 imagens reais com resoluções variadas, sendo executadas numa placa de vídeo NVIDIA RTX 6000 com 24GB de memória de vídeo e utilizando Linux Ubuntu 16.4 como sistema operacional. Diferentemente do dataset artificial, tais imagens reais não possuem imagens de referência (*ground-truth*) para a comparação final em termos quantitativos.

4 Resultados e Discussão

Nesta seção serão apresentados e analisados os resultados alcançados durante a pesquisa.

4.1 Criação do *dataset*

A criação dos *datasets* foi fundamental para o estudo pois as imagens artificiais possibilitaram que o conjunto de treino tivesse um número grande de imagens, o que é importante para os modelos de aprendizagem profunda e o conjunto de imagens reais permitiu que a validação fosse assegurada em casos aplicáveis no mundo real.

As Figuras 9, 10, 11 constituem o conjunto de dados real, sendo capturadas com um smartphone buscando reproduzir os defeitos criados pela maioria dos usuários.

- (a) Imagem real com defeito de amassado e sombra (b) Imagem real com defeito de amassado e sombra (c) Imagem real com defeito de baixa luminosidade



Figura 9 – Imagens reais capturadas para serem utilizadas como teste.

- (a) Imagem real com defeito de reflexo de luz (b) Imagem real com defeito de reflexo de luz



Figura 10 – Imagens reais capturadas para serem utilizadas como teste.

- (a) Imagem real com defeito de amassado e sombra (b) Imagem real com defeito de baixa luminosidade



Figura 11 – Imagens reais capturadas para serem utilizadas como teste.

As imagens da esquerda das Figuras 13, 12 e 14 são as imagens que foram utilizadas como base para criação das imagens defeituosas, sendo utilizadas como imagens de referência no momento de treinamento e avaliação. Enquanto as imagens da direita foram criadas artificialmente pelo algoritmo e buscam reproduzir os defeitos reais comumente criados pelos usuários.

- (a) Imagem de referência (sem defeito) (b) Imagem artificial com de reflexo de
da imagem (b) luz e mudança de temperatura



Figura 12 – Imagens artificiais criadas pelo algoritmo de geração de defeitos para serem utilizadas no treinamento.

Das 29.700 imagens artificiais geradas pelo algoritmo de geração de defeitos, dez por cento foram separadas para o conjunto de teste. As imagens demonstram a similaridade dos defeitos artificiais e reais, comprovando a utilidade do *dataset* artificial. Um ponto importante a ser ressaltado é o defeito artificial de papel amassado não apresenta a distorção em onda (*warping*) presente no mundo real, podendo não representar fielmente o defeito real, mas foi útil nos casos leves analisados neste trabalho.

- (a) Imagem de referência (sem defeito) da im- (b) Imagem artificial com defeito de papel amas-
agem (b) sado



Figura 13 – Imagens artificiais criadas pelo algoritmo de geração de defeitos para serem utilizadas no treinamento.

- (a) Imagem de referência (sem defeito) (b) Imagem artificial com defeito de mudança de temperatura



Figura 14 – Imagens artificiais criadas pelo algoritmo de geração de defeitos para serem utilizadas no treinamento.

4.2 Criação da técnica e validação final

A divisão dos conjuntos de treino e validação, assim como o total de imagens utilizadas variou de acordo com cada técnica. Na maioria dos casos, a divisão foi 60/40 (treino/validação). As métricas selecionadas foram WaDIQaM, SSIM, MSE, MAE, R^2 , MS-SSIM e PIE. Como todas as métricas selecionadas utilizam a abordagem FR de avaliação, só foi possível executar a avaliação quantitativa nos dados artificiais, por possuírem imagem de referência. A tabela 2 mostra os resultados quantitativos finais de cada métrica para a média dos valores das imagens preditas, imagens defeituosas e a porcentagem da mudança, com o intuito de verificar se as técnicas de aprimoramento foram eficazes. Ou seja, a média das imagens preditas devem ter resultado melhor do que a média das imagens defeituosas.

A métrica WaDIQaM pode trazer resultados diferentes para a mesma imagem, por isso foi realizado uma média dos valores entre as métricas para se obter os valores das imagens de referência (antes de ser melhorada). A Equação 4.1 da mudança percentual foi adicionado uma constante para tornar sempre o melhor valor positivo. Ou seja, se o valor da mudança percentual for positivo, houve melhora na qualidade da imagem. Onde A é uma variável de controle para garantir que os valores da mudança percentual (melhoria) sempre estejam positivos. Este valor é 1 para: SSIM, MS-SSIM e PIE. Este valor é -1 para todas as outras métricas.

$$\text{Mudança percentual} = A * \frac{(\text{ImagemPreditada} - \text{ImagemDefeituosa})}{\text{ImagemDefeituosa}} * 100 \quad (4.1)$$

Optou-se por tentar remover todos os defeitos simultaneamente ao invés de criar uma rede especialista para cada defeito porque foi observado durante os experimentos que ao passar a imagem repetidas vezes nas redes, a imagem vai perdendo a qualidade visual geral, pois todos os *pixels* são alterados e não apenas os que possuem o defeito. Dessa forma, as passagens sucessivas causam perda da qualidade geral.

A técnica DIDN apresentou os melhores resultados em três métricas: WaDIQaM, SSIM e R^2 . A Pix2pix teve melhor desempenho em duas métricas: MSE e MAE. A técnica proposta foi melhor em duas métricas: MS-SSIM e PIE. É interessante que MAE e MSE possuem uma relação e apontaram o mesmo vencedor. Todavia, MS-SSIM e SSIM possuem uma relação, mas apontaram vencedores diferentes. PIE e WaDIQaM realizam uma avaliação perceptual e também indicaram vencedores diferentes.

