



**UNIVERSIDADE FEDERAL DO TOCANTINS
CÂMPUS UNIVERSITÁRIO DE PALMAS
CURSO DE CIÊNCIA DA COMPUTAÇÃO**

EXPLORANDO A SUPER-RESOLUÇÃO PARA RECONHECIMENTO FACIAL

PATRICK ANDERSON MATIAS DE ARAÚJO

PALMAS (TO)

2021

PATRICK ANDERSON MATIAS DE ARAÚJO

EXPLORANDO A SUPER-RESOLUÇÃO PARA RECONHECIMENTO FACIAL

Trabalho de Conclusão de Curso II apresentado à Universidade Federal do Tocantins para obtenção do título de Bacharel em Ciência da Computação, sob a orientação do(a) Prof.(a) Dr. Eduardo Ferreira Ribeiro.

Orientador: Dr. Eduardo Ferreira Ribeiro

PALMAS (TO)

2021

PATRICK ANDERSON MATIAS DE ARAÚJO

EXPLORANDO A SUPER-RESOLUÇÃO PARA RECONHECIMENTO FACIAL

Trabalho de Conclusão de Curso II apresentado à UFT – Universidade Federal do Tocantins – Câmpus Universitário de Palmas, Curso de Ciência da Computação foi avaliado para a obtenção do título de Bacharel e aprovada em sua forma final pelo Orientador e pela Banca Examinadora.

Data de aprovação: 19 / 8 / 2021

Banca Examinadora:

Profa. Dra. Anna Paula de Sousa Parente Rodrigues

Prof. Dr. Eduardo Ferreira Ribeiro

Prof. Dr. Tanilson Dias dos Santos

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal do Tocantins

- A663e Araújo, Patrick Anderson Matias de.
Explorando a Super-Resolução para Reconhecimento Facial. / Patrick Anderson Matias de Araújo. – Palmas, TO, 2021.
46 f.
- Monografia Graduação - Universidade Federal do Tocantins – Câmpus Universitário de Palmas - Curso de Ciências da Computação, 2021.
Orientador: Eduardo Ferreira Ribeiro
1. Reconhecimento Facial. 2. Deep Learning. 3. Super-Resolução. 4. Processamentos de Imagens. I. Título

CDD 004

TODOS OS DIREITOS RESERVADOS – A reprodução total ou parcial, de qualquer forma ou por qualquer meio deste documento é autorizado desde que citada a fonte. A violação dos direitos do autor (Lei nº 9.610/98) é crime estabelecido pelo artigo 184 do Código Penal.

Elaborado pelo sistema de geração automática de ficha catalográfica da UFT com os dados fornecidos pelo(a) autor(a).

*Dedico esse trabalho a Eder
Ahmad Charaf Eddine e Luciano
de Jesus Gonçalves.*

AGRADECIMENTOS

Gostaria de agradecer a todos que me ajudaram a chegar esse momento, minha mãe Luciana Matias de Barros e minha irmã Ananda Liz Matias de Araújo. Expresso aqui meu amor por todos e meu carinho especial. Agradeço também a Eduardo Ferreira Ribeiro pela paciência e empenho em me orientar nesse processo.

RESUMO

O reconhecimento biométrico faz parte de muitos aspectos da sociedade moderna. Com a popularização dos smartphones, o reconhecimento facial ganha espaço neste ambiente de tecnologias biométricas. Acompanhada da diversidade de dispositivos de captura de imagens, de diferentes marcas e qualidades, as imagens nem sempre estarão no padrão ideal para serem reconhecidas. Este trabalho testa e compara diferentes cenários e situações para avaliar os resultados obtidos pelo reconhecimento facial em diferentes ambientes. Para isso, foi utilizado o método quantitativo de análise dos dados. No primeiro cenário, todas as imagens foram enviadas sem alterações. Em seguida essas imagens têm suas resoluções reduzidas, que pode ou não ser acompanhada de ampliação para a resolução original via interpolação bicúbica ou através do algoritmo de Super-Resolução de Imagem, essas imagens podem ser todas, ou apenas que passam por testes. Os resultados indicam que o primeiro cenário obteve o melhor desempenho, seguido do cenário onde somente as imagens de teste são alteradas. O pior desempenho ocorre quando as propriedades de todas as imagens são afetadas. Nas situações em que existe uma redução e o alargamento é opcional, a opção de alargamento tem melhor desempenho, nesse quesito o alargamento através da interpolação bicúbica tem vantagem sobre o algoritmo de Super-Resolução, a situação em que apenas ocorre a redução tem o pior desempenho.

Palavra-chave: Reconhecimento Facial. *Deep Learning*. Super-Resolução. Imagens e Vídeos. Processamento de Imagens. Visão Computacional.

ABSTRACT

Biometric recognition is part of many aspects of modern society. With the popularization of smartphones, facial recognition gains space in this environment of biometric technologies. Accompanied by the diversity of image capture devices, of different brands and qualities, the images will not always be in the ideal standard to be recognized. This work tests and compares different scenarios and situations to assess the results obtained by facial recognition in different environments. For this, the quantity method of data analysis was used. In the first scenario, all images were submitted without changes. In the following, we have the reduction of image resolution, which may or may not be followed by enlargement to the original resolution via bicubic interpolation or through the Image Super-Resolution algorithm, these images can be all, or only that undergo tests. Results indicate that the first scenario obtained the best performance, followed by only the tests images change. The worst performance occurs where the properties of all images are affected. In situations where there is a reduction and enlargement are optional, the enlargement option performs better, so the bicubic enlargement has an advantage over the ISR, the situation in which only the reduction occurs has the worst performance.

Keywords: Facial Recognition. Deep Learning. Super-Resolution. Images and Videos. Image Processing. Computer Vision.

LISTA DE FIGURAS

Figura 1 – Exemplificação de uma imagem matricial e uma imagem vetorial. Figura obtida da <i>Wikimedia Commons</i> [2021] e modificada para uso nesse trabalho.	19
Figura 2 – Combinação das cores no sistema <i>RGB</i> para obtenção de outras cores e tonalidades. Figura obtida através da <i>Wikimedia Commons</i> [2020b].	20
Figura 3 – Representação do conceito de vizinhança onde um pixel em vermelho tem vizinhos representados com a cor azul.	21
Figura 4 – Exemplo de uso da Super-Resolução. Figura obtida através da <i>Wikimedia Commons</i> [2020a].	22
Figura 5 – Relacionamento entre Inteligência Artificial, Aprendizado de Máquina e <i>Deep Learning</i>	24
Figura 6 – Exemplificação de uma Rede Neural com suas entrada e sua saída (SHAH, 2018).	25
Figura 7 – Exemplos de funções de ativação.	26
Figura 8 – Ampliação através da Interpolação Bicúbica e algoritmo SRI.	27
Figura 9 – Representação do funcionamento dos algoritmos clássicos de reconhecimento de faces.	28
Figura 10 – Representação do funcionamento do sistema proposto de reconhecimento de faces.	31
Figura 11 – Detalhamento do funcionamento do sistema proposto.	33
Figura 12 – Exemplo de uma imagem a ser reconhecida e o resultado do reconhecimento.	35
Figura 13 – Representação de uma imagem que passa pelo processo de interpolação bicúbica a uma dada resolução menor.	35
Figura 14 – Ilustração de uma imagem que é reduzida e posteriormente aumentada pelo processo de interpolação bicúbica.	36

Figura 15 – Demonstração de uma imagem que é reduzida via interpolação bicúbica e imediatamente aumentada para a resolução original através do algoritmo SRI.	37
Figura 16 – Resumo dos Resultados em número de acertos do processo de reconhecimento agrupados em diferentes configurações de testes pelos dois algoritmos de super-resolução.	41

LISTA DE TABELAS

Tabela 1 – Resultados no cenário onde todas as imagens são diminuídas de forma bicúbica.	35
Tabela 2 – Resultados no cenário onde todas as imagens são diminuídas e em seguida aumentadas de forma bicúbica.	36
Tabela 3 – Resultados no cenário onde todas as imagens são diminuídas de forma bicúbica e em seguida aumentadas usando o algoritmo SRI.	38
Tabela 4 – Resultados no cenário onde somente as imagens do diretório de teste são diminuídas.	38
Tabela 5 – Resultados no cenário onde só as imagens de teste são reduzidas e em seguida aumentadas.	39
Tabela 6 – Resultados no cenário as imagens de testes são diminuídas de forma bicúbica e em seguida aumentadas usando o algoritmo SRI.	40

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Objetivos deste Trabalho	17
2	FUNDAMENTAÇÃO TEÓRICA	18
2.1	Fundamentos de Processamentos de Imagens	18
2.2	Fundamentos de Super-Resolução	21
2.3	Fundamentos de Inteligência Artificial	23
2.4	Super-Resolução via redes neurais de aprendizado	26
2.5	Reconhecimento de Faces	27
3	MATERIAIS E MÉTODOS	30
4	RESULTADOS	34
5	CONCLUSÃO	42
	REFERÊNCIAS	44

1 INTRODUÇÃO

Devido à pandemia de Sars-CoV-2, a adoção de recursos digitais se amplificou ficando mais evidente em vários aspectos da vida cotidiana. No Brasil, por exemplo, onde se chegou a 15.632.584 casos e 436.001 mortes¹ por conta da COVID-19 após 447 dias desde o primeiro caso nacional, se fazem necessárias soluções alternativas, tendo em vista que, até o dia 18 de maio de 2021 as vacinas foram injetadas em 18,54%² da população e se encontram em falta por conta da escassez de insumos importados³.

Mesmo antes desse triste evento, os caminhos já se encontravam pavimentados: sistemas que monitoram indivíduos através de câmeras, reconhecem suas faces e atividades, ultrapassam fronteiras chinesas⁴, país conhecido por um uso soberbo de tecnologias, principalmente as que utilizam recursos biométricos empregados na vigilância da população. De maneira gratuita, dispositivos móveis, redes sociais e outros serviços têm a capacidade de reconhecer seus usuários e até seus não-utilizadores.

Independente das controvérsias atuais em que o uso de tecnologias esteve presente, o seu uso adequado pode ajudar em questões pungentes, especialmente no que diz respeito a temas como a própria pandemia, um exemplo são sistemas que conseguem identificar a temperatura corporal e impedir não portadores de máscaras que não entrem em espaços públicos⁵, impedindo assim a proliferação do vírus e, como consequência, novos infectados. Essas tecnologias dependem de conceitos como *Deep Learning*, biometria e super-resolução.

Segundo Hofbauer, Debiasi e Uhl (2019), com a popularização dos smartphones, a

¹Dados do dia 18 de maio de 2021. (Monitor do coronavírus - saúde. **Folha de São Paulo**. Disponível em: <<https://arte.folha.uol.com.br/equilibrioesaude/2020/casos-mortes-coronavirus-brasil-mundo/#/local/brasil>>. Acesso em 18 mai. 2021.)

²Dados do dia 18 de maio de 2021. (Mapa da vacinação contra Covid-19 no Brasil | Vacina. **G1 - O portal de notícias da Globo (globo.com)**. Disponível em: <<https://especiais.g1.globo.com/bemestar/vacina/2021/mapa-brasil-vacina-covid/>>. Acesso em 18 mai. 2021.)

³Vacina CoronaVac é suspensa por falta de insumos importados. (Declarações da gestão Bolsonaro contra a China afetam liberação de insumos de vacinas, diz Butantan. **Folha de São Paulo**. Disponível em: <<https://www1.folha.uol.com.br/equilibrioesaude/2021/05/declaracoes-da-gestao-bolsonaro-contra-a-china-afetam-liberacao-de-insumos-diz-butantan.shtml>>. Acesso em 18 de maio de 2021.)

⁴Um uso recente de inteligência artificial para o reconhecimento facial ocorreu recentemente no Carnaval de Salvador, que na ocasião ajudou a prender um criminoso. (OH, S. Hyeon. Reconhecimento facial ajuda a prender criminoso no Carnaval de Salvador. **Canaltech**, São Bernardo do Campo, 6 de março de 2019. Disponível em: <<https://canaltech.com.br/seguranca/reconhecimento-facial-ajuda-a-prender-criminoso-no-carnaval-de-salvador-134189/>>. Acesso em: 18 mai. 2021.)

