UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO MULTIDISCIPLINAR

FÁBIO JÚNIOR MIRANDA DOS SANTOS
VITOR GUSMÃO LOURA

Super Resolução: um estudo da *Enhanced Super Resolution GAN*

Prof. Filipe Braida do Carmo, D.Sc.
Orientador

Nova Iguaçu, Agosto de 2021

Super Resolução: um estudo da *Enhanced Super Resolution GAN*

Fábio Júnior Miranda dos Santos

Vitor Gusmão Loura

Projeto Final de Curso submetido ao Departamento de Ciência da Computação do Instituto de Multidisciplinar da Universidade Federal Rural do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Apresentado por:

_____

Fábio Júnior Miranda dos Santos

_____

Vitor Gusmão Loura

Aprovado por:

_____

Prof. Filipe Braida do Carmo, D.Sc.

_____

Prof. Bruno José Dembogurski, D.Sc.

_____

Prof. Leandro Guimaraes Marques Alvim, D.Sc.

NOVA IGUAÇU, RJ - BRASIL

Agosto de 2021

Emitido em 23/08/2021

DOCUMENTOS COMPROBATÓRIOS Nº 10991/2021 - CoordCGCC (12.28.01.00.00.98)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 23/08/2021 16:31)
BRUNO JOSE DEMBOGURSKI
PROFESSOR DO MAGISTERIO SUPERIOR
DeptCC/IM (12.28.01.00.00.83)
Matrícula: 2124964

(Assinado digitalmente em 23/08/2021 16:31)
FILIPE BRAIDA DO CARMO
PROFESSOR DO MAGISTERIO SUPERIOR
DeptCC/IM (12.28.01.00.00.83)
Matrícula: 3929524

(Assinado digitalmente em 23/08/2021 16:23)
LEANDRO GUIMARAES MARQUES ALVIM
PROFESSOR DO MAGISTERIO SUPERIOR
DeptCC/IM (12.28.01.00.00.83)
Matrícula: 1800852

(Assinado digitalmente em 23/08/2021 16:50)
FABIO JUNIOR MIRANDA DOS SANTOS
DISCENTE
Matrícula: 2014780178

(Assinado digitalmente em 23/08/2021 16:50)
VITOR GUSMÃO LOURA
DISCENTE
Matrícula: 2016780515

Para verificar a autenticidade deste documento entre em https://sipac.ufrrj.br/documentos/ informando seu número: **10991**, ano: **2021**, tipo: **DOCUMENTOS COMPROBATÓRIOS**, data de emissão: **23/08/2021** e o código de verificação: **4a8f10f497**

Agradecimentos

Fábio Júnior Miranda dos Santos

Primeiramente, gostaria de agradecer à minha família, cujos ensinamentos foram essenciais para o meu crescimento pessoal. Gratidão especial aos meus pais por terem me possibilitado a experiência de estudar em uma universidade federal num estado diferente do que sempre vivi. Um agradecimento especial ao meu avô, José Amâncio Prates Miranda, que mesmo não estando mais presente para poder ver a conclusão desta etapa da minha vida, sempre demonstrou orgulho e felicidade pelo neto que teve.

Agradeço a todos os meus professores. Em particular, ao professor de Matemática do Ensino Médio, Marcos Cetra, o qual não apenas teve a capacidade de ministrar uma ótima disciplina, mas também a transmitiu de forma prazerosa.

Ao professor e meu orientador, Filipe Braida, por ter aceitado nos guiar nesse TCC com toda sua sabedoria. Também sou grato ao professor Marcel, por ter viabilizado minha experiência como monitor, durante a qual pude vivenciar o prazer de passar o conhecimento adiante, percebendo que mais gratificante do que aprender algo novo é ver o conhecimento ser absorvido por novos alunos.

Ao meu amigo e parceiro de jogos, Vitor Loura, por me acompanhar em horas e horas de jogatinas, tanto divertidas quanto estressantes. Por último, agradeço profundamente à uma possível entidade presente nessa realidade ou ao simples destino cósmico, por essa vida valiosa na qual concluo este TCC.

Vitor Gusmão Loura

Primeiramente, agradeço às entidades cósmicas da magia e da gambiarra por manterem meus códigos funcionando.

Agradeço também a minha família, em especial, meus pais, Mônica Gusmão e Jônatas Loura, que sempre fizeram de tudo por mim e me forneceram o necessário para estar aqui hoje.

Agradeço por ter experienciado essa etapa da minha vida no Instituto Multidisciplinar da Universidade Federal Rural do Rio de Janeiro, onde, sem o mesmo, eu não teria me tornado quem sou hoje.

Meus agradecimentos ao Departamento de Ciência da Computação e, em especial, ao professor Filipe Braida por ter nos guiado da melhor forma durante esse projeto.

À Ianá Maria e sua família por terem me apoiado e acreditado em mim nos momentos difíceis.

Aos meus amigos Caio César Sampaio, Beatriz Lima, Ingrid Cardin, Victor Diniz e Guilherme Mendes que estiveram comigo durante todo esse percurso.

E em especial ao meu amigo Fábio Júnior Miranda pelas horas e horas de jogo e principalmente o qual fez esse trabalho dar certo.

Meus mais sinceros agradecimentos.

Sarani mukoe - Plus Ultra

All Might, *Boku no Hero Academia*

RESUMO

Super Resolução: um estudo da *Enhanced Super Resolution GAN*

Fábio Júnior Miranda dos Santos e Vitor Gusmão Loura

Agosto/2021

Orientador: Filipe Braida do Carmo, D.Sc.

Considerando que grande parte das informações absorvidas pelo ser humano são retidas visualmente, é cada vez mais frequente a demanda por imagens com alta resolução. Sendo assim, algoritmos de super resolução foram desenvolvidos com o intuito de suprir essa demanda. No entanto, nem todos os modelos criados foram capazes de atingir resultados satisfatórios. Em 2017, as Redes Adversárias Generativas (GAN's) foram utilizadas como alternativa para possibilitar uma super resolução com resultados superiores aos dos métodos vigentes, tais como aqueles baseados em redes neurais e processamento de imagens. Entre as diferentes GAN's, a *Enhanced Super Resolution GAN* (ESRGAN) apresentou grande eficácia na melhoria de imagem. Posto isso, propôs-se neste trabalho a pesquisa dos principais métodos que implementaram a super resolução, e a comparação desses com a ESRGAN. Dentre os resultados obtidos, a ESRGAN alcançou 34.90 na métrica de avaliação PSNR, sendo essa a melhor média entre os experimentos realizados.

ABSTRACT

Super Resolução: um estudo da *Enhanced Super Resolution GAN*

Fábio Júnior Miranda dos Santos and Vitor Gusmão Loura

Agosto/2021

Advisor: Filipe Braida do Carmo, D.Sc.

Considering that a large part of the information absorbed by the human being is retained visually, the demand for images with high resolution is increasing frequently. Therefore, super resolution algorithms were developed in order to meet this demand. However, not all models created were able to achieve satisfactory results in the process. In 2017*, the Adversary Generative Networks (GAN's) were used as an alternative to enable a super resolution with results superior to the current methods, such as those based on neural networks and image processing. Among the different GAN's, the Enhanced Super Resolution GAN (ESRGAN) showed great effectiveness in improving images. Therefore, it was proposed in this work the survey of the main methods that implemented the super resolution and the comparison of these with the ESRGAN. Among the results obtained, ESRGAN reached* 34.90 *in the PSNR evaluation metric, which is the best average among the experiments performed.*

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas e Siglas

CSAM	*Channel-Spatial Attention Module*
CNN	*Convolutional Neural Network*
CycleGAN	*Cycle-Consistent Adversarial Network*
DCNN	*Deep CNN*
DCGAN	*Deep Convolutional GAN*
EDSR	*Enhanced Deep Super Resolution Network*
ESRGAN	*Enhanced Super Resolution GAN*
GAN	*Generative Adversarial Network*
HAN	*Holistic Attention Network*
LAM	*Layer Attetion Module*
MDSR	*Multi-scale Deep Super Resolution*
MSE	*Mean Square Error*
PSNR	*Peak Signal-to-Noise Ratio*
ReLU	*Rectified Linear Unit*
RRDB	*Residual-in-Residual Dense Block*
RCAN	*Residual Channel Attention Network*
SR3	*Super-Resolution via Repeated Refinement*
SRCNN	*Super Resolution CNN*

SRGAN	*Super Resolution GAN*
SSIM	*Structural Similarity Index Measure*
SVH	Sistema Visual Humano
SVM	*Support Vector Machine*
VDSR	*Very Deep CNN*

Sumário

Capítulo 1

Introdução

Sabe-se que a busca da sociedade contemporânea por avanços tecnológicos que propiciem melhoria na qualidade de vida e otimização do tempo é contínua. Tais recursos vêm contribuindo significativamente, tanto nas diferentes áreas profissionais, quanto para o aperfeiçoamento dos mais variados tipos de entretenimento.

Ao longo das últimas décadas, diversos métodos vêm sendo desenvolvidos para atender a demanda pelo aperfeiçoamento de imagens, recorrente nas áreas supracitadas, em especial naquelas relacionadas à produção de jogos e reconhecimento facial.

Um dos recursos tecnológicos que surgiu para suprir tal necessidade, foi a super resolução, a qual passou a possibilitar a geração de imagens em alta resolução a partir de outras em baixa resolução. Esse feito foi um marco para o desenvolvimento de jogos, embora essa técnica ainda não seja adotada por todas as empresas, uma vez que muitas delas continuam optando pelo processamento das imagens até obtê-las em alta resolução, o que requer alto custo.

A super resolução é eficiente não apenas para as áreas dedicadas à criação de produtos de entretenimento, como também para o setor de reconhecimento de faces através de imagens. Dado que as aplicações atuais de reconhecimento facial manipulam, em grande parte, imagens em baixa resolução, a introdução da super resolução proporcionou progressos significativos nesse setor, implicando assim na

criação da *Face Hallucination*[1].

Sabendo-se que essa técnica foi, e continua sendo, muito explorada, diversos métodos foram desenvolvidos para atender demandas em diversas áreas. Neste trabalho será apresentado um desses métodos, o qual obteve ótimos resultados na geração de imagens, em formato digital, retratadas com a maior fidelidade possível. Essa é a função da *Enhanced Super Resolution* GAN.

Anteriormente à criação da ESRGAN, outros modelos também foram capazes de replicar essa funcionalidade. Porém, nem todos puderam balancear a viabilidade da melhoria com a performance adquirida. Esse desbalanceamento deve-se ao fato de que foram desenvolvidos modelos que atingiram bons resultados apenas após uma quantidade extrema de treinamento, enquanto outros que apresentaram rápido processamento não obtiveram resultados tão satisfatórios.

Visto que cada modelo criado apresenta diversas maneiras de obter a super resolução, propõe-se, neste trabalho, analisar a eficácia da ERSGAN em relação aos modelos concorrentes, assim como pontuar a viabilidade dos principais modelos de super resolução de imagem.

1.1 Objetivo

O objetivo deste trabalho consiste em implementar técnicas de super resolução, visando comparar o desempenho de cada técnica com a da ESRGAN. Para tal, serão pesquisados na literatura os principais modelos de técnicas de melhoria de imagem e, posteriormente às comparações, será realizada a análise da eficiência dos mesmos.

1.2 Organização do Trabalho

A organização deste trabalho é constituída por cinco capítulos, sendo eles:

- Introdução: um resumo deste trabalho será apresentado neste capítulo, consis-

[1]Técnicas que aplicam a super resolução de diversas formas, utilizando imagens em baixa resolução.

tido de contexto, motivação, problemas encontrados e proposta sugerida.