Ao analisar os tempos de predição na Tabela 3, pode-se observar que os resultados da Pix2Pix e da ComPlix2Denoise também foram semelhantes, isso deve acontecer pelo fato de uma rede variar da outra. BRDNet e CycleGAN tiveram bons tempos de predição, mas os resultados nas métricas de avaliação não foram satisfatórios. Analisando o aspecto de tempo, a ComPlix2Denoise e a Pix2Pix se mostraram interessantes porque possuem desempenho equiparável.

A análise da melhor técnica deve ser complementada de maneira qualitativa, portanto as Figuras 16, 17, 15, 18, 19 e 20 apresentam os resultados preditos pelas redes para o *dataset* de imagens artificiais com defeitos diversos. No geral, DIDN e a técnica proposta apresentam resultados mais estáveis, conseguindo pelo menos atenuar a maioria dos defeitos.

Por fim, para completar a análise qualitativa, as imagens com defeitos reais são analisadas. Nota-se uma melhoria considerável nas cores e qualidade das imagens preditas por algumas redes. As Figuras 21, 22 e 23 mostram que assim como nas imagens artificiais, os resultados variam de acordo com a rede e o tipo de defeito apresentado. Nenhuma rede pode ser considerada campeã, mas se destacam a ComPlix2Denoise, DIDN e Pix2pix.

Algumas limitações também foram encontradas em todas as técnicas, o que sugere que pode ser um problema dos dados de treinamento, isto é, no *dataset* artificial criado. A Figura 24 apresenta a principal limitação encontrada pelas técnicas, que é o fato de "queimar" algumas partes brancas da foto. Supõe-se que seja causado pelo defeito de reflexo de luz das imagens artificiais utilizadas durante o treinamento dos modelos. A diminuição da intensidade do defeito artificial de reflexo de luz pode ser uma solução para este problema.

	DIDN	PR-Net	Pix2Pix	BRDNet	CycleGAN	MS VAE	ComPlix2Denoise
(a) Pred.:	5,716	16,458	18,498	7,597	18,111	15,471	17,646
(a) Def.:	8,677	8,677	8,677	8,677	8,677	8,677	8,677
(a) Mud.:	34,3%	-90,3%	-113,8%	12,7%	-108,3%	-78,5%	-103,56%
(b) Pred.:	0,884	0,826	0,836	0,855	0,778	0,746	0,834
(b) Def.:	0,856	0,856	0,856	0,856	0,856	0,856	0,856
(b) Mud.:	3,3%	-3,4%	-2,3%	-0,1%	-9,0%	-12,8%	-2,55%
(c) Pred.:	2.062,266	2.167,544	1.116,814	2.272,745	2.684,712	1.474,513	1.226,82
(c) Def.:	4.150,907	4.150,907	4.150,907	4.150,907	4.150,907	4.150,907	4.150,907
(c) Mud.:	50,3%	47,8%	73,1%	45,2%	35,3%	64,5%	70,44%
(d)* Pred.:	16,222	16,382	11,719	17,614	18,609	13,956	12,361
(d)* Def.:	22,512	22,512	22,512	22,512	22,512	22,512	22,512
(d) Mud.:	27,9%	27,2%	47,9%	21,8%	17,3%	38,0%	45,09%
(e) Pred.:	0,895	0,879	0,891	0,885	0,884	0,884	0,889
(e) Def.:	0,909	0,909	0,909	0,909	0,909	0,909	0,909
(e) Mud.:	-1,6%	-3,3%	-2,0%	-2,7%	-2,7%	-2,7%	-2,19%
(f) Pred.:	0,934	0,923	0,942	0,901	0,912	0,920	0,943
(f) Def.:	0,922	0,922	0,922	0,922	0,922	0,922	0,922
(f) Mud.:	1,3%	0,2%	2,2%	-2,3%	-1,0%	-0,2%	2,25%
(g) Pred.:	1,102	1,112	0,894	1,376	1,680	1,302	0,861
(g) Def.:	1,438	1,438	1,438	1,438	1,438	1,438	1,438
(g) Mud.:	23,4%	22,7%	37,9%	4,3%	-16,8%	9,5%	40,15%

Tabela 2 – Tabela final comparativa das métricas. Mud: Mudança percentual. Def: Média imagens defeituosas. Pred: Média imagens preditas. Legenda das métricas: (a) WadiQaM, (b) SSIM, (c) MSE, (d) MAE, (e) R^2 , (f) MS-SSIM, (g) PIE. As células marcadas em verde indicam qual técnica obteve a melhor mudança em relação a imagem defeituosa. * 1×10^6 (valores grandes foram colocados em base 10).

	DIDN	PR-Net	Pix2Pix	BRDNet	CycleGAN	MS VAE	ComPlix2Denoise
Tempo médio (s):	10,068	8,622	3,954	1,139	3,161	17,995	4,012
Desvio padrão (s):	3,93	3,16	1,52	0,44	1,16	6,84	1,49

Tabela 3 – Comparação do tempo médio em segundos e o desvio padrão para inferência (predição) das imagens reais.