⁵Dispositivo possui sistema que consegue medir a temperatura corporal e identificar um não portador de máscara. (ZARAMELA, Luciana. COVID-19: dispositivo indica quem está com febre em empresas ou instituições. **Canaltech**, São Bernardo do Campo, 01 de junho de 2020. Disponível em: <<https://canaltech.com.br/saude/covid-19-dispositivo-indica-quem-esta-com-febre-em-empresas-ou-instituicoes/>>. Acesso em: 18 mai. 2021.)

adoção de recursos biométricos para permitir o acesso do seu usuário é observado nesses dispositivos. Complementam Uhl e Kirchgasser (2020) que o reconhecimento facial é aplicado em cenários diferentes como a vigilância, a patrulha de fronteiras e a ciência forense. Atualmente “[...] milhões de imagens já foram geradas abrindo uma gama de possibilidades para os propósitos mais diversos [...]” (Ribeiro, Uhl e Alonso-Fernandez, 2019).

Entende-se por reconhecimento facial um conjunto de técnicas/algoritmos que, através de imagens, consegue reconhecer rostos. Os dados do rosto podem ser comparados com os dados em um banco na tentativa de identificar o indivíduo com aquelas características. Reconhecimento facial realizado por máquinas são “[...] técnicas biométricas que consiste identificar padrões faciais [...]” (BISSI, 2018). Os principais algoritmos são denominados através de suas abordagens metodológicas que podem ser do tipo holístico (baseado em aparência), estrutural (baseado em características) ou híbrida (mistura do tipo holístico com estrutural). No tipo holístico, têm-se os algoritmos *Eigenfaces*, *Fisherfaces* e *TensorFaces*, no tipo estrutural, o principal algoritmo é o *Elastic Bunch Graph-Matching*.

Deep Learning ou aprendizado profundo, em português, é uma subárea da inteligência artificial. Segundo Sundararajan e Woodard (2018), essa subárea procura através de seus algoritmos imitar o modo de funcionamento do cérebro humano e sua estrutura. Uma rede neural, no contexto de *Deep Learning*, é onde a informação passa e, então, é processada, a primeira camada é a camada de entrada, enquanto a última camada é referenciada como a camada de saída. As redes neurais convolucionais (*CNN*) são uma classe contida dentro da *Deep Learning* e são muito utilizadas em vários aspectos que envolvem imagens.

A biometria, por outro lado, se preocupa em “[...] identificar pessoas usando suas características físicas ou comportamentais (SUNDARARAJAN; WOODARD, 2018)”, no sentido de detectar o portador da característica. Isso ocorre através da transformação dessas características em várias medidas estatísticas. Um exemplo pode ser o dos detectores de digitais em bancos: um aparelho que capta a leitura biométrica da mão do indivíduo e substitui a senha em operações bancárias.

Já a super-resolução “[...] refere-se ao processo de criar imagens nítidas e em alta resolução (AR) de uma imagem em baixa resolução (BR) ou de uma sequência de observações em baixa resolução” (SCHULTZ; STEVENSON, 1996). Uma imagem que passou por processos de super-resolução permite um maior detalhamento do conteúdo exibido na imagem. “O uso de *Deep Learning*, especificamente redes neurais artificiais (*RNNs*) para realizar o mapeamento entre imagens/fragmentos em BR e AR tem sido extensivamente explorado nos anos recentes” (RIBEIRO et al., 2019).

No âmbito das imagens digitais, a “[...] interpolação é, basicamente, o processo que utiliza dados conhecidos para estimar valores em pontos [...]” (DOURADO, 2014) não

conhecidos.

Nota-se que há uma simbiose entre os conceitos. Normalmente, as alterações de resolução de imagens são realizadas a partir de algoritmos específicos. No entanto, uma imagem obtida em baixa resolução pode ser transformada em uma imagem em alta resolução (super-resolução) a partir de técnicas de *Deep Learning* e submetida a um algoritmo de reconhecimento facial para detecção de características biométricas.

A super-resolução de imagens pode ser de suma importância para a área de segurança pública e nas aplicações comerciais, pois as imagens capturadas de câmeras convencionais e utilizadas nessas áreas são, segundo Xiang Xu, Wanquan Liu e Ling Li (2014), pequenas e diferem de imagens em baixa resolução reduzidas de imagens em alta resolução. Isso ocorre por diversos fatores, seja porque o objeto das imagens capturada se encontra distante, o ambiente não corrobora no quesito de iluminação, imagens que precisam sofrer reduções em suas especificidades para serem transmitidas em tempo real ou o tipo da câmera.

Existem atualmente poucas pesquisas sobre super-resolução, com isso, este projeto se propõe ao aprofundamento dos estudos de algoritmos de *Deep Learning* bem como algoritmos de super-resolução testando a integração deles para aumentar ainda mais a eficácia deles em imagens e vídeos de baixa resolução. Projeto este que pode trazer novas compreensões sobre o assunto tratado e futuramente novas aplicações em várias áreas, principalmente no que diz respeito a segurança pública.

Para o escopo dessa monografia, o foco será na testagem de diferentes cenários e situações envolvendo imagens, suas especificações e seu conteúdo formado pelos pixels, para observar e escolher uma solução que melhor se adequa aos critérios do reconhecimento facial eficaz e de melhor acurácia, o que faz o algoritmo de reconhecimento de faces identificar e reconhecer corretamente os rostos.

Em um levantamento inicial, os autores Bilgazyev et al. (2011) relatam que imagens em baixa resolução tornam o reconhecimento de faces desafiador. No entanto, os pesquisadores propuseram um método para aumentar a resolução de imagens que se diferencia de outros que buscam investigar a relação direta entre imagens de alta resolução e baixa resolução. O método dos pesquisadores estima os componentes de alta frequência que não estão presentes em outros métodos.

A pesquisa de Arashloo e Kittler (2014) discute a questão do reconhecimento invariante da postura de faces usando um *MRF*. Sigla que vem do inglês *Markov Random Field* (campo aleatório de Markov), e é conhecido por sua complexidade computacional. Para isso, foi usado o descritor *daisy* para representação de imagem facial no reconhecimento de imagem e criado uma implementação de uma unidade de processamento gráfico do processo de correspondência em multiresolução *MRF*. A eficiência tornou a abordagem *MRF* viável e facilita otimizações empíricas extensivas e estudos de avaliações.

Gunturk et al. (2003) propuseram transferir a reconstrução em super-resolução

do domínio de pixel para um espaço de face com dimensão menor. Tal abordagem tem como vantagem diminuir a complexidade computacional no que diz respeito à reconstrução em super-resolução de imagens, uma vez que não é o foco do algoritmo uma imagem visualmente melhorada e sim buscar as informações necessárias para o sistema de reconhecimento.

Prestando atenção no impacto da perda de camada de redes neurais artificiais no contexto de processamento de imagem, os autores Zhao et al. (2017) mostram outras alternativas para restauração de imagem e, de maneira específica, a importância de perdas motivadas perceptivamente quando a imagem resultante é avaliada por um observador humano. O desempenho de diferentes tipos de perdas são comparados, e uma nova função de erro diferenciável é proposta. Além disso, os autores mostram que a qualidade das imagens melhora com funções de perdas mais elaboradas.

Por fim, a pesquisa de Davis et al. (2011) descreve um sistema para reconhecimento e detecção de rostos humanos a partir de câmeras de segurança em tempo real. O sistema usa o detector de rosto *Viola-Jones* e, em seguida, extrai características locais e cria um vetor características em base de formas. Os autores levaram em consideração melhorar a performance e precisão das fases de reconhecimento e detecção.

1.1 Objetivos deste Trabalho

Neste trabalho, o foco será nos estudos dos algoritmos clássicos de super-resolução, para observar e escolher uma solução que melhor se adequa aos critérios do reconhecimento facial eficaz e de melhor acurácia, o que faz o algoritmo de reconhecimento de faces identificar e reconhecer corretamente os rostos. A partir daí o objetivo é comparar esses algoritmos com a testagem de algoritmos de super-resolução que utilizam redes neurais de aprendizado profundo (*Deep Learning*) e verificar se o custo computacional das Redes Neurais Convolucionais compensam os melhores resultados no reconhecimento final das faces.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Fundamentos de Processamentos de Imagens

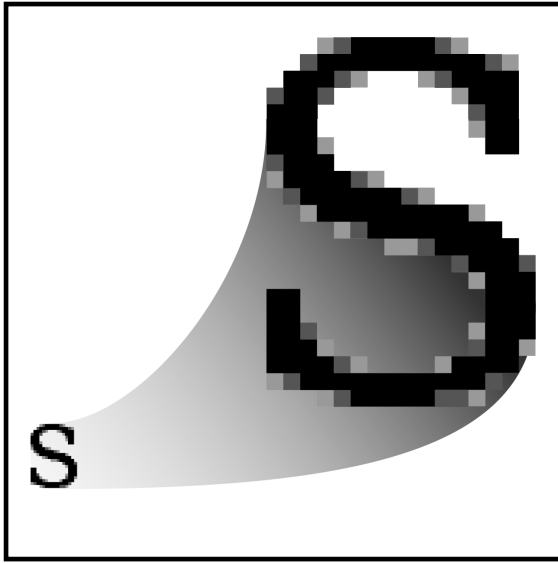
O aprimoramento dos aspectos visuais de informações de imagens para interpretação humana consiste em um dos principais interesses do processamento de imagens. Além disso, atualmente, com a popularização de dispositivos de captura de imagens, o processamento de imagem trabalha questões como “armazenamento, transmissão e representação, considerando a percepção automática por máquinas” (GONZALEZ; WOODS, 2009). Ou seja, o Processamento de Imagens consiste em processar dados em que tanto a entrada quanto a saída são imagens.

A visão humana transforma padrões luminosos detectados do ambiente externo em imagens, sendo um dos nossos mais avançados sentidos. O olho humano, órgão responsável pela função de visão, fica situado na órbita ocular dos crânios da maioria dos seres vivos. Essa esfera é formada por três camadas de tecidos, sendo a mais importante, a retina, responsável pela captação da luz e transmissão do estímulo elétrico responsável pela visão. A região fóvea se encontra no centro da retina, é nessa região onde encontra-se os cones e bastonetes, elementos chaves para a captação de luz. No caso humano há uma limitação à banda visual do espectro eletromagnético.

No caso das máquinas responsáveis pelo processamento digital de imagens, é possível cobrir quase todo espectro eletromagnético.

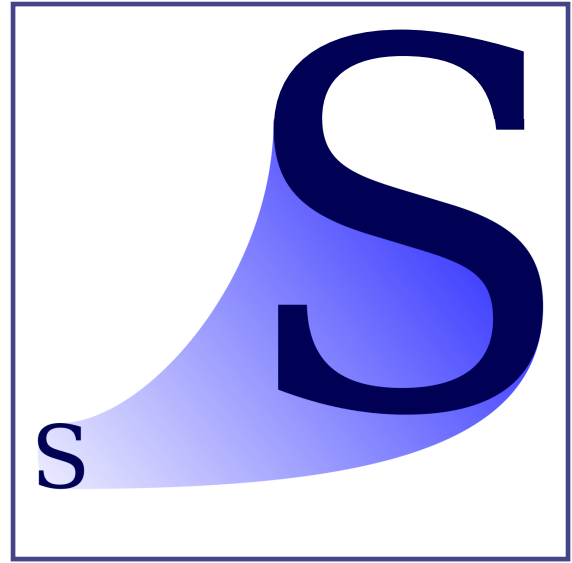
Nesses dispositivos, as imagens digitais são representadas de forma matricial ou vetorial. As imagens vetoriais consistem em gráficos que são construídos através de fórmulas matemáticas descrevendo, assim, o objeto formado, como linhas, formas, curvas e textos.

As imagens matriciais, por outro lado, também são conhecidas como *bitmaps* (mapa de bits) ou *raster*, formadas através de linhas e colunas, onde cada célula em uma linha e coluna específicas descrevem um pixel. Na Figura 1 pode ver uma comparação gráfica entre imagens matriciais (*raster*) e imagens em vetores.



Raster

GIF, JPEG, PNG



Vetor / Vector

SVG

Figura 1 – Exemplificação de uma imagem matricial e uma imagem vetorial. Figura obtida da *Wikimedia Commons* [2021] e modificada para uso nesse trabalho.

A palavra pixel se origina da junção dos termos *picture* (imagem) e *element* (elemento), isto é, elemento de imagem em português. Conforme amplia-se uma imagem matricial, ou chega-se mais próximo de um monitor de computador, pode-se ver uma composição de vários quadrados pequenos coloridos que juntos formam uma imagem completa. Cada pixel representa uma cor, essa cor é formada de três cores simples, o vermelho (*red*), verde (*green*) e azul (*blue*), que formam o sistema de cores aditivas *RGB*. Essas cores são computacionalmente representadas, cada, de 256 valores, variando do zero até o 255, os valores variam entre as diversas tonalidades, que, agrupadas juntamente, podem gerar em 16 milhões em possibilidades de cores. Abaixo na Figura 2 pode ter uma noção de como se obter diferentes cores e tonalidades através de combinações de cores no sistema *RGB*.