- Fundamentações Teóricas: neste capítulo, serão abordados os princípios teóricos de uma *Generative Adversarial Network* e os pré-requisitos essenciais para sua compreensão.

- *Enhanced Super Resolution* GAN: serão apresentadas as motivações para a elaboração deste trabalho, bem como os trabalhos relacionados que já abordaram a técnica de super resolução, e a proposta aqui realizada.

- Experimentos: o detalhamento dos diferentes modelos de super resolução será feito, tal como as métricas utilizadas para a avaliação desses, os *datasets* manipulados e os resultados obtidos através dos experimentos.

- Conclusão: este capítulo tratará do resumo deste trabalho, as considerações observadas no decorrer de seu desenvolvimento, as limitações encontradas e os possíveis trabalhos futuros.

Capítulo 2

Fundamentações Teóricas

Este capítulo descreve os conceitos básicos que serão utilizados no presente trabalho. Nele, serão detalhadas as Redes Adversárias Generativas (GAN's), assim como os seguintes conceitos prévios necessários para sua melhor compreensão: Aprendizado Supervisionado e Redes Neurais Artificiais.

2.1 Aprendizado Supervisionado

O Aprendizado Supervisionado é uma área da Inteligência Artificial que consiste em aprender o mapeamento de um grupo de variáveis de entrada, um grupo de saída, e aplicar esse mapeamento para prever saídas de dados ainda não vistos (CUNNINGHAM; CORD; DELANY, 2008). Principalmente, a ideia desse aprendizado foca na utilização de um supervisor, *i.e.* aquele cujo papel é instruir o sistema quanto aos rótulos associados dos dados de treino.

Existem dois tipos de problemas abordados no Aprendizado Supervisionado: classificação e regressão. O primeiro trabalha com previsões de dados discretos, *i.e.* dados com número finito de possíveis valores, e o segundo com dados contínuos, *i.e.* dados com número infinito de possíveis valores dentro de um intervalo. Na tabela 2.1, trata-se de um problema de classificação, onde seu objetivo é encontrar um modelo que tente prever a ocorrência ou ausência de chuva. Já a tabela 2.2, exemplifica um

problema de regressão, onde a meta é obter um modelo que estime valores de casas.

ID	CLIMA	TEMPERATURA	VENTO	CLASSE
1	ensolarado	36ºC	fraco	Não choveu
2	nublado	16ºC	forte	Choveu
3	ameno	25ºC	médio	Choveu
4	nublado	20ºC	médio	Não choveu
5	ensolarado	28ºC	médio	Choveu

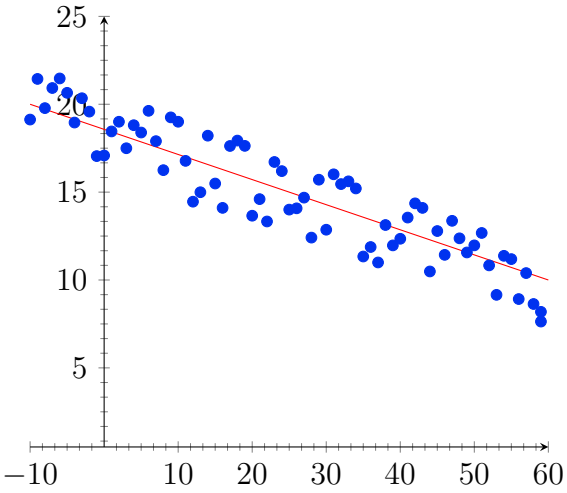
Tabela 2.1: Exemplo de classificação

ID	TAMANHO	Nº DE QUARTOS	CONDIÇÃO	VALOR
1	71m²	1	Ótimo	R$ 194.723,43
2	54m²	1	Bom	R$ 110.259,99
3	100m²	3	Ótimo	R$ 300.000,00
4	70m²	2	Ruim	R$ 120.800,00
5	200m²	4	Péssimo	R$ 352.300,05

Tabela 2.2: Exemplo de regressão

A partir desses problemas, é definida uma função que melhor mapeia os dados e os rotula com uma previsão. Tendo em mãos essa função, dados novos podem ser analisados, realizando assim o papel de classificá-los ou predizê-los corretamente. Para tal, dispõe-se de métricas de avaliação que julgam o quão próximo foi o acerto ou erro da resposta prevista pela função, *e.g.* acurácia, precisão e *recall*.

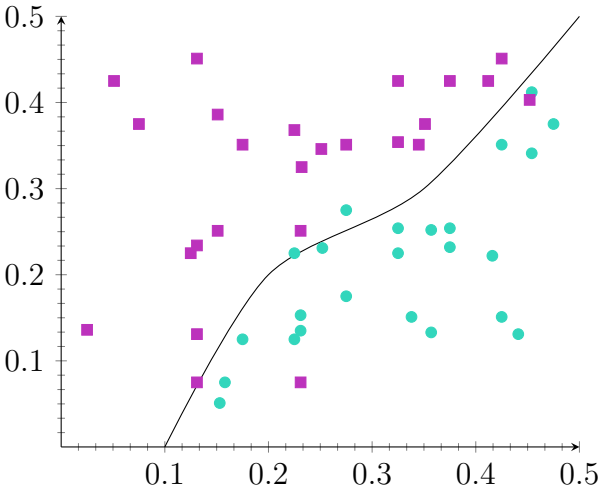
Como mencionado, o objetivo da função que mapeia os dados é classificá-los ou estimá-los, respectivamente nos problemas de classificação e regressão. No caso de regressões, um método pode ser utilizado para verificar a eficácia de seu mapeamento. Um desses métodos é o *Mean Square Error* (MSE), o qual calcula a média das diferenças entre os valores de dados reais e os preditos pela função. A figura 2.1 exemplifica o comportamento de uma função em um problema de regressão e o tratamento do MSE.



$$MSE(f, g) = \frac{1}{n}\sum_{i=1}^{n}(f_i - g_i)^2 \tag{2.1}$$

Figura 2.1: Função mapeadora em um problema de Regressão.

Distintivamente, nos problemas de classificação é necessário indicar uma função que melhor classifica os dados. Assim, métricas como acurácia, precisão e *recall* são úteis para verificar a performance dessa função. A acurácia, por exemplo, examina o fracionamento entre o número de classificações preditas corretamente e o número total de testes. A figura 2.2 apresenta melhor um caso de classificação tendo como métrica a acurácia.



$$
\begin{aligned}
Acurácia = \frac{VP + VN}{VP + FN + VN + FP} \\
onde \\
VP = Verdadeiro\ Positivo \\
VN = Verdadeiro\ Negativo \\
FP = Falso\ Positivo \\
FN = Falso\ Negativo
\end{aligned}
\tag{2.2}
$$

Figura 2.2: Função mapeadora em um problema de Classificação.

Existem diversos modelos que conseguem reproduzir esse tipo de mapeamento de dados, como: árvore de decisão, *Naive Bayes*, regressão linear, regressão logística, *Support Vector Machine* (SVM) etc. Segundo Gurney (1997), um dos modelos que é utilizado com frequência para solucionar classificações e regressões é a rede neural, a qual será mais detalhada na seção posterior.

2.2 Rede Neural

Rede Neural Artificial (ANN) é uma composição de camadas de neurônios, sendo uma de entrada, uma ou mais camadas intermediárias (ocultas) e uma final, de saída (WANG, 2003). A figura 2.3 mostra uma das típicas estruturas de uma rede neural,

onde cada círculo representa um neurônio, cada linha representa a ligação entre eles, e cada neurônio que se encontra num nível diferente da camada de entrada e de saída é representado na camada oculta.

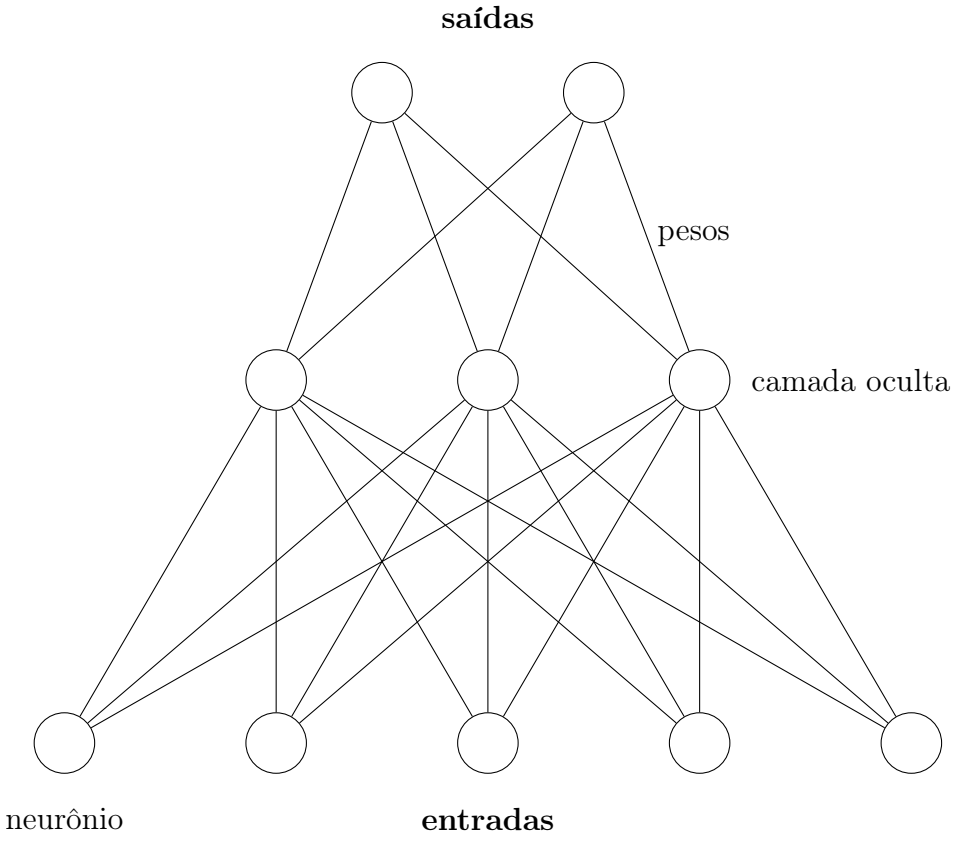


Figura 2.3: Rede neural simples.

O objetivo de uma rede neural é propagar os sinais de cada neurônio até a camada de saída, resultando assim em uma análise mais complexa comparada com a função de um único neurônio. Quanto maior for a quantidade de camadas ocultas na rede, maior será a complexidade de análise que a mesma pode oferecer. Sendo assim, uma rede neural pode ser aplicada para solucionar diversos problemas da atualidade, como: diagnósticos médicos, detecção de fraude em cartões de crédito, reconhecimento de voz etc.

A comparação entre um neurônio biológico e um neurônio matemático foi abordada outrora por Sima (1998). O primeiro com seu núcleo, dendritos e axônio, e o segundo sendo uma reformulação da função do neurônio biológico através de fórmulas matemáticas. Como representado na figura 2.4, um neurônio matemático é composto

por uma ou mais entradas equivalentes aos sinais transmitidos através dos dendritos de um neurônio biológico. Da mesma maneira, pesos sinápticos são rotulados junto a cada uma dessas entradas.

