



SUPER RESOLUÇÃO UTILIZANDO REDES ADVERSÁRIAS GERADORAS PARA ANÁLISE DE ANOMALIAS EM IMAGENS PATOLÓGICAS

Trabalho de Conclusão de Curso

Engenharia da Computação

Victor Guilherme Ferreira Trajano
Orientador: Carmelo J. A. Bastos Filho.
Coorientador: Rodrigo de Paula Monteiro.



**Universidade de Pernambuco
Escola Politécnica de Pernambuco
Graduação em Engenharia de Computação**

**VICTOR GUILHERME FERREIRA
TRAJANO**

**SUPER RESOLUÇÃO UTILIZANDO
REDES ADVERSÁRIAS GERADORAS
PARA ANÁLISE DE ANOMALIAS EM
IMAGENS PATOLÓGICAS**

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia de Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

Recife, maio de 2022.

Trajano, Victor Guilherme Ferreira

SUPER RESOLUÇÃO UTILIZANDO REDES ADVERSÁRIAS GERADORAS PARA ANÁLISE DE ANOMALIAS EM IMAGENS PATOLÓGICAS / Victor Guilherme Ferreira Trajano. – Recife - PE, 2022.

xiii, 63 f. : il. ; 29 cm.

Trabalho de Conclusão de Curso (Graduação em Engenharia de Computação) Universidade de Pernambuco, Escola Politécnica de Pernambuco, Recife, 2022.

Orientador: Prof. Dr. Carmelo José Albanez Bastos Filho.

Coorientador: Prof. Me. Rodrigo de Paula Monteiro.

Inclui referências.

1. aprendizado de máquina. 2. super-resolução. 3. Redes geradoras adversárias. I. Super resolução utilizando redes adversárias geradoras para análise de anomalias em imagens patológicas. II. Bastos Filho, Carmelo J. A. . III. Universidade de Pernambuco.

MONOGRAFIA DE FINAL DE CURSO

Avaliação Final (para o presidente da banca)*

No dia 27/5/2022, às 08h45min, reuniu-se para deliberar sobre a defesa da monografia de conclusão de curso do(a) discente **VICTOR GUILHERME FERREIRA TRAJANO**, orientado(a) pelo(a) professor(a) **CARMELO JOSE ALBANEZ BASTOS FILHO**, sob título SUPER RESOLUÇÃO UTILIZANDO REDES NEURAS ADVERSARIAIS PARA ANÁLISE DE ANOMALIAS EM IMAGENS PATOLÓGICAS, a banca composta pelos professores:

ALEXANDRE MAGNO ANDRADE MACIEL (PRESIDENTE)

CARMELO JOSE ALBANEZ BASTOS FILHO (ORIENTADOR)

Após a apresentação da monografia e discussão entre os membros da Banca, a mesma foi considerada:

Aprovada Aprovada com Restrições* Reprovada

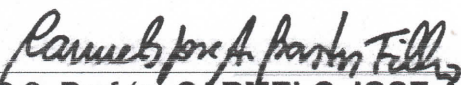
e foi-lhe atribuída nota: 10,0 (Dez)

*(Obrigatório o preenchimento do campo abaixo com comentários para o autor)

O(A) discente terá Sete dias para entrega da versão final da monografia a contar da data deste documento.

Documento assinado digitalmente
gov.br ALEXANDRE MAGNO ANDRADE MACIEL
Data: 27/05/2022 09:17:26-0300
Verifique em <https://verificador.iti.br>

AVALIADOR 1: Prof (a) **ALEXANDRE MAGNO ANDRADE MACIEL**


AVALIADOR 2: Prof (a) **CARMELO JOSE ALBANEZ BASTOS FILHO**

AVALIADOR 3: Prof (a)

* Este documento deverá ser encadernado juntamente com a monografia em versão final.

Dedico esse trabalho aos meus pais e aos meus amigos.

Agradecimentos

Gostaria de agradecer primeiramente a todos os grandes amigos que conheci durante minha graduação, sem os quais esta jornada acadêmica jamais poderia ser tão memorável.

Gostaria de agradecer também aos meus pais que sempre estiveram ao meu lado me apoiando em todos os momentos e que foram motivadores para a minha graduação. Além disso gostaria de agradecer ao Prof. Carmelo J. A. Bastos Filho por me orientar neste trabalho.

Muito obrigado a todos!

Autorização de publicação de PFC

Eu, **Victor Guilherme Ferreira Trajano** autor(a) do projeto de final de curso intitulado: **SUPER RESOLUÇÃO UTILIZANDO REDES NEURAIS ADVERSARIAS PARA ANÁLISE DE ANOMALIAS EM IMAGENS PATOLÓGICAS**; autorizo a publicação de seu conteúdo na internet nos portais da Escola Politécnica de Pernambuco e Universidade de Pernambuco.

O conteúdo do projeto de final de curso é de responsabilidade do autor.

Victor Guilherme Ferreira Trajano

Orientador(a): **Carmelo Jose Albanéz Bastos Filho**

Rodrigo de
Paula Monteiro

Assinado de forma digital por
Rodrigo de Paula Monteiro
Dados: 2022.05.30 23:58:58
-03'00'

Coorientador(a): **Rodrigo de Paula Monteiro**

Prof, de TCC: **Daniel Augusto Ribeiro Chaves**

Data: 27/5/2022

Resumo

Redes adversárias geradoras vem ganhando destaque na área aprendizado de máquina, cada vez mais surgem novos modelos capazes de resolver problemas computacionais complexos. Contudo, ainda há uma vasta área não explorada do potencial que essas redes possuem. Esse trabalho propõe o uso de um modelo de GAN (*Generative Adversarial Network*) para reconstrução de imagens de exames médicos utilizando técnicas de super-resolução, com objetivo de avaliar a performance do modelo quanto a esse tipo de imagem, utilizando métricas de qualidade para classificar o quão similar são as imagens geradas pelo algoritmo, além de uma avaliação de qualidade por profissionais e estudantes da área de biomedicina. Como resultado, obteve-se valores que indicam uma melhoria significativa na qualidade das imagens geradas e também bons resultados quanto a avaliação de qualidade.

Palavras-chave: aprendizado de máquina, super-resolução, redes geradoras adversárias, GAN.

Abstract

Generative adversarial networks have been gaining prominence in the area of machine learning, new models capable of solving complex computational problems are increasingly emerging. However, there is still a vast untapped area of potential that these networks have. This work proposes the use of a GAN (Generative Adversarial Network) model for image reconstruction of medical examinations using super-resolution techniques, in order to evaluate the performance of the model regarding this type of image, using quality metrics to classify how similar the images generated by the algorithm are, in addition to a quality assessment by professionals and students in the field of biomedicine. As a result, values were obtained that indicate a significant improvement in the quality of the generated images and also good results in terms of quality assessment.

Keywords: machine learning, super-resolution, generative adversarial networks, GAN.

Lista de Figuras

Figura 1 - Efeitos visuais da baixa resolução espacial e da baixa resolução da intensidade.....	15
Figura 2 - Arquitetura de um <i>perceptron</i>	21
Figura 3 - Exemplo de arquitetura para <i>perceptrons</i> multicamadas com uma camada escondida.....	22
Figura 4 - Funções de ativação.....	23
Figura 5 - Exemplo de convolução para pixel (de intensidade 99) em imagem com 1 canal de cor.....	27
Figura 6 - Exemplo de camada de convolução.....	28
Figura 7 - Esqueleto completo de uma CNN da forma Rede Totalmente Convolutacional.....	28
Figura 8 - Modelos discriminadores e modelos geradores.....	30
Figura 9 - Modelo original da GAN.....	31
Figura 10 - Gráfico da função de erro JG ao se adotar $JG = JD$	33
Figura 11 - Exemplo de esqueleto para uma DCGAN.....	35
Figura 12 - Ilustração dos casos mais comuns de obtenção de diferentes informações entre as imagens capturadas.....	37
Figura 13 - Visão geral da geração de dados sintéticos puros adotados no Real-ESRGAN.....	40
Figura 15 - Diagrama do funcionamento da medição do SSIM.....	43
Figura 16 - Comparação ilustrativa entre o PSNR e o SSIM.....	44
Figura 17 - Primeira pergunta do formulário de avaliação de qualidade.....	45
Figura 18 - Segunda pergunta do formulário de avaliação de qualidade.....	46
Figura 20 - Comparação entre a imagem com degradação após passagem pelo processo de redução bicúbica e a imagem gerada pelo modelo.....	47
Figura 21 - Comparação entre a imagem original e a imagem gerada pelo modelo.....	48

Figura 22 - Gráfico de barras para cada conjunto de imagens.....	48
Figura 23 - Gráfico de setores do conjunto de imagens degradadas.	50
Figura 24 - Gráfico de setores do conjunto de imagens geradas.	50
Figura 25 - Gráfico de setores do conjunto de imagens originais.....	51
Figura 26 - Gráfico de <i>box-plot</i> do número de avaliações por conjunto de imagens.	51
Figura 27 - Gráfico de setores do conjunto de imagens degradadas.	52
Figura 28 - Gráfico de setores do conjunto de imagens geradas.	53
Figura 29 - Gráfico de setores do conjunto de imagens originais.....	53

Lista de Símbolos

Notação

X	Conjunto
x	Vetor de dados
\hat{x}	Vetor de predição
θ	Parâmetros ou pesos da rede
$f_{\theta}(\cdot)$	Função parametrizada por θ
p	Distribuição de probabilidade
$x \sim p$	Vetor de dados x amostrado da distribuição p
$p(\cdot; \theta)$	Distribuição de probabilidade parametrizada por θ
$p(\cdot, \cdot)$	Distribuição de probabilidade conjunta
$p(\cdot \cdot)$	Distribuição de probabilidade condicional
$Ez[X]$	Valor esperado de X com respeito a z
c_k	Translação vertical do deslocamento da imagem de baixa resolução
M	Largura e altura da imagem de alta resolução
x	Imagem de alta resolução (vetor coluna), no domínio matricial
\hat{x}	Imagem de alta resolução estimada no domínio matricial

Lista de Siglas e Acrônimos

ANN	Rede Neural Artificial (<i>Artificial Neural Network</i>)
CNN	Rede Neural Convolutacional (<i>Convolutional Neural Network</i>)
dB	Decibéis
DCGAN	Rede Convolutacional Geradora Adversária Profunda (<i>Deep Convolutional Generative Adversarial Networks</i>)
GAN	Rede Geradora Adversária (<i>Generative Adversarial Network</i>)
GANs	Redes Adversárias Geradoras (<i>Generative adversarial networks</i>)
HR	Alta Resolução (<i>High Resolution</i>)
HVS	Sistema Visual Humano (<i>Human Visual System</i>)
IA	Inteligência Artificial
LR	Baixa Resolução (<i>Low Resolution</i>)
MLP	<i>Perceptron</i> Multicamadas (<i>Multilayer Perceptron</i>)
MSE	Erro Quadrático Médio (<i>Median Squared Error</i>)
PSNR	Relação Sinal-Ruído de Pico (<i>Peak Signal-to-Noise Ratio</i>)
RGB	Vermelho Verde Azul (<i>Red Green Blue</i>)
RM	Ressonância Magnética
SR	Super-resolução (<i>super-resolution</i>)
SRGAN	Rede Adversária Geradora de Super-Resolução (<i>Super-Resolution Generative Adversarial Network</i>)
SRGANs	Redes Adversárias Geradoras de Super-Resolução (<i>Super-Resolution Generative Adversarial Networks</i>)
SSIM	Índice de Similaridade Estrutural (<i>Structural Similarity Index</i>)
TC	Tomografia Computadorizada

Sumário

1	INTRODUÇÃO	14
1.1	Motivação	16
1.2	Objetivos	17
1.2.1	Objetivo geral	17
1.2.2	Objetivos específicos	17
1.3	Estrutura do trabalho	17
2	APRENDIZADO DE MÁQUINA	19
2.1	Redes Neurais Artificiais	19
2.1.1	Perceptron	20
2.1.2	Funções de ativação	22
2.1.3	Treinamento por Backpropagation	24
2.2	Redes Neurais Convolucionais	25
2.2.1	Camada de Convolução	26
3	REDES GERADORAS ADVERSÁRIAS	29
3.1	Modelagem Geradora	29
3.2	Modelo para Redes Geradoras Adversárias	30
3.3	Funções de Perda	32
3.3.1	Rede Discriminadora	32
3.3.2	Rede Geradora	33
3.4	Redes Convolucionais Geradoras Adversárias Profundas	34
3.5	Super-resolução de imagens	36

4	MATERIAIS E MÉTODOS	38
4.1	Análise do estado da arte	38
4.2	Escolha do modelo	38
4.3	Base de dados	40
4.4	Métricas de qualidade	41
4.4.1	PSNR	41
4.4.2	SSIM	42
4.5	Avaliação de qualidade	44
4.6	Ferramentas de desenvolvimento	46
5	RESULTADOS E DISCUSSÕES	47
5.1	Aplicação das métricas de qualidade	47
5.2	Resultados da avaliação de qualidade	48
6	CONCLUSÃO E TRABALHOS FUTUROS	54
6.1	Conclusões	54
6.2	Trabalhos futuros	55
	REFERÊNCIAS	56

1 Introdução

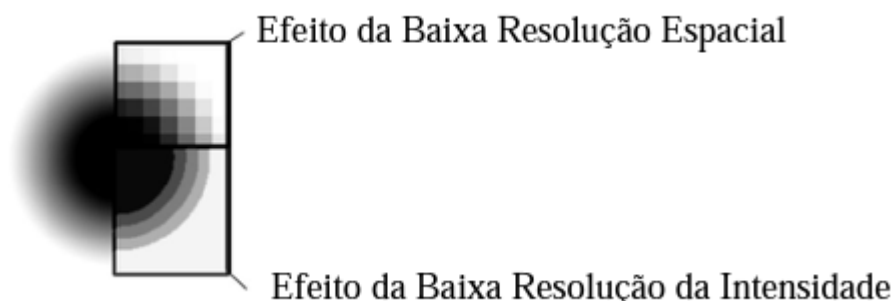
O diagnóstico por imagens tem passado por grandes transformações nos últimos dez anos, determinadas especialmente pela revolução tecnológica no campo da informática e da descoberta de novas fontes de energia para obtenção de imagens do interior do corpo humano [1]. As imagens seccionais ou tomográficas, como a tomografia computadorizada (TC) e imagens por ressonância magnética (RM), reúnem muitos princípios e qualidades da modalidade ideal, o que inclui a obtenção de imagens eficientes, precisas em relação aos aspectos anatômicos, com riqueza de detalhes tridimensionais [1].

O diagnóstico por imagem continua a ser uma das melhores maneiras de tratar pacientes sem nenhum efeito colateral prejudicial e capaz de ver o que está acontecendo dentro do corpo sem a necessidade de cirurgia ou outros procedimentos invasivos [2]. Logo, é de extrema importância que a qualidade dessas imagens seja a mais alta possível, facilitando para a análise do especialista e evitando erros na hora do diagnóstico.

A qualidade de uma imagem digital está diretamente relacionada à sua resolução. A palavra “resolução”, neste contexto, é definida como o nível de detalhe de uma representação visual [5, 14, 19]. Quanto maior a resolução de uma imagem digital, mais precisa é sua representação da imagem real.

A resolução de uma imagem digital é representada por dois parâmetros: a resolução espacial, que especifica o número de amostras, elementos ou pixels, em uma imagem, e a resolução da intensidade, ou quantização, que especifica o número de níveis possíveis para representar a intensidade em determinado pixel. A Figura 1 ilustra o efeito da baixa resolução espacial e da baixa resolução da intensidade.

Figura 1 - Efeitos visuais da baixa resolução espacial e da baixa resolução da intensidade.



Fonte: Extraído de [58].

Imagens com alta resolução (HR) são fundamentais em várias aplicações. Alguns exemplos práticos são: o diagnóstico por imagem pode detectar doenças ainda em estágios iniciais [15, 16, 19]; a astronomia pode revelar novas estrelas e galáxias nunca antes visíveis [17, 19], e o sensoriamento remoto pode ser feito a distâncias ainda maiores.