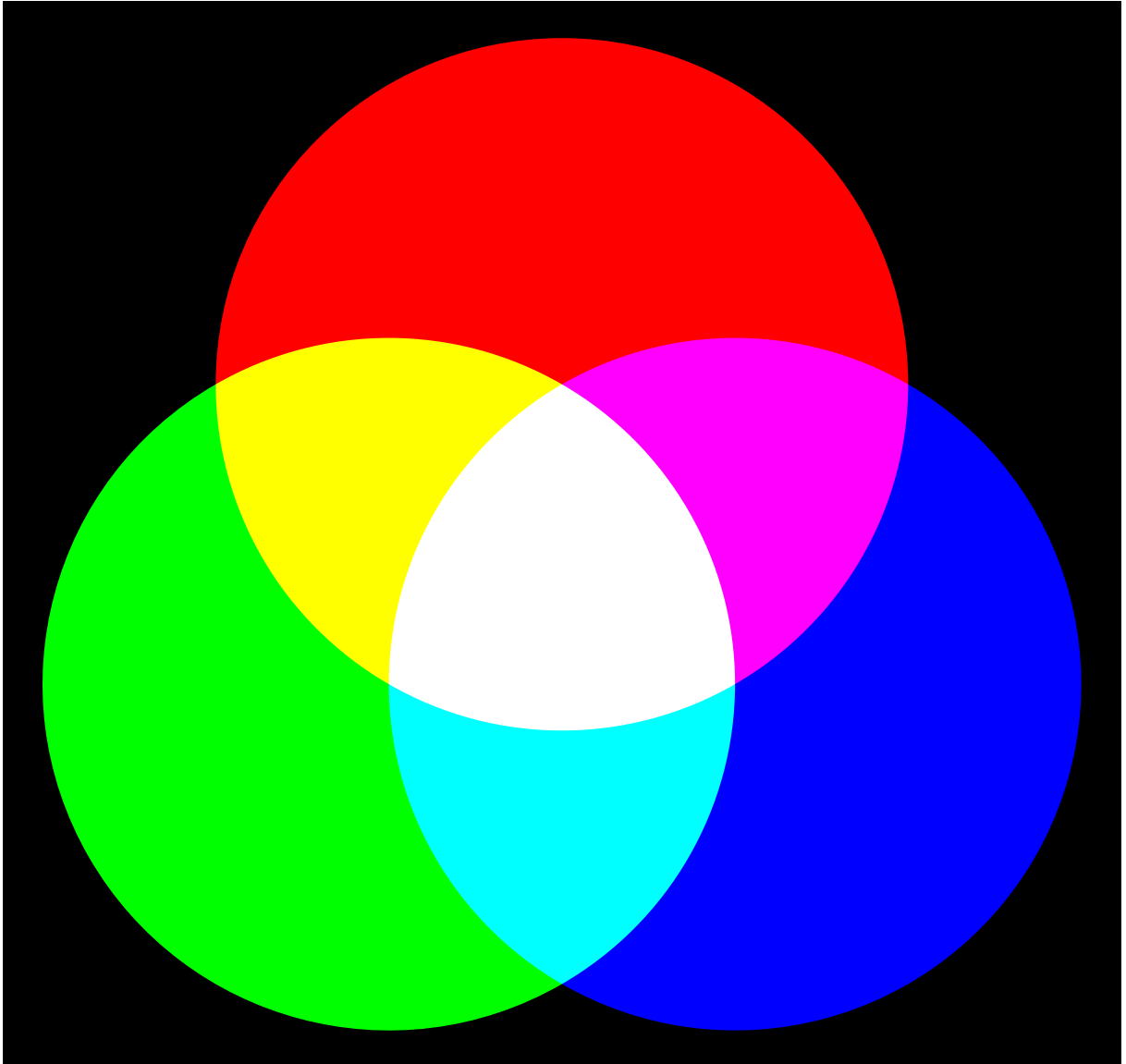


Figura 2 – Combinação das cores no sistema *RGB* para obtenção de outras cores e tonalidades. Figura obtida através da *Wikimedia Commons* [2020b].

Uma imagem de 640x480 pixels, possui 640 colunas e 480 linhas de composição, ou seja, existem 307.200 pixels ao longo da imagem, por exemplo.

Quanto maior a quantidade de pixels em uma imagem, mais informações são necessárias serem armazenadas, conseqüentemente a imagem ficará com melhor qualidade e mais definida, ao mesmo tempo em que o dispositivo para exibir ou gerar esse conteúdo precisará de um maior processamento e configurações mais robustas.

Dado um pixel de uma imagem, a vizinhança consiste em todos os pixels que circundam o pixel em questão. A Figura 3 abaixo exemplifica o conceito de vizinhança, onde o pixel vermelho contém pixels vizinhos azuis e os pixels restantes representando os pixels não vizinhos.

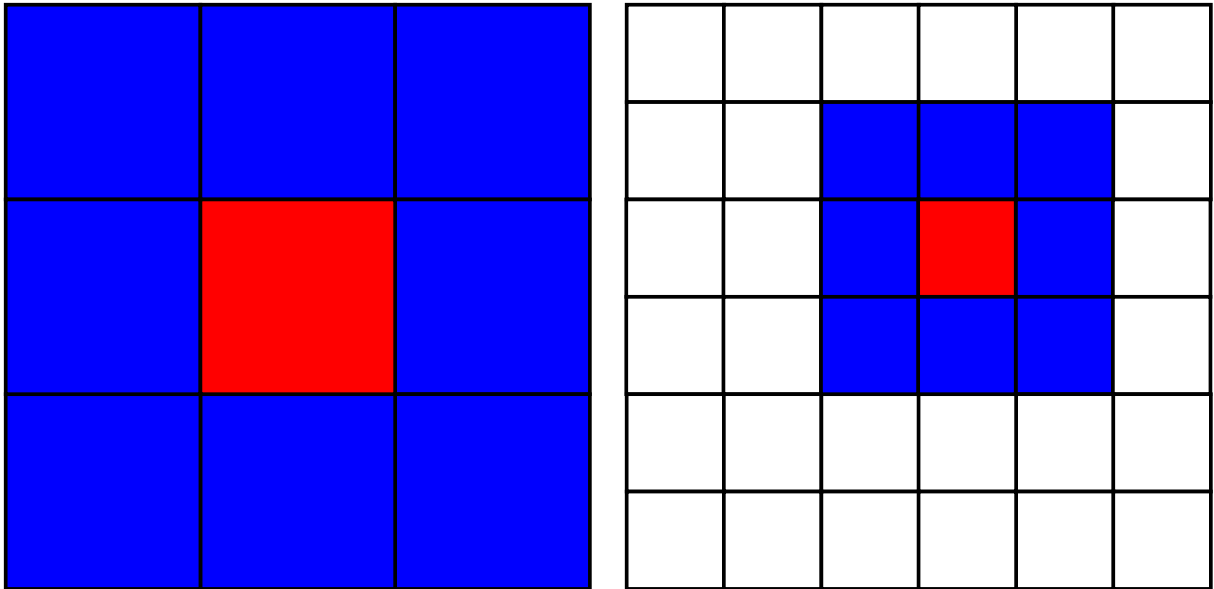


Figura 3 – Representação do conceito de vizinhança onde um pixel em vermelho tem vizinhos representados com a cor azul.

Já o processo de reamostragem de pixels é utilizado quando “[...] coordenadas da imagem processada (linha e coluna) não coincidem com aquelas da imagem original” (SPRING, 2006) alterando a quantidade de pixels originalmente, assim diminuindo ou aumentando os pixels da imagem.

Métodos de interpolação são utilizados na reamostragem sendo “efetuada por interpolação híbrida, onde a aplicação do interpolador de alocação de vizinho mais próximo e do interpolador bilinear dependem da característica local dos níveis de cinza na imagem” (SPRING, 2006) utilizando assim dados conhecidos.

2.2 Fundamentos de Super-Resolução

A Super-Resolução é uma das mais recentes áreas de pesquisa em processamento de imagens. Pode ser definida como um conjunto de técnicas que procuram aprimorar e ampliar imagens em suas características, como resolução, definição e qualidade, a partir de uma imagem com características inferiores.

O uso da Super-Resolução (SR) é de vital importância para diversas áreas, como a geografia, geologia, medicina, agropecuária, astronomia, biologia e militar. Um exemplo do uso da Super-Resolução está representado na Figura 4, aplicada na astronomia o exemplo abaixo demonstra como uma imagem foi aperfeiçoada através da técnica.

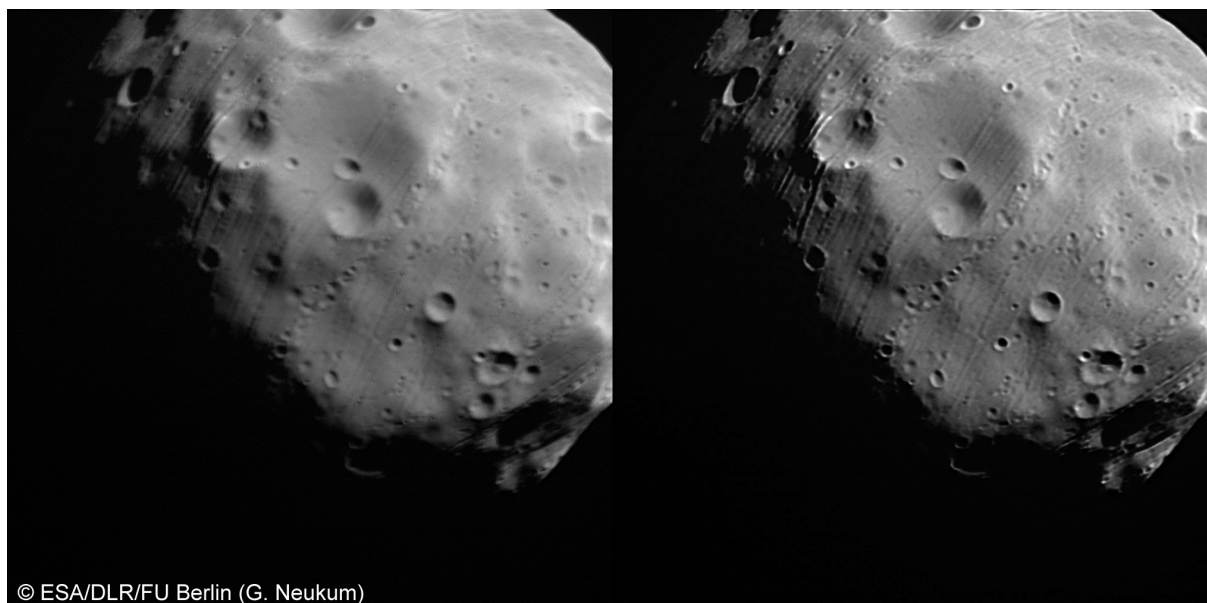


Figura 4 – Exemplo de uso da Super-Resolução. Figura obtida através da *Wikimedia Commons* [2020a].

A qualidade das câmeras digitais melhorou consideravelmente nos últimos anos com o desenvolvimento de sensores melhores impulsionado também pelo avanço das fotografias em dispositivos móveis como *smartphones*. No entanto, apesar desse desenvolvimento, há diversos fatores que tornam seu uso impeditivo, imagens com mais informações implicam maior custo para armazenamento e transmissão desses dados, além disso há o custo monetário em adquirir equipamentos melhores.

As técnicas de interpolação constituem como um dos componentes da Super-Resolução. Basicamente, com a interpolação, cria-se novos pixels a partir de pixels existentes. Apesar da Super-Resolução se utilizar da interpolação as “[...] técnicas de interpolação (vizinho mais próximo, bilinear e convolução cúbica) diferem da SR porque nas primeiras, somente uma imagem é utilizada como fonte de informação para que uma imagem de maior resolução seja gerada, diferente do que é utilizado para produzir uma imagem utilizando SR” (JR; ROSA, 2007).

Com o interpolador de alocação de vizinho mais próximo, é atribuído a mesma quantidade da tonalidade de cinza do pixel da imagem que vai passar por correções a mesma quantidade de cinza do pixel adjacente localizado na mesma posição. A quantidade de nível de cinza permanece a mesma. Nas regiões das imagens onde os valores de nível de cinza são diferentes, seu uso se mostra ideal.