Figura 15 – Comparação das imagens artificiais previstas pelas redes



(a) Original



(b) Defeituosa



(c) BRDNet



(d) ComPlix2Denoise



(e) CycleGAN



(f) DIDN



(g) MS VAE



(h) Pix2pix



(i) Relighting

Figura 16 – Comparação das imagens artificiais previstas pelas redes

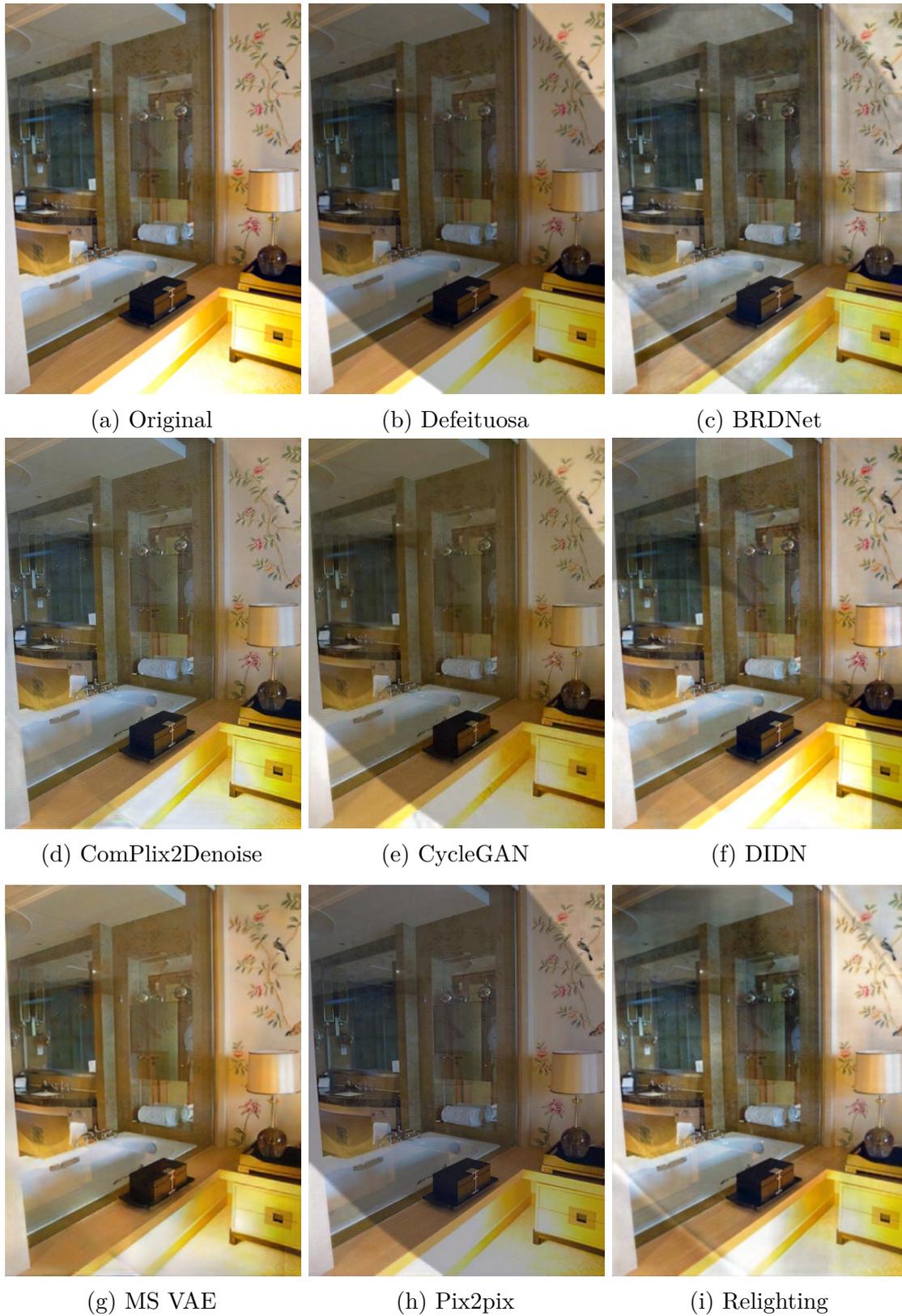


Figura 17 – Comparação das imagens artificiais previstas pelas redes

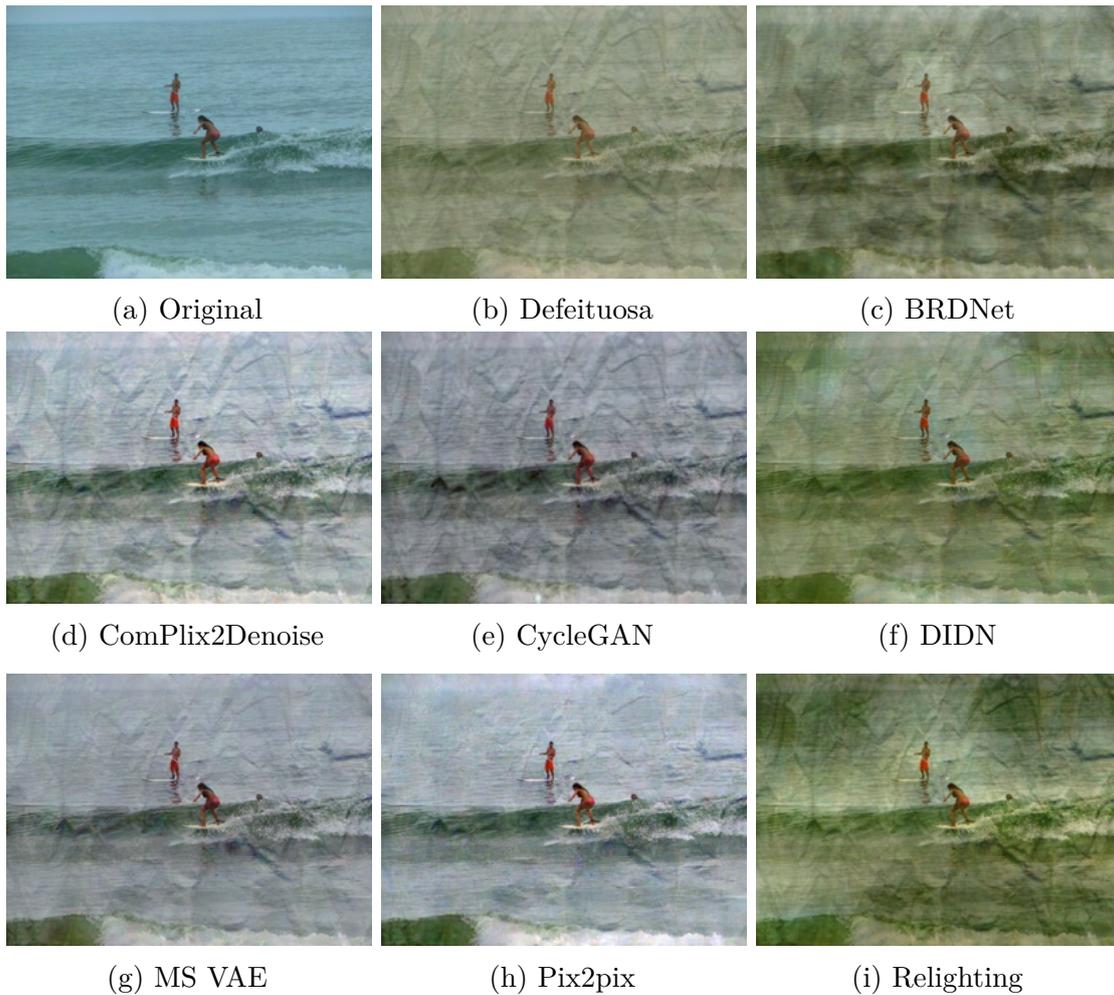


Figura 18 – Comparação das imagens artificiais previstas pelas redes

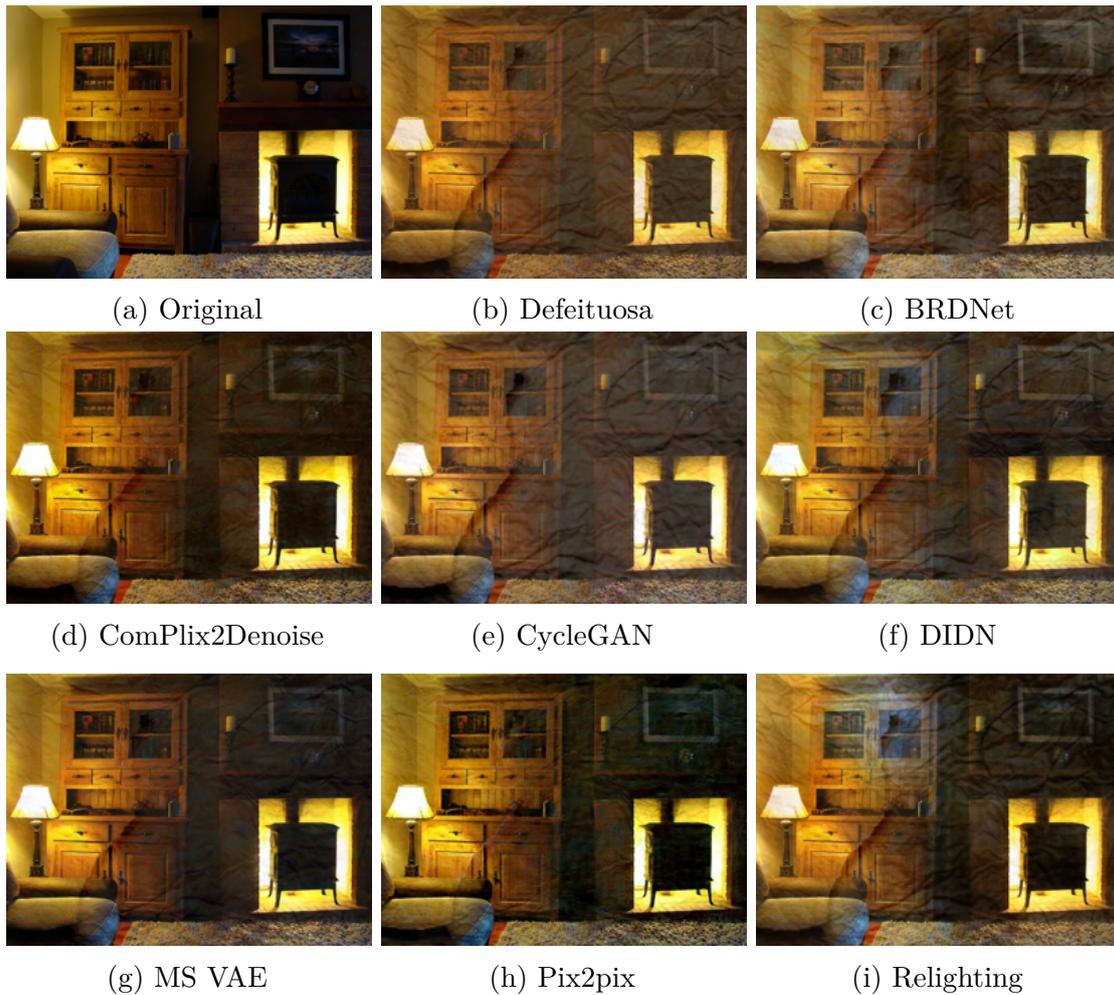


Figura 19 – Comparação das imagens artificiais previstas pelas redes



Figura 20 – Comparação das imagens artificiais previstas pelas redes

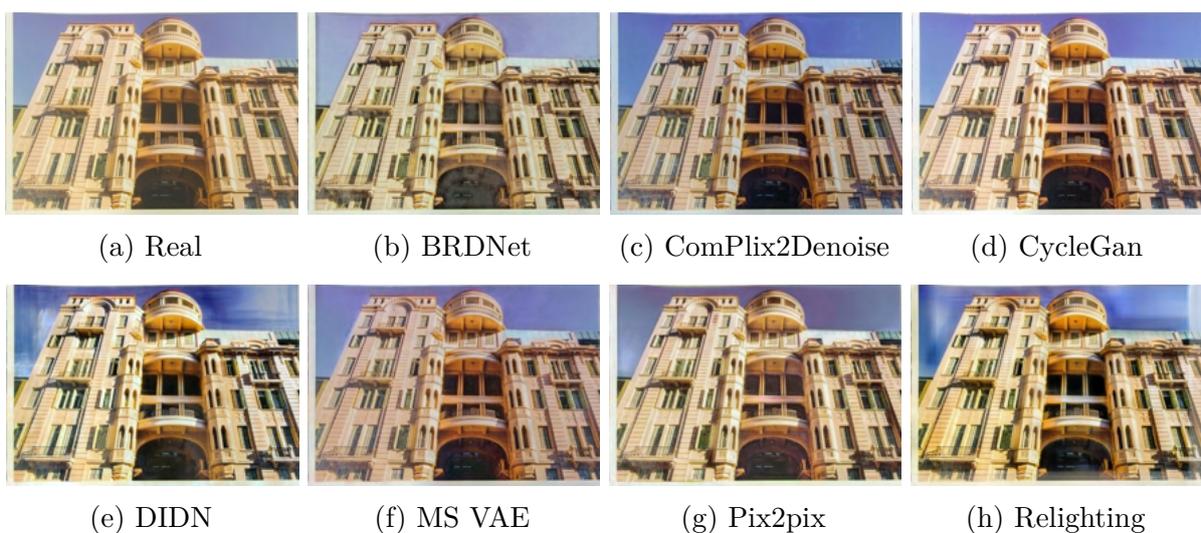


Figura 21 – Comparação das imagens previstas pelas redes

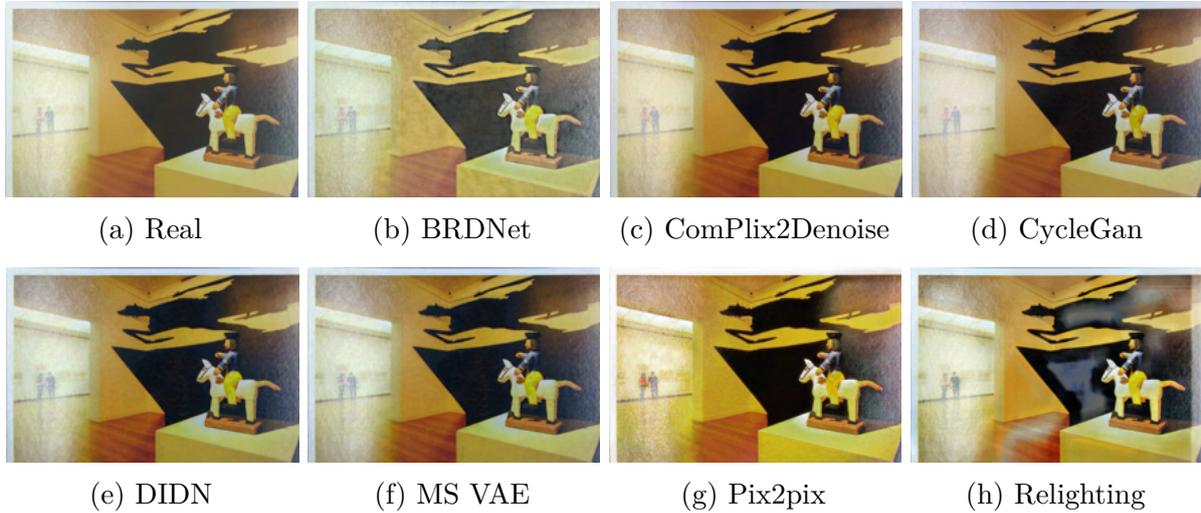


Figura 22 – Comparação das imagens previstas pelas redes

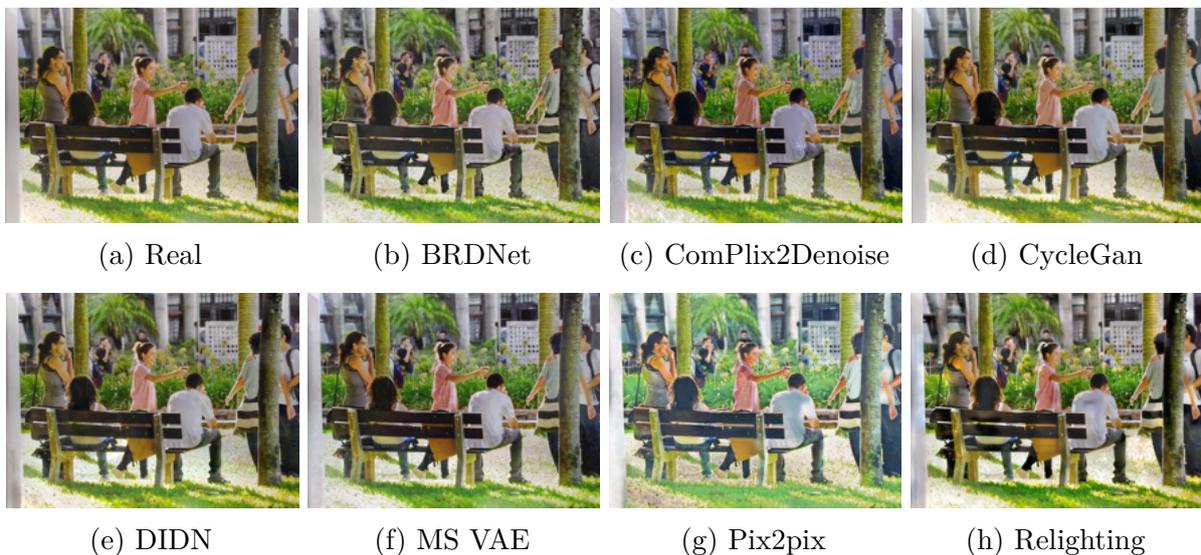


Figura 23 – Comparação das imagens previstas pelas redes



Figura 24 – Imagem real predita pela ComPlix2Denoise apresentando distorções

5 Considerações Finais

Este trabalho teve como objetivo propor uma técnica baseada em aprendizagem de máquina profunda, denominada ComPlix2Denoise, para realizar o aprimoramento de imagens digitalizadas com defeitos de sombra, reflexo de luz, papel amassado, baixa luminosidade e mudança de temperatura utilizando camadas complexas. Estes experimentos demonstraram que apesar de os modelos terem sido treinados com imagens contendo defeitos sintéticos, eles foram capazes de generalizar bem as imagens reais apresentadas, obtendo bons resultados com a maioria das técnicas exploradas.