Entretanto, nem sempre é vantajoso, ou mesmo possível, melhorar a resolução de um dispositivo de aquisição de imagens. Utilizar lentes de melhor qualidade, sensores mais sofisticados e mecanismos mais precisos pode elevar significativamente o custo de um dispositivo. Os fabricantes de sensores têm demonstrado que o aumento da densidade de células detectoras num sensor é proporcional ao aumento do ruído de aquisição e leva, conseqüentemente, ao decréscimo na qualidade da imagem capturada [18, 5, 19]. Além disso, deve-se considerar a grande quantidade de imagens já capturadas de cenários que podem não se repetir mais, nestes casos, a aplicação de processamento digital para melhorar a resolução da imagem é fundamental.

Neste trabalho são consideradas as técnicas de processamento digital chamadas de “super-resolução” (SR) [5, 4], cujo objetivo é melhorar a resolução das imagens capturadas, aprimorando os detalhes nas imagens originalmente obtidas.

O uso de técnicas de inteligência computacional que podem gerar modelos estatísticos computacionais capazes de lidar com tarefas complexas é cada vez mais

notável. As Redes Adversárias Generativas (GANs - *Generative adversarial networks*) estão atraindo um interesse crescente na comunidade de aprendizagem profunda. GANs pertencem à família de modelos generativos em aprendizado de máquina. Em comparação com outros modelos generativos, por exemplo, autoencoders variacionais, As GANs oferecem vantagens como uma capacidade de lidar com funções de densidade estimadas nítidas, gerando amostras desejadas com eficiência, eliminando viés determinístico e com boa compatibilidade com a arquitetura neural interna. Essas propriedades permitiram que as GANs usufríssem de grandes sucessos especialmente no campo da visão computacional, por exemplo, geração de imagem plausível, tradução de imagem para imagem, super-resolução (SR) de imagem, e finalização de imagem [3].

Existem técnicas de SR que utilizam GANs no processo de construção de imagens em HR, Ledig et al. [13] propuseram empregar pela primeira vez as técnicas de GAN para o problema da SR, aproveitando das características inerentes a esse modelo de redes, e assim superando as redes existentes e se tornando o *State of the Art* (SOTA) (em português, Estado da Arte). Essa nova rede é denominada de Rede Adversária Gerativa de Super-Resolução (SRGAN - *Super-Resolution Generative Adversarial Network*).

Sendo assim, este trabalho propõe a utilização de Redes Adversárias Gerativa de Super-Resolução (SRGANs - *Super-Resolution Generative Adversarial Networks*) em imagens de exames médicos facilitando a análise do especialista e gerando uma maior acessibilidade aos procedimentos, visto que, o algoritmo se propõe a realizar a melhoria das imagens com baixa qualidade, o que poderia gerar um barateamento de todo custo associado ao processo.

1.1 Motivação

O presente trabalho buscar contribuir para o uso de algoritmos de inteligência computacional aplicados a área da saúde, buscando incentivar o uso de redes adversárias geradoras como ferramenta de auxílio aos profissionais da área, com ênfase em facilitar a análise de imagens de exames médicos e também diminuir número de casos de erro durante o processo de análise. Além disso contribuir com o

barateamento do custo associados aos procedimentos de diagnóstico por imagem, fazendo com que estes procedimentos se tornem mais acessíveis a população de baixa renda, o que pode vir a gerar uma melhor qualidade de vida para as pessoas que se encontram nessa situação.

Para mais, busca-se contribuir com o avanço do uso de GANs aplicado a área de saúde e diagnóstico e incentivar pesquisa no âmbito científico a respeito deste tema, utilizando esse trabalho como base para trabalhos futuros.

1.2 Objetivos

1.2.1 Objetivo geral

O presente trabalho tem como objetivo geral avaliar o uso de SRGANs no processo de ampliação e remoção de distorções de imagens aplicado a imagens de exames médicos.

1.2.2 Objetivos específicos

Para sucesso do objetivo geral, alguns objetivos específicos são estipulados:

- Validação da hipótese a respeito do uso de redes adversárias geradoras na implementação de técnicas de super-resolução em imagens de exames médicos.
- Definição de um modelo do estado da arte que se adequa para o aumento de resolução de imagens de exames médicos.
- Validação dos resultados obtidos com profissionais e estudantes da área de biomedicina e afins.

1.3 Estrutura do trabalho

No Capítulo 2 e 3 serão apresentadas revisões a respeito dos temas que este trabalho engloba para que o leitor possa se habituar com o contexto atrelado ao trabalho, expondo um pouco sobre aprendizado de máquina, redes neurais artificiais, redes neurais convolucionais e redes geradoras adversárias. No Capítulo 4, é

detalhada toda a metodologia que foi definida para a execução do trabalho e no Capítulo 5 são apresentados as análises e os resultados propostos a respeito do tema trabalhado. No Capítulo 6 são apresentadas as conclusões e trabalhos futuros.

2 Aprendizado de Máquina

Sendo um campo de estudo da inteligência artificial, o aprendizado de máquina representa uma forma de estatística aplicada à construção de algoritmos que permitam a uma máquina aprender, isto é, melhorar seu desempenho em determinada tarefa por meio da experiência [91]. O processo de aprender sem ser explicitamente programado possibilita que o programa se desenvolva automaticamente, por meio da observação de um conjunto de dados e elaboração de padrões que motivem a máquina a tomar decisões no futuro com base nos dados previamente fornecidos, após o devido processo de treino [92].

Quando o conjunto de dados provido na entrada é acompanhado de seus respectivos rótulos de saída, tem-se um algoritmo supervisionado, o qual produz uma função de aproximação para prever valores futuros. Ao comparar os resultados obtidos com os valores corretos, o programa pode modificar seus parâmetros, por meio de um processo de treino, para melhor representar os dados escolhidos [45]. Busca-se assim aprender uma função $f: X \rightarrow Y$ dado um conjunto rotulado de treinamento $\{ \langle x_i, y_i \rangle \}$ de entradas x_i e saídas $y_i = f(x)$ [92].

Em contrapartida, algoritmos não supervisionados abordam problemas cuja estrutura dos dados de entrada não é conhecida. Não há rótulos, de forma que cabe à máquina identificar os padrões que melhor representem relações entre as variáveis fornecidas, identificando arquétipos ocultos nestas estruturas [45].

Outras categorias estão presentes ao tratar de aprendizado de máquina, como aprendizado semi-supervisionado e por reforço, porém não são relevantes ao desenvolvimento deste projeto.

2.1 Redes Neurais Artificiais

Redes Neurais Artificiais (ANN) são algoritmos capazes de usar o aprendizado de máquina em uma estrutura inspirada nos sistemas nervosos biológicos para a aproximação de funções. Tais redes são constituídas por camadas de elementos

interconectados - os neurônios - através de parâmetros ajustáveis, os quais fazem paralelo às sinapses cerebrais [93].

De acordo com Haykin [94], uma rede neural artificial se assemelha ao cérebro em dois aspectos:

1. Conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem.
2. Forças de conexão entre neurônios, chamadas de pesos sinápticos, são usadas para armazenar este conhecimento.

ANNs podem, teoricamente, aproximar qualquer função em um domínio compacto se fizerem uso de pelo menos uma camada invisível, atuando assim como aproximador universal [95]. Tal propriedade confere à rede um imenso poder representacional, uma vez que, quanto mais profundo é o modelo, maior é a sua capacidade de aproximar funções complexas. Com a adição de camadas ocultas ao sistema, cada nível transforma seus dados de entrada em uma representação mais abstrata do que o nível anterior, aprendendo assim as características ótimas [96]. Essas redes profundas, ou *deep networks*, são o estado da arte em diversos problemas de aprendizado de máquina com extração implícita de características e servem como estímulo para uma gama de diferentes arquiteturas.

2.1.1 Perceptron

Em sua forma mais simples, uma rede neural age como um único neurônio artificial, por meio de uma operação de classificação binária dada pelo *perceptron*. Atuando como um classificador linear, este neurônio mapeia um vetor de entrada $x = [x_1, x_2, \dots, x_m]$ em uma função binária y , submetendo inicialmente o vetor x a uma soma ponderada com os pesos de sinapse $\theta = [\theta_1, \theta_2, \dots, \theta_m]$, acrescida de um valor de *bias* b que apenas aumenta seu grau de liberdade, fornecendo a função soma s , conforme Equação (2.1) [97].

$$s = \sum_i x_i \theta_i + b_i$$

(2.1)

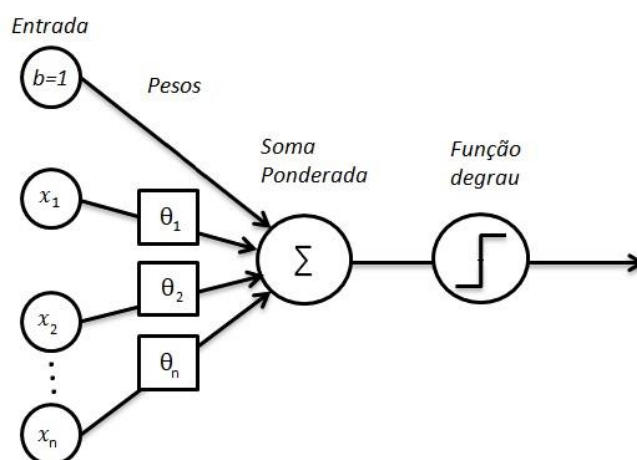
O resultado desta operação é submetido a uma função de ativação - a função degrau - mapeando s em uma saída binária conforme Equação (2.2). O *perceptron* atua, dessa forma, como um classificador linear, categorizando os dados de entrada em duas classes através de um hiperplano delimitado por $s = 0$. Caso os pesos θ associados aos respectivos valores de entrada sejam grandes o suficiente para alcançar o valor de ativação da função degrau, a sinapse é disparada e $y = 1$.

$$y = \begin{cases} 0, & \text{para } s < 0 \\ 1, & \text{para } s \geq 0 \end{cases}$$

(2.2)

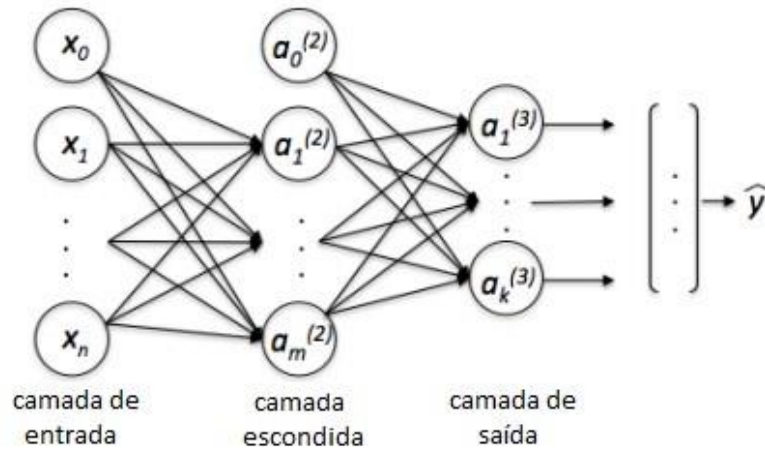
Quando dados não podem ser linearmente separados, é necessário o uso de modelos mais poderosos para realizar tais operações. Nesta situação é possível combinar *perceptrons* em uma estrutura em camadas, cada uma contendo uma quantidade diferente de neurônios. Forma-se assim uma rede neural denominada *perceptrons* multicamadas (MLP), usada no contexto deste projeto como um dos componentes que constituem uma rede geradora adversária [94]. Em um MLP, o vetor x é inserido na camada inicial cujos valores de saída servem como entradas da camada seguinte, e assim por diante, em uma estrutura totalmente conectada, na qual todos os neurônios de uma camada são ligados aos neurônios da seguinte, conforme Figura 3.

Figura 2 - Arquitetura de um *perceptron*.



Fonte: Adaptado de: Ahmet Taspinar.

Figura 3 - Exemplo de arquitetura para *perceptrons* multicamadas com uma camada escondida.



Fonte: Adaptado de: Erin LeDell.

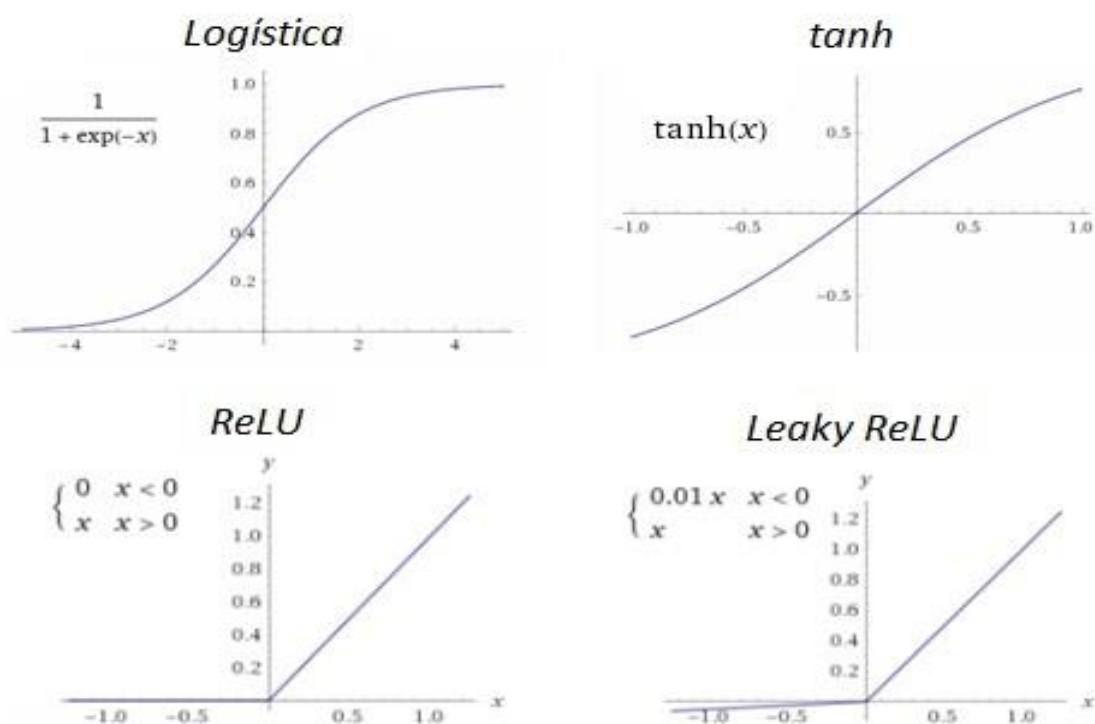
2.1.2 Funções de ativação

As funções de ativação agem na rede neural definindo o comportamento de saída de cada neurônio dado um vetor x , limitando estes valores de entrada em um intervalo definido. A função de ativação utilizada no *perceptron* é a função degrau, que fornece uma saída binária e permite apenas uma divisão linear de pontos. Em situações mais complexas e com várias classes possíveis de saída, no entanto, faz-se necessária uma função que introduza não-linearidades ao sistema, permitindo uma maior capacidade representativa do modelo [94].

As funções logística (σ), tangente hiperbólica (\tanh), ativação linear retificada ($ReLU$) e ativação linear retificada com vazamento ($Leaky ReLU$) são algumas das principais escolhas em se tratando de funções de ativação.

No escopo deste projeto, tais funções desempenham papéis específicos nas arquiteturas de redes

Figura 4 - Funções de ativação.



Fonte: Extraído de [45].

neurais utilizadas. A função logística é usada na última camada da rede, uma vez que comprime valores reais no intervalo $[0, 1]$, o que pode ser representado como a probabilidade de escolha de determinada classe, sendo assim fundamental em questões de classificação. Já a função *tanh* força os valores no intervalo $[-1, 1]$ de forma que seu centro em torno de 0 facilita o treinamento de redes em determinados casos, devido ao gradiente mais acentuado [98].

A função *ReLU* apresenta a grande vantagem de acelerar o aprendizado por apresentar gradientes facilmente calculáveis (0 ou 1, dependendo do intervalo considerado). Em se tratando do processamento de imagens, tal propriedade é relevante devido ao intenso uso computacional. *Leaky ReLU* desponta como uma variedade de tal função, permitindo a passagem de pequenos valores negativos dada por uma constante muito pequena, geralmente igual a 0,01 [99].

2.1.3 Treinamento por Backpropagation

Algoritmos de aprendizado de máquina supervisionados, sejam eles de regressão logística, Naive-Bayes, árvore de decisão ou redes neurais, agem com os mesmos dois princípios básicos: treino e predição [100]. No primeiro, os parâmetros θ do modelo são aprendidos a partir dos dados de treinamento x e y . Usando estes parâmetros é possível fazer predições durante a segunda etapa, emitindo resultados \hat{y} que devem se aproximar da saída original.