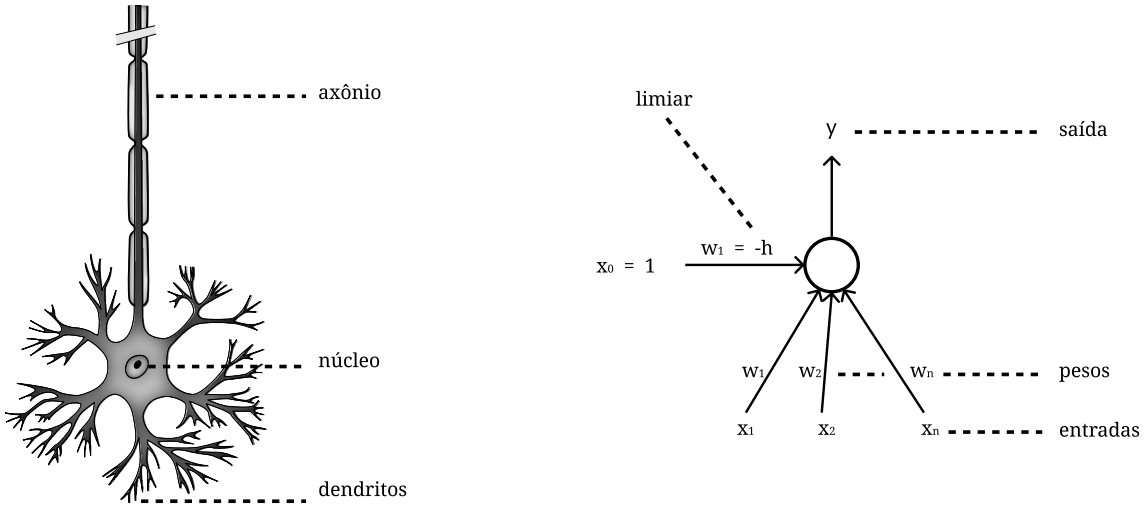


Figura 2.4: Neurônio biológico e neurônio matemático.

Matematicamente, um neurônio pode ou não ser ativado após receber um impulso elétrico do axônio. Essa ativação é calculada por meio do somatório dos produtos resultantes das entradas do neurônio e seus respectivos pesos. Caso o valor obtido nesse somatório seja maior que um limiar denotado como h, o qual pode variar de aplicação para aplicação, o neurônio é ativado pelo impulso, *i.e.* a saída do neurônio será verdadeira e, por convenção, representada pelo valor 1. Do contrário, o neurônio se mantém desativado e sua saída será 0. A fórmula 2.3 retrata essa função.

$$y = \begin{cases} 1, & f \geq h. \\ 0, & f < h. \end{cases}, \ onde \ \ f = \sum_{i=1}^{n} W_i X_i. \tag{2.3}$$

Com o comportamento dessa função, um único neurônio matemático pode fazer classificações simples como um neurônio biológico. Por exemplo: dado que os atributos de entrada do neurônio sejam características referentes ao clima do dia, *e.g.*

temperatura, umidade, velocidade do vento, um cenário de classificação, seria ativar o neurônio na previsão da ocorrência de chuva ou mantê-lo desativado na ausência. Em um outro cenário, poderia ser considerada a análise dos atributos do neurônio, tais como características de um animal e, posteriormente, a sua classificação, *e.g.* como sendo um cachorro ou um gato.

São diversos os tipos de redes neurais que foram desenvolvidos, *e.g. Feed Forward*, *Deep Feed Forward*, *Markov Chain*, *Deep Convolution Network*, entre outros. Dentre as redes com aprendizado profundo, *i.e.* aquelas que contêm quantidades altas de camadas de neurônios, a Rede Adversária Generativa (GAN) é uma das que mais obteve sucesso em áreas como processamento de linguagem natural e visão computacional (CHENG et al., 2020). Essa rede será melhor detalhada na seção a seguir.

2.3 GAN - Rede Adversária Generativa

Rede Adversária Generativa (GAN) é um modelo de rede neural que opera através de um processo adversário (GOODFELLOW et al., 2014). Nesse modelo, são treinados simultaneamente dois componentes que competem entre si: um gerador, que utiliza os dados a serem treinados, e um discriminador, que verifica a probabilidade da amostra recebida ser da saída do componente gerador ao invés dos dados de treino.

De acordo com Creswell et al. (2018), diferentemente dos modelos comuns de rede neural supervisionados, a GAN tem a grande vantagem de tirar proveito de quantidades amplas de imagens não rotuladas e conseguir manter um comportamento do aprendizado supervisionado, além do seu alto potencial de aprendizado profundo.

Simultaneamente, Cheng et al. (2020) aponta o fato da GAN não depender de dados de entrada bem distribuídos, uma vez que outros modelos geradores necessitam dessa distribuição para obter um bom resultado. Em contrapartida, devido a essa mesma questão, as amostras geradas pela GAN não tendem a focar em dados de treino específicos.

Dessa forma, serão abordados nesta seção, tópicos pertinentes ao detalhamento de uma GAN assim como suas aplicações, a estrutura mais trivial que o modelo pode possuir e o comportamento dos componentes dessa estrutura.

2.3.1 Aplicações

A primeira aplicação de GAN foi realizada por Goodfellow et al. (2014), na qual foram testadas algumas bases de dados, e os resultados comparados com os obtidos por outros modelos. Uma das bases foi o MNIST (figura 2.5), onde dados de dígitos manuscritos são processados. Outra base testada foi a *Toronto Face Database* (TFD) (figura 2.6), na qual expressões faciais são processadas.

Figura 2.5: Comparação de processamento na base de dados MNIST. (GOODFELLOW et al., 2014)

Figura 2.6: Comparação de processamento na base de dados TFD. (GOODFELLOW et al., 2014)

Após tais experimentos, Goodfellow et al. (2014) constataram que os resultados obtidos pela GAN foram superiores aos modelos concorrentes. A partir dessa análise, uma gama de possibilidades de novas aplicações surgiram, como por exemplo a *Deep Convolutional GANs* (DCGAN), que alterou a arquitetura da GAN convencional para estabilizar o treinamento dos dados.

Atualmente, há diversos tipos de aplicações que utilizam *GAN's*. De acordo com Gui et al. (2020), em 2018 foram publicados 11.800 artigos relacionados a esse tema e suas aplicações. Eles vão desde o processamento de imagens, dados sequenciais até utilidades em áreas médicas, *e.g* melhorar a resolução de uma tomografia para uma análise mais eficaz.

Pelo fato desse modelo ser amplamente utilizado em inúmeras áreas e aplicações, muitos outros tipos foram criados, tendo como inspiração o original. Desenvolvido por Ledig et al. (2017), um deles é a *Super Resolution GAN (SRGAN)*, que consiste em estimar uma imagem de alta resolução a partir de uma outra com baixa resolução.

Figura 2.7: Aplicação da SRGAN para uma imagem qualquer.

Para atender as necessidades da fiscalização de trânsito, a *SRGAN* foi adaptada para melhorar a resolução de uma foto da placa de um veículo registrada por câmeras diferentes. Levando em consideração que os registros das câmeras de segurança costumam ter várias questões que impossibilitam uma boa análise da placa do veículo, *e.g.* baixa resolução, interferência da iluminação do local, constantes borrões na imagem devido à velocidade do veículo e outros. A GAN produzida por Liu et al.

(2017), auxiliou na recuperação do conteúdo dessas placas que muitas vezes não pode ser reconhecido a olho nu. A imagem 2.8 demonstra o resultado desses conteúdos recuperados.

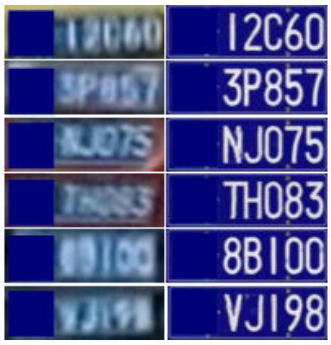


Figura 2.8: SRGAN adaptada para recuperar conteúdos de imagens de placa de automóveis. (LIU et al., 2017)

Outro modelo elaborado foi a *Cycle-Consistent Adversarial Networks* (CycleGAN) de Zhu et al. (2020). Neste trabalho, dado um par de imagens, é possível converter características de uma na outra e vice-versa. Na ilustração 2.9, a técnica foi executada para transformar uma fotografia de uma paisagem real em imagens com traços de pintores famosos como Monet, Van Gogh, Cezanne, entre outros. Essa transformação também pode ser adotada para transformar objetos, animais e situações similares, *e.g* transformar uma laranja numa maçã e transformar uma paisagem de verão na mesma paisagem, porém no inverno.

Figura 2.9: CycleGAN aplicada em tipos de pinturas. (ZHU et al., 2020)

Até mesmo na área da ilustração, a GAN é utilizada. Su et al. (2020) projetaram um modelo que extraísse as características de um rosto humano e gerasse um personagem típico de mangás com traços equivalentes ao original. O processo de

transformação ocorre inicialmente com a extração dos aspectos da foto e, posteriormente, a conversão desses aspectos humanos em atributos de mangás equivalentes. Em paralelo a esses passos, é realizada a conversão do formato do rosto para formatos comuns presentes em quadrinhos japoneses. Exemplos dos resultados obtidos com esse processo são apresentados na imagem 2.10.

Figura 2.10: Imagens modificadas através da MangaGAN. (SU et al., 2020)

Considerando que, frequentemente, é desenvolvido um novo tipo de GAN para cada aplicação diferente, existe uma gama de outras adaptações além das supracitadas. Desse modo, o objetivo desse estudo limita-se a analisar e confrontar modelos relacionados à melhoria da resolução de imagens como a *Enhanced Super Resolution GAN* (ESRGAN), versão otimizada da SRGAN.

2.3.2 Estrutura de uma GAN

Como citado na introdução, um modelo GAN é definido por dois componentes: um gerador e um discriminador. Esses componentes tratam-se de redes neurais independentes e uma analogia da competição entre eles pode ser feita para auxiliar no entendimento. Utilizando dados visuais, essa analogia consiste em comparar um deles com um falsificador de artes e o outro com um especialista (CRESWELL et al., 2018).

Sendo o falsificador nessa comparação, o componente gerador gera artes falsificadas

com o intuito de criar artes realistas. O discriminador, por sua vez, recebe artes tanto reais quanto geradas pelo falsificador e procura saber diferenciar as artes reais das falsas. Ambos componentes competem simultaneamente e se atualizam a fim de desempenharem seu papel com mais eficiência. Sendo assim, a figura 2.11 representa visualmente a estrutura de uma GAN.

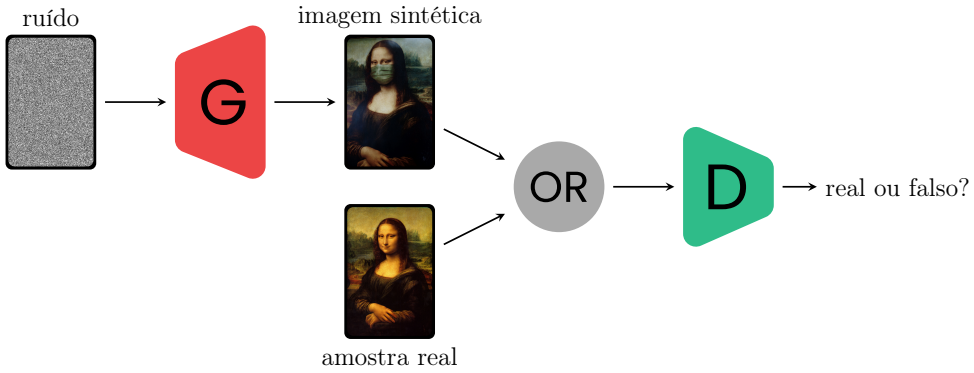


Figura 2.11: Representação da primeira GAN desenvolvida.

Ainda utilizando a analogia dos dados da GAN serem representados por artes, o componente gerador funciona da seguinte forma: ele gera constantemente amostras de arte falsas. Primeiramente, a amostra é gerada com parâmetros de entradas aleatórios. Após a primeira iteração da GAN, o gerador é treinado e seus pesos, atualizados.