O interpolador bilinear usa os valores dos quatro pixels vizinhos para determinar o valor de nível de cinza de determinado pixel de uma imagem a ser corrigida. São usados os pixels localizados a leste, oeste, norte e sul de um dado pixel, assim, o pixel tem suas características determinadas levando em consideração seus pixels vizinhos. Diferentemente do interpolador de alocação vizinho mais próximo, o uso desse interpolador é aplicado em

regiões da imagem onde os níveis de cinza dos pixels não são diferentes.

Já a interpolação bicúbica se utiliza de todos os 16 pixels (4x4 pixels vizinhos) mais próximos de um pixel para realizar sua estimativa. Uma vez que eles estão em vários intervalos do pixel não especificado, os pixels próximos recebem uma ponderação maior no cálculo. Os algoritmos de redimensionamento que utilizam a lógica de interpolação bicúbica realizam ampliação ou redução das imagens. Quando comparada à interpolação Bilinear e Vizinho Mais Próximo, a interpolação Bicúbica é mais eficiente e produz imagens mais nítidas, por isso, muitos programas de manipulação e edição de imagens implementam a interpolação Bicúbica.

2.3 Fundamentos de Inteligência Artificial

Inteligência Artificial é uma “[...] área de pesquisa que investiga formas de habilitar o computador a realizar tarefas nas quais, até o momento, o ser humano tem um melhor desempenho” (RICH; KNIGHT; CALERO, 1994).

Dentro da Inteligência Artificial (IA) tem-se uma subárea chamada de *Machine Learning*, ou Aprendizado de Máquina em português, onde os computadores adquirem a capacidade de compreender de acordo com resultados esperados por meio de correlações de diferentes tipos de dados.

O *KNN* (*K-Nearest Neighbors*) é um dos mais famosos e simples algoritmos utilizados quando o assunto é aprendizado de máquina. O algoritmo de reconhecimento de faces utilizado se aproveita do *KNN* para realizar o seu trabalho. O *KNN* é usado para classificar objetos baseado em um conjunto de dados de treinamento que mais se aproximam em um espaço de características. É preciso de dados de exemplos para a treinamento, definir uma medida para estimar a distância entre os dados de treinamento e determinar o valor k , que é a quantidade de vizinhos mais próximos que serão levados em consideração pelo algoritmo.

Principal órgão do sistema nervoso central, o cérebro humano é responsável pela ação e conduta do corpo humano, raciocínio e pensamento. São extremamente complexos, estima-se que há 86 bilhões de neurônios¹.

Como mencionado na introdução deste trabalho, *Deep Learning* é uma subárea da Inteligência Artificial, contida no âmbito do Aprendizado de Máquina. Além disso, faz o treino de sistemas computacionais para realizar tarefas humanas, como identificar padrões, imagens, reconhecer fala e fazer previsões. O foco é configurar medida simples de dados e treinar computadores para serem capazes de aprenderem sem auxílio externo

¹Imaginava-se que a quantidade de neurônios que o cérebro humano possuía giravam em torno de 10 bilhões, mas essa estimativa estava errada. (D’ANGELO, Helô. Cérebro humano tem 16 tipos diferentes de neurônios, revela estudo. **Super Interessante**. São Paulo, 27 de junho de 2016. Disponível em: <<https://super.abril.com.br/saude/cerebro-humano-tem-16-tipos-diferentes-de-neuronios-revela-estudo/>>. Acesso em 05 ago. 2021.)

a partir do reconhecimento de padrões em vários níveis de processamento, em vez de agrupar dados para serem executados a partir de equações definidas.

Na Figura 5 temos uma representação da relação entre Inteligência Artificial, Aprendizado de Máquina e *Deep Learning*.

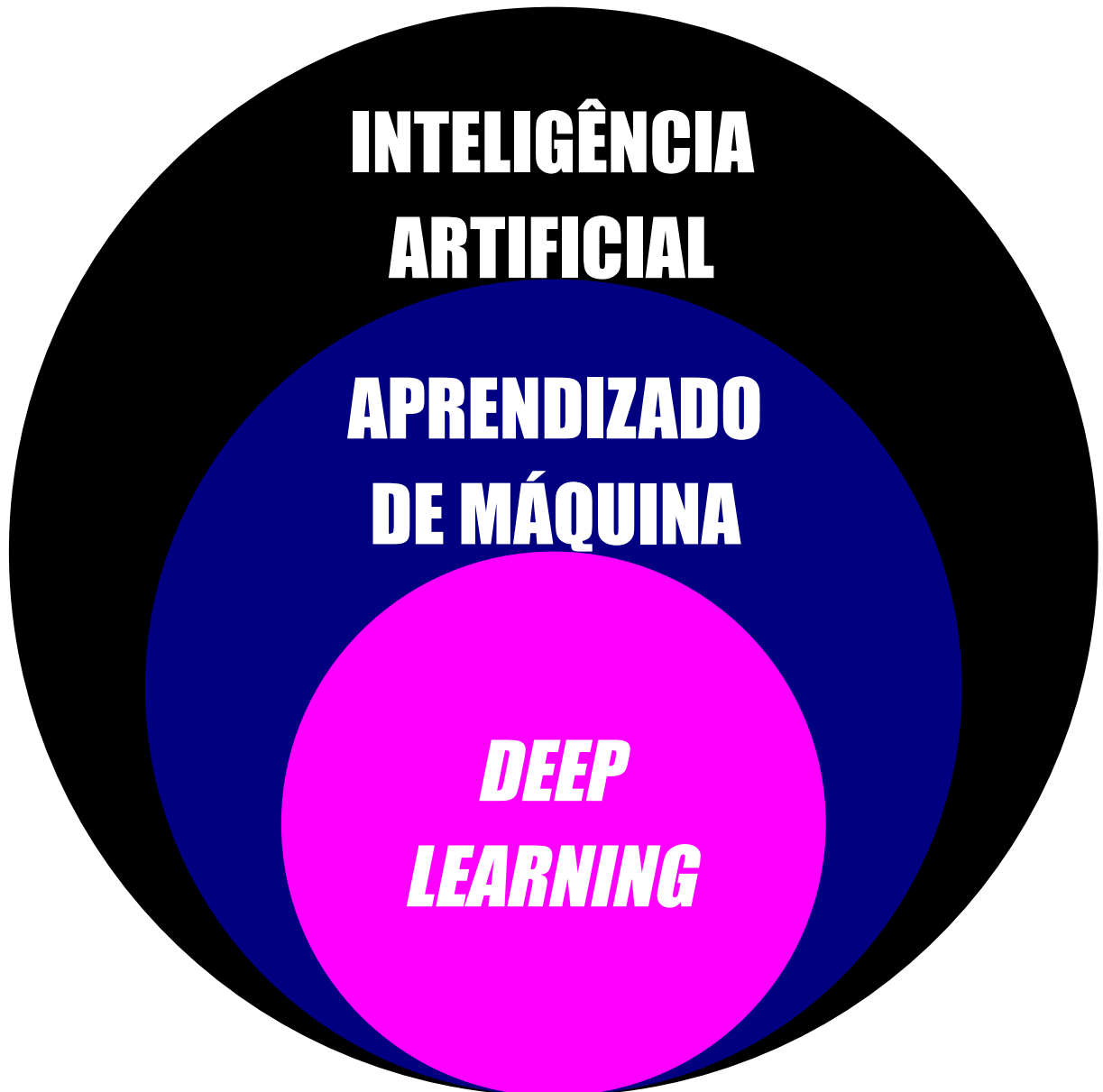


Figura 5 – Relacionamento entre Inteligência Artificial, Aprendizado de Máquina e *Deep Learning*

Redes Neurais é um ramo de *Machine Learning* e são modelos computacionais com nós inter-relacionados inspirados nos neurônios do cérebro. São utilizadas para desenvolver sistemas de inteligência artificial. Através de algoritmos elas são capazes de realizar aprendizado de máquina, reconhecer padrões.

As características das Redes Neurais, inclui-se sua capacidade de grande paralelismo, o que permite maior eficácia de processamento, seu poder de realizar o armazena-

mento associativo de informações, representação de domínios complexos e não lineares, generalizar informações insuficientes além de serem modelos ajustáveis e treináveis. Como entradas, as redes neurais funcionam similarmente aos estímulos e impulsos enviados aos neurônios. No caso entra uma série de variáveis de entrada em cada neurônio, após cálculos serem realizados é retornado uma saída, como pode ser notado na Figura 6.

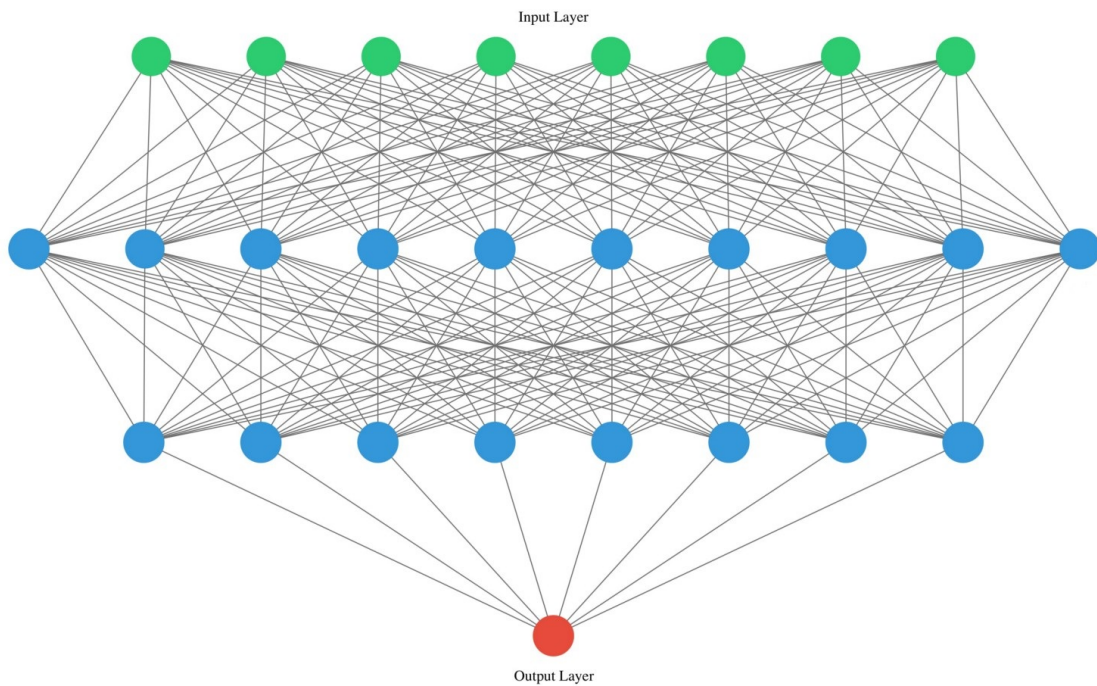


Figura 6 – Exemplicação de uma Rede Neural com suas entrada e sua saída (SHAH, 2018).

Além disso, pode-se aplicar uma função, nomeada de função de ativação aos valores de entrada, o que deriva no nível de ativação do neurônio, que pode ser chamado também de saída do neurônio. Entre essas funções possíveis, tem-se, entre as mais comuns, a *Hard Limiter*, *Threshold Logic* e a *Sigmoid* que estão representadas graficamente na Figura 7.

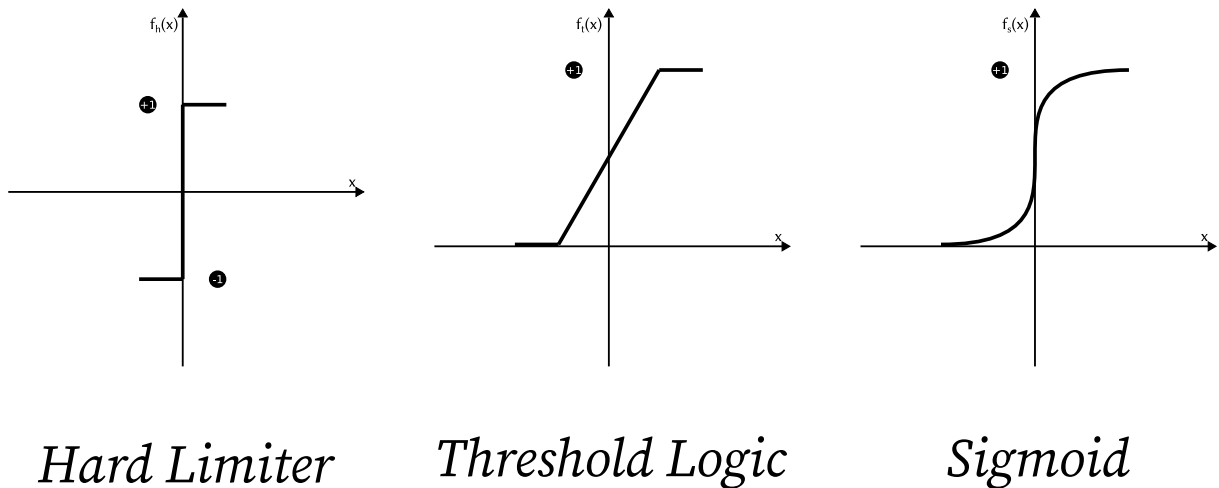


Figura 7 – Exemplos de funções de ativação.