Em redes neurais, a etapa de treino é chamada de *backpropagation*, ou propagação retroativa de erro. Nesta etapa tenta minimizar-se este erro, uma função de perda H parametrizada por θ que mede a diferença entre y e \hat{y} , por meio de um método iterativo denominado descida de gradiente.

Ao iniciar-se o processo de predição de uma rede com seus parâmetros distribuídos de forma aleatória, o vetor de entrada x levará a uma saída \hat{y} que difere do vetor y desejado, indicando assim um valor elevado para a função de perda, que pode assumir diversas formas, dependendo do tipo de rede trabalhada.

Redes supervisionadas de classificação usualmente têm como função de perda a função de entropia-cruzada [101], dada por

$$H(y, \hat{y}) = - \sum_i y_i \log(\hat{y}_i) \tag{2.3}$$

em que y e \hat{y} indicam as distribuições "real" e "estimada", respectivamente. Mede-se assim a distância entre esses dois vetores discretos de probabilidades, estimando o quanto a distribuição estimada se aproxima do real em um universo com i classes distintas.

Como a função de perda depende dos parâmetros da rede, o método do gradiente pode ser aplicado para minimizar H de forma iterativa, em que, a cada passo, θ é atualizado na direção negativa do gradiente, correspondente à direção de máximo declive:

$$\nabla H = \left(\frac{\partial H}{\partial \theta_1}, \dots, \frac{\partial H}{\partial \theta_i} \right) \quad (2.4)$$

$$\theta_i := \theta_i - \alpha \frac{\partial H}{\partial \theta_i} \quad (2.5)$$

O termo α recebe o nome de taxa de aprendizado e define o tamanho de cada passo tomado no processo. Para uma rede neural profunda, os pesos das camadas ocultas devem ser atualizados de acordo com os valores dos pesos provenientes de seu neurônio mais à direita, o que pode ser visto ao tomar-se a regra da cadeia na Equação (2.4), de modo a propagar o erro inicial desde a saída até os nós iniciais da rede.

Uma vez que os parâmetros θ tenham sido atualizados, entra-se novamente na etapa de predição em busca de satisfazer $\hat{y} = y$.

Atualmente os métodos de treinamento evoluíram e muitas outras abordagens mais eficientes e que lidam com problemas de treinamento foram propostas, e.g. ADAM [18], RPROP [46], etc.

2.2 Redes Neurais Convolucionais

Um tipo específico de rede profunda recebe o nome de rede neural convolucional (CNN), desenvolvida especialmente para reconhecer formas bidimensionais com um elevado grau de invariância à translação, redimensionamento, inclinação e outras formas de distorção [102]. CNNs têm estruturas muito semelhantes às ANNs ordinárias, compostas por neurônios ponderados por parâmetros modificáveis que recebem um vetor de entrada e realizam operações não-lineares em busca de classificá-lo corretamente. No entanto, CNNs fazem uso da convolução, operação matemática entre duas funções - no âmbito do processamento de imagens, estas equivalem a uma imagem e a um filtro - que altera a intensidade de um pixel para refletir as intensidades dos pixels que o circundam [45].

Várias arquiteturas de redes convolucionais foram propostas ao longo dos anos, porém neste projeto faz-se uso da *All Convolutional Network* - Rede Totalmente Convolucional [103], cujo esqueleto consiste em repetidas camadas de convolução sobrepostas, seguidas por algumas camadas totalmente conectadas.

2.2.1 Camada de Convolução

A principal proposta da convolução é a de extrair características da imagem de entrada, por meio de pequenas matrizes (usualmente, não maiores que 5x5 [103]) denominadas filtros ou *kernels*, os quais percorrem toda a imagem em um processo de convolução, uma soma ponderada que expressa como as características da figura são modificadas pelo filtro. Para isso, cada pixel da imagem original é pareado com o filtro, incluindo também os pixels de fronteira, como vê-se na Figura 5. Cada elemento do *kernel* é então multiplicado pelo par correspondente, somando-se os produtos para gerar o novo valor de saída [45].

Considerando uma imagem 3D, a convolução entre o pixel p e o filtro f pode ser vista na Equação (2.6), a qual se assemelha muito à Equação (2.1) para o cálculo dos parâmetros θ de uma rede neural. Tal propriedade não é por acaso, uma vez que o filtro da CNN é composto por parâmetros de sinapse treináveis da mesma forma que para um MLP, porém agora em uma disposição tridimensional, organizada em uma estrutura de pesos compartilhados e não mais totalmente conectados, já que os filtros são replicados ao longo de toda a imagem de entrada [45].

$$\text{conv}(p * f)[i, j] = \sum_{k_1=1}^K \sum_{k_2=1}^K \sum_{c=1}^C p[i - k_1, j - k_2, c] f[c, k_1, k_2] \quad (2.6)$$

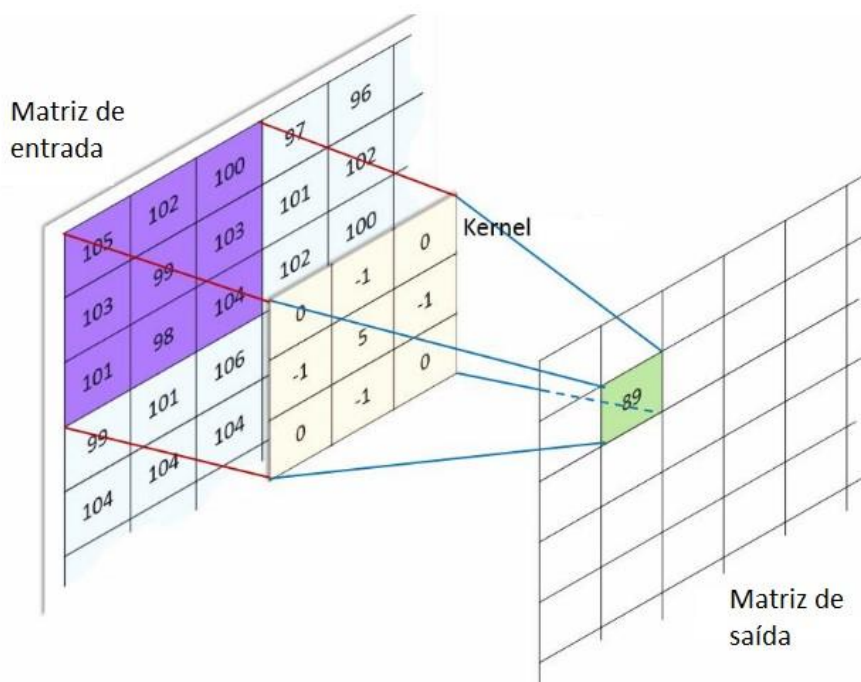
As dimensões da imagem de saída após uma camada de convolução dependem de alguns parâmetros específicos:

- **número de filtros utilizados:** para um caso com n filtros, a imagem C_{out} apresenta n canais de cores.
- **zero-padding:** representa a adição de zeros (pixels nulos) às bordas da imagem de entrada, normalmente para permitir que C_{out} tenha as mesmas dimensões espaciais que C_{in} .

- **stride**: controla quantos pixels são pulados a cada produto escalar no processo de convolução, regulando também as dimensões espaciais de C_{out} . Para $stride = 2$, por exemplo, são pulados 2 pixels na Equação (2.6).

CNNs clássicas fazem uso de camadas de pooling intercaladas com as camadas de convolução descritas, uma técnica que busca redimensionar C_{out1} para uma imagem C_{out2} com resolução mais baixa, buscando achar uma representação ideal com uma quantidade menor de parâmetros. Na Rede Totalmente Convolutiva, essa abordagem é substituída pelo simples uso de strides > 1 [45].

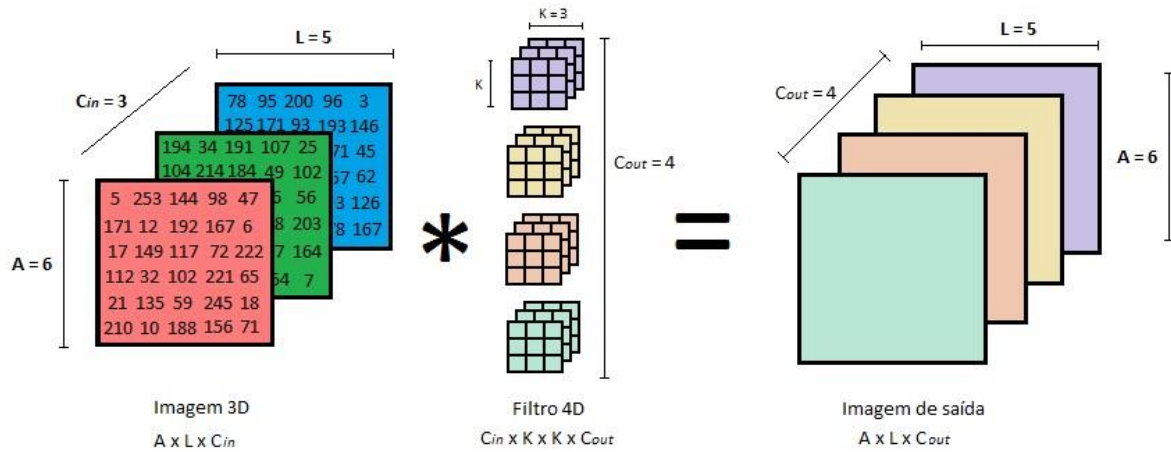
Figura 5 - Exemplo de convolução para pixel (de intensidade 99) em imagem com 1 canal de cor



Fonte: Adaptado de: Machine Learning Guru.

Este pixel e os demais ao seu arredor são multiplicados pelos elementos correspondentes do filtro e então somados, gerando um valor de saída. O processo se repete para todos os pixels da imagem.

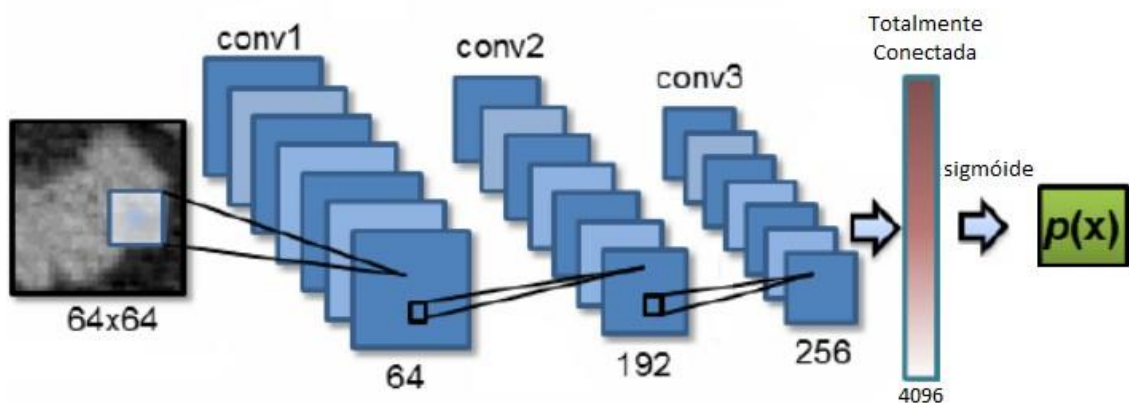
Figura 6 - Exemplo de camada de convolução.



Fonte: Extraído de [58].

Uma imagem de entrada 3D, com 6 pixels de altura, 5 de largura e 3 canais de cores, sofre convolução com um filtro 4D, de altura e largura $K = 3$, gerando uma imagem de saída com o mesmo número de canais C_{out} que a quantidade de filtros utilizada, 4. Este processo se repete várias vezes ao longo da rede convolucional.

Figura 7 - Esqueleto completo de uma CNN da forma Rede Totalmente Convolucional.



Fonte: Adaptado de: Le Lu.

3 Redes Geradoras Adversárias

3.1 Modelagem Geradora

A aprendizagem profunda ou *deep learning* constitui área de pesquisa da inteligência artificial que faz uso de diversas camadas não-lineares para descobrir padrões que modelem corretamente uma distribuição de probabilidade, usando de diferentes níveis de abstração como visto no capítulo anterior.

Os principais estudos na aprendizagem profunda têm seu foco em modelos discriminadores, os quais buscam mapear dados de entrada multidimensionais x em uma classe de saída y por meio da probabilidade condicional $p(y|x;\theta)$. Um *perceptron*, por exemplo, modela $p(y|x;\theta)$ como $\hat{y}_\theta(x) = g(\theta^T x)$ em que g é a função degrau. As ANNs e CNNs vistas até aqui também se encaixam nessa categoria, de forma a encontrar apenas uma fronteira que separe corretamente os dados de entrada em classes [105].

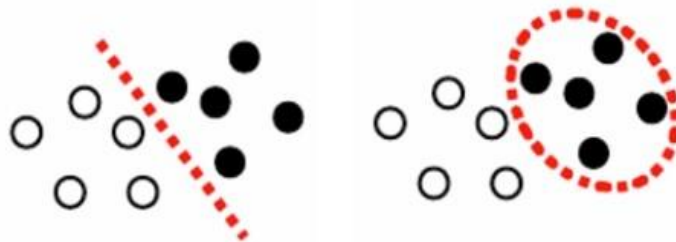
Considere um problema de classificação no qual deseja-se diferenciar entre gatos ($y = 0$) e cães ($y = 1$), com base em alguma característica dos animais. Dado os pares de treinamento (x,y) , o algoritmo tenta encontrar uma linha ou curva (a fronteira de decisão) que separe os cães e gatos, como na Figura 8. Em seguida, averigua em qual lado da fronteira o próximo exemplo está localizado e o classifica de acordo. Tais algoritmos tentam assim responder à pergunta "dado estas entradas qual a probabilidade de se ter esta classe como a mais provável?".

Modelos geradores, por outro lado, tentam aprender a probabilidade conjunta $p(x,y)$, a qual pode ser usada em modelos de classificação através da conversão para $p(x|y)$. Usando do mesmo exemplo em que y denota cães (1) e gatos (0), $p(x|y = 1)$ modela a distribuição das características dos cães $p(x|y = 0)$ modela a dos gatos. O modelo gerador pode, em seguida, usar estes valores para encontrar $p(y|x)$ indiretamente pelo teorema de Bayes (Equação 3.1), servindo como ferramenta para problemas de classificação.

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \quad (3.1)$$

O verdadeiro potencial de modelos geradores, no entanto, está em sua capacidade de entender estruturas dos dados de entrada e criar novos pares prováveis (x,y) em um processo de geração, à medida em que foca nas classes em busca de assimilar suas características, e não apenas a fronteira que as separa. Este processo pode ocorrer de duas maneiras, ambos seguindo uma estimativa por máxima verossimilhança: modelos geradores explícitos, os quais definem uma distribuição p_{model} que se aproxima da distribuição de amostras p_{data} e modelos geradores implícitos, que aprendem uma função que faça amostragens de p_{model} sem defini-la diretamente. Redes geradoras adversárias, o foco deste trabalho, se encaixam no segundo grupo.

Figura 8 - Modelos discriminadores e modelos geradores.



Fonte: Adaptado de: Nigel Goddard.

3.2 Modelo para Redes Geradoras Adversárias

Para estimar modelos geradores, foi criada uma arquitetura que treina duas redes simultaneamente, em processo adversário entre um modelo discriminador D e um gerador G chamado de redes geradoras adversárias (GAN). Nesta estrutura, o objetivo do gerador é criar amostras que sejam idealmente indistinguíveis de amostras reais, com o nível de semelhança entre elas sendo medido pelo discriminador. No caso específico em que os modelos são representados por redes neurais de

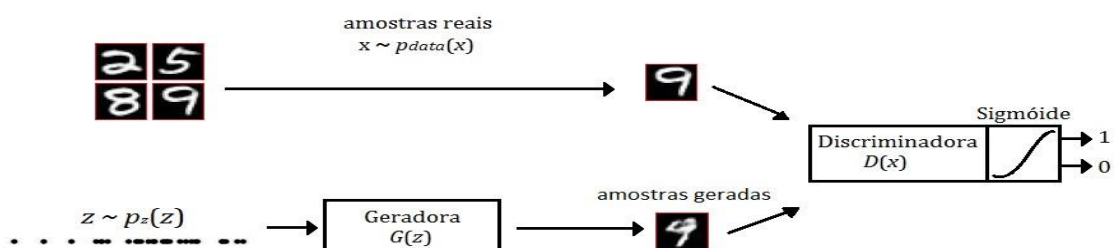
perceptrons multicamadas (MLP), ambos podem ser treinados com o uso do algoritmo *backpropagation* [104].

