Universidade Federal do Piauí
Centro de Ciências da Natureza
Programa de Pós-Graduação em Ciência da Computação

Aumento de dados para reconhecimento facial com transferência de aprendizado de CNNs

Valeska de Sousa Uchôa

Teresina-PI, Agosto de 2019

Valeska de Sousa Uchôa

Aumento de dados para reconhecimento facial com transferência de aprendizado de CNNs

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFPI (área de concentração: Sistemas de Computação), como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação

Orientador: Kelson Rômulo Teixeira Aires

Coorientador: Rodrigo de Melo Souza Verass

Teresina-PI

Agosto de 2019

Valeska de Sousa Uchôa
Aumento de dados para reconhecimento facial com transferência de aprendizado de CNNs/ Valeska de Sousa Uchôa. – Teresina-PI, Agosto de 2019-
44 p. : il. (algumas color.) ; 30 cm.

Orientador: Kelson Rômulo Teixeira Aires

Dissertação (Mestrado) – Universidade Federal do Piauí – UFPI
Centro de Ciências da Natureza
Programa de Pós-Graduação em Ciência da Computação, Agosto de 2019.

1. Reconhecimento facial. 2. Aumento de dados. I. Orientador. II. Universidade Federal do Piauí. III. Aumento de dados para reconhecimento facial com transferência de aprendizado de CNNs.

CDU 02:141:005.7

"Aumento de dados para reconhecimento facial com transferência de aprendizado de CNNs"

VALESKA DE SOUSA UCHÔA

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Natureza da Universidade Federal do Piauí, como parte integrante dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

Aprovada por:

Prof. Kelson Rômulo Teixeira Aires
(Presidente da Banca Examinadora)

Prof. Rodrigo de Melo Souza Veras
(Examinador Interno)

Prof. Laurindo de Sousa Britto Neto
(Examinador Interno)

Prof. Anselmo Cardoso de Paiva
(Examinador Externo à Instituição)

Teresina, 09 de agosto de 2019

Aos meus pais Lúcia e Edvaldo
por sempre estarem comigo em todos os momentos.

Agradecimentos

Agradeço a Deus.

Agradeço aos meus pais, Lúcia e Edvaldo por tudo que fizeram para que eu tivesse a oportunidade de estar aqui.

Agradeço ao meu orientador, Kelson Aires, por todos os conselhos, pela ajuda e pela compreensão nos últimos anos.

Ao meu namorado Juninho, ao meu irmão Lucas, minha prima Talessa e aos meus amigos Felipe, Renato, Lucas, Luis, Paulo e Clara e tantos outros por tornarem mais fácil viver essa jornada.

Aos professores do Departamento de Computação da UFPI pelo conhecimento compartilhado.

À CAPES pelo apoio financeiro para realização deste trabalho de pesquisa.

"The world always seems brighter
when you've just made something that wasn't there before."
(Neil Gaiman)

Resumo

O reconhecimento facial é uma tarefa desafiadora de Visão Computacional. Nesta dissertação, é proposto um método para reconhecimento de faces aplicando aumento de dados e transferência de aprendizado de Redes Neurais Convolucionais (CNNs) pré-treinadas. O foco é analisar o poder do aumento de dados para melhorar o desempenho do sistema. Foram extraídas características das imagens para o treinamento de classificadores usando a CNN VGG-Face. Para uma base de imagens de entrada, são aplicadas várias transformações gerando 12 versões diferentes da base de imagens de entrada, para avaliar qual combinação produz melhores resultados. Experimentos foram realizados usando o aumento de dados na base Labeled Faces in the Wild (LFW). Os testes mostraram que a acurácia para essa base chegou a 98.43%. Também foi criada uma base de dados proprietária composta por imagens de 12 indivíduos. Para essa base a melhor acurácia foi de 95.41%. A melhoria dos resultados com o método proposto leva a inferir que o aumento de dados é um passo essencial para a tarefa de reconhecimento facial. No entanto, como a operação de aumento que mais contribui com a melhora dos resultados não é a mesma para as duas bases de entrada é necessário realizar esse estudo para cada aplicação.

Palavras-chaves: reconhecimento facial, aumento de dados, transferência de aprendizado, redes neurais convolucionais.

Abstract

Face recognition is a challenging Computer Vision task. In this dissertation, we propose a method for face recognition by applying data augmentation and transfer learning in pre-trained Convolutional Neural Networks (CNNs). Our main focus is to analyze the power of data augmentation towards improving system accuracy. We have trained classifiers with extracted features from the VGG-Face CNN. For a given input dataset, we applied several transformations to generate 12 different versions of the datasets used to evaluate which combination produces better results. We ran experiments using data augmentation on the Labeled Faces in the Wild (LFW) dataset. The experiments with LFW reached accuracy of 98.43%. We also created a proprietary dataset composed of 12 subjects. For the proprietary dataset, the best accuracy obtained was 95.41%. The improvement of results with the proposed method leads to infer that data augmentation is an essential step for the facial recognition task. However, since the augmentation operation that contributes the most to the results improvement is not the same for the two input datasets it is necessary to conduct this study for each application.

Keywords: face recognition, data aumentation, transfer learning, convolutional neural networks.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

CFTV	Circuito Fechado de Televisão
HSV	*Hue-Saturation-Value*
ILSRC	*ImageNet Larges-Scale Visual Recognition Challenge*
LFW	*Labeled Faces in the Wild*
KNN	*K-Nearest Neighbor*
RGB	*Red-green-blue*
SVM	*Support Vector Machine*
UFD	*UFPI Faces Database*

Sumário

1 Introdução

Reconhecimento facial em ambientes não controlados é uma tarefa de Visão Computacional de grande aplicabilidade, especialmente em sistemas de vigilância. O objetivo do reconhecimento facial é reconhecer a identidade de um indivíduo pertencente a uma base de dados dada uma imagem (ZHAO et al., 2003).

Uma etapa de detecção de face pode ser bastante útil em um sistema de reconhecimento facial. Detecção de face consiste em localizar uma face humana em uma determinada cena. Uma vez que a face é detectada, o sistema consulta a base de dados e identifica a qual classe a face pertence.

A tarefa de reconhecimento facial pode ser dividida em dois grandes grupos, as aplicações de identificação e de verificação. A identificação consiste em descobrir a identidade de de uma face – trata-se de um problema um-para-muitos. Por outro lado, a técnica de verificação tenta confirmar a identidade de amostra de face – é um problema um-para-um (VEL; AEBERHARD, 1999). Neste trabalho, o foco é a tarefa de identificação.

Apesar dos avanços realizados nesse campo, ainda há desafios a serem superados nas tarefas de detecção e reconhecimento de face. Os maiores obstáculos são pose, escala, mudanças de iluminação, plano de fundo, oclusão etc. O sistema adquire robustez quando essas dificuldades são contornadas, uma vez que mudanças na orientação, expressão facial e iluminação podem alterar a aparência do indivíduo a ser reconhecido.