Os resultados se complementam para o alcance dos objetivos específicos propostos no início do trabalho. O *dataset* artificial comprova sua qualidade quando os modelos apesar de treinados no *dataset* artificial, conseguem remover ou atenuar os defeitos reais. Logo, o algoritmo de geração de defeitos e o *dataset* artificial apresentados neste trabalho são importantes marcos no campo da melhoria de imagens digitalizadas visando a correção de múltiplos defeitos.

A técnica desenvolvida demonstra sua qualidade ao apresentar resultados promissores, disputando diretamente com as melhores técnicas presentes no estado da arte atualmente tanto na avaliação quantitativa quanto qualitativa. A introdução de camadas complexas para a remoção de defeitos que vão além do ruído Gaussiano estabelece novas possibilidades de pesquisa e de desenvolvimento de redes com números complexos para outras tarefas de visão computacional.

Os resultados qualitativos e quantitativos completam o terceiro objetivo específico e só foram possíveis graças a qualidade dos resultados obtidos nos objetivos específicos anteriores. O atingimento dos três objetivos específicos garantiu que o objetivo geral fosse atingido.

Em pesquisas futuras, sugere-se que seja aprimorado o modelo de rede neural generativa adversária (GANs) para a remoção dos defeitos e solução das limitações encontradas, pois foi observado que as GANs obtiveram os melhores resultados. Observa-se também que o *dataset* real deve ser expandido para uma melhor utilização no teste e possível utilização no treino de modelos futuros. Além disso, o papel amassado pode conter distorção (*warping*) em um cenário real. A não-simulação dessa distorção pode ter viés que podem denegrir a performance dos modelos em cenários reais. A avaliação do quanto que a distorção não-simulada prejudica a performance do modelo em casos reais (papéis amassados reais) também pode ser estudada posteriormente.

Referências

- AGGARWAL, A.; MITTAL, M.; BATTINENI, G. Generative adversarial network: An overview of theory and applications. **International Journal of Information Management Data Insights**, Elsevier, p. 100004, 2021. Citado na página 22.
- AKSOY, Y.; KIM, C.; KELLNHOFER, P.; PARIS, S.; ELGHARIB, M.; POLLEFEYS, M.; MATUSIK, W. A dataset of flash and ambient illumination pairs from the crowd. In: **Proc. ECCV**. [S.l.: s.n.], 2018. Citado na página 37.
- ANWAR, S.; KHAN, S.; BARNES, N. A deep journey into super-resolution: A survey. **ACM Computing Surveys (CSUR)**, ACM New York, NY, USA, v. 53, n. 3, p. 1–34, 2020. Citado na página 26.
- AZIMI, K.; MICHEJDA, J. **Cellular telephone based document scanner**. [S.l.]: Google Patents, 2006. US Patent App. 10/986,811. Citado na página 12.