Residual Dense Network (*RND*), ou Rede Densa Residual em português, extrai grandes recursos locais por meio de camadas convolucionais densamente conectadas. *RND* é composto por quatro elementos, *shallow feature extraction net* (*SFENet*), *residual dense blocks* (*RDBs*), *dense feature fusion* (*DFE*) e a *up-sampling net* (*UPNet*). O algoritmo de super-resolução tratado nessa monografia implementa a *RND*.

Por fim, ainda no tópico de *Deep Learning*, há as Redes Neurais Convolucionais ou *Convolutional Neural Network* (*CNN*). O foco das *CNNs* reside em questões que envolvem imagens, principalmente a análise visual. Quando comparado com outros algoritmos de classificação de imagens, as *CNNs* usam pouco pré-processamento, o que significa que a rede consegue assimilar e otimizar os filtros a partir do aprendizado automatizado. Em outros algoritmos, não há essa otimização, então o fato de não precisar de conhecimento prévio e nem auxílio humano na extração dos recursos torna a *CNN* vantajosa.

2.4 Super-Resolução via redes neurais de aprendizado

Recentemente com os avanços da Inteligência Artificial e seus subcampos como o Aprendizado de Máquina, a área de *Deep Learning* vem presenciando as conquistas das Redes Neurais Convolucionais Profundas (*CNNs*). De modo específico, quando o assunto é Processamento de Imagens (PI) as *CNNs* vem acompanhadas das Redes Generativas Adversárias (*GANs*). O processo de reconstruir uma imagem em resolução alta a partir de uma imagem em resolução menor se constitui um dos desafios atuais da PI sendo objeto de estudo e tratado nesse estudo.

Um dos objetivos desse trabalho é a geração de imagens a partir de um algoritmo de super-resolução. A partir de imagens que tiveram suas qualidades e especificações inferiorizadas, o algoritmo de super-resolução, a saber *Image Super-Resolution* (*ISR*) ou Super-Resolução de Imagens (*SRI*), surge para reconstruir essas imagens em alta resolução. Em seguida, essas imagens passarão pelo algoritmo de reconhecimento de faces

para averiguar se o uso da super-resolução altera a eficácia do reconhecimento facial. Na Figura 8 pode ver uma comparação da interpolação bicúbica e o algoritmo SRI quando ampliam uma dada imagem diminuída.



Figura 8 – Ampliação através da Interpolação Bicúbica e algoritmo SRI.

O algoritmo SRI se utiliza da Interface de Programação de Aplicações ou em inglês *Application Programming Interface (API) Keras*.

A *API Keras* oferece uma interface para redes neurais artificiais na linguagem de programação *Python*. Criada para permitir experimentação rápida com redes de *Deep Learning*, ficou conhecida por ser extensível, modular e amigável ao usuário. A sua atuação também abrange e oferece uma interface para a biblioteca *TensorFlow*.

Diferentes implementações de Redes Densas Residuais para super-resolução de imagem única (SISR) são implementadas a partir da *API Keras*. As Redes Densas Residuais já foram abordadas em tópicos anteriores.

As redes utilizadas pelo algoritmo SRI incluem a Rede Densa Residual em superescala descrita em *Residual Dense Network for Image Super-Resolution* (ZHANG et al., 2018), o Superescalonamento Residual na Rede Densa Residual descrito em *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks* (WANG et al., 2018), uma versão de múltiplas saídas da rede *Keras VGG19* para extração de recursos profundos usados na perda de percepção e uma rede discriminadora personalizada com base na descrita em *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network* (LEDIG et al., 2016).

2.5 Reconhecimento de Faces

A biometria explora as características intrínsecas do ser humano para sua identificação. Basicamente, existem três tipos de modalidades em biometria: média (“*soft*”), alta (“*hard*”) e escondida (“*hidden*”). As técnicas biométricas de fiabilidade alta são as mais clássicas como por exemplo as faces, impressões digitais ou assinaturas. As de fiabilidade média são aquelas características mais usuais relacionadas às faces como por exemplo à cor

da pele, tamanho e cor dos cabelos, ou às relacionadas ao corpo que podem ser facilmente modificadas como por exemplo, peso, altura e acessórios. As características escondidas também chamadas de características intrínsecas são aquelas relacionadas à dados médicos como por exemplos imagens radiográficas, sinais de radiofrequência (MRI) dentre outros (GONZALEZ-SOSA et al., 2018).

Vários são os algoritmos utilizados para extrair características das faces para seu possível reconhecimento, tais algoritmos identificam e recolhem medidas de pontos importantes como as bordas da boca, dos olhos, narinas, extraindo as medidas dos mesmos bem como as relações entre essas áreas. Essas medidas são então armazenadas em um vetor que será armazenado em um banco de dados como se fosse a assinatura do indivíduo para uma futura comparação. Quando todo o banco de dados estiver pronto, ao se apresentar uma nova face extraída para o sistema, o mesmo irá fazer a comparação (“*matching*”) das características extraídas com aquelas do banco de dados e, existindo uma correspondência, irá calcular a probabilidade de semelhança entre a consulta e a sua correspondência (SAVVIDES; KUMAR; KHOSLA, 2004).

Os algoritmos clássicos funcionam de modo que as imagens capturadas passam por um processo de detecção de faces. Por fim, são comparados os dados da detecção de faces com os dados em um banco de dados na tentativa de identificar os sujeitos da imagem em questão como observado na Figura 9.

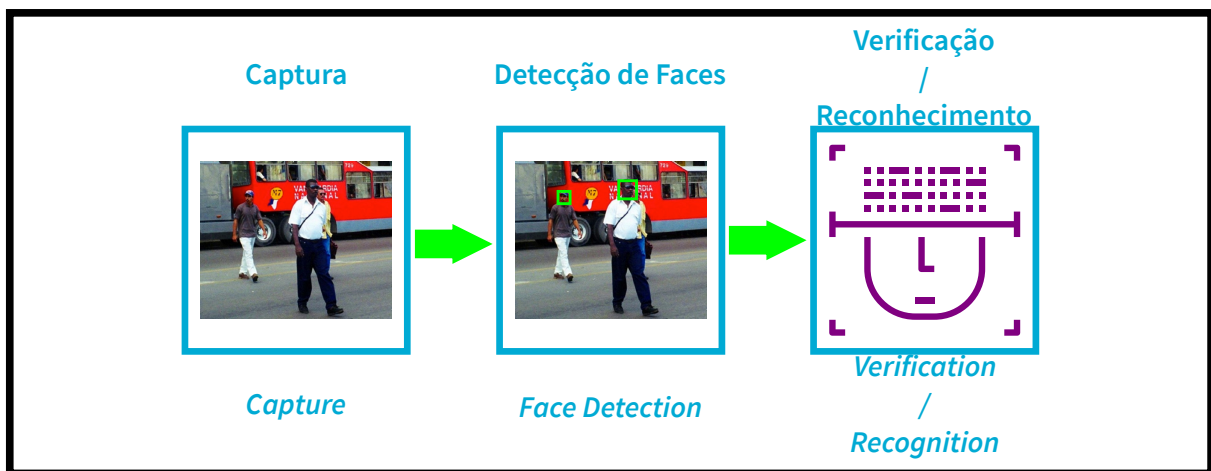


Figura 9 – Representação do funcionamento dos algoritmos clássicos de reconhecimento de faces.

O algoritmo de reconhecimento facial utilizado, *Face Recognition*, construído em *Python*, reconhece e manipula faces que aparecem em imagens com o apoio de *Deep Learning*. Especificamente falando da manipulação, é possível obter a localização e delinear partes de rostos exibidos em imagens, como o olho, nariz, boca e queixo. Também é possível utilizar o algoritmo para reconhecimento em tempo real.

Para o correto uso das funcionalidades do algoritmo é necessário prover uma pasta com, pelo menos, uma imagem de pessoas para o sistema ter uma noção de quem se

trata para o processo de treino, os nomes dos arquivos das imagens devem fornecer uma identificação de quem está na imagem. Em seguida é necessária uma segunda pasta com as imagens de pessoas que se deseja identificar.

Após esse processo é preciso invocar o procedimento *train*, onde através das imagens em pasta dedicada o sistema será treinado para identificar os rostos, por meio do classificador *KNN*.

Em seguida o classificador *KNN* será usado acompanhado das imagens a serem reconhecidas no procedimento *predict*. Nessa função é onde ocorre o reconhecimento em si, também pode ser definido o limiar de distância para classificação de faces, quanto maior o limiar, maior a possibilidade de ocorrência de erros na identificação.

No fim, são representados visualmente os resultados da identificação a partir da função *show_prediction_labels_on_image*, onde o rosto será delimitado através de um retângulo azul com uma faixa identificado o rosto textualmente.

3 MATERIAIS E MÉTODOS

Como dito no início deste trabalho, o objetivo desse estudo é verificar se o uso da super-resolução em imagens de baixa qualidade pode melhorar a acurácia de um algoritmo de reconhecimento facial. Nesse trabalho, foi utilizado como ferramenta de pesquisa o método quantitativo de análise de dados com suporte do algoritmo de reconhecimento *Face Recognition*, esse algoritmo utiliza a linguagem de programação *Python* e a biblioteca *dlib*. A escolha de uma solução que se adequa aos critérios do reconhecimento facial eficaz e de melhor acerto fez com que esse algoritmo se tornasse preferência no escopo desse estudo, na própria documentação do algoritmo é mencionado que o modelo possui 99,38% de precisão¹ para a identificação correta de rostos.

Para que os testes sejam realizados, um banco de dados bastante utilizado na literatura faz-se necessário para o funcionamento ideal do algoritmo, visto que são necessárias imagens contendo faces.

O banco de dados de rostos chamado *Georgia Tech face database* foi usado para o treino e testagem, as 750 imagens contidas no banco são de 50 pessoas diferentes, cada pessoa com 15 imagens de resolução (640x480 pixels), totalizando 128MB de imagens. Essas imagens foram obtidas em sessões de fotográficas entre os dias 01/06/1999 e 15/11/1999, nas dependências do *Center for Signal and Image Processing* no *Georgia Institute of Technology*. “As imagens mostram rostos frontais e/ou inclinados com diferentes expressões faciais, condições de iluminação e escala” (NEFIAN, 1999).

Essas imagens foram utilizadas mudando suas propriedades, alteradas suas resoluções, e, em seguida só reduzidas ou reduzidas e aumentadas, ou de forma bicúbica, ou utilizando um algoritmo de super-resolução, para comparar diferentes cenários e avaliar de forma quantitativa essas situações diversas. As imagens sofrem então redução apenas ou redução e ampliação, a redução ocorre de forma bicúbica, a ampliação acontece ou de forma bicúbica, ou usando um algoritmo de super-resolução.

O sistema proposto, representado na Figura 10, adiciona uma etapa ao processo de reconhecimento exemplificado na Fundamentação Teórica, que é a aplicação do processo de super-resolução nas imagens antes de seu reconhecimento na tentativa de aperfeiçoar a qualidade da imagem e, conseqüentemente, sofisticar a acurácia do sistema de reconhecimento.

¹Documentação do algoritmo *Face Recognition*. (GEITGEY, Adam. **Face Recognition**: ageitgey/face_recognition. 2017. Disponível em: <https://github.com/ageitgey/face_recognition>. Acesso em 1 fev. 2021.)