Abordagens clássicas para resolver o problema de reconhecimento facial são *Eigenfaces* (TURK; PENTLAND, 1991) e *Fisherfaces* (BELHUMEUR; HESPANHA; KRIEGMAN, 1997). No entanto, esses métodos são conhecidamente sensíveis a variações de iluminação, expressão facial e pose, que normalmente ocorrem em cenas não controladas (ODIL; OBILEN, 2014). Métodos que são mais robustos a essas mudanças são aqueles que usam Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNN) (PARKHI; VEDALDI; ZISSERMAN, 2015a). No entanto, abordagens de aprendizado profundo utilizando CNN requerem um número grande de amostras para o treinamento. Quando o número de amostras não é o suficiente para treinar uma CNN é possível usar o poder de uma CNN pré-treinada por meio da técnica de transferência de aprendizado (YOSINSKI et al., 2014). Outra forma de contornar o problema de poucas amostras disponíveis é a utilização de técnicas de aumento de dados, que consiste em gerar novas amostras sintéticas a partir dos dados originais (HOWARD, 2013).

1.1 Motivação

A motivação para a realização deste projeto é que o reconhecimento facial tem uma variedade de possíveis aplicações, dentre elas, vigilância automática (Wheeler; Weiss; Tu, 2010; WANG, 2013), autenticação (DELAC; GRGIC, 2004; GALBALLY; MARCEL; FIERREZ, 2014), interação humano-computador (BARTLETT et al., 2003; RAUTARAY; AGRAWAL, 2015), realidade aumentada (AJANKI et al., 2011) etc.

A autenticação por meio de reconhecimento facial é uma alternativa mais atraente em comparação com outros métodos por se tratar de uma forma não invasiva de realizar esse procedimento.

Aplicações de vigilância automática são um importante uso da tecnologia de reconhecimento de face que se tornaram possíveis pela popularização de Circuito fechado de televisão (CFTV).

Circuito fechado de televisão é uma tecnologia baseada no monitoramento de ambientes por meio de câmeras analógicas ou digitais que se difundiu bastante nos últimos anos em residências, empresas, estabelecimentos comerciais e condomínios. Trata-se da colocação de câmeras em pontos estratégicos do ambiente a ser monitorado. Com isso, as imagens captadas são transmitidas para um ou mais pontos de visualização, contribuindo para a diminuição de crimes nesses locais. As câmeras podem usar essa tecnologia para rastrear rostos humanos e manter o controle das pessoas autorizadas em um determinado local. No entanto, grande parte desses sistemas exige a presença de uma pessoa monitorando as imagens. Um sistema de reconhecimento facial automático poderia dispensar essa necessidade.

1.2 Objetivos

O objetivo principal deste trabalho é propor e avaliar combinações de operações de aumento de dados para um sistema de reconhecimento facial. Para atingir esse objetivo geral, têm-se os seguintes objetivos específicos:

- Criar uma base de dados contendo imagens para treino e vídeo para teste;
- Selecionar operações de aumento de dados;
- Avaliar a melhor combinação de operações de aumento de dados para o problema de reconhecimento facial;
- Avaliar o classificador que melhor se adequa ao reconhecimento de faces.

1.3 Organização do Texto

Esta proposta está dividida em 6 (seis) capítulos.

O Capítulo 2 contém trabalhos relacionados que abordam os tópicos de aprendizado profundo, transferência de aprendizagem e aumento de dados.

Em seguida, no Capítulo 3, têm-se materiais e métodos, onde são apresentadas ferramentas utilizadas no desenvolvimento desta pesquisa, como classificadores, base de imagens e métricas de avaliação.

A descrição do sistema proposto se encontra no Capítulo 4, onde são detalhadas as etapas que compõem a proposta de solução para o problema de reconhecimento facial. Ainda nesse capítulo, são mostradas as operações de aumento de dados e como são usadas para treinar os classificadores.

Resultados dos experimentos e a discussão dos mesmo podem ser vistos no Capítulo 5. Finalmente, o Capítulo 6 discute conclusões que podem ser tiradas a partir dos resultados e as possibilidades de continuidade desta pesquisa.

2 Trabalhos relacionados

Neste capítulo, são abordados trabalhos que procuram resolver o problema de reconhecimento facial, trabalhos que utilizam a transferência de aprendizado em problemas de classificação de imagens, além de estratégias de aumento de dados usadas na literatura.

2.1 Aprendizado profundo

Recentemente, técnicas de aprendizado profundo vêm sendo utilizadas como solução para o problema de reconhecimento facial com bons resultados (PARKHI; VEDALDI; ZISSERMAN, 2015a) (SUN; WANG; TANG, 2014). Esses métodos têm uma imagem como entrada em uma rede neural convolucional. Eles utilizam filtros convolucionais para produzir representações para essa imagem que possibilite a classificação de múltiplas classes, em que cada classe representa um indivíduo da base de dados.

Parkhi et al. (PARKHI; VEDALDI; ZISSERMAN, 2015b) criaram uma arquitetura chamada VGG-Face, uma rede neural convolucional que consiste de 37 camadas neurais e é dividida em 11 blocos, dos quais 8 são convolucionais e 3 são completamente conectadas. Cada um desses blocos contém um operador linear seguido por uma ou mais não-linearidades. Eles aplicaram algumas tranformações para expandir a base de dados que usaram, mas como Herrmann, Willersinn e Beyerer (HERRMANN; WILLERSINN; BEYERER, 2016) apontaram, eles não avaliaram as contribuições das operações realizadas.

Wang et al. (WANG et al., 2017) apresentou um sistema de reconhecimento facial baseado em vídeo usando o *fine-tuning* de CNN. Os autores coletaram e rotularam automaticamente um novo conjunto de dados de vídeos de vigilância. O conjunto de dados é usado para treinar um método para reconhecimento de face em vídeos de vigilância do mundo real, ajustando a CNN VGG-Face. O procedimento de coleta de dados e rotulação é dividido em quatro etapas: geração de conjunto de dados, purificação dentro de cada classe, purificação entre classes e filtragem. Apenas as camadas totalmente conectadas são ajustadas com o novo conjunto de dados para tornar o modelo mais adequado ao seu domínio da aplicação. Em seu conjunto de dados de teste, o modelo de face do VGG após o ajuste fino atinge uma taxa de reconhecimento de 92.1%.

A primeira abordagem baseada em CNN para resolver o problema de reconhecimento de face de vídeo foi proposta por Ding e Tao (DING; TAO, 2018). Os autores abordam o problema apresentando representações faciais robustas contra borrões. Durante o treinamento, dois fluxos de dados são fornecidos. O primeiro é composto por imagens estáticas. O segundo fluxo é composto por quadros de vídeo simulados. A simulação ocorre

a partir da aplicação de um desfoque artificial aleatório às imagens do primeiro fluxo. A proposta é chamada de *CNN of the trunk-extension (TBE-CNN)* e inclui uma rede de troncos e várias redes de ramos. A rede de troncos aprende as representações faciais para imagens holísticas e cada rede de ramos aprende as representações para os *patches* de imagem. Os experimentos são conduzidos em três bancos de dados de face de vídeo em grande escala publicamente disponíveis. A abordagem mostrou taxas de verificação de 96.12% e 99, 33% para os conjuntos de dados PaSC (Beveridge et al., 2013) e COX Face (HUANG et al., 2015), respectivamente. Para o banco de dados do YouTube Faces in the Wild (Wolf; Hassner; Maoz, 2011), esse método atingiu uma acurácia de 94.96 % e um erro padrão de 0.31.