O treino desse componente ocorre no fim de cada iteração, sendo modificado em relação ao desempenho do discriminador em saber distinguir se a imagem recebida é real ou falsa. Os pesos do gerador atualizam-se com a finalidade de gerar artes da forma mais realista possível, dificultando assim o papel do discriminador em evidenciá-las.

No caso do discriminador, ele é treinado inicialmente com uma amostra de arte real rotulada como verdadeira para que os seus parâmetros de pesos sejam bem definidos. Posteriormente, ele recebe tanto imagens reais quanto imagens falsas geradas pelo componente gerador, o qual, de início, as rotula como falsas.

A atualização de seus pesos ocorre após cada iteração, analisando o resultado obtido pela rede, através da verificação dos acertos e dos erros em relação à entrada recebida. Esse processo se repete com o intuito do discriminador minimizar os erros

de suas respostas.

Tendo em vista a estrutura da GAN e o comportamento dos seus componentes, Goodfellow et al. (2014) mencionaram que embora a GAN tenha vantagens relevantes em comparação a outros modelos, uma grande desvantagem está presente na sua funcionalidade: o modelo não irá gerar bons resultados caso um dos componentes seja mais treinado do que o outro.

Na hipótese do gerador ser mais treinado, a distribuição da saída do componente discriminador, *i.e.* valores no intervalo de 0 a 1 convencionalmente rotulados como falso e verdadeiro, respectivamente, tenderá a 0.5, tornando confusa a análise do discriminador. Esse evento pode resultar na situação em que o discriminador avalie como verdadeiras imagens ruins falsificadas pelo gerador.

No caso do discriminador ser mais treinado, a distribuição de sua saída tenderá a 0, ou seja, mesmo as imagens falsificadas que sejam muito próximas às reais, serão identificadas como falsas. Como o intuito da rede adversária não é a identificação absoluta do que seria real ou falso, mas sim a geração de amostras falsas muito semelhantes às reais, então esse cenário é prejudicial à funcionalidade do modelo.

Desse modo, a solução encontrada por Goodfellow et al. (2014) para contornar essa desvantagem foi treinar os componentes sincronizadamente para que nenhum dos dois seja mais esperto do que o outro. Outras soluções foram apresentadas para sanar a instabilidade do aprendizado dos componentes adversários, como proposto por Arjovsky, Chintala e Bottou (2017) e Wei et al. (2018).

Capítulo 3

Enhanced Super Resolution GAN

Super resolução é o processo através do qual várias imagens de um objeto em baixa resolução são combinadas para formar uma única imagem em alta resolução desse mesmo objeto (WALT, 2010). De acordo com Chaudhuri (2002), são várias as aplicações em que é necessária a geração de uma imagem em alta resolução, uma vez que a maioria das informações recebidas por um ser humano são visuais.

Segundo Heng e Dongdong (2015), os métodos de super resolução são categorizados da seguinte forma: baseados em interpolação, em reconstrução e em aprendizado. Métodos baseados em interpolação são definidos por um processo que estima novos *pixels* através da interpolação de *pixels* fornecidos (HENG; DONGDONG, 2015). Esse método utiliza a forma mais trivial de aumentar a resolução de uma imagem, além de possuir uma complexidade mínima de computação e manter procedimentos simples de cálculo. Existem três tipos de interpolação clássica: *nearest neighbor*, *bilinear* e *bicubic*.

Quanto aos métodos baseados em reconstrução, o principal objetivo é impor uma restrição linear em imagens de alta resolução reconstruídas. Para atingir esse objetivo, é aplicado o processo de modelagem da degradação de uma imagem (HENG; DONGDONG, 2015). Métodos baseados em reconstrução focam, sobretudo, em como adquirir o modelo de observação avançado, obtido através da solução do problema de super resolução, utilizando modelos de degradação de imagem, *e.g. Iterative Back*

Projection e *Maximum a Posterior* (HENG; DONGDONG, 2015).

Um ponto de pesquisa muito estudado foram os métodos de super resolução baseados em aprendizado. Freeman, Pasztor e Carmichael (2000) foram os primeiros a propor tal método. A ideia principal desse método foi estudar o mapeamento da relação entre a imagem em baixa resolução com a imagem em alta resolução e, posteriormente, reconstruir a imagem alvo utilizando essa relação.

De modo geral, métodos baseados em aprendizado dividem primeiramente a imagem em blocos e constroem a biblioteca de amostras de imagens em baixa resolução e alta resolução. Portanto, o modelo consegue aprender a relação do bloco de imagens em baixa resolução com o bloco de alta resolução. Por fim, o modelo utiliza essa relação para reconstruir uma imagem em alta resolução utilizando uma em baixa resolução.

Tendo em vista os vários modelos de super resolução, serão pontuadas neste capítulo as principais motivações para a utilização dessa técnica, como também os trabalhos relacionados que visaram solucionar a necessidade de melhoria da qualidade de imagens.

Considerando a variedade de modelos de redes neurais desenvolvidos e, em especial, as redes para atender a necessidade da melhoria de uma imagem, propõe-se nesse trabalho, através de experimentos, comparar a ESRGAN com os principais métodos de super resolução encontrados, sendo eles: *Nearest Neighbour*, *Hamming*, *Bilinear*, *Bicubic*, SRCNN, EDSR e SRGAN.

Ao final, particularidades da *Enhanced Super Resolution GAN* (ESRGAN) serão apresentadas.

3.1 Motivação

Com o rápido desenvolvimento dos *smart phones*, a obtenção de imagens e vídeos de melhor qualidade vem se tornando cada vez mais almejada (HENG; DONGDONG, 2015). Além disso, como mais de 80% das informações absorvidas pelo ser humano

são retidas pelo Sistema Visual Humano (SVH), a procura por imagens *high definition* (HD) é muito frequente.

Heng e Dongdong (2015) também apontam que, embora essa procura seja grande, as imagens desses dispositivos precisam ser armazenadas e exibidas em baixa resolução, uma vez que há limitações como capacidade de armazenamento e largura de banda. Até o momento, um dos meios mais viáveis que atende essa demanda, é exibir imagens em HD baseadas em imagens com baixa resolução.

Analogamente à evolução dos *smart phones*, dispositivos como televisões *smart* e monitores de vídeo de alta frequência também instigam grande procura por exibição de imagens HD. Sabe-se que, atualmente, são produzidos conteúdos em larga escala para esses dispositivos, como por exemplo: filmes e principalmente séries, transmitidos através de canais de *streaming*, *e.g. Netflix*[1], *Prime Video*[2] e *Disney Plus*[3].

Da mesma forma, a resolução de imagem em jogos eletrônicos vem sido explorada com frequência a fim de obter visuais mais agradáveis. Sendo assim, foi desenvolvida a *AMD FidelityFX Super Resolution* (FSR)[4], na qual é possível detectar bordas de uma imagem origem e recriá-las em alta resolução, a partir da super resolução. A aplicação dessa técnica na área de jogos é essencial para a comunidade *gamer*, uma vez que a qualidade visual dos jogos é aumentada sem a necessidade de adquirir uma placa de vídeo melhor.

A super resolução é almejada não apenas no âmbito de entretenimento, como também na área médica. Machado e Souki (2004) pontuaram que diagnósticos e planejamentos odontológicos evoluíram consideravelmente com a introdução de imagens dentárias de alta resolução. Assim como Zhu, Yang e Lio (2019) mostraram que imagens já obtidas por tomografias podem ser melhoradas e, dessa forma, auxiliar em análises minuciosas em pontos de lesão da imagem.

De fato, a demanda por super resolução de imagem é cada vez mais recorrente em diversas áreas. Porém, dependendo da área, não há a necessidade de uma melhoria

[1] <https://www.netflix.com/>
[2] <https://www.primevideo.com/>
[3] <https://www.disneyplus.com/>
[4] <https://www.amd.com/pt/technologies/radeon-software-fidelityfx-super-resolution>

tão precisa quanto em outras. Além disso, nem todos os métodos que implementam essa técnica são viáveis para qualquer tipo de problema. Portanto, ao decorrer desse capítulo, serão apresentados os principais métodos de super resolução de imagem e suas áreas de aplicações.

3.2 Trabalhos Relacionados

Nesta seção, serão destacados, cronológica e compiladamente, trabalhos que visaram a técnica de *Super Resolution* utilizando métodos distintos como: processo de interpolação e, principalmente, abordagens baseadas em aprendizado com foco em aprendizado profundo, *e.g.* redes residuais, redes convolucionais e GAN's. Em especial, o enfoque será nas técnicas que utilizaram GAN's.

A função de interpolação *bicubic*, a qual estimou os valores contínuos intermediários dos dados discretos de uma fotografia que foi representada como uma amostra discreta da realidade contínua, tem mais precisão que o algoritmo *nearest-neighbor* ou o método de interpolação *bilinear*. O método *bicubic* provou-se mais eficiente do que outros métodos de interpolação (KEYS, 1981).

Desenvolvido por Dong et al. (2015), *Super Resolution Convolutional Neural Network* (SRCNN) foi o primeiro modelo de *Convolutional Neural Network* (CNN) proposto para a super resolução. O mapeamento proposto por essa rede recebe como entrada uma imagem em baixa resolução e a retorna em alta resolução. Porém, diferentemente de modelos tradicionais que lidam com cada componente separadamente, esse processo otimiza todas as camadas de forma conjunta. Ainda que essa CNN possua uma estrutura leve, a mesma consegue alcançar alta qualidade e alta velocidade para uso prático *online*.

Inspirada na VGG-net de Simonyan e Zisserman (2015), foi construída a *Very Deep Convolutional Neural Network* (VDSR): "um modelo mais profundo, que atinge maior acurácia e apresenta um total de 20 camadas" (KIM; LEE; LEE, 2016), o qual obteve grande desempenho em relação à SRCNN clássica. Tendo em vista esse modelo, ao realizar uma grande cascata de filtros numa rede profunda, foram

exploradas com eficiência informações contextuais de largas regiões da imagem.

Entretanto, por ser uma rede muito profunda, a velocidade de convergência acaba se tornando um problema crítico durante o treinamento. Para solucionar tal problema, um método de treino simples e eficiente foi implementado, utilizando aprendizado somente de resíduos e uma taxa de aprendizado extremamente alta. Como resultado, foi apresentada uma melhor performance na acurácia em relação aos modelos concorrentes e um aperfeiçoamento visual notável nas imagens obtidas.

Com o desenvolvimento de novas *Deep Convolutional Neural Networks* (DCNN's), o avanço na área da super resolução teve um grande progresso. Em particular, as técnicas de aprendizado residual exibiram alta performance. Devido a esse fato, Lim et al. (2017) criou a *Enhanced Deep Super Resolution Network* (EDSR), a qual apresentou performance superior aos seus predecessores.

Seu desempenho foi significativo devido à remoção de módulos desnecessários presentes em uma CNN convencional, *e.g.* camadas de *Batch Normalization* e camada final com função de ativação do tipo *Rectified Linear Units* (ReLU). Também foi proposto nesse trabalho um método novo de treino e o sistema *Multi-scale Deep Super Resoluiton* (MDSR), o qual pode reconstruir imagens em alta resolução com diferentes fatores de escala em um único modelo.

Um mecanismo que desempenhou um papel crucial na tarefa de super resolução foi a extração de características informativas. A *Channel Attention* de Woo et al. (2018) se provou eficiente na preservação de características ricas em informação de cada camada da rede. Também foi definida como pilar da rede a *Residual Channel Attention Network* (RCAN) implementada por Zhang et al. (2018).