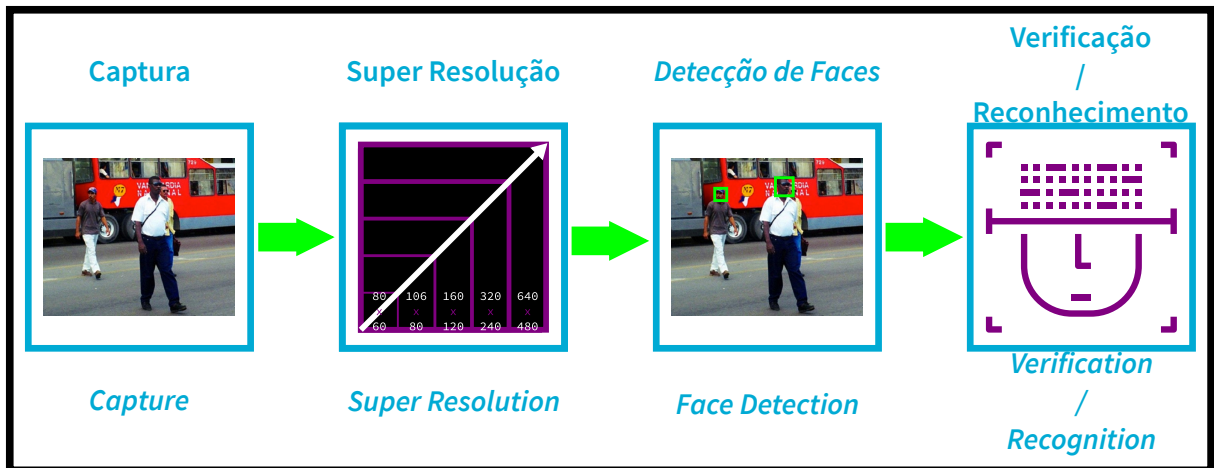


Figura 10 – Representação do funcionamento do sistema proposto de reconhecimento de faces.

Para que os testes possam ser realizados de maneira controlada, o algoritmo escolhido precisa, como entrada, de imagens que treinarão o sistema de reconhecimento de faces, além de imagens de testes que serão identificadas pelo sistema. No âmbito do treinamento das imagens, elas se encontram em uma pasta exclusiva, onde há subpastas dedicadas, cada, para um único indivíduo. Já no que se refere às imagens de testes, elas também se encontram em uma pasta exclusiva, não contendo subpastas.

Com essa divisão feita, o algoritmo inicia o processo de treinamento e, quando finalizada essa parte, começa a detecção de faces e reconhecimentos das faces identificadas.

Inicialmente, as imagens foram treinadas e testadas as imagens de acordo com as especificações originais, sendo mantida suas resoluções.

Em seguida, foram geradas através do sistema novos diretórios, onde cada destino contém as mesmas imagens do banco de dados de faces, sendo as imagens contidas em cada ficheiro de menor resolução. Cada pasta representa uma resolução menor que a resolução original, consequência da divisão da resolução original, por 2, 4, 6 e 8, resultando nas resoluções 320x240, 160x120, 106x80 e 80x60, respectivamente. Isso significa que, tanto as imagens que estão no diretório de testagem, quanto as imagens que estão no diretório de treinamento, terão as mesmas resoluções. Mas, é possível que somente as imagens que se encontram no diretório de testagem sofram esse processo de diminuição de resolução, sendo mantida a resolução e as propriedades originais nas imagens do diretório de treinamento.

As imagens que sofrem o processo de redução de sua resolução podem voltar para a sua resolução original, através da interpolação bicúbica ou do algoritmo *Image Super-Resolution (ISR)*.

O processo de interpolação bicúbica é bem conhecido por ser tradicionalmente implementado nas aplicações de manipulação de imagens, onde a imagem é suprimida de detalhes tornando-a borrada dependendo do quão a imagem é reduzida.

Já as imagens submetidas através do algoritmo SRI (*ISR*) passam por um proces-

samento mais elaborado para recuperação de detalhes em imagens em baixa resolução. O projeto SRI “[...] contém uma implementação *Keras* de diferentes Redes Residuais Densas (*Residual Dense Networks*) para Super-Resolução de Imagem Única bem como *scripts* para treinar essas redes usando conteúdo e componentes de perda adversária” (Francesco Cardinale et al., 2018, *on-line*). A escolha dessa solução se baseia em sua disponibilidade, tendo em vista o quão custoso são as outras soluções o que implica no ajuste das configurações ideais para a execução da solução, instalação de bibliotecas etc. No âmbito da super-resolução, temos o que são denominados como modelos, diferentes implementações para se obter a super-resolução, nesse trabalho são utilizados os modelos *GANs*, *PSNR* e *Noise-Cancel*.

Essas mudanças nas propriedades da imagem correspondem aos diferentes cenários para monitorar o comportamento do algoritmo de reconhecimento de faces. Em cenários reais, a imagem capturada pode não ser a melhor ou há um indivíduo que se encontra a uma distância considerável mesmo em imagens com resoluções e propriedades consideráveis. Essas técnicas contribuem para a identificação tornando os resultados mais significativos. Na Figura 11 abaixo, é possível observar com mais detalhes o funcionamento do sistema proposto.

A Figura 11 ilustra os testes divididos pelo tamanho das imagens de entradas do sistema, para testar os níveis de reconstrução das imagens, quanto maior o fator de redução (2, 4, 6 e 8) menor a qualidade das imagens. Após reduzidas, as imagens podem ser testadas de três formas: em seu tamanho original, reamostradas por interpolação bicúbica e reamostradas utilizando a Rede Neural Convolutiva. Além disso todos os testes descritos acima foram repetidos utilizando as imagens de treino do sistema de reconhecimento sem serem redimensionadas, simulando um ambiente em que a captura das imagens pode ser controlada, porém o reconhecimento pode ser realizado utilizando câmeras de baixa resolução.

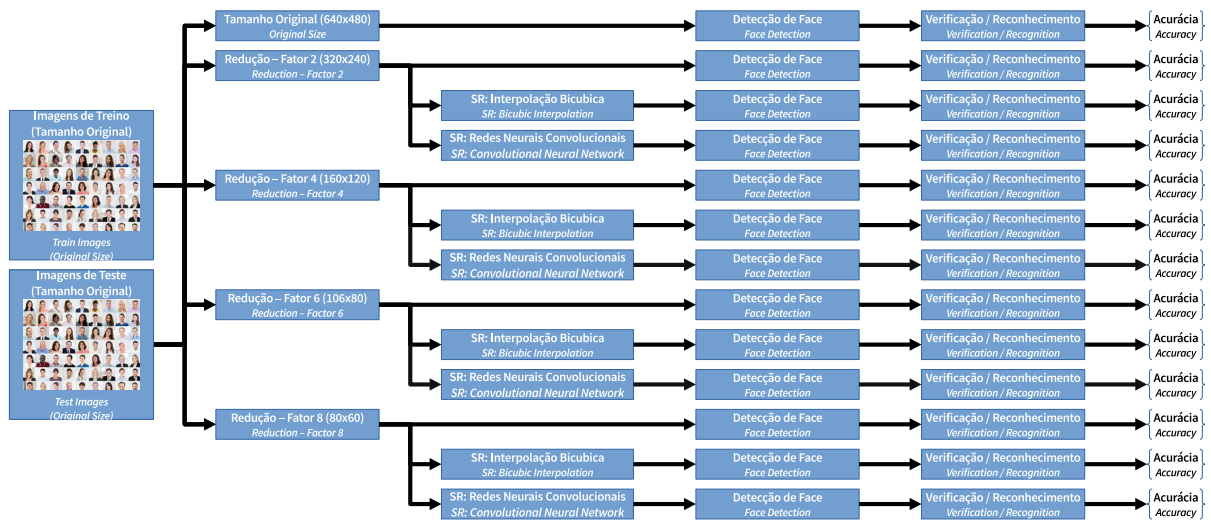


Figura 11 – Detalhamento do funcionamento do sistema proposto.

O computador utilizado para a execução dos testes possui como configuração 8GB de memória RAM, 64 bits e processador Intel®Core™ i7-7700 CPU @ 3.60GHz de 4 núcleos, 8 threads e 8MB de cache. Especificamente para a geração de imagens em super-resolução devido ao alto custo computacional, tornou-se inviável utilizar o computador mencionado, entretanto foi utilizado o serviço *Google Colab* que permite terceirizar o processamento o que tornou necessário mandar as imagens que seriam realizadas super-resolução para a nuvem proprietária da *Google*, a saber *Google Drive*.

4 RESULTADOS

Nessa secção, serão mostrados os resultados de todos os cenários e situações mencionados na parte anterior de modo mais detalhado e de acordo com suas especificidades abaixo:

- Conjunto de imagens de teste e treino de tamanho original;
- Redução da resolução original (640x480) pelos fatores de 2 (320x240), 4 (160x120), 6 (106x80) e 8 (80x60):
 - Conjuntos de imagens de teste e treino:
 - * reduzidos por interpolação bicúbica;
 - * reduzidos e em seguida aumentados para a resolução original em seguida por interpolação bicúbica;
 - * reduzidos por interpolação bicúbica e em seguida aumentados para a resolução original através do algoritmo *ISR* com os modelos *GANs*, *PSNR* e *Noise-Cancel*.
 - Somente as imagens do diretório de testes sendo mantida as propriedades originais das imagens do diretório de treino:
 - * Redução das imagens do diretório de testes por interpolação bicúbica;
 - * Redução e em seguida ampliação das imagens do diretório de testes para a resolução original em seguida por interpolação bicúbica;
 - * Redução por interpolação bicúbica e em seguida ampliação das imagens do diretório de testes para a resolução original através do algoritmo *ISR* com os modelos *GANs*, *PSNR* e *Noise-Cancel*.

Com as imagens sem nenhuma alteração nas suas especificações originais, no primeiro cenário, no que diz respeito à treinamento e testagem, foi obtido 100% de acerto nas identificações das faces, não resultando em nenhum erro. Ou seja, todas as 550 imagens foram identificadas com sucesso. A Figura 12 exemplifica o resultado obtido de uma dada imagem de entrada.



Figura 12 – Exemplo de uma imagem a ser reconhecida e o resultado do reconhecimento.

Em seguida, é alterado o cenário, mudando as resoluções das imagens através da divisão da resolução original por 2, 4, 6 e 8, conforme mostrado na Figura 13, os resultados variam.



Figura 13 – Representação de uma imagem que passa pelo processo de interpolação bicúbica a uma dada resolução menor.

Tabela 1 – Resultados no cenário onde todas as imagens são diminuídas de forma bicúbica.

Configuração		Resultados	
Fator	Resolução	Acertos	Erros
2	320x240	548	2
4	160x120	543	7
6	106x80	17	553
8	80x60	0	550

Inicialmente, na resolução 320x240, resultado da resolução original dividido por 2 há 548 imagens que são identificadas acertadamente, resultando em 99,64% do total de

550 imagens, restando 0,36% das imagens que resultam em erro, ou seja, duas imagens. Já pelo fator 4 (160x120), temos 543 imagens que resultam em acerto (98,73%), restando sete imagens que resultam em erro (1,27%). No que diz respeito ao fator 6 (106x80), há 17 imagens que são identificadas corretamente, ou seja, somente 3,09% das imagens resultam em acerto enquanto a maioria 96,91% das imagens resultam em erro (533 imagens). Por fim, nenhuma imagem é acertada, sendo todas as imagens resultantes de erro.

Diferentemente da situação anterior, onde somente era diminuída a resolução das imagens, traz-se também uma situação em que as imagens são diminuídas e, através da interpolação bicúbica, são aumentadas para a resolução original, perdendo-se assim as propriedades originais. A Figura 14 ilustra esse processo.