2.2 Transferência de aprendizado

Treinar uma rede neural convolucional e otimizar os milhões de parâmetros de uma arquitetura de aprendizado profundo requer centenas de amostras e GPUs, e ainda assim leva um tempo considerável até que o treinamento seja completo. Esses recursos nem sempre estão disponíveis para pequenos grupos de pesquisa. Para contornar esse problema, pesquisadores utilizam duas abordagens de transferência de aprendizado, que usam informações de uma CNN treinada anteriormente para um propósito diferente (YOSINSKI et al., 2014) (TAIGMAN et al., 2014). A primeira consiste em fazer um ajuste fino nos parâmetros da rede por meio de retropropagação. Essa abordagem é indicada quando a transferência de aprendizado é aplicada a um problema com uma grande base de dados. Outra possibilidade de transferir aprendizado é preservar os pesos originais da arquitetura, utilizar a rede neural para extrair características e usá-las para treinar um classificador. Este trabalho segue a segunda abordagem porque a base de dados original contém poucas amostras.

2.3 Aumento de dados

Uma técnica que vem sendo largamente utilizada por vários pesquisadores em classificação de imagens é aumento de dados (PEREZ; WANG, 2017; HOWARD, 2013). Aumento de dados é uma estratégia que consiste na utilização de métodos para gerar dados sintéticos a partir de uma base de dados. A seguir apresentamos alguns trabalhos que aplicam esses métodos em seus sistemas de classificação.

Perez e Wang (PEREZ; WANG, 2017) fazem uso de uma série de técnicas de aumento de dados e avaliam qual produz melhores resultados em seus experimentos. Eles trabalharam com três base de dados, cada uma com duas classes, sendo duas delas subconjuntos da base *tiny-imagenet-200*: *dog vs. cat* e *dog vs. goldfiish*; e a terceira base é um subconjunto da MNIST: 0's vs. 8's. Os autores usam três formas de aumento de

dads. A primeira é a forma tradicional, em que imagens são transformadas por meio de translação, rotação, escala, espelhamento e mudanças de matiz. A segunda abordagem é aumento com Redes Adversárias Geradoras (Generative Adversarial Networks - GAN) em que cada imagem tem um estilo transferido para a mesma de um conjunto de seis estilos diferentes (*Cezanne*, *Enhance*, *Monet*, *Ukiyoe*, *Van Gogh* e *Winter*). A terceira é um aumento de dados neural em que, durante o treinamento, tomam duas imagens da mesma classe como entrada de uma rede convolucional que retorna uma camada (interpretada como imagem). Essa imagem modificada passa por uma segunda rede conhecida como rede de classificação. Nesse método, as imagens de teste são usadas como entradas somente na segunda rede. Os experimentos mostraram que para o problema *dog vs. goldfish* a rede neural com aumento de dados atingiu acurácia de 91.5% contra 85.5% sem aumento. O problema *dog vs. cat* apresentou melhores resultados com o aumento de dados tradicional com acurácia de 77.5%. O aumento mostrou pouca melhoria para o subconjunto de MNIST, indo de 97.2% a 97.5% de acurácia.

Krizhevsky, Sutskever e Hinton (KRIZHEVSKY; SUTSKEVER; HINTON, 2012a) realizaram reconhecimento de objeto em um subconjunto do ImageNet (DENG et al., 2009) conhecido como ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), que contém 1000 classes, com aproximadamente 1000 images por classe. Os autores tentaram reduzir sobreajuste aplicando técnicas de aumento de dados. A primeira abordagem consiste em fazer translações e reflexão horizontal. Isso significa que para cada imagem de treino são extraídos fragmentos aleatórios e suas versões espelhadas. Na fase de teste, a rede extrai cinco fragmentos das imagens e suas reflexões e prediz a que classe o objeto pertence ao fazer uma média das predições para os dez fragmentos. A segunda abordagem altera as intensidades dos canais RGB nas imagens utilizando *Principal Component Analysis* (PCA) (PEARSON, 1901). Esse método se baseia no fato de que a identidade do objeto é invariante a mudanças da iluminação. Eles propuseram uma arquitetura que contém oito camadas com pesos das quais cinco são convolucionais e três são completamente conectadas. Essa arquitetura atingiu taxas de erro top-1 e top-5[1]. de 37.5% e 17.0%, respectivamente, na base ILSVRC-2010.

Esse último trabalho inspirou o trabalho apresentado por Howard (HOWARD, 2013). Ele trabalha em um subconjunto da ImageNet e propõe uma arquitetura CNN similar. A diferença entre os dois trabalhos é que as camadas completamente conectadas nessa arquitetura são duas vezes maiores que na arquitetura apresentada por Krizhevsky, Sutskever and Hinton (KRIZHEVSKY; SUTSKEVER; HINTON, 2012a). O autor usa várias transformações na imagem para gerar novas imagens para treino e teste. As modifi-

[1] Taxa de erro Top-1 e Top-5 são medidas comumente utilizadas na avaliação do *benchmark* ILSVRC. Taxa de erro Top-1 mostra a porcentagem de vezes que o classificador não apontou a classe correta com maior probabilidade, enquanto a Taxa de erro Top-5 se refere às vezes em que o classificador não incluiu a classe correta entre as cinco com maior probabilidade

cações aplicadas nas imagens são as mesmas do trabalho anterior adicionadas a outras que extendem a invariância a translação e cor. A primeira seleciona um corte aleatório da imagem original, similar ao que foi feito no trabalho anterior, com a diferença que nessa proposta o menor lado da imagem é redimensionado para 256 pixels em vez do maior. Isso é feito para prevenir perda de informação. As outras transformações são mudanças no contraste, brilho e cor ao escolher um número aleatório entre 0.5 e 1.5 para o nível de mudança. Além disso, em tempo de teste, predições são feitas em 5 translações, 2 espelhamentos, 3 escalas e 3 perspectivas. Dessa forma, são 90 predições no total. No entanto, um algoritmo guloso é utilizado para escolher um subconjunto de 10 e 15 predições que apresenta resultados similares a utilizar as 90. O sistema final de classificação foi composto de 10 redes neurais contendo 5 modelos básicos e 5 de alta resolução e conseguiu uma taxa de erro top-5 de 13.6%.

Os trabalhos anteriormente citados mostram as vantagens de aumento de dados em problemas de classificação de objetos de um modo geral. No entanto, existem estudos que propõem técnicas voltadas para o reconhecimento de face, que é o foco deste trabalho.

Alguns métodos realizam transformações a nível de características, como o trabalho proposto por Leng, Yu, QIN (LENG; YU; QIN, 2017). Os autores geram amostras extras ao adicionar ruído nas características extraídas das amostras originais da base de dados. Para avaliar essa estratégia, os autores treinaram uma Máquina de Vetores de Suporte (*Support Vector Machine* - SVM) (CORTES; VAPNIK, 1995) com a base de dados *Labeled Faces in the Wild* (LFW) (HUANG et al., 2007).

Ao invés de fazer transformações a nível de características, alguns métodos populares de aumento de dados propõem manipulações diretas das imagens. Masi et al. (MASI et al., 2016) descreveram abordagens de aumento de dados para o reconhecimento facial que sintetizam imagens ao aplicar variações de posição, forma e expressões na base de dados. O método desses autores foi testado no *benchmark* IARPA Janus (Whitelam et al., 2017) e na base LFW e atingiram resultados de estado-da-arte.