- BANNIGIDAD, P.; GUDADA, C. Restoration of degraded kannada handwritten paper inscriptions (hastaprati) using image enhancement techniques. In: IEEE. **2017 International Conference on Computer Communication and Informatics (ICCCI)**. [S.l.], 2017. p. 1–6. Citado 2 vezes nas páginas 29 and 33.
- BOSSE, S.; MANIRY, D.; MÜLLER, K.-R.; WIEGAND, T.; SAMEK, W. Deep neural networks for no-reference and full-reference image quality assessment. **IEEE Transactions on Image Processing**, IEEE, v. 27, n. 1, p. 206–219, 2017. Citado na página 26.
- BULAT, A.; YANG, J.; TZIMIROPOULOS, G. To learn image super-resolution, use a gan to learn how to do image degradation first. In: **Proceedings of the European conference on computer vision (ECCV)**. [S.l.: s.n.], 2018. p. 185–200. Citado na página 26.
- CHEN, C.; CHEN, Q.; XU, J.; KOLTUN, V. Learning to see in the dark. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2018. p. 3291–3300. Citado na página 37.
- CHEN, Y.-S.; WANG, Y.-C.; KAO, M.-H.; CHUANG, Y.-Y. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2018. p. 6306–6314. Citado 2 vezes nas páginas 29 and 33.
- CHIANG, J.-S.; HSIA, C.-H.; TU, H.-Y.; GIANG, H. T. H.; LIN, T.-Y. Adaptive image enhancement method for document. In: IEEE. **2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)**. [S.l.], 2017. p. 417–420. Citado na página 35.
- CRESWELL, J. W.; CLARK, V. L. P. **Pesquisa de Métodos Mistos-: Série Métodos de Pesquisa**. [S.l.]: Penso Editora, 2015. Citado na página 16.