Contudo, a *Channel Attention* trata cada camada de convolução como sendo um processo diferente, o que faz com que haja a perca da correlação entre diferentes camadas. Para solucionar tal problema, Niu et al. (2020) elaboraram a *Holistic Attention Network* (HAN), um modelo que consiste de dois módulos: o *Layer Attetion Module* (LAM) e o *Channel-Spatial Attention Module* (CSAM). Ambos os módulos servem como interdependência holística entre as camadas, os canais de cor e as

posições.

Em suma, a LAM enfatiza adaptativamente as características hierárquicas por considerar a correlação entre as diferentes camadas da rede. Enquanto isso, a CSAM modela explicitamente as interdependências das características espaciais e de canais de cor, podendo assim assimilar as características inter-canal e intra-canal. Como resultado desse processo, a rede proposta tem uma habilidade discriminativa maior que as demais.

Apesar de descobertas contínuas no aprimoramento da acurácia e da velocidade de super resolução em imagens únicas, fazendo uso de CNN's mais rápidas e profundas, um problema permaneceu não resolvido: como recuperar os menores detalhes da textura quando, especificamente, é feita super resolução com fatores de grandes escalas?

Ainda que métodos de super resolução fundamentados em otimização possuam um alto *Peak Signal-to-Noise Ratio* (PSNR), muitas vezes faltam detalhes de alta frequência e as imagens resultam num visual perceptivelmente insatisfatório, no sentido de que os modelos falharam em retratar fielmente os detalhes em alta resolução.

Ledig et al. (2017) apresentaram a SRGAN, uma GAN para super resolução. Para o desenvolvimento desse modelo, foi proposta uma função de *perceptual loss*, a qual é composta de um *loss* adversário e um *loss* de conteúdo. O primeiro levou a solução para o campo de imagens naturais utilizando a rede discriminadora para diferenciar as imagens geradas das originais. Já o segundo, teve o propósito de alcançar similaridade perceptual da imagem ao invés de uma similaridade a nível de *pixels*.

Em Saharia et al. (2021), foi idealizada uma nova abordagem para a super resolução de imagem, chamada de *Super-Resolution via Repeated Refinement* (SR3). A ideia desse modelo é funcionar através de refinamentos repetidos. Sua função adapta outros modelos do tipo *denoising diffusion probabilistic*, citados em Ho, Jain e Abbeel (2020), com intuito de gerar imagens condicionais e desempenhar a super

resolução através de um processo iterativo de remoção de ruído estocástico.

O processo de geração desse novo modelo começa com um ruído Gaussiano puro e, iterativamente, refina a imagem ruidosa com o uso de uma *U-Net*[5] treinada em remoção de vários níveis de ruído. Como métrica de análise, uma avaliação com humanos foi feita com imagens em super resolução geradas numa escala de *8x*.

O resultado dessa avaliação foi a obtenção de uma taxa de *50%* de enganação, sendo que, métodos baseados em GAN's comparados à *SR3*, como os desenvolvidos em Menon et al. (2020) e Chen et al. (2017), só conseguiram atingir, no máximo, uma taxa de *34%* (SAHARIA et al., 2021). Sendo assim, a *SR3* exibiu alta performance em super resolução utilizando diferentes valores de escala em rostos e imagens naturais.

3.3 *Enhanced Super Resolution* GAN (ESRGAN)

Mesmo considerando a alta capacidade da SRGAN de gerar texturas realistas em super resolução de uma imagem, seus resultados ainda apresentaram alguns artefatos indesejados, que serão discutidos na seção 4.4. Com o objetivo de solucionar essa questão e assim melhorar ainda mais a qualidade visual de uma imagem, Wang et al. (2018) criaram a *Enhanced SRGAN* (ESRGAN), na qual incluíram o estudo de três componentes chaves de seu antecessor: a arquitetura da rede, a *loss* adversária e a *perceptual loss*.

A partir desse estudo, a ESRGAN aperfeiçoou esses componentes. Em sua essência, o novo modelo introduziu o *Residual-in-Residual Dense Block* (RRDB) como composição básica da rede, sem a utilização de camadas de *Batch Normalization*. Aproveitando o proposto por Jolicoeur-Martineau (2018), foi aplicada a ideia de fazer com que o discriminador previsse uma realidade aproximada ao invés do valor absoluto.

Por último, a *perceptual loss* foi aprimorada com o propósito de tirar proveito das

[5]Rede Neural de convolução para a segmentação de imagem biomédica (RONNEBERGER; FISCHER; BROX, 2015)

características antes da função de ativação, o que passou a oferecer maior supervisão quanto à consistência do brilho e à recuperação da textura da imagem. A figura 3.1 apresenta a diferença entre a arquitetura da SRGAN e da ESRGAN.

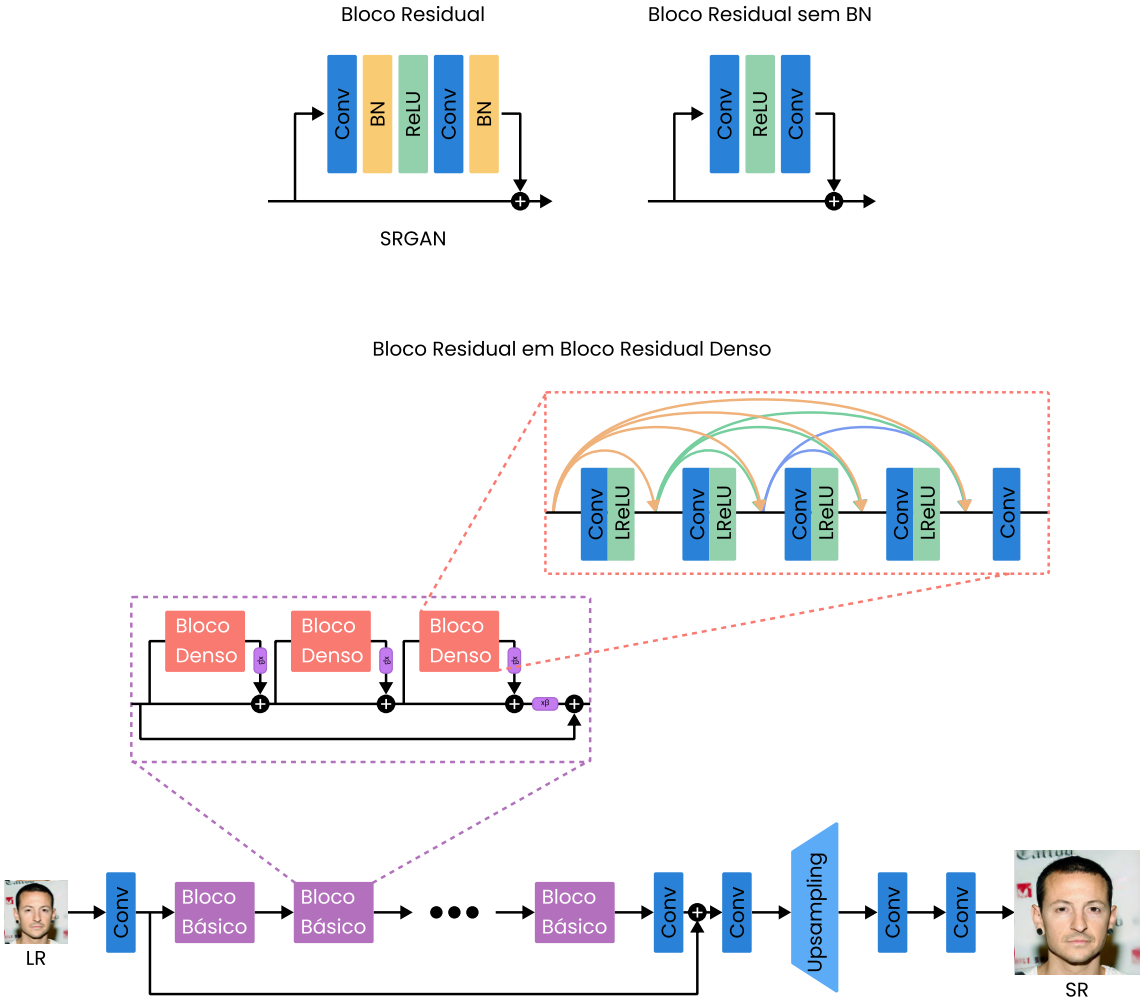


Figura 3.1: Arquitetura da rede ESRGAN. (WANG et al., 2018)

Com esses aprimoramentos, a proposta da ESRGAN alcançou consistentemente uma qualidade visual melhor, proporcionando assim texturas naturais e realísticas comparadas à SRGAN.

Visto que, modelos de super resolução baseados em GAN's buscam atingir performance superior a modelos com outras arquiteturas, serão realizados no próximo capítulo, experimentos comparando a ESRGAN com alguns dos modelos que implementaram essa técnica.

Capítulo 4

Experimentos

Neste capítulo, serão descritos os experimentos realizados, as metodologias e bases de dados neles utilizadas, assim como feito no trabalho de Saharia et al. (2021). Nesses experimentos, foram ainda aplicadas práticas que, de acordo com Wang et al. (2018), evitam o enviesamento dos testes, como por exemplo: utilizar bases de dados distintas para treinos e para testes, assim como selecionar bases que contenham alta variação de atributos.

4.1 Base de Dados

A primeira base de dados manipulada foi a *Flickr-Faces-HQ* (figura 4.1) de Karras, Laine e Aila (2019), a qual contém um conjunto de 70.000 imagens de faces humanas em resolução 1024×1024 *pixels*. Nessa base, características específicas são identificadas como: alta variação em termos de idade, etnia e planos de fundo. A diversidade dessa base auxilia o gerador dos modelos de GAN's na produção de resultados mais realistas (WANG et al., 2018) e, tal como feito na comparação do trabalho de Saharia et al. (2021), utilizaremos essa base para treinar os modelos de redes neurais citados neste capítulo.

Figura 4.1: Imagens presentes na Flickr-Faces-HQ (KARRAS; LAINE; AILA, 2019)

Ainda seguindo o experimento de Saharia et al. (2021), utilizaremos para teste, tanto nas redes neurais quanto nos métodos de processamento de imagens e nas GAN's, a base *CelebA-HQ* (figura 4.2) encontrada em Karras et al. (2017), a qual detém 30.000 imagens de faces de diferentes celebridades, também em resolução 1024×1024 *pixels*.

Figura 4.2: Imagens presentes na CelebA-HQ (KARRAS et al., 2017)

4.2 Métricas

A fim de avaliar os experimentos realizados neste capítulo, serão fundamentadas duas métricas comumente utilizadas na avaliação da qualidade de imagens, uma vez que qualquer processo aplicado numa imagem pode causar perda de qualidade ou de informação da mesma (HORé; ZIOU, 2010).