Herrmann, Willersinn and Beyerer (HERRMANN; WILLERSINN; BEYERER, 2016) usaram técnicas de aumento de dados como corte, espelhamento, rotação, borramento, adição de ruído, compressão e redimensionamento. Eles avaliaram o impacto dessas estratégias na acurácia do sistema. Os experimentos mostraram que usando esses métodos aumentaram acurácia para todas as bases testadas. Por exemplo, para a base MSRA (ZHANG et al., 2012), a acurácia foi de 53.8% para 60.3% com o aumento de dados.

Outro trabalho que propôs aumento de dados para o reconhecimento facial é o apresentado por Hu et al. (HU et al., 2016). Propuseram sintetizar dados para expandir uma base de dados tornando-a grande o suficiente para treinar redes neurais convolucionais profundas. Os autores sintetizaram imagens de um sujeito "virtual"c compostas de partes

de face detectadas automaticamente (olhos, nariz, boca) de outros dois indivíduos a e b da base. Os testes foram feitos com a base de dados LFW.

Xu et al. (XU et al., 2016) apresentaram um método de gerar imagens de faces aproximadamente simétricas. Para produzir as imagens resultantes as originais foram divididas ao meio. A metade direita é espelhada e tratada como a nova metade esquerda. Cada metade é armazenada em um vetor coluna: z_1 é o vetor da direita e z_2 representa a metade da esquerda. O vetor esquerdo armazena a metade esquerda da matriz de trás para frente. O objetivo é que os dois vetores se tornem aproximadamente iguais. Para que esse objetivo seja alcançado, ambos os vetores têm seus valores atualizados por meio do algoritmo gradiente descendente. Quando os valores ótimos para z_1 e z_2 são encontrados a imagem é recomposta unindo os dois vetores. O método é usado como uma etapa de pré-processamento nas imagens de treino e de teste. Quatro bases de dados foram usadas para avaliar o método: Face Recognition Dataset (FERET) (PHILLIPS et al., 2000), The ORL Database of Faces (SAMARIA; HARTER, 1994), Georgia Tech Face Database e Labeled Faces in the Wild (LFW). Para cada base de dados, os autores treinaram um classificador K-Vizinhos mais próximos (KNN) (FIX; JR, 1951) e os melhores resultados atingidos foram respectivamente, 80%, 91.5%, 63.6% e 31.1%.

2.4 Reconhecimento facial com uma amostra

Esta seção aborda trabalhos que se propõem a resolver o problema de reconhecimento facial com uma amostra por classe (*Single Sample Face Recognition*).

Zhuang (Zhuang et al., 2013) propuseram um novo algoritmo de reconhecimento facial para resolver esse problema com base em uma estrutura de classificação em representação esparsa. O algoritmo é robusto ao desalinhamento de imagem e à corrupção de pixels. Para compensar a perda de informação de iluminação, normalmente fornecida por várias imagens de treinamento, é introduzida uma técnica de transferência de iluminação esparsa (SIT – *Sparse Illumination Transfer*). Os algoritmos SIT buscam exemplos adicionais de iluminação de imagens faciais de uma ou mais classes e formam um dicionário de iluminação.O dicionário de iluminação é construído a partir da base de dados YaleB (GEORGHIADES; BELHUMEUR; KRIEGMAN, 2001). Para treinamento e teste do algoritmo de reconhecimento foi usada a base CMU Multi-PIE (GROSS et al., 2010). Nessa base a acurácia atingida chegou a 93.6%.

Outro trabalho combina transferência de aprendizado de CNN e expansão de amostras no espaço de características (Min; Xu; Cui, 2019). O método treina uma rede neural convolucional em uma base de dados com múltiplas amostras e utiliza os parâmetros aprendidos da rede para o problema de uma amostra. As características extraídas da penúltima camada passam pelo processo de expansão de características *K-Class Feature*

Transfer. A rede foi treinada com as imagens da base CASIA-WebFace (YI et al., 2014). O método foi testado com subconjuntos das bases LFW, ORL (SAMARIA; HARTER, 1994) e FERET conseguindo acurácias de 98.8%, 97.7% e 93.16%, respectivamente.

Liu utilizaram uma abordagem baseada em *bag of features* (Fei-Fei; Perona, 2005) para o reconhecimento facial com uma amostra por classe. O método é chamado *KNN collaborative coding based Bag-of-Feature* (MKCC-BoF). São extraídos descritores locais das imagens e o dicionário visual é obtido por meio da *clusterização* com K-Means (LLOYD, 1982). Foi projetado um esquema de codificação colaborativa com KNN de múltiplos estágios para projetar características locais no espaço semântico. Para descrever informação espacial as características codificadas passam por um processo de *pooling* que gera um histograma com as palavras visuais. Após, um classificador SVM linear foi treinado com as características resultantes. Os autores obtiveram acurácia de 94.97% na base AR Face (CVC technical report, 1998).

2.5 Considerações finais

Ao contrário de alguns sistemas encontrados na literatura, muitas aplicações reais enfrentam o problema de poucas amostras. Portanto, é importante adotar estratégias para aumentar o número de amostras e obter bons resultados de classificação. Além disso, os autores que aplicam o aumento de dados não investigam exatamente quais operações têm impacto mais positivo nos resultados. Assim, este trabalho propõe analisar de perto a influência do aumento de dados nos resultados de reconhecimento facial. Para isso, avaliamos o impacto de alguns procedimentos de aumento de dados em um sistema de reconhecimento de faces com extração de características CNN.

Ainda que este trabalho use aumento de dados para tornar maior uma base de dados pequena, o tamanho não é grande o suficiente para treinar uma rede neural convolucional do início. Então, a estratégia é tirar vantagem de modelos pré-treinados utilizando transferência de aprendizado.

Neste trabalho é apresentado um método de reconhecimento facial que usa transferência de aprendizado utilizando o modelo pré-treinado VGG-Face como um extrator de característica. Também foram analisadas que transformações mais contribuem para aumentar os valores de acurácia. Além dos testes com uma base de dados original, experimentos também foram realizados com a base de dados LFW.

3 Materiais e métodos

Neste capítulo são apresentadas as ferramentas utilizadas no desenvolvimento desta pesquisa. São elas: a arquitetura de rede VGG-Face; e os classificadores *Support Vector Machine* (SVM), K-Vizinhos mais Próximos (KNN) e Bagged Trees. Além disso, são dados mais detalhes sobre as bases de imagens de treino e a base de vídeo de teste, bem como as métricas utilizadas na avaliação.

3.1 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNNs) são redes neurais artificiais profundas que podem ser usadas na classificação de imagens (GOODFELLOW; BENGIO; COURVILLE, 2016). As redes convolucionais ingerem e processam imagens como tensores. Tensores são matrizes de números com várias dimensões. As imagens são percebidas como volumes com dimensões correspondendo à largura e à altura da imagem e aos canais do espaço de cores, por exemplo, RGB. Essas redes são um tipo especial de *multilayer perceptron* que possui camadas especiais. São elas: camadas convolucionais, camadas de *pooling*, camadas totalmente conectadas e camadas de não-linearidades. Um esquema genérico de rede neural convolucional é ilustrado na Figura 1.