Para tal, define-se uma distribuição conhecida p_z da qual são amostradas variáveis latentes de ruído z , representado por $z \sim p_z$. Usando uma rede neural $G(z, \theta_g)$ parametrizada por θ_g que tem como entrada o vetor z , é possível treiná-la para criar amostras de sua distribuição geradora implícita p_g . A rede discriminadora $D(x, \theta_d)$ tem θ_d como parâmetros e recebe como entrada imagens x_i amostradas de uma distribuição desconhecida $p_{data}(x)$ ou geradas por $G(z)$, retornando a probabilidade de que a imagem em questão seja proveniente de p_{data} . $D(x)$ dessa forma equivale a valores próximos a 1 quando a imagem vem de p_{data} e próximos a 0 quando a imagem é falsa, vinda de p_g , arquitetura vista na imagem 3.2. No caso ideal correspondente ao Equilíbrio de Nash do sistema, tem-se $p_g = p_{data}$.

Vê-se assim que a rede discriminadora atua de forma supervisionada (com rótulos) para auxiliar a rede geradora, em um processo não-supervisionado. Um ponto interessante que surge como consequência desta configuração é que $G(z)$ têm como ferramenta para aprimorar suas amostras apenas o *feedback* obtido por $D(x)$, sem nunca ter acesso direto às amostras reais.

Uma analogia usada por Goodfellow [104] para esclarecer o funcionamento de GANs é comparar a geradora a um falsificador, que gera notas falsas enquanto que a discriminadora tem papel análogo à polícia, tentando diferenciar as falsificações dos originais. A competição entre os dois leva ambos a melhorarem seus métodos de falsificação e identificação, respectivamente, repetindo o processo até que as notas reais e falsas sejam indistinguíveis com $D(x) = 1/2$.

Figura 9 - Modelo original da GAN.



Fonte: extraído de [45].

3.3 Funções de Perda

3.3.1 Rede Discriminadora

Redes supervisionadas de classificação comumente fazem uso da função de entropia-cruzada dada na Equação (2.1) como função de perda, considerando um sistema com i diferentes classes. A rede discriminadora atua com apenas 2 classes (imagens reais e geradas) de forma a otimizar uma função treinável $D: \mathbb{R}^n \rightarrow [0, 1]$ em que D indica a probabilidade de algum ponto $x_i \in \mathbb{R}^n$ pertencer à primeira classe. Neste caso, a função de perda do discriminador pode ser simplificada pela entropia cruzada binária. Tomando $(x_i, y_i) \in (\mathbb{R}^n, \{0, 1\})$ como os pares de dados de entrada em D e considerando um dataset de N imagens, tal função pode ser representada por

$$H((x_i, y_i)_{i=1}^N, D) = - \sum_{i=1}^N y_i \log D(x_i) - \sum_{i=1}^N (1 - y_i) \log(1 - D(x_i)) \quad (3.2)$$

A equação acima apresenta dois termos distintos, e um deles sempre será reduzido a 0 de acordo com o valor de y_i . $D(x_i)$ indica a probabilidade que uma imagem x_i seja real, já a probabilidade que a imagem tenha sido gerada é dada por $(1 - D(x_i))$. Os valores de D , dessa forma, devem estar contidos no intervalo $[0, 1]$, o que é possível com o uso da função logística na última camada do discriminador.

Ao trabalhar com GANs é costumeiro estabelecer algumas particularidades sobre os dados utilizados. Sabe-se que as imagens x_i provêm necessariamente de duas fontes possíveis: $x_i \sim p_{data}$ para a distribuição real ou $x_i = G(z)$ com $z \sim p_{gerador}$, para a distribuição do gerador. Considerando o caso ideal no qual cada fonte provê metade dos dados e substituindo os somatórios por expectativas para uma conotação probabilística a Equação (3.2) passa a representar a função de perda J^D do discriminador [106], dada por

$$J^D(\theta_d, \theta_g) = H((x_i, y_i)_{i=1}^\infty, D) = -\frac{1}{2} \mathbb{E}_{x \sim p_{data}} [\log D(x)] - \frac{1}{2} \mathbb{E}_z [\log(1 - D(G(z)))] \quad (3.3)$$

Ao minimizar a entropia cruzada binária em respeito aos parâmetros de rede θ_d e θ_g , o discriminador tem como missão diferenciar corretamente as imagens reais das geradas. Este processo é dado pelo método do gradiente, descrito no capítulo anterior.

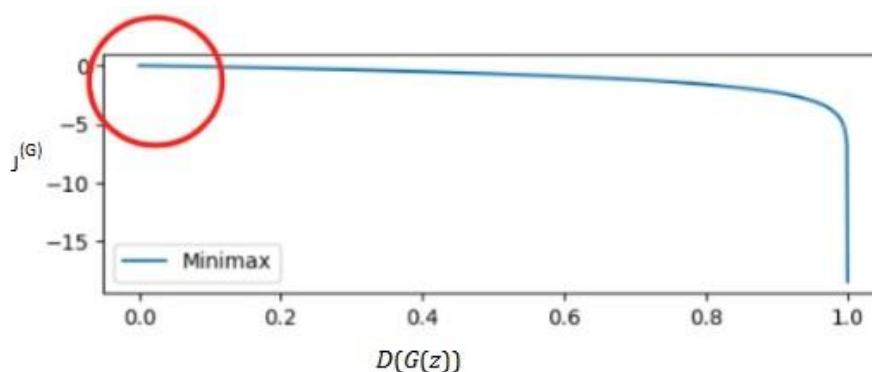
3.3.2 Rede Geradora

A rede geradora deve maximizar sua habilidade em enganar a Discriminadora, cenário no qual sua função de perda pode ser dada, teoricamente, por

$$J^G = -J^D \quad (3.4)$$

Nesta situação tem-se um caso ideal descrito na teoria dos jogos como *minimax* ou jogo de soma-zero, o qual, como o nome já diz, ocorre quando a soma dos custos de todos os jogadores é sempre zero. Na prática, porém, a escolha desta J^G não resulta em bons resultados [106] uma vez que inicialmente as amostras geradas por $G(z)$ são facilmente distinguíveis por $D(x)$ das amostras originais, situação em que $D(G(z))$ se aproxima de 0 e tem-se o problema de desaparecimento dos gradientes. Nesse caso, as derivadas da função de perda, calculadas pelo método de descida de gradiente para otimizá-la, atingem valores próximos de zero e os parâmetros de rede não são atualizados [106], caso do gráfico da figura 10.

Figura 10 - Gráfico da função de erro J^G ao se adotar $J^G = J^D$.



Fonte: Extraído de Lazy Programmer.

Na porção destacada, o gradiente da curva atinge valores muito pequenos, levando ao problema de desaparecimento de gradientes e impossibilidade de atualização de parâmetros via *backpropagation* [93].

Para solucionar tal problema, adota-se uma função heurística para a perda do gerador, que não mais tenta maximizar o erro de entropia cruzada do discriminador, mas sim diminuir sua própria entropia por meio da inversão de rótulos. $G(z)$ neste caso interpreta as imagens reais como 0 e as imagens geradas como 1, levando à nova função de perda J^G , vista na Equação (3.5).

$$J^G = -\frac{1}{2}\mathbb{E}_z[\log D(G(z))] \quad (3.5)$$

Da mesma forma que o discriminador, o gerador também apresenta função de perda diferenciável e pode ser treinado por meio de *backpropagation*. J^G e J^D no cenário proposto deixam de representar um jogo de soma-zero.

3.4 Redes Convolucionais Geradoras Adversárias Profundas

GANs convolucionais profundas (DCGAN) foram inicialmente sugeridas como modelo alternativo para melhor estabilizar o treinamento de redes adversárias [107]. Embora GANs já apresentassem camadas profundas e de convolução antes das DCGANs, seu nome é usado para referir a um tipo específico de arquitetura, que faz uso de algumas diretrizes especiais:

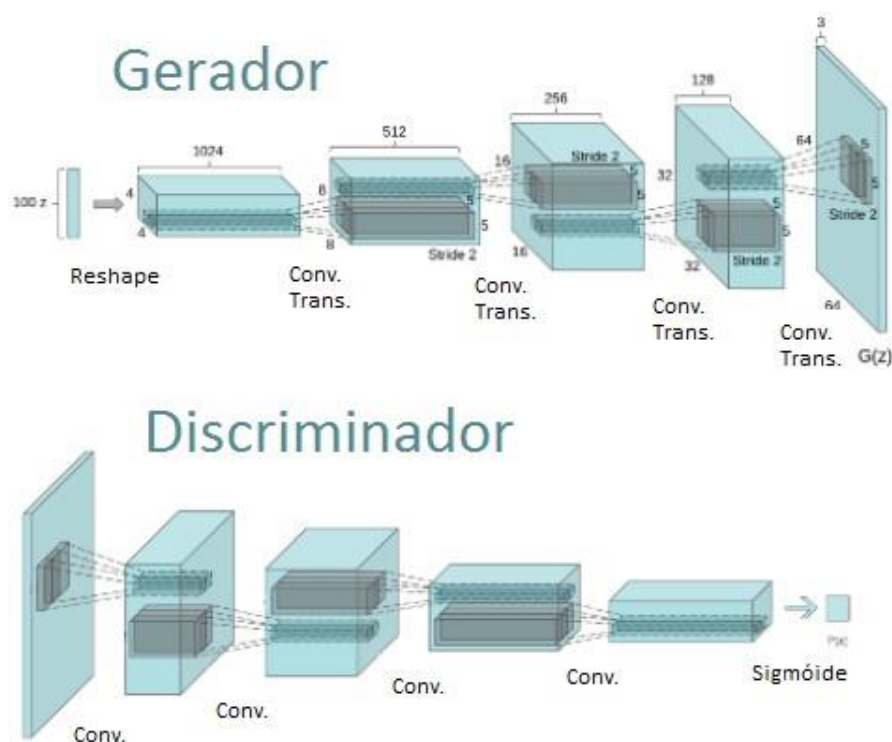
1. Uso de *Batch Normalization* - normalização em lote, técnica que consiste no redimensionamento dos dados na entrada de todas as camadas da rede, permitindo o uso de taxas de aprendizado maiores.
2. Uso da rede totalmente convolucional como ferramenta para simplificar o processo computacional, acelerando o processo de treinamento. Para aumentar a resolução da imagem, é usado um tipo especial de convolução denominado convolução transposta.
3. Utilização do algoritmo especial para método do gradiente denominado *Adam* [18].

4. Eliminação de camadas totalmente conectadas no gerador.
5. Aplicar a função de ativação *ReLU* em todas as camadas do gerador, com exceção da última a qual faz uso da função tangente hiperbólica *tanh*.
6. Aplicar a função de ativação *LeakyReLU* em todas as camadas do discriminador, com exceção da última a qual faz uso da função logística.

Essas instruções são aplicadas à rede discriminadora para formar uma rede convolucional semelhante à da figura 7, recebendo uma imagem x e fornecendo a probabilidade dela vir de $pdata$.

A rede geradora efetua pequenas modificações neste processo, uma vez que sua camada inicial recebe um vetor z ao invés de uma imagem, aplicando sobre este um algoritmo de redimensionamento (*reshape*) para torná-lo tridimensional. Para aumentar a resolução da imagem são usadas camadas de convolução transposta, representadas por convoluções com $stride < 1$. Isso pode ser entendido como a adição de pixels nulos entre os pixels originais da imagem antes de sua passagem pelo filtro, dilatando-a.

Figura 11 - Exemplo de esqueleto para uma DCGAN.



Fonte: Adaptado de [107].

3.5 Super-resolução de imagens

As técnicas de super-resolução proporcionam uma forma de obter imagens com resolução superior à fornecida pelo dispositivo de aquisição [60, 61, 62, 63, 64]. Estas técnicas removem distorções causadas durante o processo de aquisição, realçam características importantes nas imagens, como as bordas dos objetos, e, ainda, recuperam detalhes pela utilização da informação contida em um grande conjunto de imagens capturadas.

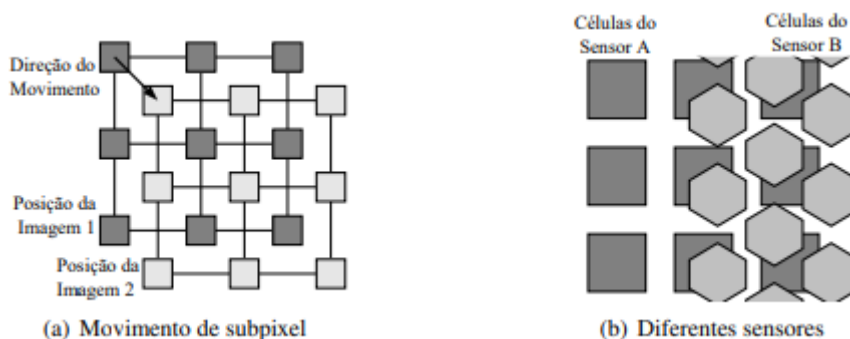
A super-resolução teve origem na óptica, na qual se busca recriar as frequências além do limite de difração, empregando extrapolação das altas frequências espaciais [62]. De acordo com [65], ainda na física, pesquisadores demonstraram que era possível recuperar a resolução espacial utilizando a informação distribuída no tempo, ou seja, a informação contida em um conjunto de imagens capturadas ao longo do tempo [66, 67, 68].

A obtenção da super-resolução através da utilização de um conjunto de imagens é uma das formas mais exploradas atualmente [63, 64, 60, 61]. Assume-se que as imagens capturadas, chamadas de imagens de baixa resolução (LR - *low-resolution*), são altamente correlacionadas. Porém, uma pequena diferença de informação presente em cada imagem capturada proporciona uma nova informação para a recuperação dos detalhes na imagem HR. Os exemplos mais comuns de obtenção dessa diferença de informação são:

- o movimento de subpixel [61], que ocorre quando o movimento entre as imagens é fracionário com respeito à grade de baixa resolução;
- as diferentes distorções na aquisição [69, 70], que ocorrem quando, por exemplo, a aquisição é feita por sensores diferentes ou por câmeras diferentes.

A Figura 12 ilustra como a posição relativa do arranjo de células durante o processo de aquisição proporciona uma observação única da imagem HR. Neste trabalho será enfatizada a super-resolução obtida por movimento de subpixel, embora a formulação inclua as duas possibilidades.

Figura 12 - Ilustração dos casos mais comuns de obtenção de diferentes informações entre as imagens capturadas.



Fonte: Extraído de [58].

As técnicas de super-resolução diferenciam-se de técnicas simples de ampliação digital de imagem, como as técnicas de interpolação [71, 72], principalmente porque estas não tentam remover distorções da aquisição. As técnicas de super-resolução têm uma formulação semelhante à das técnicas de restauração de imagem [73], porém são diferentes quanto aos objetivos. A restauração de imagem não busca aumentar a resolução, mas, apenas, remover as distorções que degradaram a imagem. A super-resolução pode ser vista como uma técnica mista de restauração e interpolação, pois proporciona ampliação e remoção de distorções.

4 Materiais e métodos

4.1 Análise do estado da arte

Inicialmente foram realizados estudos a respeito do estado da arte das redes adversárias geradoras com foco em entender sobre os modelos atuais de GANs e suas variações, foram analisados os trabalhos [22, 23, 24, 25, 26, 27]. Com isso foi voltada a atenção para um ramo específico de rede adversarias geradoras, as SRGANs.

Após analisar diversos trabalhos e executar várias implementações de modelos, foi feita uma análise macroscópica dos resultados obtidos nos modelos [41, 42, 43, 44] ao se utilizar a mesma imagem de entrada, o modelo [44] apresentou mais clareza nos detalhes e contornos e uma maior fidelidade quando comparada a imagem resultante gerada pela GAN com a imagem de entrada, além disso o modelo foi treinado com diversas imagens que representam o mundo real o que gerou uma maior capacidade de generalização do modelo diante de quaisquer imagens de entrada além da sua arquitetura robusta, permitindo avaliar e tratar diversos tipos de distorções que podem vim a ser introduzidas nas imagens.