Culjak, I.; Abram, D.; Pribanic, T.; Dzapo, H.; Cifrek, M. A brief introduction to opencv. In: **2012 Proceedings of the 35th International Convention MIPRO**. [S.l.: s.n.], 2012. p. 1725–1730. Citado na página [13](#).

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A.; BENGIO, Y. **Deep learning**. [S.l.]: MIT press Cambridge, 2016. v. 1. Citado 5 vezes nas páginas [5](#), [20](#), [21](#), [22](#), and [23](#).

GOODFELLOW, I. J.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. **Generative Adversarial Networks**. 2014. Citado na página [21](#).

GUO, R.; DAI, Q.; HOIEM, D. Single-image shadow detection and removal using paired regions. In: **2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011**. [S.l.]: IEEE Computer Society, 2011. (Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition), p. 2033–2040. ISBN 9781457703942. Citado na página [37](#).

GUO, S.; YAN, Z.; ZHANG, K.; ZUO, W.; ZHANG, L. Toward convolutional blind denoising of real photographs. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2019. p. 1712–1722. Citado 2 vezes nas páginas [29](#) and [33](#).

HAYKIN, S. S. et al. **Neural networks and learning machines/Simon Haykin**. [S.l.]: New York: Prentice Hall,, 2009. Citado 5 vezes nas páginas [5](#), [17](#), [18](#), [19](#), and [20](#).

HUA, W.; XIA, Y. Low-light image enhancement based on joint generative adversarial network and image quality assessment. In: IEEE. **2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)**. [S.l.], 2018. p. 1–6. Citado 2 vezes nas páginas [28](#) and [32](#).

IGNATOV, A.; KOBYSHEV, N.; TIMOFTE, R.; VANHOEY, K.; GOOL, L. V. Dslr-quality photos on mobile devices with deep convolutional networks. In: **Proceedings of the IEEE International Conference on Computer Vision**. [S.l.: s.n.], 2017. p. 3277–3285. Citado 3 vezes nas páginas [30](#), [34](#), and [35](#).

IGNATOV, A.; KOBYSHEV, N.; TIMOFTE, R.; VANHOEY, K.; GOOL, L. V. Wespe: weakly supervised photo enhancer for digital cameras. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops**. [S.l.: s.n.], 2018. p. 691–700. Citado 2 vezes nas páginas [30](#) and [35](#).

ISOLA, P.; ZHU, J.-Y.; ZHOU, T.; EFROS, A. A. Image-to-image translation with conditional adversarial networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 1125–1134. Citado 5 vezes nas páginas [31](#), [36](#), [42](#), [43](#), and [45](#).

KE, X.; LIN, W.; CHEN, G.; CHEN, Q.; QI, X.; MA, J. Edllie-net: Enhanced deep convolutional networks for low-light image enhancement. In: IEEE. **2020 IEEE 5th**

International Conference on Image, Vision and Computing (ICIVC). [S.l.], 2020. p. 59–68. Citado 2 vezes nas páginas 28 and 32.

KULKARNI, M.; KAKAD, S.; MEHRA, R.; MEHTA, B. Denoising documents using image processing for digital restoration. In: SWAIN, D.; PATTNAIK, P. K.; GUPTA, P. K. (Ed.). **Machine Learning and Information Processing**. Singapore: Springer Singapore, 2020. p. 287–295. ISBN 978-981-15-1884-3. Citado na página 13.

LI, C.; YANG, Y.; HE, K.; LIN, S.; HOPCROFT, J. E. Single image reflection removal through cascaded refinement. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2020. Citado na página 37.

LI, X.; ZHANG, B.; SANDER, P. V.; LIAO, J. Blind geometric distortion correction on images through deep learning. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2019. p. 4855–4864. Citado 3 vezes nas páginas 12, 30, and 35.

LIN, T.-Y.; MAIRE, M.; BELONGIE, S.; HAYS, J.; PERONA, P.; RAMANAN, D.; DOLLÁR, P.; ZITNICK, C. L. Microsoft coco: Common objects in context. In: SPRINGER. **European conference on computer vision**. [S.l.], 2014. p. 740–755. Citado na página 41.

LIN, Y.-H.; CHEN, W.-C.; CHUANG, Y.-Y. Bedsr-net: A deep shadow removal network from a single document image. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2020. p. 12905–12914. Citado 2 vezes nas páginas 30 and 34.

LOH, Y. P.; CHAN, C. S. Getting to know low-light images with the exclusively dark dataset. **Computer Vision and Image Understanding**, Elsevier, v. 178, p. 30–42, 2019. Citado na página 37.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, n. 4, p. 115–133, 1943. Citado na página 16.

NEUKERMANS, A. P.; DOWNING, J. P.; SLATER, T. G. **Handheld document scanner**. [S.l.]: Google Patents, 2000. US Patent 6,064,779. Citado na página 11.

O'MAHONY, N.; CAMPBELL, S.; CARVALHO, A.; HARAPANAHALLI, S.; HERNANDEZ, G. V.; KRPALKOVA, L.; RIORDAN, D.; WALSH, J. Deep learning vs. traditional computer vision. In: SPRINGER. **Science and Information Conference**. [S.l.], 2019. p. 128–144. Citado na página 13.

PRASHNANI, E.; CAI, H.; MOSTOFI, Y.; SEN, P. Pieapp: Perceptual image-error assessment through pairwise preference. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2018. p. 1808–1817. Citado 2 vezes nas páginas 26 and 27.