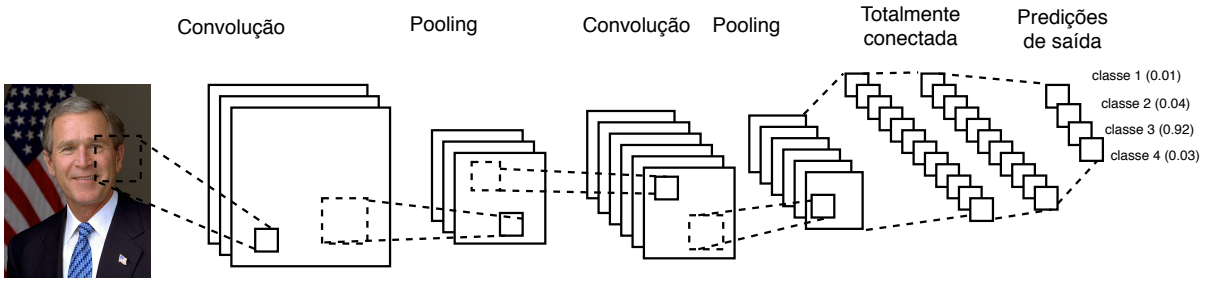


Figura 1 – Modelo genérico de rede neural convolucional. Imagem de face proveniente da base LFW.

3.1.1 Camada convolucional

Esse tipo de camada executa um produto escalar entre duas matrizes, em que uma matriz é o conjunto de parâmetros que podem ser aprendidos, também conhecido como *kernel*, e a outra matriz é a parte restrita do campo receptivo. O kernel é espacialmente menor que uma imagem, mas é mais profundo. Isso significa que, se a imagem é composta por três canais (RGB), a altura e a largura do kernel serão espacialmente pequenas, mas a profundidade se estende aos três canais.

Durante a fase *feedforward*, o kernel desliza pela altura e largura da imagem produzindo a representação da imagem do campo receptivo. Isso produz uma representação bidimensional da imagem conhecida como um mapa de ativação que fornece a resposta do kernel em cada posição espacial da imagem. O tamanho do passo dado pelo kernel é chamado de *stride*.

3.1.2 Camada de pooling

A camada de pooling substitui a saída da rede em determinados locais. Isso ajuda a reduzir o tamanho espacial da representação, o que diminui a quantidade necessária de computação e o número de pesos. A operação de *pooling* é processada em cada fatia da representação individualmente.

Existem várias funções de *pooling*, como a média da vizinhança, a norma L2 da vizinhança e uma média ponderada com base na distância do pixel central. No entanto, o processo mais popular é o *max pooling*, que informa a saída máxima da vizinhança.

3.1.3 Camada totalmente conectada

Neurônios nessa camada têm conectividade total com todos os neurônios na camada anterior e seguinte, como ocorre em uma rede neural tradicional. É por isso que pode ser calculado como de costume por uma multiplicação de matrizes seguida por um *bias offset*. A camada totalmente conectada ajuda a mapear a representação entre a entrada e a saída.

3.1.4 Camada de não-linearidade

Como a convolução é uma operação linear e as imagens não são lineares, as camadas de não linearidade são frequentemente colocadas diretamente após a camada convolucional para introduzir a não linearidade no mapa de ativação. Existem vários tipos de operações não lineares, sendo as mais populares: Sigmoide, Tangente hiperbólica (Tanh) e *Rectified Linear Unit* (ReLU). A sigmoide transforma valores reais para o intervalo [0,1]. Tanh coloca os valores entre o intervalo [-1,1]. Já a ReLU retorna $f(k) = max(0, k)$.

3.2 Arquitetura VGG-Face

PARKHI; VEDALDI; ZISSERMAN treinaram uma arquitetura voltada para o reconhecimento facial. Os autores construíram uma base de faces com mais de 2 milhões de imagens.

A arquitetura é composta por 11 blocos, cada um contendo um operador linear seguido por uma ou mais não-linearidades como *ReLU* e *max-pooling*. Os oito primeiros blocos são considerados convolucionais, pois o operador linear é um banco de filtros lineares

(convolução linear). Os últimos três blocos são chamados completamente conectados (*Fully Connected* - FC); eles são camadas convolucionais, com a diferença de que o tamanho dos filtros corresponde ao tamanho da entrada, de forma que cada filtro detecta característica da imagem inteira. Todas as camadas de convolução são seguidas por uma camada *ReLU*. No entanto, elas não incluem o operador *Local Response Normalization*.

3.3 Classificadores

Um dos objetivos deste trabalho é analisar que classificador é mais indicado para resolver o problema de reconhecimento facial com transferência de aprendizado. Para isso, foram realizados experimentos envolvendo três classificadores. Nesta seção, descrevem-se os princípios básicos dos classificadores *Support Vector Machine* (SVM) (CORTES; VAPNIK, 1995), K-Vizinhos mais Próximos (KNN) (FIX; JR, 1951) e *Bagged Trees* (BREIMAN, 1996) assim como as configurações utilizadas nos experimentos. Todos os classificadores usados são provenientes da ferramenta *scikit-learn*, que compreende uma série de algoritmos de aprendizagem implementados e disponibilizados para a programação com a linguagem Python (PEDREGOSA et al., 2011). Os classificadores utilizados foram selecionados por serem de fácil uso com a ferramenta.

3.3.1 *Support Vector Machine*

Support Vector Machine é um algoritmo de aprendizagem binário não-probabilístico. Nesse algoritmo, cada amostra de dado é representada como um ponto no espaço n-dimensional (em que n é o número de características usadas na representação do problema), com o valor de cada característica sendo o valor de uma determinada coordenada. Então, realiza-se a classificação encontrando o hiperplano que melhor diferencia as duas classes.

Vetores de Suporte são simplesmente as coordenadas da observação individual. O objetivo é maximizar as distâncias entre o ponto de dados mais próximo para decidir o hiperplano adequado. Outro motivo para selecionar o hiperplano com maior margem é a robustez. Ao selecionar um hiperplano com baixa margem, há uma grande chance de classificação errada no futuro.

Nem sempre, os hiperplanos são facilmente encontrados. Para resolver esse problema existem as funções *kernel*. Essas funções tomam um espaço de entrada dimensional pequeno e o transformam em um espaço dimensional grande, ou seja, convertem um problema não separável em um problema separável. É principalmente útil no problema de separação não linear.

Como o SVM é um classificador binário e o problema de reconhecimento facial é um problema de múltiplas classes, é necessário o uso de estratégias que possibilitem a

classificação (ALLWEIN; SCHAPIRE; SINGER, 2000). Essas estratégias são: Um-Contra-Um (*One-vs-One*) e Um-Contra-Todos (*One-vs-All*).

A estratégia Um-Contra-Um cria e treina $K(K-1)/2$ classificadores binários para um problema multiclasse de K classes. Na etapa de teste, aplica-se um esquema de votação: todos os $K(K-1)/2$ classificadores binários treinados predizem a saída para a amostra de teste. A classe predita um maior número de vezes é a classe selecionada pela combinação de classificadores.

A abordagem de um Um-Contra-Todos envolve o treinamento de um único classificador por classe, em que as amostras dessa classe são amostras positivas e todas as outras amostras são tidas como negativas. Essa estratégia requer que os classificadores base produzam como saída um valor real ao invés de apenas o rótulo da classe, uma vez que esses podem levar a ambiguidades, no caso em que várias classes são previstas para uma única amostra.