4.2 Escolha do modelo

O modelo escolhido foi o Real-ESRGAN [44], este modelo estende o poderoso ESRGAN [39] para restaurar imagens LR gerais do mundo real sintetizando pares de treino com uma degradação mais prática, este modelo se propõe a tratar degradações complexas geralmente advindas de combinações complicadas de diferentes processos de degradação. Wang [44] traz como exemplo “quando tiramos uma foto com nossos celulares, as fotos podem ter várias degradações, como desfoque da câmera, ruído do sensor, nitidez de artefatos e compressão JPEG. Em seguida, fazemos algumas edições e fazer upload para um aplicativo de mídia social, que apresenta mais compressão e ruídos imprevisíveis. O processo acima torna-se mais complicado quando a imagem é compartilhada várias vezes na Internet.”

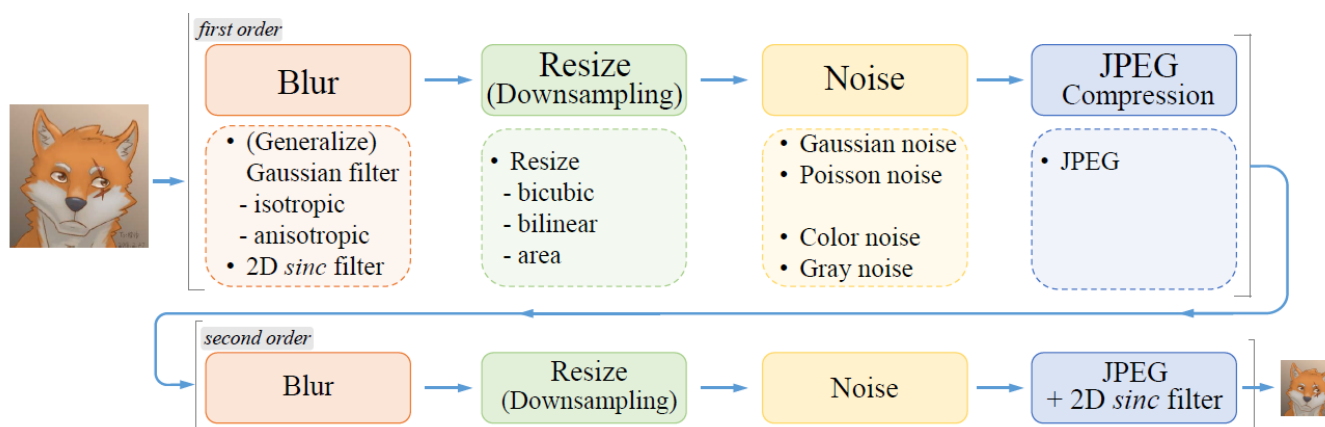
Sendo assim, foi estendida o clássico modelo de degradação de “primeira ordem” para modelagem de degradação de “alta ordem” para degradações do mundo real, ou seja, as degradações são modeladas com vários processos de degradação repetidos, cada um processo sendo o modelo clássico de degradação. Empiricamente, adotou-se um processo de degradação de segunda ordem para um bom equilíbrio entre simplicidade e eficácia.

Além disso, como o espaço de degradação é muito maior que o do ESRGAN, o treinamento também se tornou mais complexo, o discriminador requer uma capacidade mais poderosa para discriminar realismo de saídas de treinamento complexas, enquanto o feedback gradiente do discriminador precisa ser mais preciso para aprimoramento de detalhes locais. Portanto, foi melhorado o discriminador de estilo VGG do ESRGAN para um Projeto U-Net [38, 40, 37]. A estrutura U-Net e degradações complicadas geraram um aumento na instabilidade do treinamento.

Assim, foi empregado a normalização espectral de regularização [36, 38] para estabilizar a dinâmica de treinamento. Com essas melhorias dedicadas, tornou-se possível treinar facilmente o Real-ESRGAN e alcançar um bom equilíbrio de aprimoramento de detalhes locais e supressão de artefatos.

Em resumo, o modelo, 1) propõe-se um processo de degradação para modelar degradações práticas, e a utilização de filtros *sinc* para modelar o toque comum e o *overshoot* de artefatos. 2) Empregasse várias modificações essenciais (por exemplo, discriminador *U-Net* com normalização espectral) para aumentar a capacidade discriminadora e estabilizar o treinamento dinâmica. 3) Real-ESRGAN foi treinado com dados sintéticos puros sendo capaz de restaurar a maioria das imagens do mundo real e alcançar melhor desempenho visual do que os modelos anteriores, tornando-o mais prático em aplicações do mundo real [44].

Na figura 13, pode-se visualizar como o modelo utiliza um processo de degradação de segunda ordem para modelar degradações mais práticas, onde cada processo de degradação adota o modelo de degradação clássico. As opções detalhadas para desfoque, redimensionamento, ruído e compactação JPEG são listadas. Também foi empregado o filtro *sinc* para sintetizar artefatos de toque e *overshoot*.

Figura 13 - Visão geral da geração de dados sintéticos puros adotados no Real-ESRGAN.

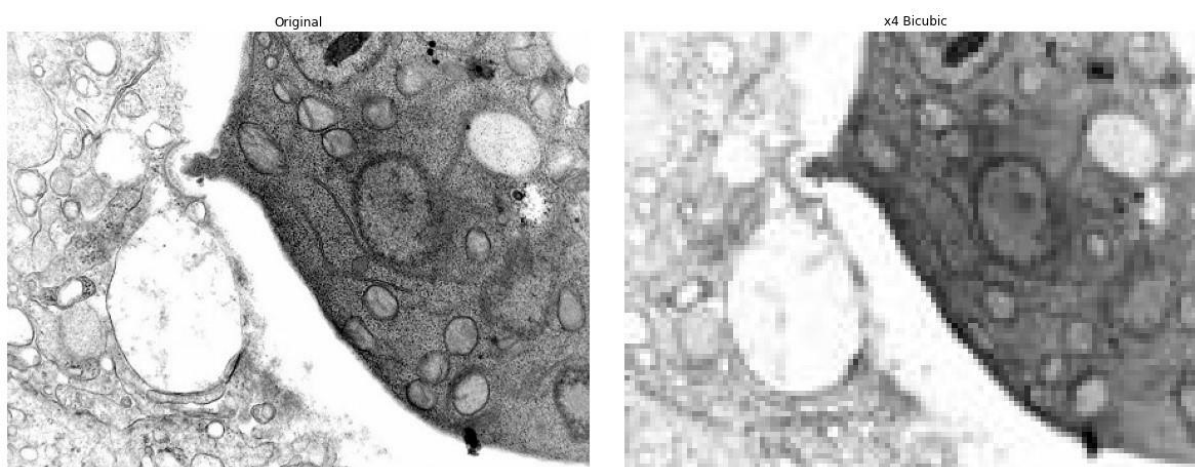
Fonte: Extraído de [44].

4.3 Base de dados

A base de imagens utilizadas para teste no trabalho foi retirada do site didático do Departamento de Anatomia Patológica da FCM-UNICAMP, hospedado no Centro de Computação da UNICAMP em servidor exclusivo e próprio do Departamento. Foram utilizadas imagens de microscopia eletrônica da subseção de Paracoccidioidomicose linfonodal, disponível em [20], esta subseção tem como foco principal a análise de fungos.

Foram utilizadas cerca de 30 imagens da base de dados na execução e avaliação do modelo, para simular casos reais, as imagens originais de alta qualidade (HR) passaram por alguns processos de degradação natural como tráfego de internet e compressão, além disso foram submetidas a uma redução de escala utilizando o processo de redução bicúbica (*Bicubic Downsampling*) que consiste em usar uma média ponderada para decidir a cor do pixel. Esse método geralmente reduz a resolução melhor do que a redução média no que diz respeito à qualidade das imagens reduzidas. A redução bicúbica é o método mais preciso, porém levando tempo para produzir as gradações tonais mais suaves.

Figura 14 - A figura acima mostra a imagem original de alta resolução (HR) ao lado da mesma imagem após passar pelo processo de redução bicúbica.



Fonte: o autor.

4.4 Métricas de qualidade

O objetivo das métricas de qualidade é comparar a imagem HR original e a HR estimada. A escolha da métrica está relacionada à sua equivalência quanto à percepção humana. Em outras palavras, deseja-se quantificar a comparação entre as imagens da mesma forma que a visão humana as compara, qualificando-as quanto à similaridade existente entre elas [21]. Dois tipos diferentes de métrica são utilizados neste trabalho: PSNR (Relação Sinal-Ruído de Pico, do inglês, *Peak Signal-to-Noise Ratio*) e SSIM (Índice de Similaridade Estrutural, do inglês, *Structural Similarity Index*).

4.4.1 PSNR

Utilizado em muitos trabalhos nas áreas de compressão de sinal, imagem e vídeo, a PSNR é simples de calcular, mas por outro lado, não traduz numericamente a percepção humana [29]. Define-se o erro quadrático médio (MSE) entre a imagem HR estimada \hat{x} e a imagem HR original x , aqui nas suas formas lexicográficas,

$$MSE(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{M^2} \sum_{i=1}^{M^2} [\mathbf{x}(i) - \hat{\mathbf{x}}(i)]^2, \quad (1.1)$$

que é uma medida de diferença entre pixels das duas imagens. Quanto mais próximo de zero for o valor do MSE, mais parecidas são as duas imagens, teoricamente.

O PSNR é dado por

$$PSNR(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \left(\frac{MAX_{\mathbf{x}}^2}{MSE} \right), \quad (1.2)$$

e aumenta à medida em que as duas imagens testadas são mais parecidas uma da outra. Devido à presença do logaritmo na equação, o valor é dado em decibéis (dB). Além disso, tal métrica não tem valor absoluto, servindo apenas para comparar métodos de compressão de perda geralmente ou, como é o caso deste trabalho, métodos de reconstrução. Essa é a métrica utilizada em diversos trabalhos, tais como [30, 31, 32, 33].

4.4.2 SSIM

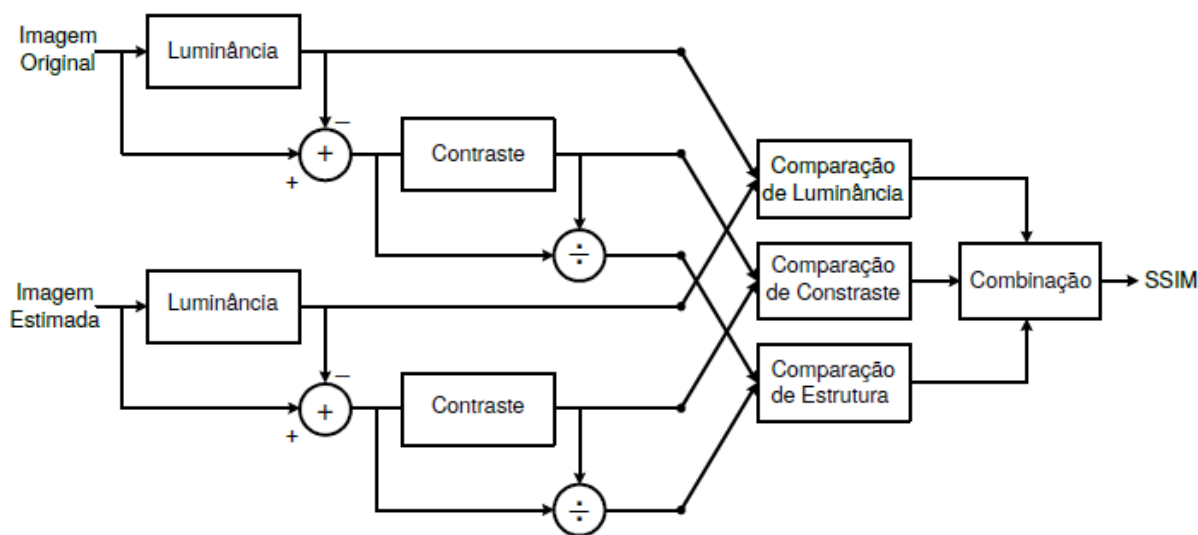
Durante as três últimas décadas, muitos trabalhos estiveram focados em desenvolver uma forma de avaliação de qualidade de imagens que aproveite características conhecidas do sistema visual humano (HVS - *human visual system*) [34]. A maioria dos modelos desenvolvidos tem como estratégia modificar a medida MSE de forma que os erros sejam penalizados de acordo com sua visibilidade. Wang [34] desenvolve uma medida de similaridade estrutural que compara padrões locais de intensidade que são normalizados pelo luminância e pelo contraste, ilustrada na Figura 15 e definida por

$$SSIM(\mathbf{x}, \hat{\mathbf{x}}) = \frac{(2\mu_{\mathbf{x}}\mu_{\hat{\mathbf{x}}} + c_1)(2\sigma_{\mathbf{x}\hat{\mathbf{x}}} + c_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\hat{\mathbf{x}}}^2 + c_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\hat{\mathbf{x}}}^2 + c_2)}, \quad (1.3)$$

onde $\mu_{\hat{\mathbf{x}}}$ e $\mu_{\mathbf{x}}$ são a média de \mathbf{x} e $\hat{\mathbf{x}}$, respectivamente. Além disso, $\sigma_{\mathbf{x}}$ e $\sigma_{\hat{\mathbf{x}}}$ representam as variâncias de \mathbf{x} e $\hat{\mathbf{x}}$, respectivamente; $\sigma_{\mathbf{x}\hat{\mathbf{x}}}$, a covariância entre os dois sinais e c_1 e c_2 são constantes que estabilizam a divisão, no caso de um denominador próximo de zero.

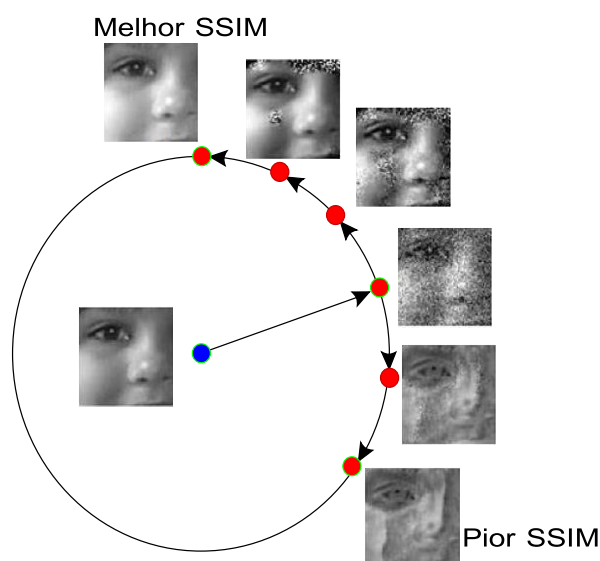
O valor do SSIM (Structural Similarity Index) varia entre -1 e 1, mas geralmente, se encontra entre 0 e 1. Assume-se 1 o valor resultante quando duas imagens idênticas são comparadas, e 0 quando duas imagens completamente diferentes são dadas como entrada.

Figura 15 - Diagrama do funcionamento da medição do SSIM.



Fonte: Adaptado de [34].

A Figura 16 ilustra a comparação entre o MSE e o SSIM. O raio da circunferência representa o valor MSE (portanto, o mesmo PSNR), ou seja, as imagens em volta do círculo têm o mesmo MSE, quando comparadas com a imagem em seu interior. Por inspeção visual, observa-se que a imagem no topo, à esquerda, é igual à referência e imagem na parte inferior, à direita da figura, é muito diferente da referência. O MSE, entretanto, não é capaz de traduzir tais aspectos, uma vez que, de acordo com seu valor, todas as imagens do lado exterior do círculo têm a mesma semelhança com a imagem no interior. Por outro lado, a imagem superior tem o maior SSIM e é, de fato, a imagem mais parecida com a referência. Observe que todas as imagens na circunferência do círculo possuem a mesma PSNR em relação à original.

Figura 16 - Comparação ilustrativa entre o PSNR e o SSIM.

Fonte: Adaptado de [35].