A primeira métrica aplicada é a *Peak Signal-to-Noise Ratio* (PSNR), para a qual, dada uma imagem referência f e uma imagem teste g, ambas de tamanho $M \times N$, calcula-se a relação de qualidade entre elas (HORé; ZIOU, 2010). O valor obtido na PSNR tende ao infinito conforme o resultado do MSE tende a zero, demonstrando assim que quanto maior o valor resultado da PSNR, maior a relação de qualidade entre as duas imagens. No entanto, quanto menor for esse valor, maior a diferença encontrada nessa relação. A fórmula 4.1 representa o cálculo da PSNR.

$$\begin{gathered} PSNR(f,g) = 10log_{10}(255^2/MSE(f,g)) \\ onde \\ MSE(f,g) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(f_{ij} - g_{ij})^2 \end{gathered} \tag{4.1}$$

Uma outra abordagem é a *Structural Similarity Index Measure* (SSIM) criada por Wang et al. (2004), a qual também calcula a similaridade entre duas imagens. Porém, diferentemente da métrica PSNR, essa não utiliza métodos com somatórios para calcular a semelhança entre imagens, mas sim qualquer distorção da imagem através da combinação de três fatores: *loss of correlation*, *luminance distortion* e *contrast distortion*, conforme representados na fórmula 4.2.

$$
\begin{aligned}
SSIM(f,g) &= l(f,g)c(f,g)s(f,g) \\
&\quad onde \\
l(f,g) &= \frac{2\mu_f\mu_g + C_1}{\mu_f^2 + \mu_g^2 + C_1} \\
c(f,g) &= \frac{2\sigma_f\sigma_g + C_2}{\sigma_f^2\sigma_g^2 + C_2} \\
s(f,g) &= \frac{\sigma_{fg} + C_3}{\sigma_f\sigma_g + C_3}
\end{aligned}
\tag{4.2}
$$

O termo l da equação representa a comparação da luminosidade, a qual mede a proximidade entre as duas imagens. De acordo com Horé e Ziou (2010), o valor desse termo é máximo e igual a 1 somente quando $\mu_f = \mu_g$. O segundo termo caracteriza a proximidade do contraste entre as imagens. Da mesma forma, seu valor é máximo e igual a 1 somente quando $\sigma_f = \sigma_g$. Por fim, o terceiro termo mede o coeficiente de correlação entre essas imagens.

O intervalo de valores possíveis obtidos pela SSIM se encontra entre $[0, 1]$, onde 0 indica a ausência de correlação entre as imagens e 1 simboliza que $f = g$. Além disso, constantes positivas C_1, C_2 e C_3 servem de auxílio para evitar denominadores nulos (HORé; ZIOU, 2010).

A partir dessas métricas, serão realizados experimentos com as bases de dados da seção 4.1, utilizando as metodologias apontadas na seção 4.3.

4.3 Metodologia

Nesta seção, será apontada a metodologia desse trabalho, a qual será baseada em metodologias usualmente utilizadas nos trabalhos de super resolução citados na seção 3.2.

Definimos, por convenção, em todos os experimentos, a super resolução de imagem na escala de fator $4\times$, *i.e.* gerando imagens com resolução quatro vezes maior em

relação à origem. Além disso, fixamos que a resolução de origem será de 64 *pixels* e as imagens geradas terão 256 *pixels*. Nos modelos de redes neurais e nas GAN's, utilizamos apenas uma imagem por iteração na etapa de treino.

Com intuito de fazer uso das metodologias apresentadas nos trabalhos citados no capítulo 3, preservamos alguns dados experimentais em certos modelos, como por exemplo: bases de dados utilizadas, quantidade de épocas no treinamento e no pré-treinamento, tempo referente a essas etapas, taxa de aprendizado da rede neural, taxa de decaimento do aprendizado e número de blocos de convolução.

Uma outra estratégia adotada neste trabalho foi obter a baixa resolução das imagens da base de dados para treino. Para isso, foi aplicada a técnica *downsampling*, *i.e.* redução da resolução de imagens, utilizando o algoritmo *Bicubic*. Além dessa técnica, foi realizado o espelhamento vertical e horizontal das imagens da base, assim como feito por Wang et al. (2018) e Lim et al. (2017), com o intuito de aumentar a quantidade de desses.

Outra abordagem da metodologia foi manter os hiperparâmetros definidos nos trabalhos comparados, sendo alguns deles: número de blocos das redes neurais, quantidade de camadas presentes nessas redes, filtros das camadas de convolução e o tamanho do *kernel* dessas camadas.

Inicialmente, utilizamos o mesmo modelo da SRCNN apresentada no trabalho de Dong et al. (2015). Nesse modelo, foi realizado um treino de 200.000 iterações. Fixamos uma taxa de aprendizado no valor de 10^{-4}, assim como a definição de somente 3 camadas na rede neural e o otimizador desse modelo definido como sendo o algoritmo *Adam*[1].

Implementamos a rede EDSR como proposto por Lim et al. (2017), na qual foi aplicado um treino de 300.000 iterações e uma taxa de aprendizado inicial no valor de 10^{-4}. Posteriormente, essa taxa foi decaída pela metade após a iteração 200.000. Também foram inseridos 32 blocos de convolução e utilizado como otimizador, o *Adam*.

[1] algoritmo baseado em gradiente de primeira ordem de funções objetivas estocásticas, baseado em estimativas adaptativas de momentos de ordem inferior (Kingma & Ba, 2014)

A SRGAN adotada nesse trabalho foi implementada de acordo com Ledig et al. (2017). Inicialmente, um pré-treino do componente gerador foi realizado com o valor de 200.000 iterações. O treino da rede completa compreendeu 60.000 iterações, levando em consideração que o treinamento da rede no intervalo de 50 a 100 mil iterações produz apenas pequenas mudanças visuais de melhoria (LEDIG et al., 2017). Foram ainda utilizadas nesse experimento, as funções de custo mencionadas na seção 3.2.

Outra GAN utilizada nesse experimento foi a ESRGAN de Wang et al. (2018). Nela, definimos parâmetros relevantes tanto para o pré-treino do componente gerador, quanto para a rede em si. Foram realizadas 500.000 iterações de pré-treino no gerador, sendo definida a taxa de aprendizado de 2×10^{-4} e a cada 200.000, o decaimento dessa taxa pela metade.

Por fim, após essa etapa, treinamos a rede completa em 20.000 iterações, tendo sido estabelecidas as taxas de aprendizado do componente gerador e do discriminador com os valores de 10^{-4} e 4×10^{-4}, respectivamente. No caso do discriminador, também decaímos essas taxas pela metade nas iterações: 50.000, 10.0000, 20.0000 e 30.0000.

Vale ressaltar que, como em Wang et al. (2018), foi utilizada para o treino da ESRGAN, a base de dados *Flickr-Faces-HQ*, uma vez que, de acordo com o esse trabalho, a sua alta variação de dados gera melhores resultados na avaliação da métrica PSNR e nos efeitos qualitativos.

Com base nas metodologias aplicadas nos pré-treinos, treinos e testes, serão apresentados na próxima seção, os resultados obtidos nos experimentos e informações visuais para a melhor compreensão dos mesmos.

4.4 Resultados

Nesta seção, apresentaremos os resultados obtidos com os testes de cada modelo de super resolução experimentado, após a aplicação das metodologias supracitadas. Para isso, serão descritas as análises qualitativas e quantitativas desses testes.

4.4.1 Análise Quantitativa

Utilizando as métricas apresentadas na seção 4.2, foram obtidos nesse experimento os seguintes resultados na comparação da ESRGAN com outros método de super resolução:

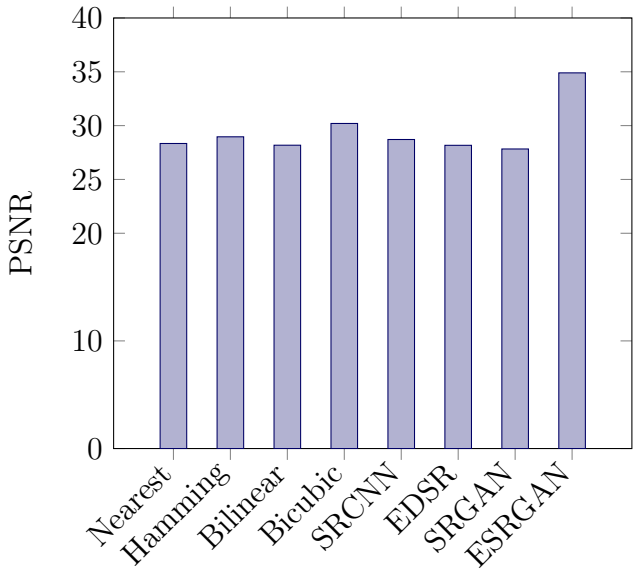


Figura 4.3: Média de testes utilizando a métrica PSNR

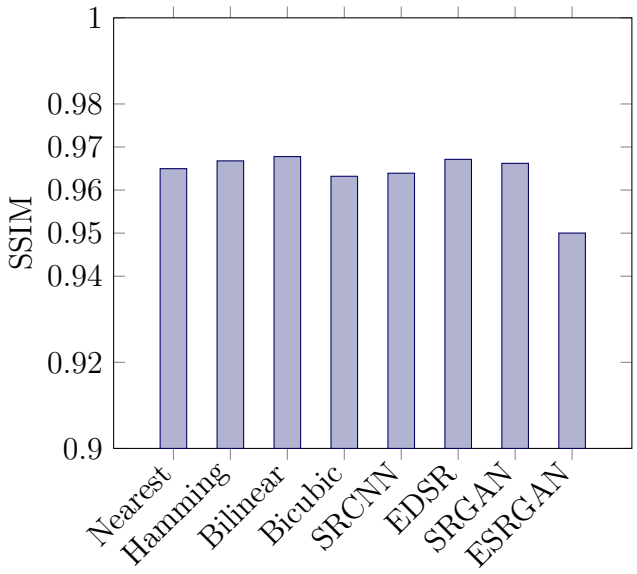


Figura 4.4: Média de testes utilizando a métrica SSIM

A partir dos resultados obtidos, foi possível verificar que dentre os modelos de super resolução analisados, a ESRGAN apresentou valor superior na comparação da

média da métrica de avaliação PSNR. Em contrapartida, foi obtido valor inferior na métrica SSIM. Assim como na comparação realizada no trabalho de Wang et al. (2018), existe a possibilidade desses resultados quantitativos não ultrapassarem os dos modelos comparados. Apesar de não ter atingido melhores resultados numéricos na média de testes, foi possível verificar que os resultados visuais da ESRGAN, apresentados a seguir, são perceptivelmente melhores, comparados aos dos outros modelos analisados.

4.4.2 Análise Qualitativa

Nessa subseção, serão relatados e exibidos os principais padrões visuais encontrados durante os experimentos dos modelos comparados.

Um dos pontos mais notável foi a geração de artefatos indesejáveis (figura 4.5) por parte do modelo SRGAN durante o treinamento e, consequentemente, nos testes. Vale lembrar que a proposta de Wang et al. (2018) foi justamente melhorar a arquitetura da SRGAN, a fim de evitar essa inconveniência. Esse objetivo foi satisfatoriamente atingido pela ESRGAN, visto que nenhum artefato foi encontrado durante o treinamento e testes realizados em seus experimentos.

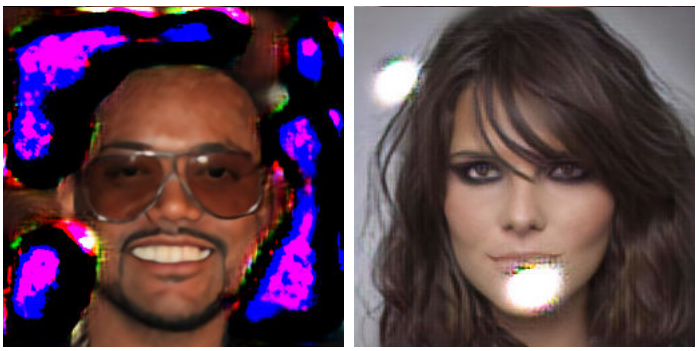


Figura 4.5: Artefatos indesejados gerados durante o pré-treino e treino da SRGAN.

Dentre os casos de experimentos que geraram esses artefatos, convém mencionar alguns padrões encontrados. Na figura 4.6, foi possível observar que a maior parte dos artefatos gerados na etapa de teste, foi localizada em pontos da imagem que continham alta luminosidade, como por exemplo: queixo, bochechas e testa.