3.3.2 K-Vizinhos mais Próximos

K-Vizinhos mais Próximos é um classificador não-paramétrico e baseado por instância (FIX; JR, 1951). Trata-se de um algoritmo que objetiva encontrar $h(x)$ sem fazer presunções sobre a distribuição de x e que não aprende um modelo explícito mas sim compara novos exemplos com as instâncias em memória.

O algoritmo funciona essencialmente por meio de uma votação entre as k instâncias mais similares a uma instância nova. A similaridade é calculada por alguma medida de distância entre dois pontos. A medida utilizada na configuração dos experimentos deste trabalho foi a distância Euclidiana, que é definida pela Equação 3.1,

$$d(x, x') = \sqrt{(x_1 - x'_1)^2 + (x_2 - x'_2)^2 + ... + (x_n - x')^2_n}, \tag{3.1}$$

em que x e x' são os pontos entre os quais se deseja calcular a distância.

Dado um inteiro positivo k, uma observação nova x e a medida de similaridade d, o classificador KNN percorre toda a base de dados calculando a distância d entre x e cada observação de treinamento. Cria-se então o conjunto A com os k pontos da base de treino mais próximos da amostra x. Após, é estimada a probabilidade condicional para cada classe, ou seja, a fração de pontos em A com o mesmo rótulo da classe, como apresentado na Equação 3.2,

$$P(y = j|X = x) = \frac{1}{K} \sum_{i \in A} I(y^{(i)} = j). \tag{3.2}$$

Dessa forma, a amostra x é classificada como a classe com maior probabilidade.

3.3.3 *Bagged Trees*

Bagged Trees (*Bagged* vem de *Bootstrap Aggregation*) são um tipo de classificador *ensemble*. Um método *ensemble* é uma técnica que combina as previsões de vários algoritmos de aprendizado de máquina para fazer previsões mais precisas do que um modelo individual.

Bootstrap Aggregation é um procedimento que pode ser usado para reduzir a variação para aqueles algoritmos com alta variância. Um algoritmo que possui alta variância são árvores de decisão (BREIMAN, 1996).

As árvores de decisão são sensíveis aos dados específicos sobre os quais são treinados. Se os dados de treinamento ou a ordem em que são apresentados forem alterados, a árvore de decisão resultante pode ser bem diferente.

O processo se inicia com a criação de conjuntos de treinamento que são produzidos por amostragem aleatória com substituição do conjunto original. Na amostragem com substituição, algumas observações podem ser repetidas em cada novo conjunto de dados de treinamento. Uma árvore de decisão é treinada para cada subconjunto. Para predizer a classe a que uma nova instância pertence, ela é analisada por todas as árvores e a classe que aparecer mais vezes é escolhida como a saída.

Com o método *Bagged Trees* não há preocupação com o sobreajuste dos dados de treinamento. Por essa razão e por eficiência, as árvores de decisão individuais são profundas (poucas amostras de treinamento em cada nó de folhas da árvore) e as árvores não são podadas. Essas árvores terão alta variância e baixa *bias*. Isto é importante para caracterizar os submodelos ao combinar previsões usando o método *bagging*.

3.3.4 Detector Viola-Jones

Para detectar faces foi utilizado o método proposto por Viola e Jones (VIOLA; JONES, 2004). Esse método usa três tipos de características simples reminiscentes das funções de Haar. Os tipos são: características de dois retângulos, de três retângulos e de quatro retângulos. A Figura 2 mostra essas características.

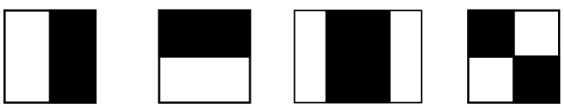


Figura 2 – Exemplos de *haar features*

O cálculo das características é um processo computacionalmente custoso. Para contornar isso, os autores introduziram o conceito de imagem integral, um método fácil de calcular essas características. A imagem integral de uma dada coordenada é a soma dos

pixels acima e à esquerda da coordenada, como especificado em,

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y'), \tag{3.3}$$

onde ii(x, y) é a imagem integral e i(x, y) é a imagem original.

Uma vez que as características tenham sido computadas, é usado um variante do AdaBoost (SCHAPIRE; SINGER, 1999) para selecionar características para treinar o classificador. O método usa uma cascada de classificadores fracos para construir um classificador forte. As imagens que passam por toda a cascata e chegam ao estágio final são classificadas como face.

3.4 Bases de dados

Para a realização dos experimentos e avaliação do método proposto foram utilizadas essencialmente três bases de dados, que serão detalhadas nesta seção. As bases são: *Labeled Faces in the Wild* (LFW) (HUANG et al., 2007) e uma original com doze classes;

3.4.1 Base *Labeled Faces in the Wild*

A base *Labeled Faces in the Wild* foi criada para estimular a pesquisa de reconhecimento facial (identificação e verificação) em ambientes não controlados (HUANG et al., 2007). Um dos objetivos da criação da base é prover protocolos para o uso da base que permita comparações justas. A base contém 13233 imagens de 5749 pessoas coletadas na internet. O número de imagens varia de um indivíduo para outro, mas 1680 pessoas na base têm mais de uma imagem. A LFW foi escolhida para os experimentos por ser bastante utilizada na literatura e permitir a comparação deste trabalho com outros.

Na página da base, e em um estudo feito pelos autores da base, encontra-se um levantamento de resultados obtidos por vários outros autores separados por diferentes protocolos. Os protocolos são divididos entre *image-restricted* e *unrestricted*. O protocolo *image-restricted* está relacionado ao problema de verificação, em que o objetivo é verificar se um par de imagens pertence à mesma classe. Nesse protocolo não é permitido inferir novos pares a partir de pares existentes. Por exemplo, se os pares (x, y) e (w, z) são de imagens de George W. Bush, não se pode utilizar a informação de que (x, z) são imagens da mesma pessoa. Já o protocolo *unrestricted* não tem essa ressalva.

Além disso os trabalhos podem utilizar ou não dados não provenientes de outras fontes, uma vez que imagens da base são encontradas na internet. Para a realização deste trabalho, utilizou-se o detector de faces Viola-Jones para cortar as imagens de modo que somente a região correspondente à face esteja presente na imagem. Caso sejam usados dados rotulados têm-se a categoria *Labeled Outside Data*. Mas há ainda a possibilidade

de usar dados não rotulados, como anotações de localização de *features* ou segmentação. O trabalho proposto se encaixa no protocolo *Unrestricted with Label-Free Outside Data* pois apesar de não usar diretamente imagem rotuladas no treinamento utiliza um modelo de rede pré-treinada com imagens presentes na base LFW (a base de treinamento da arquitetura VGG-Face contém classes da base LFW), o que configura como informações extras de forma indireta. A Tabela 1 mostra um resumo dos protocolos de utilização da base.