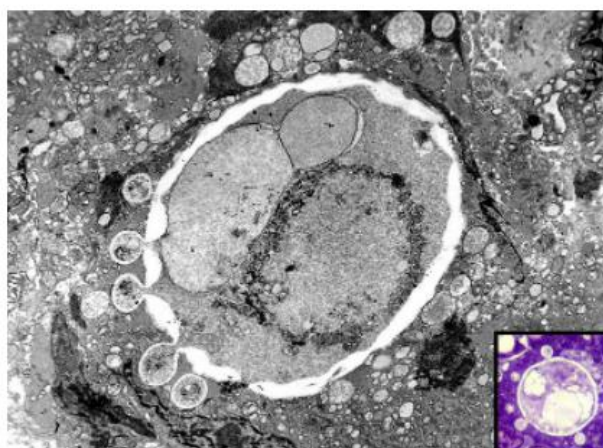
Ainda que o PSNR seja amplamente utilizado, esse método é focado em encontrar diferenças entre pixels, enquanto o SSIM é baseado em variações na estrutura. Entretanto, a mente humana não processa informações simplesmente como números ou divisões tão pequenas quanto pixels. Ao olhar para uma imagem, o usuário busca informações completas, como formas e bordas, que se traduzem como estrutura. Utilizando o SSIM é possível comparar as imagens de forma semelhante à abordagem do olho humano. Nesse sentido, neste trabalho conjectura-se que a mudança da métrica PSNR para SSIM possibilita um aperfeiçoamento nos resultados de super-resolução existentes. Espera-se obter imagens estimadas mais próximas à expectativa do usuário [21].

4.5 Avaliação de qualidade

Além do uso das métricas de qualidade para a avaliação de qualidade das imagens geradas pelo modelo, foi também elaborado um formulário eletrônico disponibilizado por meio de um *link* do Google Forms para que as imagens pudessem ser avaliadas por profissionais da área de saúde e afins de forma cega (sem que os profissionais saibam em quais imagens algoritmo foi utilizado).

No formulário foi requisitado que os participantes classificassem um grupo de 90 imagens entre as opções: Excelente, Muito Boa, Boa, Razoável, Ruim, Muito Ruim e Péssima. Cada seção do formulário continha 3 imagens da mesma figura, onde uma era a imagem original, uma a imagem gerada pelo modelo e outra a imagem original após passar pelo processo de redução bicúbica, nenhum participante sabia previamente qual das imagens havia passado por processos computacionais e a ordem das imagens foi trocada a cada mudança de seção do formulário. Foi solicitado aos participantes que, de forma geral, classificassem cada imagem em termos de qualidade para análise biológica e a visualização de detalhes bem delineados como podemos ver na figura 17.

Figura 17 - Primeira pergunta do formulário de avaliação de qualidade.



- Excelente
- Muito Boa
- Boa
- Razoável
- Ruim
- Muito Ruim
- Péssima

Fonte: o autor.

Além dessa, foi feita outra pergunta em relação a cada imagem avaliada. Como podemos ver na figura 18, os participantes tiveram que escolher entre as opções: Sim e Não, se conseguiam ver elementos importantes do ponto de vista biológico na imagem.

Figura 18 - Segunda pergunta do formulário de avaliação de qualidade.

Você consegue ver elementos importantes do ponto de vista biológico? *

Sim

Não

Fonte: o autor.

No total foram obtidas 8 respostas, participaram da pesquisa 4 biomédicos(as), 1 médico(a) e 1 professor(a) e 2 estudantes de biomedicina.

4.6 Ferramentas de desenvolvimento

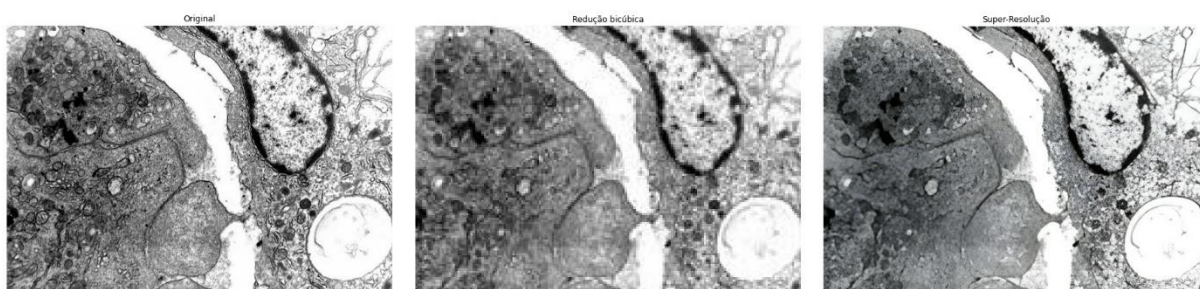
Foram utilizadas como ferramentas durante o desenvolvimento do trabalho *Google colab* juntamente com a linguagem de programação Python, além disso foram utilizadas as bibliotecas do *SckitLearn*, *Tensorflow*, *Matplotlib*, *Seaborn* e *OpenCV*.

5 Resultados e discussões

5.1 Aplicação das métricas de qualidade

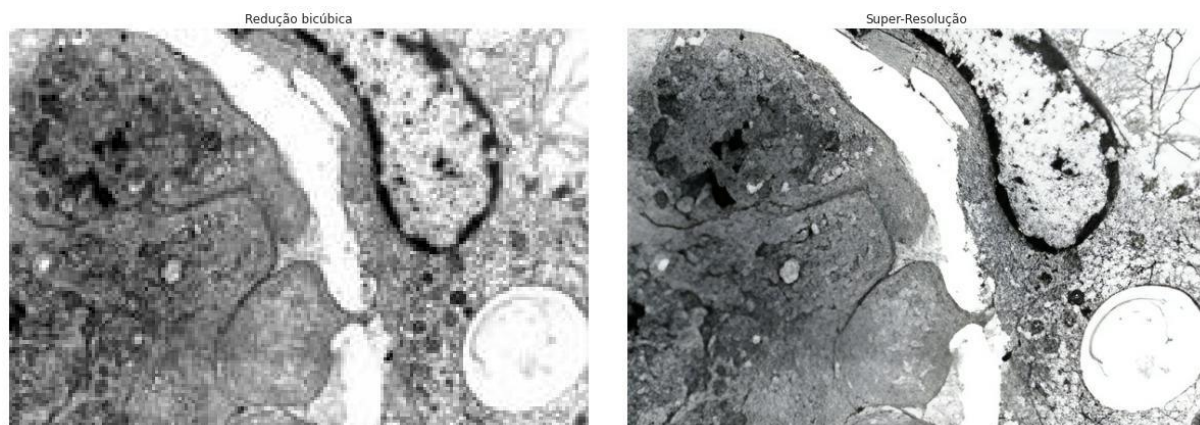
Foram utilizadas 43 imagens da base de teste para avaliação dos resultados do modelo, as figuras 19, 20 e 21 exibem alguns desses resultados.

Figura 19 - Comparação entre a imagem original, a imagem com degradação após passagem pelo processo de redução bicúbica e a imagem gerada pelo modelo.

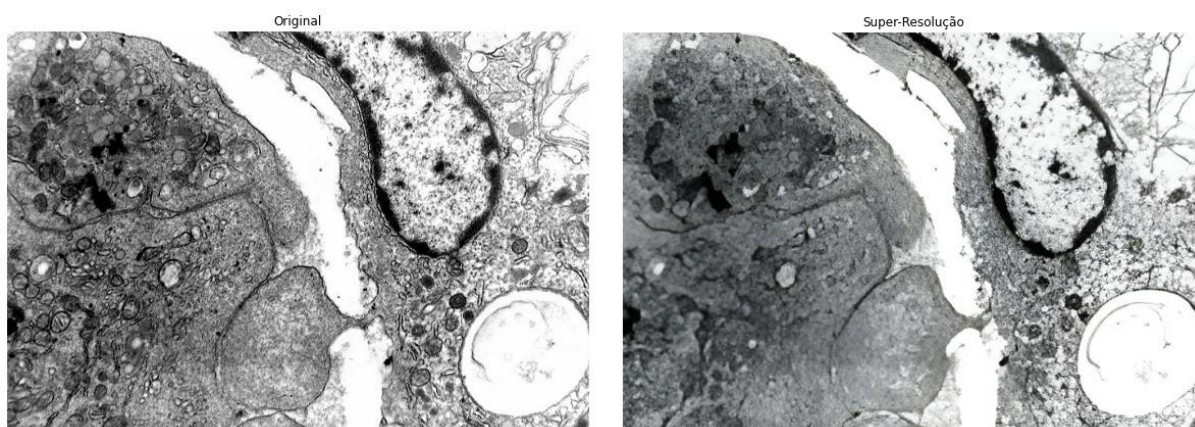


Fonte: o autor.

Figura 20 - Comparação entre a imagem com degradação após passagem pelo processo de redução bicúbica e a imagem gerada pelo modelo.



Fonte: o autor.

Figura 21 - Comparação entre a imagem original e a imagem gerada pelo modelo.

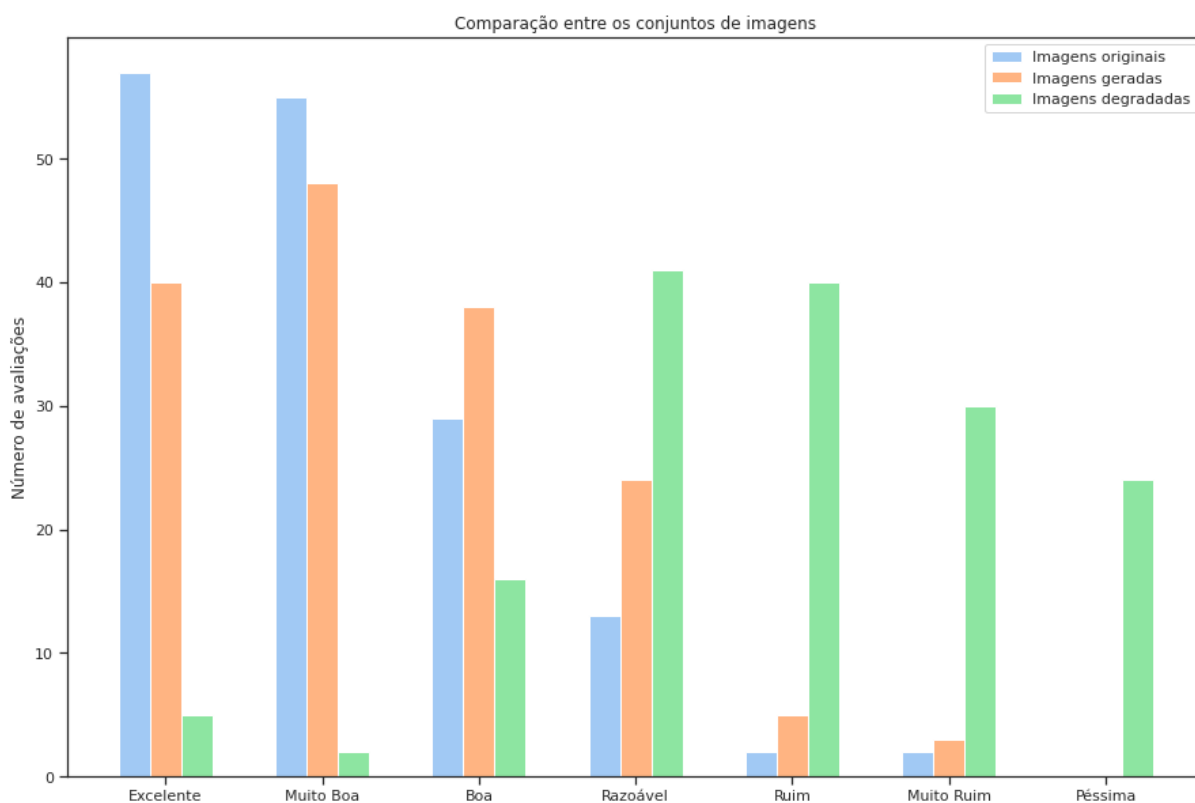
Fonte: o autor.

Dada as imagens resultantes, foi calculado o SSIM e PSNR médio da comparação entre cada imagem do conjunto das imagens originais e sua respectiva no conjunto de imagens geradas, foram obtidos os valores de SSIM: 0,896 com desvio padrão de 0,031 e PSNR: 26,59 com desvio padrão de 2,221, o que podem ser considerados bons resultados.

5.2 Resultados da avaliação de qualidade

Após a obtenção dos resultados da pesquisa, foi realizado um tratamento nos dados de forma a estrutura-los melhor para geração de gráficos e uma melhor visualização dos resultados, foram gerados 2 tipos de gráficos, na figura 22 podemos visualizar um gráfico de barras verticais, onde cada barra representa o número de avaliações que aquele conjunto de imagens recebeu em relação a sua classificação.

As imagens foram divididas em três grandes conjuntos, o conjunto das imagens originais extraídas diretamente da base de dados foram denotadas como “Imagens originais” em azul, o conjunto de imagens geradas como resultado pela GAN como “Imagens geradas” em laranja e o conjunto de imagens que sofreu redução bicúbica como “Imagens degradadas” em verde. Na parte inferior do gráfico temos as possíveis categorias que uma imagem pode ser classificada, sendo elas: Excelente, Muito Boa, Boa, Razoável, Ruim, Muito Ruim, Péssima.

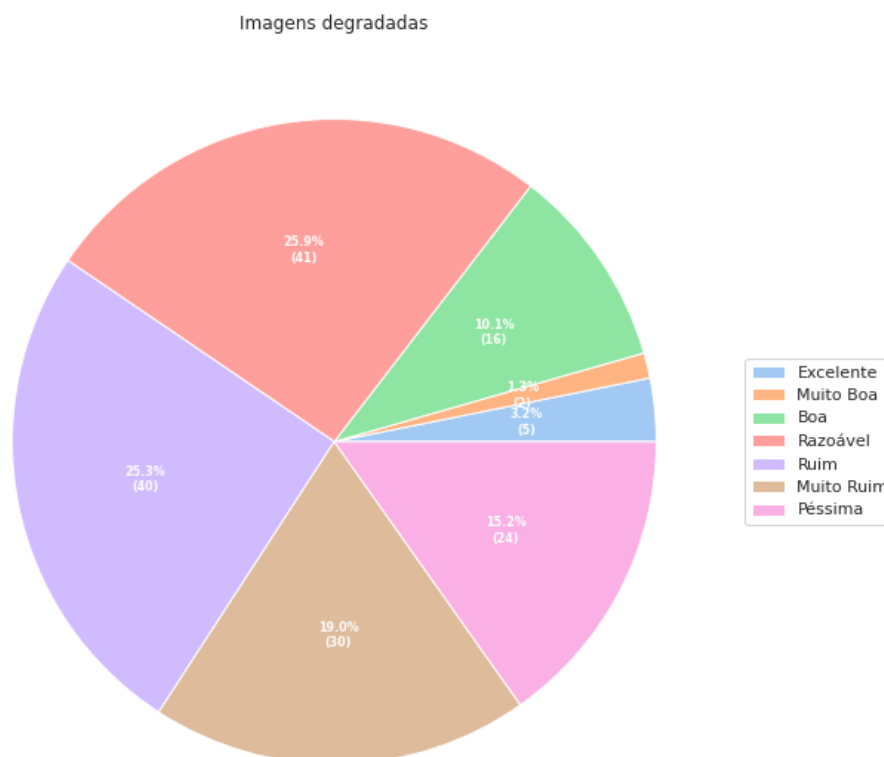
Figura 22 - Gráfico de barras para cada conjunto de imagens.

Fonte: o autor.

Com isso, podemos comparar o número de avaliações em cada categoria para cada conjunto de imagens e assim pode-se notar uma alta similaridade no número de avaliações entre os conjuntos das imagens originais e imagens geradas, ou seja, as imagens geradas pela GAN mantiveram aspectos visuais de tal forma que, foram tão bem avaliadas quanto as imagens originais. Além disso, é possível visualizar também as mudanças ocorridas entre o conjunto das imagens geradas e o conjunto das imagens degradadas, há uma expressiva mudança na avaliação das imagens degradadas após passarem pelo processo de super-resolução aplicado pela GAN, o que denota ainda mais uma boa performance do modelo, esse comportamento também pode ser visto nas figuras 23, 24 e 25.