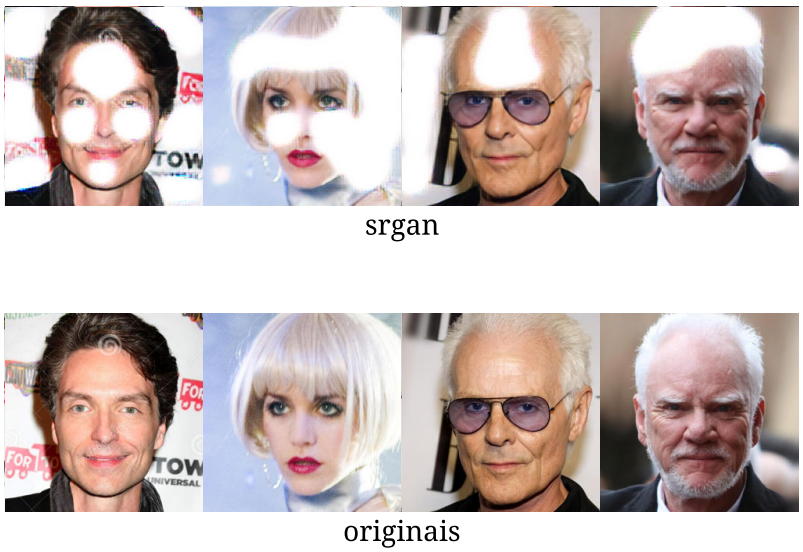


Figura 4.6: Artefatos gerados em pontos com alta luminosidade.

Outro padrão encontrado com frequência, ainda nos testes da SRGAN, foi a presença de alta saturação[2] em grande parte das imagens que apresentaram artefatos indesejados, ao passo que esses artefatos foram encontrados em poucas imagens com baixa saturação. A figura 4.7 mostra alguns desses padrões relacionados à saturação.

Figura 4.7: Influência da saturação na geração de artefatos.

Foi possível notar ainda, que a maior parte dos testes realizados com imagens de faces de celebridades negras não apresentaram artefatos e, quando ocorreram, foram

[2]Intensidade de uma cor. Quanto maior a saturação de uma cor, mais vívida ela é, e quanto menor a saturação de uma cor, mais próxima ela fica do cinza.

falhas mínimas, assim como evidenciado na figura 4.8.

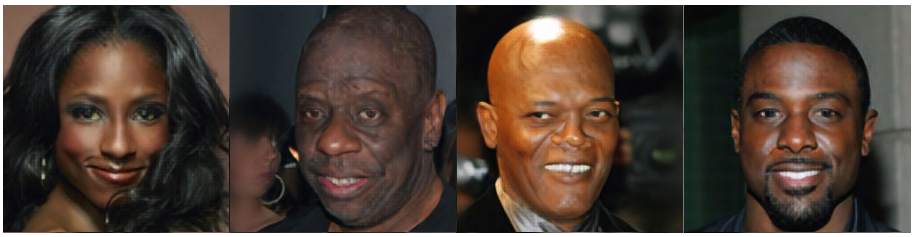


Figura 4.8: Testes de super resolução da SRGAN em imagens de faces negras.

Diferentemente da etapa de teste da SRGAN, em alguns casos durante o pré-treinamento, foram identificados artefatos, principalmente nas bordas das imagens (figura 4.9), isolando, por vezes, as faces dos planos de fundo.

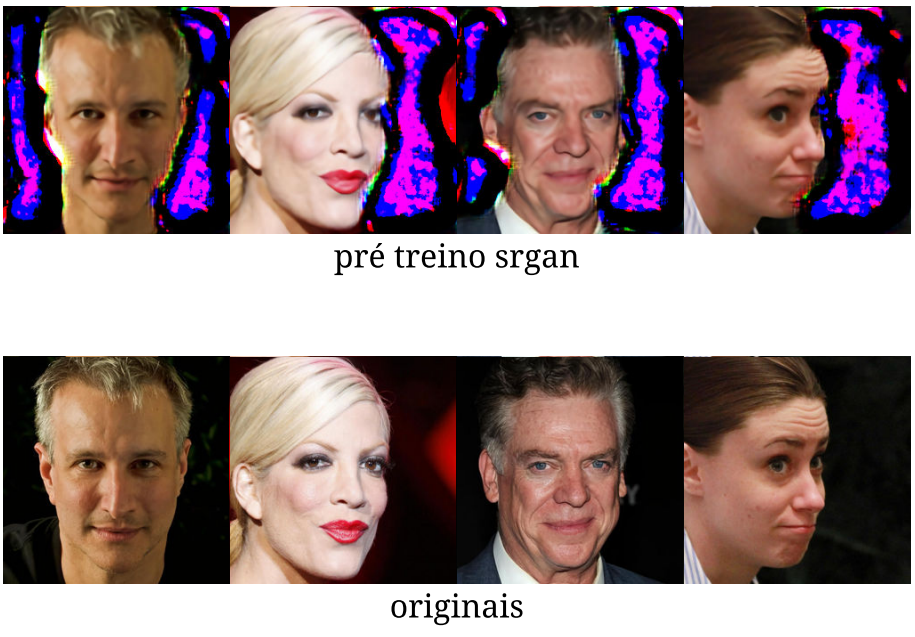


Figura 4.9: Artefatos gerados durante o pré-treinamento da SRGAN.

Na comparação dos resultados visuais de todos os modelos experimentados, foi perceptível a diferença da qualidade das imagens após a aplicação da super resolução utilizando GAN's, uma vez que sua proposta nessa área é justamente se sobressair na comparação com os modelos concorrentes. As figuras 4.10, 4.11 e 4.12 refletem essa discrepância.

Figura 4.10: Exemplo 1 das imagens obtidas através dos modelos citados neste trabalho.

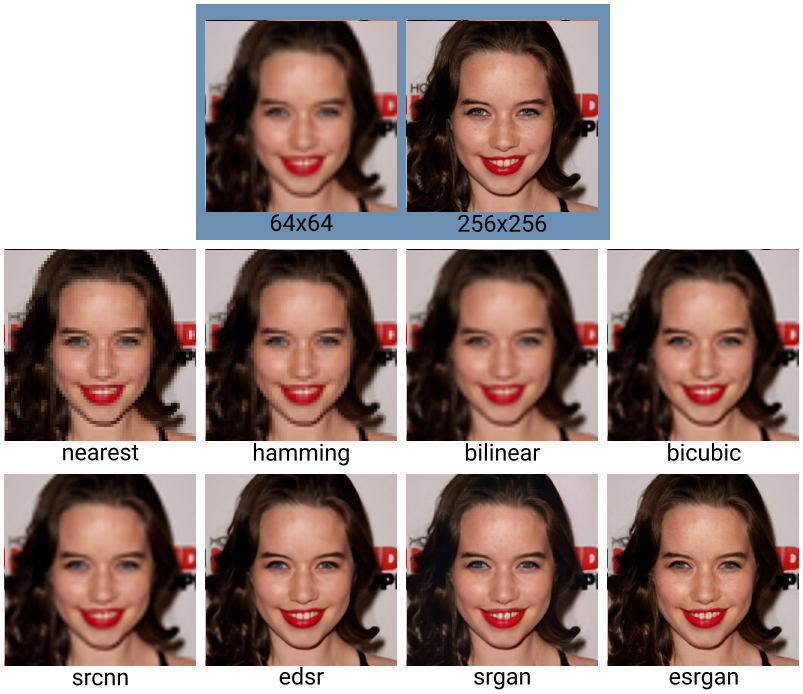


Figura 4.11: Exemplo 2 das imagens obtidas através dos modelos citados neste trabalho.

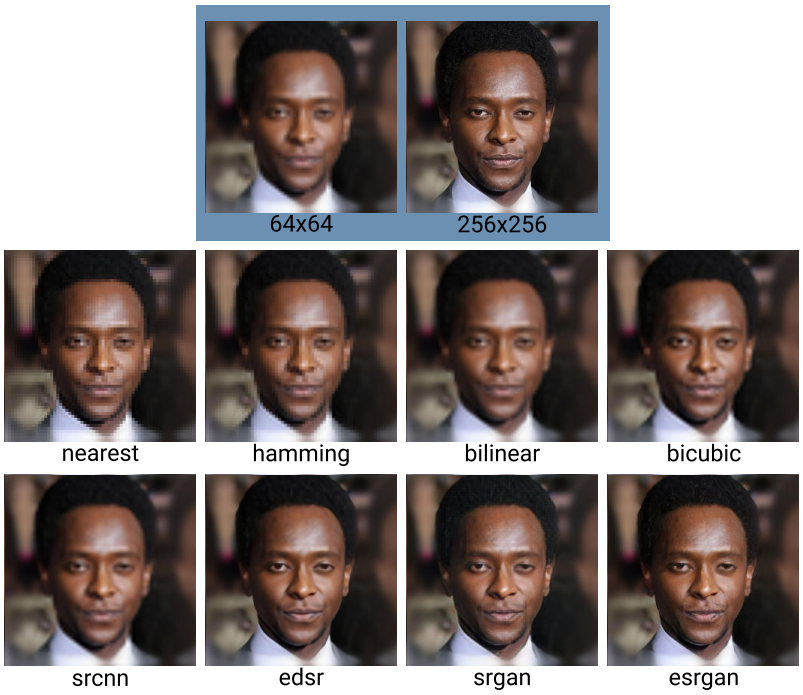


Figura 4.12: Exemplo 3 das imagens obtidas através dos modelos citados neste trabalho.

Por fim, embora os modelos baseados em GAN's tenham superado os outros, ainda foram identificadas algumas falhas em seus testes. Conforme apresentado nas figuras 4.13 e 4.14, as GAN's utilizadas neste trabalho não conseguiram representar detalhes minuciosos, como o brilho dos olhos. Além dessa ausência de detalhes, foram notadas falhas na representação de certas características faciais.

srgan

originais

Figura 4.13: Exemplo 1 de falhas encontradas nos testes da SRGAN.

esrgan

originais

Figura 4.14: Exemplo 2 de falhas encontradas nos testes das ESRGAN.

Capítulo 5

Conclusão

Neste capítulo, serão pontuadas as considerações finais referentes a este trabalho, assim como as limitações nele encontradas e a reflexão sobre possíveis trabalhos futuros que venham a ter como base as análises realizadas no capítulo 4.

5.1 Considerações finais

Levando em consideração o fato da super resolução de imagem ser um processo que está em constante aperfeiçoamento ao longo das últimas décadas, foi proposto neste trabalho o levantamento e a comparação dos principais métodos que desenvolveram essa função. Nesse levantamento, foram identificados tipos de modelos promissores baseados em redes neurais e, em especial, os baseados em GAN's.

Após identificarmos que a ESRGAN, sendo um método baseado em Redes Adversárias Generativas, apresenta grande desempenho em relação a outros modelos, realizamos uma análise qualitativa e quantitativa dos modelos, a fim de verificar a viabilidade de suas aplicações.

Como parte das considerações finais, também é importante pontuar que, nos trabalhos de super resolução baseados em GAN's, *e.g.* SRGAN e ESRGAN, não foram encontrados estudos mais profundos que justificassem alguns padrões verificados nas análises visuais dos testes. Essa ausência foi um limitante para concluirmos com

propriedade o porquê desses padrões.

Por fim, assim como citado por Ledig et al. (2017) e Wang et al. (2018), também consideramos que os resultados quantitativos dos experimentos presentes na seção 4.4.1 não refletem a eficácia das GAN's em representar detalhes de texturas mais realistas.

5.2 Limitações e trabalhos futuros

São duas as principais limitações encontradas na realização deste trabalho e ambas referem-se ao processo de treinamento dos experimentos. A primeira deve-se ao fato de que, tanto os modelos baseados em redes neurais quanto aqueles baseados em GAN, por mais que os últimos tenham gerado resultados satisfatórios, ambos necessitam de muito tempo na etapa de treino para atingirem seus potenciais. Tal fato exige o uso de memórias com alta capacidade de processamento para compensar o tempo investido.