QUAN, Y.; CHEN, Y.; SHAO, Y.; TENG, H.; XU, Y.; JI, H. Image denoising using complex-valued deep cnn. **Pattern Recognition**, v. 111, p. 107639, 2021. ISSN 0031-3203. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0031320320304428>>. Citado 3 vezes nas páginas 13, 24, and 43.

RANI, N. S.; JAIN, A. S.; KIRAN, H. A unified preprocessing technique for enhancement of degraded document images. In: SPRINGER. **International Conference on ISMAC in Computational Vision and Bio-Engineering**. [S.l.], 2018. p. 221–233. Citado 2 vezes nas páginas 29 and 33.

SHARMA, S. Activation functions in neural networks. **towards data science**, v. 6, 2017. Citado na página 17.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. **The journal of machine learning research**, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014. Citado na página 20.

SUN, T.; BARRON, J. T.; TSAI, Y.-T.; XU, Z.; YU, X.; FYFFE, G.; RHEMANN, C.; BUSCH, J.; DEBEVEC, P. E.; RAMAMOORTHI, R. Single image portrait relighting. **ACM Trans. Graph.**, v. 38, n. 4, p. 79–1, 2019. Citado 3 vezes nas páginas 31, 36, and 45.

TIAN, C.; XU, Y.; ZUO, W. Image denoising using deep cnn with batch renormalization. **Neural Networks**, v. 121, p. 461 – 473, 2020. ISSN 0893-6080. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0893608019302394>>. Citado 3 vezes nas páginas 31, 37, and 45.

TIMOFTE, R.; GU, S.; WU, J.; GOOL, L. V. Ntire 2018 challenge on single image super-resolution: Methods and results. In: **Proceedings of the IEEE conference on computer vision and pattern recognition workshops**. [S.l.: s.n.], 2018. p. 852–863. Citado na página 26.

VICENTE, T. F. Y.; HOU, L.; YU, C.-P.; HOAI, M.; SAMARAS, D. Large-scale training of shadow detectors with noisily-annotated shadow examples. In: LEIBE, B.; MATAS, J.; SEBE, N.; WELLING, M. (Ed.). **Computer Vision – ECCV 2016**. Cham: Springer International Publishing, 2016. p. 816–832. ISBN 978-3-319-46466-4. Citado na página 37.

VU, T.; NGUYEN, C. V.; PHAM, T. X.; LUU, T. M.; YOO, C. D. Fast and efficient image quality enhancement via desubpixel convolutional neural networks. In: **Proceedings of the European Conference on Computer Vision (ECCV)**. [S.l.: s.n.], 2018. p. 0–0. Citado 2 vezes nas páginas 30 and 35.

WAN, Z.; ZHANG, B.; CHEN, D.; ZHANG, P.; CHEN, D.; LIAO, J.; WEN, F. Bringing old photos back to life. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2020. p. 2747–2757. Citado 3 vezes nas páginas 29, 33, and 45.

WAN, Z.; ZHANG, B.; CHEN, D.; ZHANG, P.; CHEN, D.; LIAO, J.; WEN, F. Bringing old photos back to life. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2020. Citado na página 36.

WANG, L.-W.; LIU, Z.-S.; SIU, W.-C.; LUN, D. P.-K. Deep lightening network for low-light image enhancement. In: IEEE. **2020 IEEE International Symposium on Circuits and Systems (ISCAS)**. [S.l.], 2020. p. 1–5. Citado 2 vezes nas páginas 28 and 32.

Wang, W.; Wei, C.; Yang, W.; Liu, J. Gladnet: Low-light enhancement network with global awareness. In: **2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)**. [S.l.: s.n.], 2018. p. 751–755. Citado 2 vezes nas páginas 28 and 32.

WANG, W.; WU, X.; YUAN, X.; GAO, Z. An experiment-based review of low-light image enhancement methods. **IEEE Access**, IEEE, v. 8, p. 87884–87917, 2020. Citado 2 vezes nas páginas 29 and 33.

WANG, Z.; BOVIK, A. C.; SHEIKH, H. R.; SIMONCELLI, E. P. et al. Image quality assessment: from error visibility to structural similarity. **IEEE transactions on image processing**, v. 13, n. 4, p. 600–612, 2004. Citado na página 25.

WANG, Z.; SIMONCELLI, E. P.; BOVIK, A. C. Multiscale structural similarity for image quality assessment. In: IEEE. **The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003**. [S.l.], 2003. v. 2, p. 1398–1402. Citado na página 26.

WILLIAMS, C. Scanning digitizers: automating the input of drawings. **Computer-Aided Design**, Elsevier, v. 11, n. 4, p. 227–230, 1979. Citado na página 11.

YANG, W.; WANG, S.; FANG, Y.; WANG, Y.; LIU, J. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2020. Citado 2 vezes nas páginas 28 and 32.

YANG, Y.; MA, W.; ZHENG, Y.; CAI, J.-F.; XU, W. Fast single image reflection suppression via convex optimization. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2019. p. 8141–8149. Citado 2 vezes nas páginas 30 and 34.

YU, S.; PARK, B.; JEONG, J. Deep iterative down-up cnn for image denoising. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops**. [S.l.: s.n.], 2019. p. 0–0. Citado 3 vezes nas páginas 31, 35, and 45.

ZENG, J.; PANG, J.; SUN, W.; CHEUNG, G. Deep graph laplacian regularization for robust denoising of real images. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops**. [S.l.: s.n.], 2019. p. 0–0. Citado 3 vezes nas páginas 13, 29, and 33.

ZHU, J.-Y.; PARK, T.; ISOLA, P.; EFROS, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2017. p. 2223–2232. Citado 3 vezes nas páginas [31](#), [36](#), and [45](#).