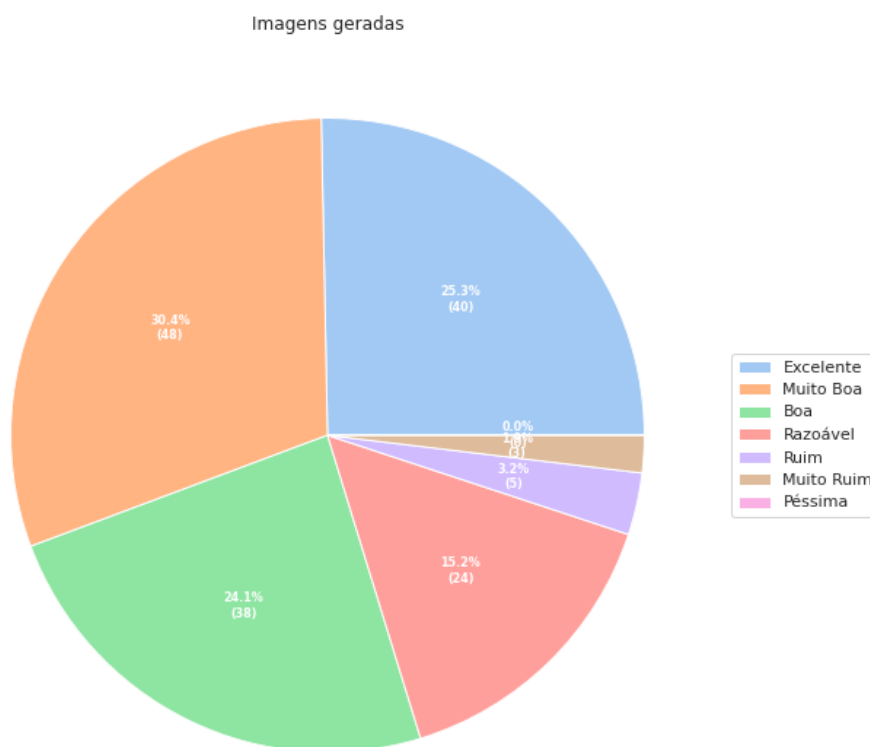
É possível avaliar também as medidas de tendência central em relação ao número de avaliações recebidas por conjunto de imagens no gráfico de *box-plot* na figura 26.

Figura 23 - Gráfico de setores do conjunto de imagens degradadas.



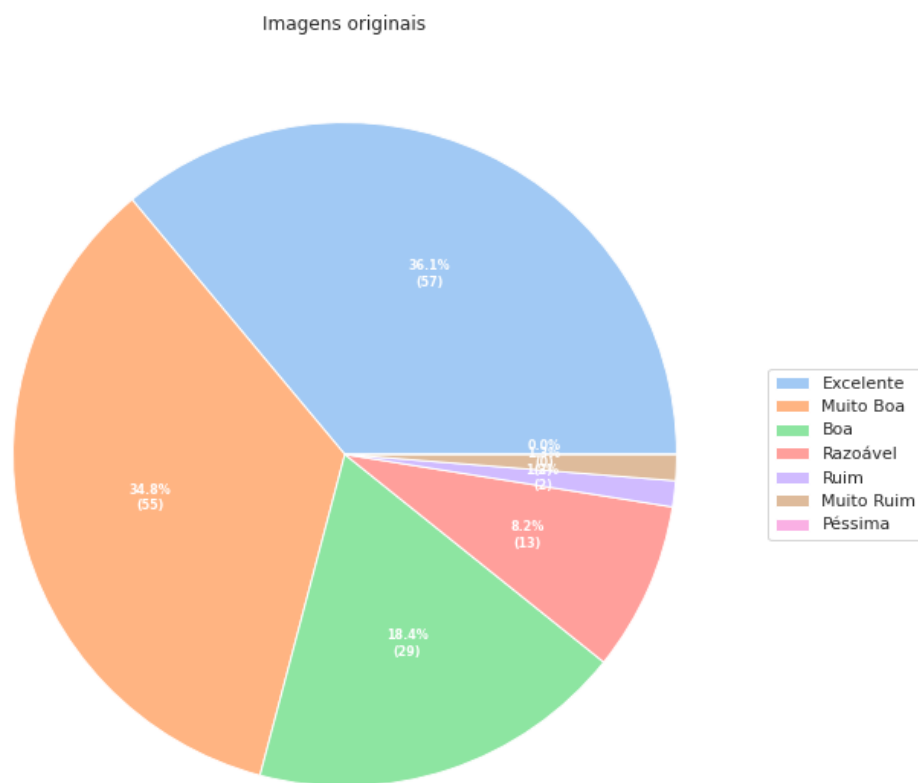
Fonte: o autor.

Figura 24 - Gráfico de setores do conjunto de imagens geradas.



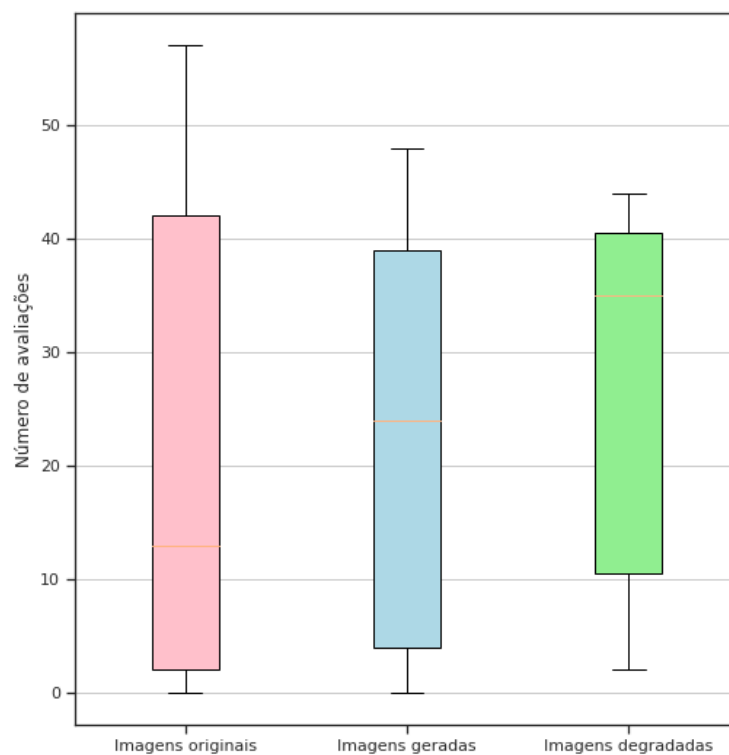
Fonte: o autor.

Figura 25 - Gráfico de setores do conjunto de imagens originais.



Fonte: o autor.

Figura 26 - Gráfico de *box-plot* do número de avaliações por conjunto de imagens.



Fonte: o autor.

Além disso, foi também solicitado aos participantes da pesquisa que respondessem se, para cada imagem, era possível visualizar elementos importantes do ponto de vista biológico na imagem. Podemos também visualizar um comportamento semelhantes aos das figuras anteriores, ao comparamos as figura 27 e 28 podemos ver uma clara diferença percentual das avaliações para mesmas figuras antes e depois de passarem pelo processo de super-resolução aplicado pela GAN, houve um aumento expressivo no percentual de respostas “sim” de 34,2% para 74,7%, ou seja, imagens degradadas onde não era possível que os avaliadores conseguissem visualizar elementos importantes do ponto de vista biológico na imagem, após passarem pelo processamento do algoritmo, foram consideradas aptas em relação ao questionamento feito.

Já em relação a figura 29, podemos novamente notar uma expressiva similaridade percentual em relação as avaliações na figura 28, mostrando que as imagens geradas pela GAN foram tão bem avaliadas, em relação a segunda pergunta do formulário de avaliação de qualidade, quanto as imagens originais, com uma diferença percentual de apenas 19%.

Figura 27 - Gráfico de setores do conjunto de imagens degradadas.

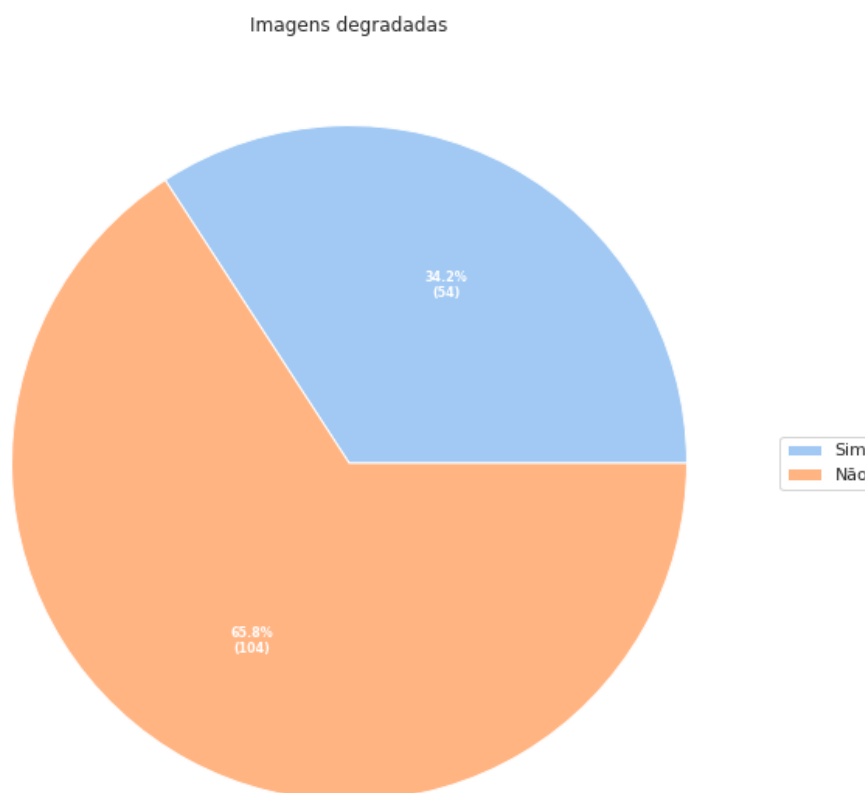
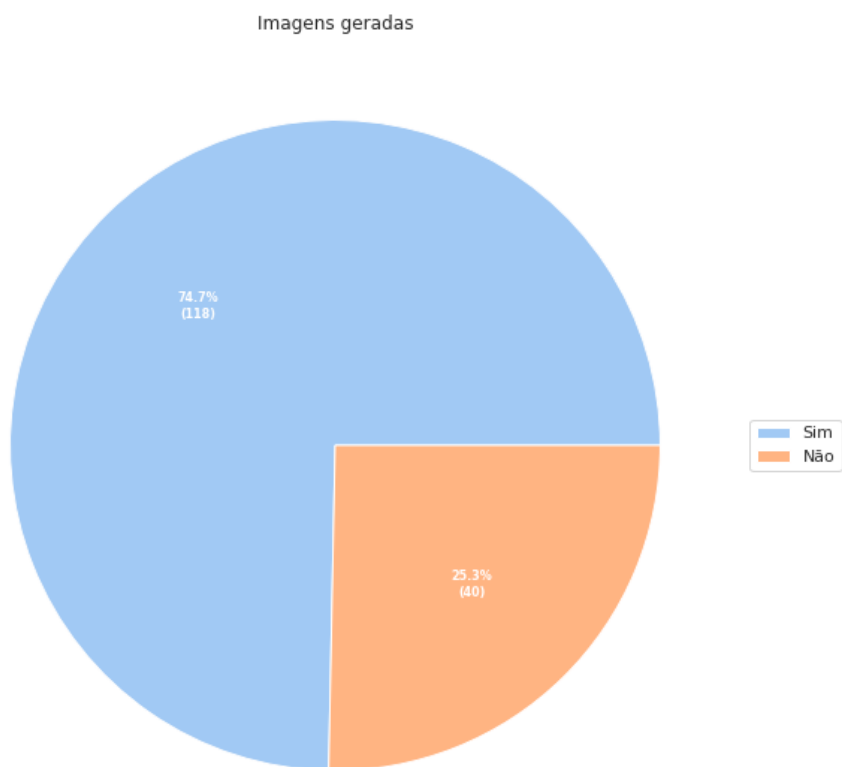
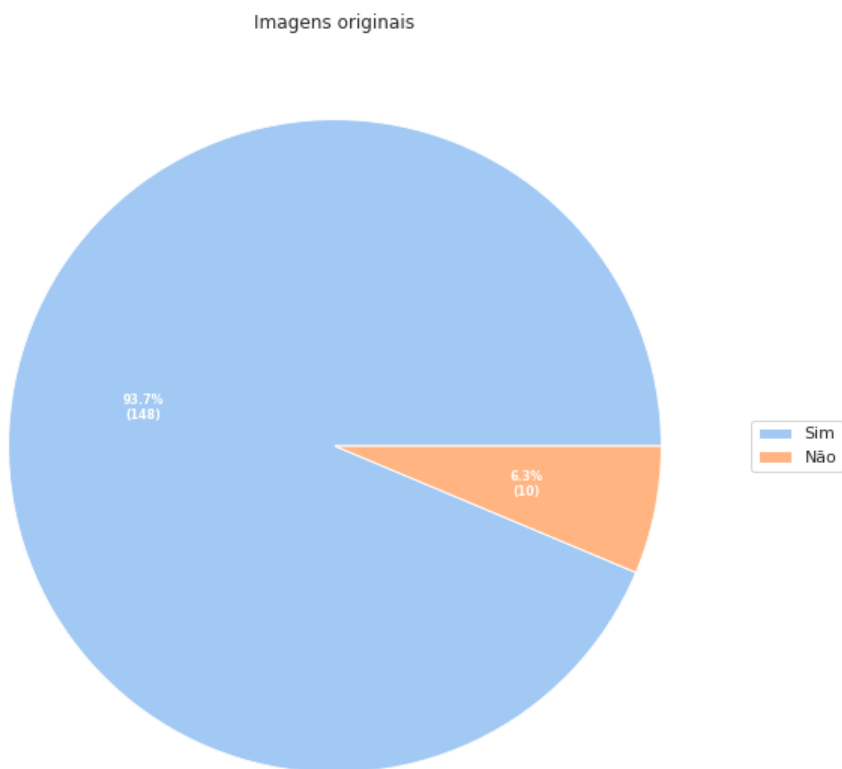


Figura 28 - Gráfico de setores do conjunto de imagens geradas.



Fonte: o autor.

Figura 29 - Gráfico de setores do conjunto de imagens originais



Fonte: o autor

6 Conclusão e Trabalhos futuros

Neste capítulo são apresentadas as conclusões do presente trabalho e as recomendações para a continuidade dos trabalhos nesta área de estudo.

6.1 Conclusões

O presente trabalho teve como objetivo geral avaliar o uso de SRGANs no processo de ampliação e remoção de distorções de imagens aplicado a imagens de exames médicos.

Com relação as dificuldades para o desenvolvimento do trabalho, a escolha da base de dados de teste foi uma etapa difícil do estudo. A falta de uma base de dados aberta e de fácil acesso, com tipificação das imagens por tipo de patologia e formatação adequada foram os principais obstáculos encontrados.

Os resultados desse estudo mostram que a aplicação de técnicas de super-resolução utilizando redes adversárias geradoras melhoraram significativamente a qualidade de imagens de exames médicos, os valores encontrados em relação as métricas de qualidade definidas no estudo se mostraram bons quanto a avaliação da similaridade entre as imagens originais e as imagens geradas pelo modelo. Além disso, foram obtidos também bons resultados em relação a avaliação por profissionais e estudantes da área de saúde, demonstrando que a imagens geradas pela GAN mantiveram detalhes e aspectos visuais em relação as imagens originais, na análise visual por esses profissionais e estudantes.

É valido destacar também, a boa performance do algoritmo em relação ao tratamento e remoção das degradações sofridas pelas imagens que passaram pelo processo de redução bicúbica utilizadas como entrada no modelo, como mencionado na seção 5.2 desse trabalho, as imagens geradas a partir das imagens degradadas foram tão bem avaliadas quanto as imagens originais, o que mostra a capacidade da GAN de remoção de degradações.

Finalizando, o presente trabalho abre caminho para que estudantes e pesquisadores possam elaborar e desenvolver novos estudos em relação ao tema.

6.2 Trabalhos futuros

Como foi dito anteriormente, a escolha da base de dados de teste foi uma etapa difícil do estudo.

Em função da indisponibilidade de algumas informações e do tempo para conclusão deste trabalho, recomenda-se para trabalhos futuros a incorporação ao presente modelo de bases de dados mais diversas com um maior número de imagens e também de diferentes segmentações da área de saúde.

Por fim, sugere-se também a incorporação de uma análise comparativa entre os diferentes modelos presentes no estado da arte, de forma a validar qualitativamente as saídas geradas e com isso definir qual a melhor escolha do modelo a ser utilizado.