Tabela 1 – Resumo dos protocolos na utilização da base LFW

Protocolo	**Permitida Informação de Identidade das imagens da LFW**	**Permitidas Imagens fora da LFW**	**Permitidas imagens rotuladas fora da LFW**
Image-Restricted, No Outside Data	não	não	não
Unrestricted, No Outside Data	sim	não	não
Image-Restricted, Label-Free Outside Data	não	sim	não
Unrestricted, Label-Free Outside Data	sim	sim	não
Unrestricted with Labeled Outside Data	sim	sim	sim

3.4.2 Base *UFPI Faces Database*

A base UFPI Faces Database (UFD) é uma base proprietária que foi construída com a colaboração de doze voluntários. Os voluntários são membros de grupos de pesquisa da Universidade Federal do Piauí (UFPI). A captura de imagens se deu em um mesmo dia, horário e condições de iluminação para todos os voluntários. As imagems foram capturadas em um ambiente interno, um laboratório da universidade.

Oito imagens de cada pessoa foram capturadas. A área da imagem correspondente à face foi cortada utilizando o detector de faces Viola-Jones após uma etapa de detecção de pele baseada em limiarização. Portanto, a base original contém 96 imagens de face. Uma amostra de cada indivíduo na base original de dados pode ser vista na Figura 3.

3.4.3 Base de uma amostra por pessoa

A base original foi proposta levando em conta o problema de reconhecimento de faces de modo geral. No entanto, há um subproblema em que é disponibilizada somente uma amostra por classe (TAN et al., 2006). Métodos de reconhecimento baseados em aparência têm sua performance bastante afetada pela quantidade de amostras disponíveis para treinamento (JAIN; CHANDRASEKARAN, 1982). Para os experimentos com esse problema foi uilizada uma base original em que cada uma das classes tem somente uma imagem de face. As imagens foram escolhidas aleatoriamente dentre as capturadas para construir a base proprietária de imagens.

Figura 3 – Exemplos de imagens dos 12 individuos da base de dados original.

3.5 Métricas de avaliação

Os resultados dos experimentos realizados são apresentados por meio da acurácia obtida. Para que se calcule a acurácia, é necessário saber os valores da matriz de confusão.

A matriz de confusão é criada a partir dos rótulos preditos em comparação com o rótulo real da amostra (FAWCETT, 2006). A matriz é gerada com base em quatro valores:

- Verdadeiro Positivo (VP), são os objetos da classe X que foram classificados corretamente como pertencentes a classe X;
- Verdadeiro Negativo (VN), são os objetos que não pertencem a classe X e foram classificados corretamente como não pertencentes a classe X;
- Falso Positivo (FP), são os objetos que não pertencem a classe X e foram classificados incorretamente como pertencentes a classe X;
- Falso Negativo (FN), são os objetos da classe X que foram classificados incorretamente como não pertencentes a classe X.

A acurácia é a porcentagem de casos corretamente classificados em um conjunto de teste. É definida pela Equação 3.4,

$$Acurácia = \frac{VP + VN}{VP + FP + VN + FN}. \tag{3.4}$$

Para os experimentos feitos com a base LFW, utilizou-se o método de validação cruzada com *k-folds*. Nesse método, todo o conjunto de dados é dividido em k grupos mutuamente exclusivos $(f_1, f_2, ..., f_k)$. Para i de 1 a k, cada f_i é mantido como conjunto de teste e os $k-1$ folds restantes são utilizados para treinar o modelo. A acurácia de cada modelo é calculada por meio da validação dos resultados preditos. Assim, a acurácia final é obtida a partir da média dos *folds*. A média é definida conforme a Equação 3.5,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{3.5}$$

em que n é a quantidade de *folds*.

Outra medida utilizada na análise dos resultados é o erro padrão, que por sua vez é dado em função do desvio padrão. O erro padrão é uma medida que ajuda a verificar a confiabilidade da média amostral calculada. Essa métrica foi selecionada para avaliar os resultados por ser utilizada nos trabalhos que usam o benchmark LFW. Dessa forma é possível a comparação com outros trabalhos utilizando a mesma base.

O desvio padrão é dado pela Equação 3.6,

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})}. \tag{3.6}$$

O erro padrão, por sua é calculado por meio da Equação 3.7,

$$S_E = \frac{\sigma}{\sqrt{n}}. \tag{3.7}$$

4 Método Proposto

Este trabalho propõe operações de aumento de dados para melhorar resultados de classificação em um sistema de reconhecimento facial. Esses procedimentos são realizados para gerar várias bases de dados usadas para treinar os modelos de classificação. A proposta consiste de transformações lineares de aumento de dados para expandir uma base de dados dada como entrada. As manipulações de imagem foram escolhidas levando em conta condições que podem ocorrer em cenas de reconhecimento facial não-controladas. As bases de dados geradas são usadas para treinar três classificadores com características extraídas da arquitetura VGG-Face. Um dos objetivos é avaliar quais operações são mais adequadas para o problema de reconhecimento facial. Além disso procura-se criar um método de reconhecimento facial com bons resultados utilizando técnicas de aumento de dados. Para isso, são apresentadas operações de aumento de dados simples mas eficientes para reconhecimento. Para testar o método, foram realizados experimentos com a base LFW. Além disso, experimentos com a base proprietária de dados também foram feitos.

A Figura 4 mostra um fluxograma contendo todas as etapas do sistema proposto. O método é essencialmente dividido em preparação de dados e reconhecimento facial. A fase de preparação de dados consiste das etapas de aumento de dados, extração de características e treinamento do classificador. A classificação inclui detecção de pele, detecção de face, extração de características e classificação. As seções seguintes dão mais detalhes sobre esses estágios.

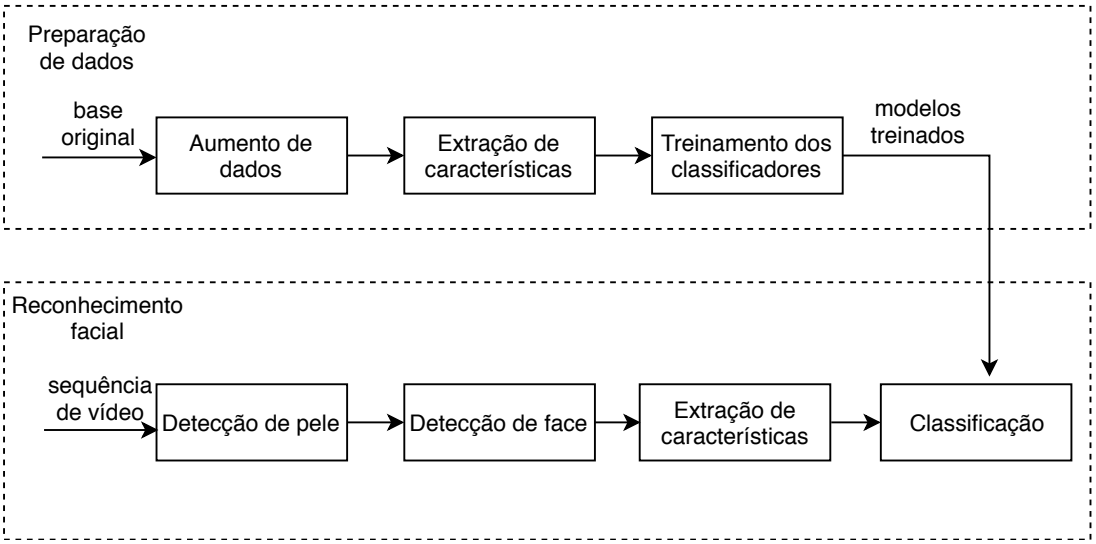


Figura 4 – Visão geral do sistema.