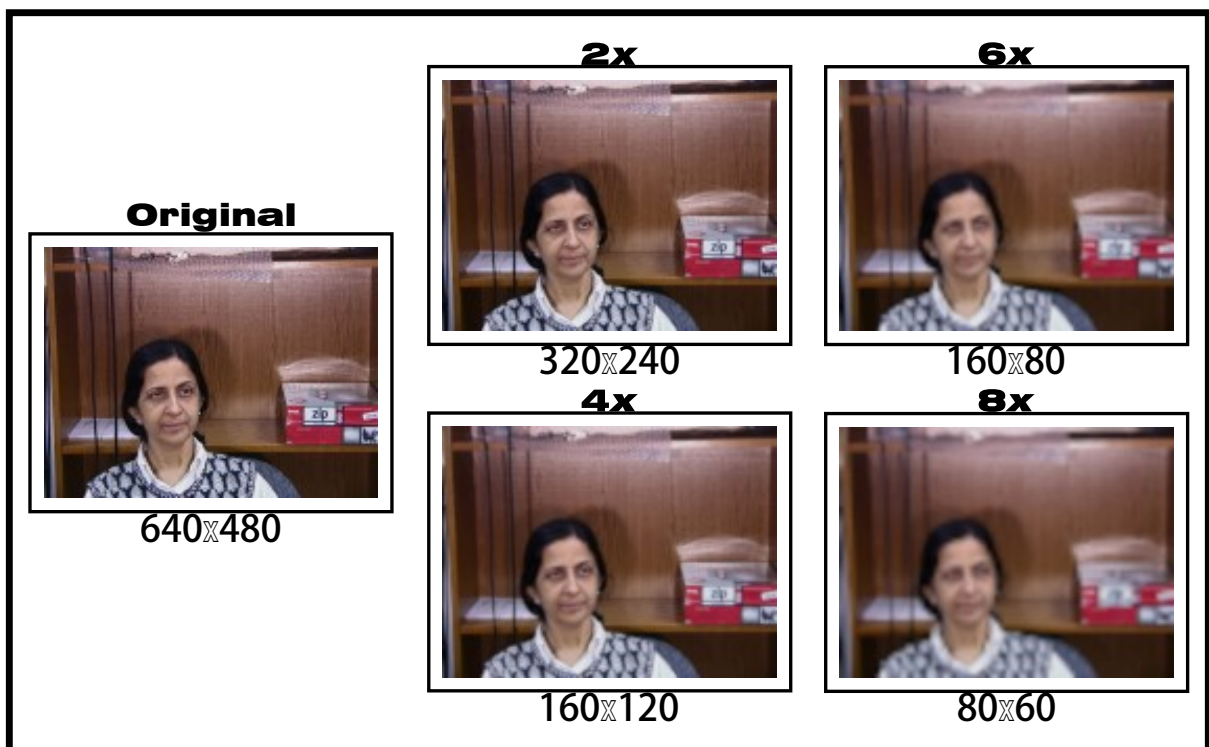


Figura 14 – Ilustração de uma imagem que é reduzida e posteriormente aumentada pelo processo de interpolação bicúbica.

Tabela 2 – Resultados no cenário onde todas as imagens são diminuídas e em seguida aumentadas de forma bicúbica.

Configuração		Resultados	
Fator	Resolução	Acertos	Erros
2	320x240	550	0
4	160x120	546	4
6	106x80	497	53
8	80x60	370	180

Nesse contexto, na resolução 320x240 (fator 2), temos 100% de acerto do total das 550 imagens. Pelo fator 4 (160x120) vê-se que 546 (99,27%) das imagens resultaram em acerto e somente quatro imagens (0,73%) em erro. Assim pelo fator 6 (106x80), têm-se que 90,36% (497 imagens) das imagens resultam em acerto enquanto 9,64% (53 imagens) resultam em erro. Por fim, há o fator 8 (80x60) onde 67,27% (370 imagens) das imagens são acertadas e 32,73% (180 imagens) resultam em erro.

Em comparação com as imagens que não sofrem a ampliação via interpolação bicúbica (como visto na Tabela 1), pode ser visto um progresso. Conforme os fatores aumentam, também aumentam os resultados das imagens que são ampliadas em comparação com as imagens que somente são diminuídas. Pelo fator 2 há uma melhora de 0,36%, no fator 4 a diferença é de 0,54%, a melhor performance ocorre no fator 6 com uma melhora de 87,27%, já no fator 8 há 67,27% de elevação.

Agora temos a introdução do algoritmo SRI. Nessa situação, as imagens são diminuídas de forma bicúbica e em seguida aumentadas para a resolução original com o algoritmo SRI utilizando os modelos *GANs*, *PSNR* e *Noise-Cancel*. A Figura 15 caracteriza esse processo.

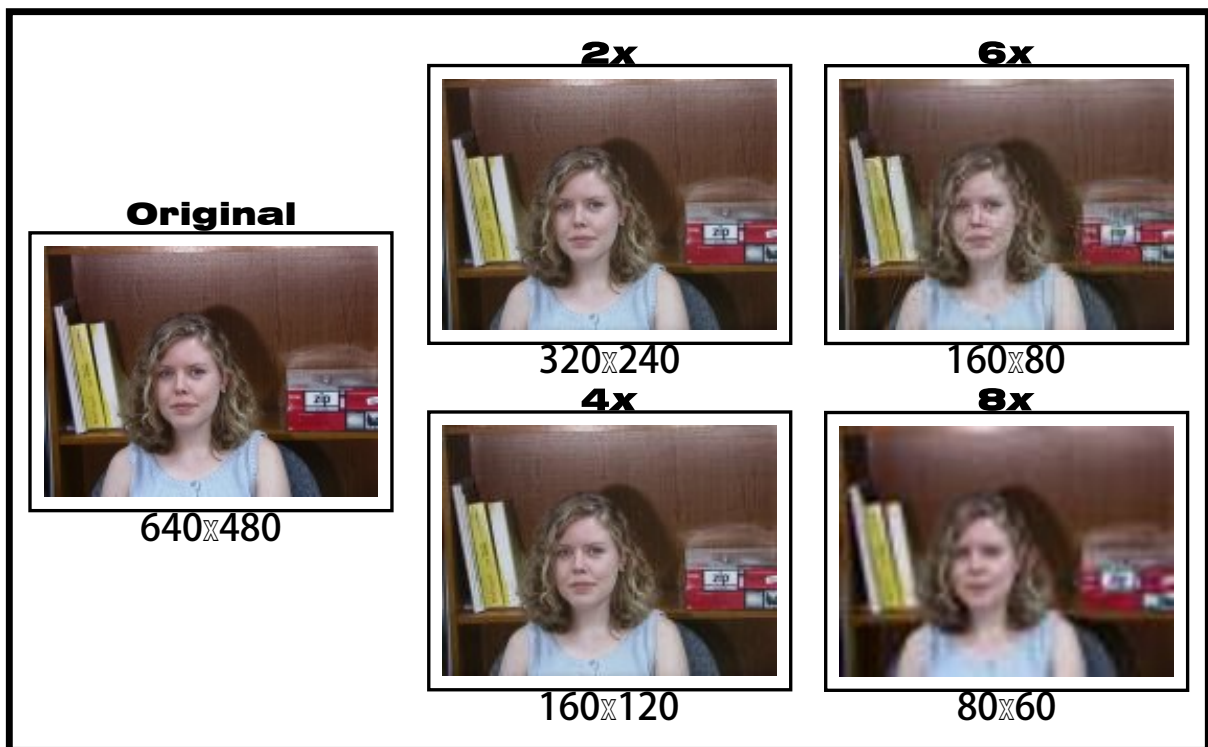


Figura 15 – Demonstração de uma imagem que é reduzida via interpolação bicúbica e imediatamente aumentada para a resolução original através do algoritmo SRI.

Tabela 3 – Resultados no cenário onde todas as imagens são diminuídas de forma bicúbica e em seguida aumentadas usando o algoritmo SRI.

Configuração		Resultados	
Fator	Resolução	Acertos	Erros
2	320x240	550	0
4	160x120	531	19
6	106x80	370	130
8	80x60	181	369

Como pode ser visto na Tabela 3, os resultados nesse cenário com a resolução 320x240 (fator 2), temos 100% de acerto do total das 550 imagens. Pelo fator 4 (160x120), vê-se que 531 (96,5%) das imagens resultaram em acerto e 19 imagens (3,5%) em erro. No fator 6 (106x80) têm-se que 67,27% (370 imagens) das imagens resultam em acerto enquanto 23,63% (130 imagens) resultam em erro. Por último, há o fator 8 (80x60) onde 32,9% (181 imagens) das imagens são acertadas e 67,1% (369 imagens) resultam em erro.

Finalmente, no último cenário é alterada somente a resolução das imagens que se situam no diretório de testes, mantendo as propriedades das imagens no diretório de treino, isso se aproxima da realidade e na maioria das aplicações. Em ambientes reais as imagens são controladas e capturadas inicialmente quando uma pessoa é cadastrada no banco de dados, portanto essas imagens precisam estar em boas resoluções.

Tabela 4 – Resultados no cenário onde somente as imagens do diretório de teste são diminuídas.

Configuração		Resultados	
Fator	Resolução	Acertos	Erros
2	320x240	550	0
4	160x120	549	1
6	106x80	110	440
8	80x60	0	550

Conforme visto na Tabela 4, começando pelo fator 2 de resolução 320x240, se mantém os resultados conseguidos através da resolução original, há 100% de acertos no total das 550 imagens. Pelo fator 4 de resolução 160x120, os resultados mudam sutilmente, há um erro o que corresponde 0,2% das 550 imagens. Assim 99,98% das imagens resultam em acerto, o que condiz em 549 imagens. Já no fator 6 de resolução 106x80, há uma piora considerável no que diz respeito aos acertos resultando em 80% de erros (440 imagens) e 20% (110 imagens) de acertos. Por fim, com o fator 8 de 80x60 há o pior desempenho, não resultado em nenhum acerto e com 100% (550 imagens) de imagens identificadas erroneamente. Em comparação com a Tabela 1 os resultados nesse cenário significam uma

melhoria. Também significa que imagens em alta qualidade e resolução para a treinamento tendem a melhorar os resultados, mas para imagens em qualidades muito baixas nenhuma situação é mostrada como ideal. Pelo fator 2 é mostrado uma melhora de 0,36%, já no fator 4 a diferença é de 1,09%, a melhor performance ocorre no fator 6 com 16,9% e finalmente no fator 8 não há melhora, em nenhum dos casos as imagens foram reconhecidas.

Similarmente, a situação anterior, agora além de diminuir a resolução, através da interpolação bicúbica, as imagens são aumentadas para a resolução original, perdendo-se, assim, as propriedades originais.

Tabela 5 – Resultados no cenário onde só as imagens de teste são reduzidas e em seguida aumentadas.

Configuração		Resultados	
Fator	Resolução	Acertos	Erros
2	320x240	550	0
4	160x120	545	5
6	106x80	527	23
8	80x60	399	151

Como visto na Tabela 5, inicialmente, no fator 2 (320x240), há 100% de acerto (550 imagens) e nenhum erro. Seguindo pelo fator 4 (160x120), há 99,1% de acerto (545 imagens) e cinco imagens resultantes de erro. Já pelo fator 6, as imagens em erros aumentam, resultando em 4,2% das imagens (23 imagens) com 95,8% das imagens identificadas corretamente. Por fim, com o fator 8, temos que 72,5% das imagens resultam em acertos (399) e 27,5% das imagens (151 imagens) condiz em erros. A comparação entre a Tabela 2 e a Tabela 5 onde as mesmas satisfazem condições similares, a Tabela 2 exibe uma performance melhor. Pelo fator 2 ambas as situações retratadas mostram que todas as imagens são reconhecidas. No fator 4 a Tabela 2 leva vantagem por 0,18%. Já no fator 6 a diferença sobe para 5,45%. Por último, no fator 8 a variação fica em 5,27%. Isso significa que o uso de imagens em alta qualidade no treinamento não melhora a performance em comparação com as imagens que são diminuídas e em seguida ampliadas.

Na última situação desse cenário, temos as imagens dos diretórios de testes e treinos reduzidas e aumentadas através do algoritmo SRI utilizando os modelos *GANs*, *PSNR* e *Noise-Cancel*.

Tabela 6 – Resultados no cenário as imagens de testes são diminuídas de forma bicúbica e em seguida aumentadas usando o algoritmo SRI.

Configuração		Resultados	
Fator	Resolução	Acertos	Erros
2	320x240	550	0
4	160x120	544	6
6	106x80	432	118
8	80x60	234	316

A apuração nessa configuração retorna no fator 2 100% de acerto. No fator 4 temos 544 acertos (98,9%) e 6 erros (1,1%). Com o fator 6 o computo aponta 432 acertos (78,55%) e 118 erros (21,45%). Ultimamente no fator 8 234 há acertos (42,55%) e 316 erros (57,45%). Comparando a Tabela 3 com a Tabela 6 mudanças significativas são mostradas em favor da Tabela 6. Pelo fator 2 ambas as tabelas reconhecem todas as imagens. No fator 4 a Tabela 6 mostra uma vantagem de 2,36%. Já no fator 6 a diferença é de 11,27%. Finalmente no fator 8 é mostrado uma melhora de 9,63%. Isso significa que no que diz respeito ao processo de treinamento, quando as imagens são bicubicamente diminuídas e em seguida ampliadas usando o algoritmo SRI os resultados são melhores do que todas as imagens que passam por esse mesmo processo. Mas esses resultados são superados pelos obtidos através da ampliação via interpolação bicúbica.