Referências

- [1] RIBEIRO-ROTTA, R. F. **Técnicas tomográficas aplicadas à Ortodontia: a evolução do diagnóstico por imagens**. Revista Dental Press de Ortodontia e Ortopedia Facial, v. 9, n. 5, p. 102–156, out. 2004.
- [2] **Entenda o que é diagnóstico por imagem e os tipos + comuns**. Disponível em: <<https://eigierdiagnosticos.com.br/blog/o-que-e-diagnostico-imagem/>>. Acesso em: 8 maio. 2022.
- [3] Zhengwei Wang, Qi She, Tomas E. Ward , **Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy**, 2020.
- [4] PARK, S. C.; PARK, M. K.; KANG, M. G. **Super-resolution image reconstruction: a technical overview**. IEEE Signal Processing Magazine, v. 20, n. 3, p. 21–36, maio 2003.
- [5] CHAUDHURI, S. **Super-Resolution Imaging**. Kluwer, 2001. (The Kluwer International Series in Engineering and Computer Science).
- [6] BERTERO, M.; BOCCACCI, P. **Super-resolution in computational imaging**. Micron, v. 34, n. 6, p. 265–273, out. 2003.
- [7] BORMAN, S.; STEVENSON, R. L. **Spatial resolution enhancement of low-resolution image sequences**. A comprehensive review with directions for future research. 1998.
- [8] BORMAN, S.; STEVENSON, R. L. **Super-resolution from image sequences- a review**. In: Proc. 1998 Midwest Symp. Circuits and Systems. 1999. p. 374–378.
- [9] NGUYEN, N. X. **Numerical Algorithms for Image Super-resolution**. Tese (Doutorado) — Stanford University, jul. 2000.
- [10] LUKOSZ, W. **Optical systems with resolving power exceeding the classical limit**. Journal of the Optical Society of America, v. 56, n. 11, p. 1463–1472, 1966.
- [11] LUKOSZ, W. **Optical systems with resolving power exceeding the classical limit ii**. Journal of the Optical Society of America, v. 57, n. 7, p. 932–941, 1967.
- [12] TSAI, R. Y.; HUANG, T. S. **Multiframe image restoration and registration**. **Advances in Computer Vision and Image Processing**, v. 1, p. 317–339, 1984.
- [13] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang e W. Shi, **Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network**, 2017. arXiv: 1609.04802 [cs.CV].

- [14] GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. Addison-Wesley, 1992.
- [15] QUISTGAARD, J. U. **Signal acquisition and processing in medical diagnostic ultrasound**. IEEE Signal Processing Magazine, v. 14, n. 1, p. 67–74, jan. 1997.
- [16] OLLINGER, J. M.; FESSLER, J. A. **Positron-emission tomography**. IEEE Signal Processing Magazine, v. 14, n. 1, p. 43–55, jan. 1997.
- [17] MOLINA, R. et al. **Image restoration in astronomy**. IEEE Signal Processing Magazine, v. 18, n. 2, p. 11–29, mar. 2001.
- [18] BARRETT, H.H.; MYERS, K.J. **Foundations of Image Science**. : John Wiley & Sons, 2004. (Wiley Series in Pure and Applied Optics).
- [58] ZIBETTI, M. V. W.; MAYER, J. **Super-Resolução Regularizada e Simultânea de Seqüência de Imagens**. Journal of Communication and Information Systems, v. 21, n. 2, p. 45–57, 30 ago. 2006.
- [20] **Anatpat-UNICAMP**. Disponível em: <<http://anatpat.unicamp.br/biinflparacoco5.html>>. Acesso em: 9 maio. 2022.
- [21] TECNOLÓGICO, C. **Universidade Federal do Espírito Santo O Uso do SSIM Como Critério de Convergência em Abordagens Bayesianas Variacionais de Superresolução Multiframe**. [s.l.: s.n.]. Disponível em: <<http://repositorio.ufes.br/handle/10/9659>>.
- [22] JABBAR, A.; LI, X.; OMAR, B. **A Survey on Generative Adversarial Networks: Variants, Applications, and Training**. ACM Computing Surveys, v. 54, n. 8, p. 1–49, 30 nov. 2022.
- [23] WANG, Z.; SHE, Q.; E. WARD, T. **Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy**. ACM Computing Surveys, v. 4, 15 jun. 2020.
- [24] CRESWELL, A. et al. **Generative Adversarial Networks: An Overview**. IEEE Signal Processing Magazine, v. 35, n. 1, p. 53–65, jan. 2018.
- [25] WANG, K. et al. **Generative adversarial networks: introduction and outlook**. IEEE/CAA Journal of Automatica Sinica, v. 4, n. 4, p. 588–598, 2017.
- [26] YI, X.; WALIA, E.; BABYN, P. **Generative adversarial network in medical imaging: A review**. Medical Image Analysis, v. 58, p. 101552, dez. 2019.

- [27] FU, K. et al. **Image Super-Resolution Based on Generative Adversarial Networks: A Brief Review**. *Computers, Materials & Continua*, v. 64, n. 3, p. 1977–1997, 2020.
- [29] SALOMON, D. **Data compression: the complete reference**. [S.l.]: Springer Science & Business Media, 2004.
- [30] VILLENA, S. **Superresolución y Reconstrucción Bayesiana de Imágenes a Partir de Imágenes de Baja Resolución Rotadas y Desplazadas. Combinación de Modelos**. Tese (Doutorado) — Universidad de Granada, December 2011. Disponível em: <<http://digibug.ugr.es/bitstream/10481/20549/1/20702504.pdf>>.
- [31] VILLENA, S.; VEGA, M.; BABACAN, S. D.; MOLINA, R.; KATSAGGELOS, A. K. **Bayesian combination of sparse and non-sparse priors in image super resolution**. *Digital Signal Processing*, Elsevier, v. 23, n. 2, p. 530–541, 2013. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1051200412002400>>.
- [32] VILLENA, S.; VEGA, M.; MOLINA, R.; KATSAGGELOS, A. K. **Bayesian superresolution image reconstruction using an l1 prior**. In: IEEE. *Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on*. [S.l.], 2009. p. 152–157.
- [33] BABACAN, S. D.; MOLINA, R.; KATSAGGELOS, A. K. **Variational bayesian super resolution**. *Image Processing, IEEE Transactions on*, IEEE, v. 20, n. 4, p. 984–999, 2011.
- [34] WANG, Z.; BOVIK, A. C.; SHEIKH, H. R.; SIMONCELLI, E. P. **Image quality assessment: from error visibility to structural similarity**. *Image Processing, IEEE Transactions on*, IEEE, v. 13, n. 4, p. 600–612, 2004.
- [35] WANG, Z.; BOVIK, A. **Mean squared error: love it or leave it? A new look at signal fidelity measures**, *IEEE Signal Processing Magazine*, v. 26, n. 1, p. 98–117, 2009.
- [36] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. **Spectral normalization for generative adversarial networks**. In *ICLR*, 2018.

- [37] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: **Convolutional networks for biomedical image segmentation**. In International Conference on Medical image computing and computer-assisted intervention. Springer, 2015.
- [38] Edgar Schonfeld, Bernt Schiele, and Anna Khoreva. **A u-net based discriminator for generative adversarial networks**. In CVPR, 2020.
- [39] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. **Esrgan: Enhanced super-resolution generative adversarial networks**. In ECCVW, 2018.
- [40] Yitong Yan, Chuangchuang Liu, Changyou Chen, Xianfang Sun, Longcun Jin, Peng Xinyi, and Xiang Zhou. **Finegrained attention and feature-sharing generative adversarial networks for single image super-resolution**. IEEE Transactions on Multimedia, 2021.
- [41] TensorFlow Hub. Disponível em: <<https://tfhub.dev/captain-pool/esrgan-tf2/1>>. Acesso em: 12 maio. 2022.
- [42] WANG, X. **Towards Real-World Blind Face Restoration with Generative Facial Prior**. 11 jan. 2021.
- [43] WANG, X. **ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks**. 17 set. 2018.
- [44] WANG, X. et al. **Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data**. Tencent ARC Lab; Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, 17 ago. 2021.
- [45] MARIZ, L. DE F. **REDES GERADORAS ADVERSÁRIAS EM GERAÇÃO DE IMAGENS**. UNIVERSIDADE DE BRASÍLIA, 5 dez. 2018.
- [46] Martin Riedmiller und Heinrich Braun: **Rprop - A Fast Adaptive Learning Algorithm**. Proceedings of the International Symposium on Computer and Information Science VII, 1992
- [47] H. R. Sheikh and A. C. Bovik, “**Image information and visual quality**,” IEEE Transactions on Image Processing, vol.15, no. 2, pp. 430–444, 2006.

- [48] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “**FSIM: A feature similarity index for image quality assessment**,” *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.
- [49] D. P. Kingma and J. Ba, “**Adam: a method for stochastic optimization**,” 2014, <https://arxiv.org/abs/1412.6980>.
- [50] E. Agustsson and R. Timofte, “**NTIRE 2017 challenge on single image super-resolution: dataset and study**,” in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1122–1131, IEEE, Honolulu, HI, USA, July 2017.
- [51] R. Timofte, E. Agustsson, L. V. Gool et al., “**NTIRE 2017 challenge on single image super-resolution: methods and results**,” in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1110–1121, IEEE, Honolulu, HI, USA, July 2017.
- [52] R. Timofte, R. Rothe, and L. V. Gool, “**Seven ways to improve example-based single image super resolution**,” in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1865–1873, IEEE, Las Vegas, NV, USA, June 2016.
- [53] M. Bevilacqua, A. Roumy, C. Guillemot, and M. A. Morel, “**Low-complexity single-image super-resolution based on nonnegative neighbor embedding**,” in *Proceedings of the 2012 British Machine Vision Conference*, pp. 135:1–135:10, Surrey, UK, September 2012.
- [54] R. Zeyde, M. Elad, and M. Protter, “**On single image scale-up using sparse-representations**,” in *Proceedings of the 2010 International Conference on Curves and Surfaces*, pp. 711–730, Avignon, France, June 2010.
- [55] D. Martin, C. Fowlkes, D. Tal, and J. Malik, “**A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics**,” in *Proceedings of the 8th IEEE International Conference on Computer Vision*, pp. 416–423, IEEE, Vancouver, Canada, July 2001.
- [56] J. B. Huang, A. Singh, and N. Ahuja, “**Single image superresolution from transformed self-exemplars**,” in *Proceedings of the 2015 IEEE Conference on*

Computer Vision and Pattern Recognition (CVPR), pp. 5197–5206, IEEE, Boston, MA, USA, June 2015.

[57] T. T. Dang, A. Beghdadi, and M. C. Larabi, “**Visual coherence metric for evaluation of color image restoration**,” in Proceedings of the 2013 Colour and Visual Computing Symposium (CVCS), pp. 1–6, IEEE, Gjøvik, Norway, September 2013.

[58] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, “**Learning a noreference quality metric for single-image super-resolution**,” *Computer Vision and Image Understanding*, vol.158, pp.1–16, 2017.

[59] A. Mittal, R. Soundararajan, and A. C. Bovik, “**Making a “completely blind” image quality analyzer**,” *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[60] CHAUDHURI, S. **Super-Resolution Imaging**.: Kluwer, 2001. (The Kluwer International Series in Engineering and Computer Science).

[61] PARK, S. C.; PARK, M. K.; KANG, M. G. **Super-resolution image reconstruction: a technical overview**. *IEEE Signal Processing Magazine*, v. 20, n. 3, p. 21–36, maio 2003.

[62] BERTERO, M.; BOCCACCI, P. **Super-resolution in computational imaging**. *Micron*, v. 34, n. 6, p. 265–273, out. 2003.

[63] BORMAN, S.; STEVENSON, R. L. **Spatial resolution enhancement of low-resolution image sequences. A comprehensive review with directions for future research**. 1998.

[64] BORMAN, S.; STEVENSON, R. L. **Super-resolution from image sequences - a review**. In: Proc. 1998 Midwest Symp. Circuits and Systems. 1999. p. 374–378.

[65] NGUYEN, N. X. **Numerical Algorithms for Image Superresolution**. Tese (Doutorado) — Stanford University, jul. 2000.

[66] LUKOSZ, W. **Optical systems with resolving power exceeding the classical limit**. *Jornal of the Optical Society of America*, v. 56, n. 11, p. 1463–1472, 1966.

- [67] LUKOSZ, W. **Optical systems with resolving power exceeding the classical limit ii.** *Jornal of the Optical Society of America*, v. 57, n. 7, p. 932–941, 1967.
- [68] TSAI, R. Y.; HUANG, T. S. **Multiframe image restoration and registration.** *Advances in Computer Vision and Image Processing*, v. 1, p. 317–339, 1984.
- [69] ELAD, M.; FEUER, A. **Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images.** *IEEE Transactions on Image Processing*, v. 6, n. 12, p. 1646–1658, dez. 1997.
- [70] RAJAN, D.; CHAUDHURI, S.; JOSHI, M. V. **Multi-objective super resolution: concepts and examples.** *IEEE Signal Processing Magazine*, v. 20, n. 3, p. 49–61, maio 2003.
- [71] THÉVENAZ, P.; BLU, T.; UNSER, M. **Handbook of medical imaging, processing and analysis.** In: San Diego Ca, USA: Academic Press, 2000. cap. Image Interpolation and Resampling, p. 393–420.
- [72] THÉVENAZ, P.; BLU, T.; UNSER, M. **Interpolation revisited.** *IEEE Transactions on Medical Imaging*, v. 19, n. 7, p. 739–758, jul. 2000.
- [73] BANHAM, M. R.; KATSAGGELOS, A. K. **Digital image restoration.** *IEEE Signal Processing Magazine*, v. 14, n. 2, p. 24–41, mar. 1997.
- [91] T. M. Mitchell, **Machine Learning.** McGraw-Hill, 1997.
- [92] **The discipline of machine learning.** Carnegie Mellon University, School of Computer Science, Machine Learning Department Pittsburgh, PA, 2006, vol. 9.
- [93] S. B. Maind, P. Wankar et al., “**Research paper on basic of artificial neural network,**” *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 2, no. 1, pp. 96–100, 2014.
- [94] S. S. Haykin, S. S. Haykin, S. S. Haykin, and S. S. Haykin, **Neural networks and learning machines.** Pearson Upper Saddle River, 2009, vol. 3.
- [95] G. Cybenko, “**Approximation by superpositions of a sigmoidal function,**” *Mathematics of control, signals and systems*, vol. 2, no. 4, pp. 303–314, 1989.
- [96] Y. LeCun, Y. Bengio, and G. Hinton, “**Deep learning,**” *nature*, vol. 521, no. 7553, p. 436, 2015.

- [97] F. Rosenblatt, “**The perceptron: a probabilistic model for information storage and organization in the brain.**” *Psychological review*, vol. 65, no. 6, p. 386, 1958.
- [98] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, “**Efficient backprop,**” in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 9–48.
- [99] A. L. Maas, A. Y. Hannun, and A. Y. Ng, “**Rectifier nonlinearities improve neural network acoustic models,**” in *Proc. icml*, vol. 30, no. 1, 2013, p. 3.
- [100] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, “**Supervised machine learning: A review of classification techniques,**” *Emerging artificial intelligence applications in computer engineering*, vol. 160, pp. 3–24, 2007.
- [101] K. Janocha and W. M. Czarnecki, “**On loss functions for deep neural networks in classification,**” *arXiv preprint arXiv:1702.05659*, 2017.
- [102] Y. LeCun, Y. Bengio et al., “**Convolutional networks for images, speech, and time series,**” *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.
- [103] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, “**Striving for simplicity: The all convolutional net,**” *arXiv preprint arXiv:1412.6806*, 2014.
- [104] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “**Generative adversarial nets,**” in *Advances in neural information processing systems*, 2014, pp. 2672–2680.
- [105] A. Ng, “*Generative learning algorithms,*” 2008.
- [106] I. Goodfellow, “**Nips 2016 tutorial: Generative adversarial networks,**” *arXiv preprint ar- Xiv:1701.00160*, 2016.
- [107] A. Radford, L. Metz, and S. Chintala, “**Unsupervised representation learning with deep convolutional generative adversarial networks,**” *arXiv preprint arXiv:1511.06434*, 2015.
- [108] D. P. Kingma and J. Ba, “**Adam: A method for stochastic optimization,**” *arXiv preprint arXiv:1412.6980*, 2014.