Além disso, a geração de imagens através da *Central Process Unit* (CPU) mostrou-se inviável, uma vez que seu processamento é limitado. E, mesmo que a utilização da *Graphics Processing Unit* (GPU) torne o processo extremamente mais rápido, ainda se faz necessária uma quantidade alta de memória para executar modelos com muitas camadas, principalmente quando aumenta-se as escalas e tamanhos das imagens na super resolução.

A fim de contornar a limitação relacionada ao aumento de escalas e resolução das imagens, foi aplicada neste trabalho uma técnica para reduzir o uso de memória. Essa técnica constitui-se em: dividir uma imagem origem de resolução x em múltiplos quadrantes; aplicar a super resolução individual nesses quadrantes utilizando o fator de escala y; uni-los com intuito de obter a imagem final com resolução y vezes maior do que a da imagem origem (figura 5.1).

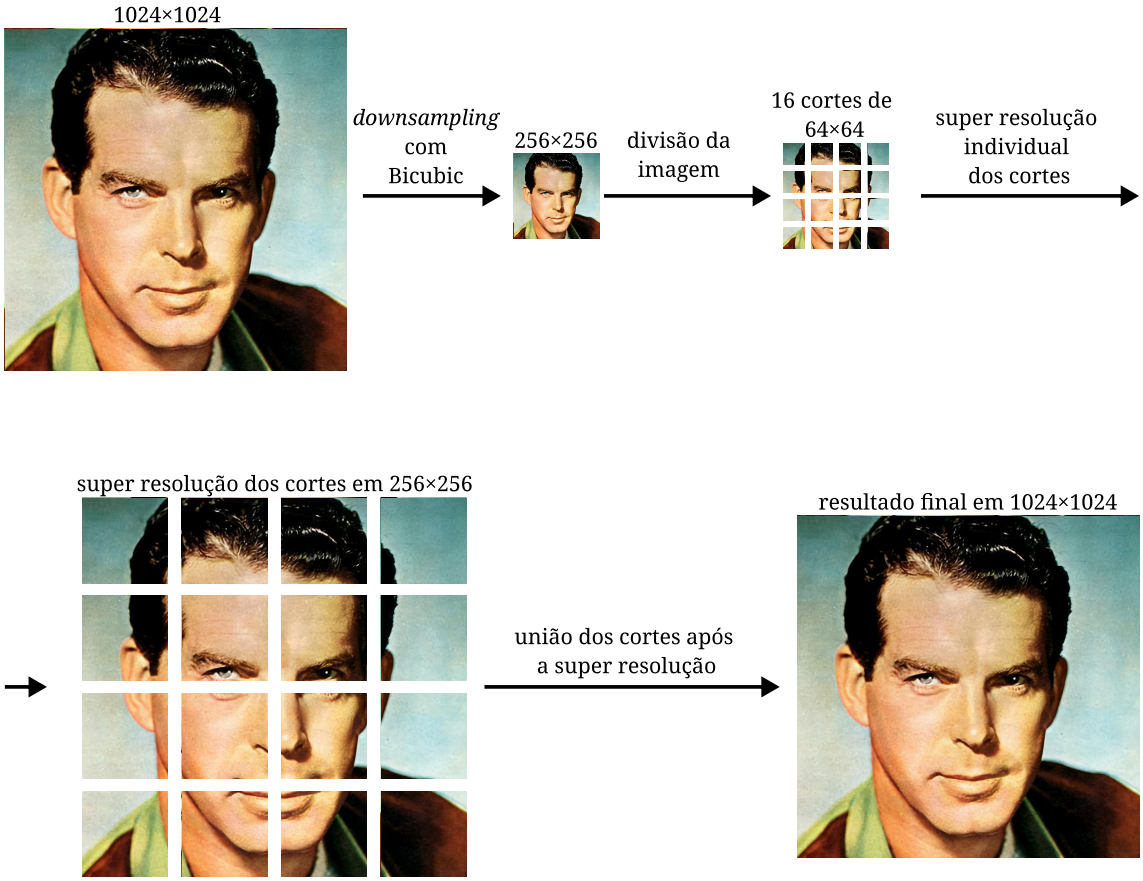


Figura 5.1: Técnica de divisão e conquista aplicada neste trabalho.

A segunda principal limitação está relacionada ao ambiente de execução dos modelos. Devido ao fato das versões de tecnologias e *frameworks* serem atualizadas frequentemente, foi necessário reduzi-las a fim de executar os modelos baseados em GAN's, uma vez que esses foram desenvolvidos há alguns anos e por isso seus códigos não são compatíveis com as versões atuais.

Quanto aos métodos de processamento de imagens comparados, uma limitação identificada foi a dificuldade em encontrar modelos atuais, razão pela qual as comparações realizadas neste trabalho utilizaram métodos antigos baseados em interpolação, os quais não obtiveram tantos avanços após serem desenvolvidos.

Outra questão limitante está ligada à quantidade de tempo que se requer nas etapas de treinamento para que sejam realizadas as análises quantitativas e, principalmente, as qualitativas. A mesma limitação ocorre quando se utiliza diferentes

bases de treino, uma vez que essas costumam conter uma quantidade vasta de dados na aplicação de super resolução. Tais fatos impedem que um bom estudo sobre métodos de super resolução seja realizado em curto ou médio prazo.

Pontuamos a seguir, temas para o desenvolvimento de eventuais trabalhos futuros sobre métodos de super resolução, sendo os estudos de comparações entre outros modelos de super resolução, um dos que consideramos mais necessários, uma vez que novos modelos são desenvolvidos frequentemente, *e.g.* SR3 de Saharia et al. (2021) e Real-ESRGAN, extensão da ESRGAN implementada por Wang et al. (2021).

Outras sugestões seriam: ampliar este estudo experimental utilizando diversas outras bases de dados para treino e teste; expandir a escala de super resolução; aumentar a resolução das imagens experimentadas. No caso de um experimento com diversas outras bases, ainda que o processo demande longo prazo, a sua realização pode contribuir para a constatação do motivo pelo qual os padrões encontrados na seção 4.4.2 foram tão evidentes.

Tal como feito por Wang et al. (2021) na criação da Real-ESRGAN, os quais inseriram a U-Net na estrutura da ESRGAN, outra abordagem de estudo seria identificar características úteis, presentes em outros tipos de modelos de redes neurais, visando introduzi-las na arquitetura de um modelo de super resolução já existente.

Por fim, considerando que a inovação dos modelos de super resolução é constituída, em sua maioria, pela identificação de falhas em modelos antecessores e, consequentemente, a inovação desses, consideramos produtiva a prática dessas mesmas inovações em GAN's atuais, visto que foi esse o feito por Wang et al. (2018) na criação da ESRGAN, aperfeiçoando o modelo da SRGAN.

Referências

ARJOVSKY, M.; CHINTALA, S.; BOTTOU, L. *Wasserstein GAN*. 2017.

CHAUDHURI, S. (Ed.). *Super-Resolution Imaging*. [S.l.]: Springer US, 2002.

CHEN, Y. et al. *FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors*. 2017.

CHENG, J. et al. *Generative Adversarial Networks: A Literature Review*. 2020.

CRESWELL, A. et al. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, Institute of Electrical and Electronics Engineers (IEEE), v. 35, n. 1, p. 53–65, Jan 2018. ISSN 1053-5888. Disponível em: <http://dx.doi.org/10.1109/MSP.2017.2765202>.

CUNNINGHAM, P.; CORD, M.; DELANY, S. J. *Machine Learning Techniques for Multimedia*. [S.l.]: Springer, Berlin, Heidelberg, 2008.

DONG, C. et al. *Image Super-Resolution Using Deep Convolutional Networks*. 2015.

FREEMAN, W.; PASZTOR, E.; CARMICHAEL, O. Learning low-level vision. *International Journal of Computer Vision*, Springer Netherlands, v. 40, n. 1, p. 25–47, out. 2000. ISSN 0920-5691.

GOODFELLOW, I. et al. Generative adversarial nets. In: GHAHRAMANI, Z. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2014. v. 27. Disponível em: <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>.

GUI, J. et al. A review on generative adversarial networks: Algorithms, theory, and applications. *CoRR*, abs/2001.06937, 2020. Disponível em: <http://dblp.uni-trier.de/db/journals/corr/corr2001.html#abs-2001-06937>.

GURNEY, K. *An Introduction to Neural Networks*. [S.l.]: Taylor & Francis, Inc., 1997. ISBN 978-1-85728-673-1.

HENG, L.; DONGDONG, H. *A Short Survey of Image Super Resolution Algorithms*. 2015.

HO, J.; JAIN, A.; ABBEEL, P. *Denoising Diffusion Probabilistic Models*. 2020.

HORé, A.; ZIOU, D. Image quality metrics: Psnr vs. ssim. In: *2010 20th International Conference on Pattern Recognition*. [S.l.: s.n.], 2010. p. 2366–2369.

JOLICOEUR-MARTINEAU, A. *The relativistic discriminator: a key element missing from standard GAN*. 2018.

KARRAS, T. et al. Progressive growing of gans for improved quality, stability, and variation. *CoRR*, abs/1710.10196, 2017. Disponível em: <http://arxiv.org/abs/1710.10196>.

KARRAS, T.; LAINE, S.; AILA, T. *A Style-Based Generator Architecture for Generative Adversarial Networks*. 2019.

KEYS, R. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 29, n. 6, p. 1153–1160, 1981.

KIM, J.; LEE, J. K.; LEE, K. M. Accurate image super-resolution using very deep convolutional networks. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 1646–1654.

LEDIG, C. et al. *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*. 2017.

LIM, B. et al. *Enhanced Deep Residual Networks for Single Image Super-Resolution*. 2017.

LIU, W. et al. *Beyond Human-level License Plate Super-resolution with Progressive Vehicle Search and Domain Priori GAN*. 2017.

MACHADO, A.; SOUKI, B. Simplificando a obtenção e a utilização de imagens digitais: Scanners e câmeras digitais. *Revista Dental Press De Ortodontia E Ortopedia Facial*, v. 9, 08 2004.

MENON, S. et al. *PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models*. 2020.

NIU, B. et al. *Single Image Super-Resolution via a Holistic Attention Network*. 2020.

RONNEBERGER, O.; FISCHER, P.; BROX, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015.

SAHARIA, C. et al. *Image Super-Resolution via Iterative Refinement*. 2021.

SIMA, J. *Introduction to Neural Networks*. 1998.

SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015.

SU, H. et al. *MangaGAN: Unpaired Photo-to-Manga Translation Based on The Methodology of Manga Drawing*. 2020.

WALT, S. van der. Super-resolution imaging. 01 2010.

WANG, S.-C. *Interdisciplinary Computing in Java Programming*. [S.l.]: Springer, Boston, MA, 2003.

WANG, X. et al. *Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data*. 2021.

WANG, X. et al. *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks*. 2018.

WANG, Z. et al. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, v. 13, n. 4, p. 600–612, 2004.

WEI, X. et al. *Improving the Improved Training of Wasserstein GANs: A Consistency Term and Its Dual Effect*. 2018.

WOO, S. et al. *CBAM: Convolutional Block Attention Module*. 2018.

ZHANG, Y. et al. *Image Super-Resolution Using Very Deep Residual Channel Attention Networks*. 2018.

ZHU, J.; YANG, G.; LIO, P. *How Can We Make GAN Perform Better in Single Medical Image Super-Resolution? A Lesion Focused Multi-Scale Approach*. 2019.

ZHU, J.-Y. et al. *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*. 2020.