A Figura 16 apresenta o resumo dos principais resultados pelo número de acertos do reconhecimento de faces comparando as quatro principais configurações: reamostrando as imagens de teste e treino utilizando a interpolação bicúbica (a) e a super-resolução via *CNN* (b), e reamostrando apenas as imagens de testes utilizando a interpolação bicúbica (c) e a super-resolução via *CNN* (d) para os quatro fatores (2, 4, 6 e 8). Nota-se que em todas as configurações, reduzindo a imagem pela metade do tamanho o número de acertos continua o mesmo, ou seja 100 por cento. A medida que a resolução começa a diminuir, o número de acertos também diminui. Em todos os casos a super-resolução utilizando a interpolação bicúbica teve uma performance superior a *CNN* principalmente para o fator mais baixo quando a imagem esta quase toda desfocada.

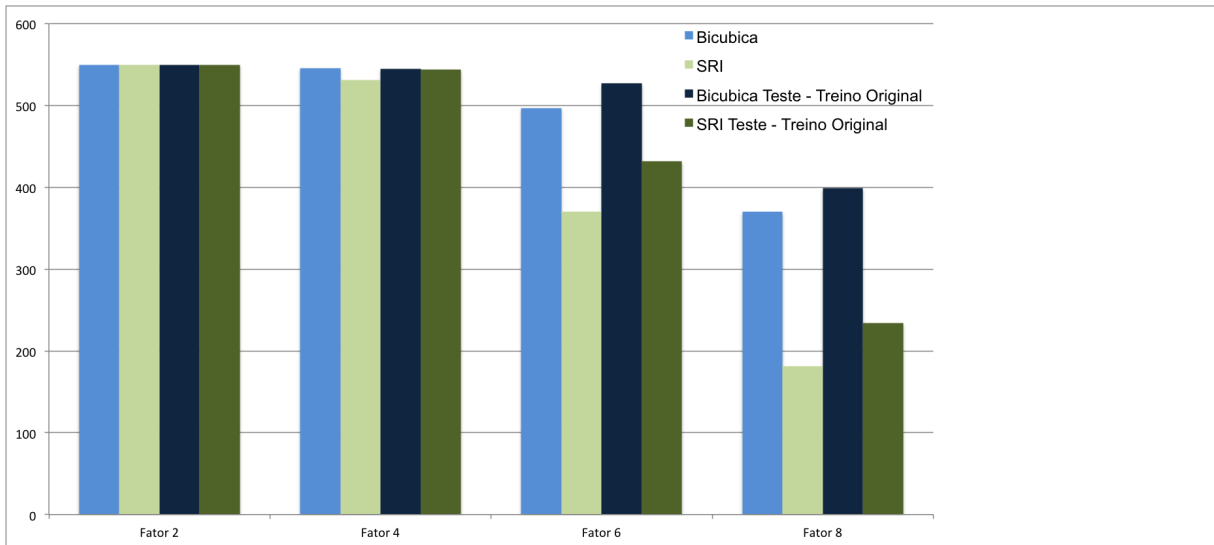


Figura 16 – Resumo dos Resultados em número de acertos do processo de reconhecimento agrupados em diferentes configurações de testes pelos dois algoritmos de super-resolução.

5 CONCLUSÃO

Dados os resultados em diferentes cenários e situações, é possível observar que a conjuntura ideal é onde as imagens estão com suas configurações ajustadas para resoluções altas, é onde se obtém os melhores resultados. No entanto, em situações cotidianas, não é sempre possível e nem provável contar com cenários onde as imagens se encontram em qualidades e resoluções excelentes.

Com isso em vista, mesmo com imagens em resoluções baixas, o processo de aumentar artificialmente a resolução melhora, consideravelmente, os resultados e o desempenho do algoritmo em questão no que diz respeito a capacidade de acerto dos rostos identificados.

Os dados apontam que, com as imagens que são treinadas em alta resolução, mesmo com as imagens que serão identificadas em resolução inferior, os resultados são superiores do que os cenários em que os treinos ocorrem com imagens em resoluções inferiores, o que é o caso de muitas aplicações utilizadas cotidianamente e em cenários reais como o da segurança pública/privada.

Havia uma expectativa que a super-resolução iria melhorar e aperfeiçoar os resultados no que diz respeito a acurácia, no entanto o uso do algoritmo deixou a desejar e não trouxe resultados concretos nesse quesito. Apesar que visualmente conforme as imagens que passam por super-resolução, desde que tenham resoluções não muito baixas, mostram um maior detalhamento, isso não torna necessariamente o processo de reconhecimento facial mais significativo e eficaz no que diz respeito ao quesito de desempenho e resultados.

A interpolação bicúbica utilizada para aumentar de modo artificial as imagens para as resoluções originais com perdas de detalhes, tem como consequência resultados consideravelmente superiores e, portanto, nesse contexto específico a sua utilização é fundamentada.

Mesmo sem nenhuma nova informação, a interpolação utiliza os próprios pixels originais da imagem no seu processo de cálculo. Os resultados favoráveis que foram obtidos nos procedimentos que usam a interpolação podem estar relacionados ao fato que o algoritmo de reconhecimento facial escolhido é baseado em detalhes finos. O algoritmo de *Deep Learning* gera vários artefatos que podem confundir esse algoritmo de reconhecimento utilizado.

Em trabalhos futuros, outros algoritmos de reconhecimento podem ser testados, assim como outras redes neurais para comparação dos resultados e sua viabilidade prática. A interpolação bicúbica em comparação com *Deep Learning* pode ser uma alternativa mais barata, rápida e factível para sistemas de baixo custo que geram imagens em qualidades e resoluções inferiores que é a realidade da maioria das aplicações, melhorando a acurácia

dos resultados de maneira considerável.

REFERÊNCIAS

- AL., F. C. et. **ISR**. 2018. <<https://github.com/idealo/image-super-resolution>>.
- ARASHLOO, S. R.; KITTLER, J. Fast pose invariant face recognition using super coupled multiresolution markov random fields on a GPU. **Pattern Recognit. Lett.**, v. 48, p. 49–59, 2014. Disponível em: <<https://doi.org/10.1016/j.patrec.2014.05.017>>.
- BILGAZYEV, E. et al. Improved face recognition using super-resolution. In: **2011 IEEE International Joint Conference on Biometrics, IJCB 2011, Washington, DC, USA, October 11-13, 2011**. IEEE Computer Society, 2011. p. 1–7. Disponível em: <<https://doi.org/10.1109/IJCB.2011.6117554>>.
- BISSI, T. D. **Reconhecimento facial com os algoritmos eigenfaces e fisherfaces**. 41 f. Monografia (Projeto de Conclusão de Curso) — Universidade Federal de Uberlândia, Uberlândia, 2018.
- COMMONS, W. **File:Phobos in super-resolution – before and after image correction ESA204381.jpg — Wikimedia Commons, the free media repository**. 2020. [Online; accessed 5-August-2021]. Disponível em: <https://commons.wikimedia.org/w/index.php?title=File:Phobos_in_super-resolution_%E2%80%93_before_and_after_image_correction_ESA204381.jpg&oldid=448429979>.
- COMMONS, W. **File:Synthese+.svg — Wikimedia Commons, the free media repository**. 2020. [Online; accessed 5-August-2021]. Disponível em: <<https://commons.wikimedia.org/w/index.php?title=File:Synthese%2B.svg&oldid=465993711>>.
- COMMONS, W. **File:Bitmap VS SVG.svg — Wikimedia Commons, the free media repository**. 2021. [Online; accessed 5-August-2021]. Disponível em: <https://commons.wikimedia.org/w/index.php?title=File:Bitmap_VS_SVG.svg&oldid=575149276>.
- DAVIS, M.; POPOV, S.; SURLEA, C. Real-time face recognition from surveillance video. In: ZHANG, J. et al. (Ed.). **Intelligent Video Event Analysis and Understanding**. [s.n.], 2011, (Studies in Computational Intelligence, v. 332). p. 155–194. Disponível em: <https://doi.org/10.1007/978-3-642-17554-1_8>.
- DOURADO, W. B. **Avaliação de Técnicas de Interpolação de Imagens Digitais**. Dissertação (Master dissertation) — Universidade Estadual Paulista (UNESP), Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Tecnologia, Presidente Prudente, ago. 2014. Disponível em: <<http://hdl.handle.net/11449/115892>>.
- GONZALEZ, R. C.; WOODS, R. C. **Processamento Digital De Imagens**. São Paulo, Brasil: ADDISON WESLEY BRA, 2009. ISBN 9788576054016.
- GONZALEZ-SOSA, E. et al. Facial soft biometrics for recognition in the wild: Recent works, annotation, and cots evaluation. **IEEE Transactions on Information Forensics and Security**, v. 13, n. 8, p. 2001–2014, 2018.

GUNTURK, B. K. et al. Eigenface-domain super-resolution for face recognition. **IEEE Trans. Image Processing**, v. 12, n. 5, p. 597–606, 2003. Disponível em: <<https://doi.org/10.1109/TIP.2003.811513>>.

HOFBAUER, H.; DEBIASI, L.; UHL, A. Mobile face recognition systems: Exploring presentation attack vulnerability and usability. In: **2019 International Conference on Biometrics (ICB)**. [S.l.: s.n.], 2019. p. 1–7.

JR, M. T.; ROSA, A. S. Super-resolução de imagens cbers 2. **XIII Simpósio Brasileiro de Sensoriamento Remoto**, p. 1197–1204, 2007. Disponível em: <<http://martes.sid.inpe.br/col/dpi.inpe.br/sbsr@80/2006/11.16.00.47/doc/1197-1204.pdf>>.

KIRCHGASSER, S. et al. Is warping-based cancellable biometrics (still) sensible for face recognition? In: **2020 IEEE International Joint Conference on Biometrics (IJCB)**. [S.l.: s.n.], 2020. p. 1–9.

LEDIG, C. et al. Photo-realistic single image super-resolution using a generative adversarial network. **CoRR**, abs/1609.04802, 2016. Disponível em: <<http://arxiv.org/abs/1609.04802>>.

NEFIAN, A. V. Georgia tech face database. 1999. Disponível em: <http://www.anefian.com/face_reco.htm>.

RIBEIRO, E.; UHL, A.; ALONSO-FERNANDEZ, F. Iris super-resolution using cnns: is photo-realism important to iris recognition? **IET Biom.**, v. 8, n. 1, p. 69–78, 2019. Disponível em: <<https://doi.org/10.1049/iet-bmt.2018.5146>>.

RICH, E.; KNIGHT, K.; CALERO, P. **Inteligencia artificial**. McGraw-Hill, 1994. ISBN 9788448118587. Disponível em: <<https://books.google.com.br/books?id=yunOPQAACAAJ>>.

SAVVIDES, M.; KUMAR, B. V.; KHOSLA, P. Cancelable biometric filters for face recognition. In: **Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004**. [S.l.: s.n.], 2004. v. 3, p. 922–925 Vol.3.

SCHULTZ, R. R.; STEVENSON, R. L. Extraction of high-resolution frames from video sequences. **IEEE Trans. Image Processing**, v. 5, n. 6, p. 996–1011, 1996. Disponível em: <<https://doi.org/10.1109/83.503915>>.

SHAH, A. **Visualizing Artificial Neural Networks (ANNs) with just One Line of Code**. 2018. Disponível em: <<https://towardsdatascience.com/visualizing-artificial-neural-networks-anns-with-just-one-line-of-code-b4233607209e>>.

SPRING. 2006. Disponível em: <http://www.dpi.inpe.br/spring/portugues/tutorial/registro_img.html>.

SUNDARARAJAN, K.; WOODARD, D. L. Deep learning for biometrics: A survey. **ACM Comput. Surv.**, v. 51, n. 3, p. 65:1–65:34, 2018. Disponível em: <<https://doi.org/10.1145/3190618>>.

WANG, X. et al. ESRGAN: enhanced super-resolution generative adversarial networks. **CoRR**, abs/1809.00219, 2018. Disponível em: <<http://arxiv.org/abs/1809.00219>>.

XU, X.; LIU, W.; LI, L. Low resolution face recognition in surveillance systems. **Journal of Computer and Communications**, v. 02, p. 70–77, 01 2014. Disponível em: <<https://doi.org/10.4236/jcc.2014.22013>>.

ZHANG, Y. et al. Residual dense network for image super-resolution. **CoRR**, abs/1802.08797, 2018. Disponível em: <<http://arxiv.org/abs/1802.08797>>.

ZHAO, H. et al. Loss functions for image restoration with neural networks. **IEEE Trans. Computational Imaging**, v. 3, n. 1, p. 47–57, 2017. Disponível em: <<https://doi.org/10.1109/TCI.2016.2644865>>.