4.1 Preparação de dados

O sistema inicia com o aumento de dados, em que várias bases de dados sintéticas são criadas com a aplicação das operações propostas nas imagens originais. O objetivo é descobrir se o classificador treinado, usando as bases geradas como combinação de transformações simples, produz bons resultados de acurácia. Nesta subseção, as operações usadas para gerar as bases de dados sintéticas são descritas. Além disso, é mostrado como as bases são usadas para treinar os classificadores.

4.1.1 Aumento de dados

Um dos objetivos é avaliar o poder do aumento de dados no problema de reconhecimento facial. Para conseguir isso, foram criadas várias sub-bases de dados com a aplicação de algumas operações de aumento de dados que serão mais detalhadas nessa subseção. Cada uma dessas bases derivadas foi usada separadamente para treinar três classificadores com as saídas obtidas da rede VGG-Face. Os modelos gerados ao final da fase de treino são usados para classificar as imagens detectadas em uma sequência de vídeo. Essa seção explica como gerar bases de dados e utilizá-las para treinar o sistema de reconhecimento.

Busca-se descobrir se um número pequeno de simples transformações é capaz de gerar bases de dados expandidass que produzam bons resultados de classificação de face. Para atingir o objetivo, foram aplicadas várias operações de aumento de dados para expandir a base. Isso provê mais dados para melhoras a fase de treinamento. Ao final dessa fase, têm-se 12 bases de dados expandidas por meio dessas operações (1 original e 11 derivadas). Duas configurações de cada operação foram realizadas nas imagens originais. Nesta subseção, essas operações assim como suas combinações são descritas com detalhes. As operações são Brilho, Contraste e Saturação. Essas transformações foram escolhidas levando em conta situações que podem ocorrer em cenas reais, que é o foco deste trabalho. Outras operações de imagem como borramento e rotação foram testadas em experimentos iniciais e resultados preliminares indicaram que essas operações escolhidas eram suficientes para apresentar bons resultados.

- **Brilho:** Mudanças de iluminação são um processo bastante comum em vídeos em tempo real. Para simular esse processo, operações que mudam o brilho da imagem foram aplicadas. Duas variações de brilho foram usadas para cada imagem original, duas aumentando em 20% e 50% os valores do canal V do espaço de cores HSV. Esse espaço de cores foi escolhido porque tem informação de brilho isolada no canal V. Essas porcentagens foram escolhidas para gerar imagens diferentes das originais, mas de forma que a mudança não fosse exagerada.

- **Contraste:** Diferentes condições de captura resultam em diferentes contrastes na

imagem e isso pode dificultar a tarefa de reconhecimento. Essas mudanças podem ser simuladas ao alterar os valores de contraste. Então, as imagens tiveram os contraste aumentado e diminuído em 20% por meio de operações no espaço de cores HSV. Essa variação de porcentagem foi escolhida para simular diferentes condições mas sem muita perda de informação.

- **Saturação:** Iluminação pobre e más condições do sensor da câmera podem fazer com que a saturação da imagem seja muito alta ou muito baixa. Para simular diferentes condições de saturação e gerar novas imagens, foram aplicadas operações para alterar a saturação das imagens por meio de mudanças no canal S do espaço HSV. A saturação das imagens foi aumentada e diminuída em 20%. Essas configurações foram escolhidas porque valores mais altos causariam mudanças de cor extremas que provavelmente não ocorreriam em situações reais.

4.1.2 Combinando as operações

Foram utilizadas três operações de aumento de dados e é investigado qual o melhor jeito de combinar as imagens resultantes de cada operação com as das outras.

Esta subseção descreve como cada uma das configurações dos experimentos foi elaborada totalizando 12 versões diferentes para a mesma base(a de entrada e 11 com aumento de dados). A primeira configuração é a base original; as outras são todas as combinações possíveis das 3 operações de aumento descritas na Subseção 4.1.1. Portanto, existe uma base como resultados das combinações $\binom{3}{k}$ onde $k = \{1, 2, 3\}$. Isso define que operação comporá uma base aumentada. Cada uma dessas combinações, em que $k \geq 2$, gera duas bases: uma em que as imagens resultantes de uma operação são simplesmente agrupadas em uma mesma base com as resultantes de outras transformações; e outra em que a operação é aplicada nas imagens resultantes de outras operações. No primeiro caso, o número de imagens em uma base cresce polinomialmente enquanto no segundo caso cresce exponencialmente.

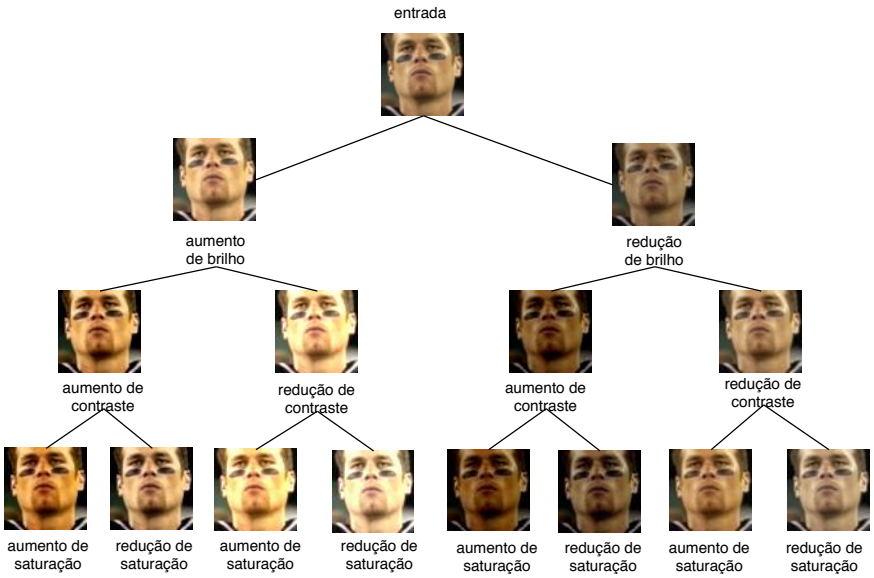
O número de imagens por base de dados criada dado um número de operações de aumento para cada valor k é especificado na Tabela 2. Por exemplo, com $k = 3$ uma das bases de dados expandidas é Brilho, Contraste e Saturação Combinados, o que significa que as operações de brilho foram aplicadas nas imagens originais. Depois disso, as imagens resultantes foram submetidas às operações de Contraste gerando imagens que por sua vez tiveram a saturação modificadas. As imagens geradas no final desse processo (tanto para o caso combinado quanto para o não-combinado) são unidas com as imagens originais de entrada, criando-se uma base de dados contendo $8n$ imagens de face (sendo n o número de amostras na base de entrada). Outra base aumentada para o mesmo valor k e as mesmas operações é composto das imagens resultantes de cada transformação nas imagens originais unidas com estas. Fazendo um total de $7n$ imagens. É importante notar que a ordem

em que as operações são aplicadas não é relevante, já que todas as operações usadas são lineares. A Figura 5 mostra parte das imagens resultantes dos métodos combinado e não-combinado para uma imagem usando as três operações citadas.

Tabela 2 – Número de imagens por base de dados criada dado um número k de operações para o método combinado e não-combinado.

# **Operação (k)**	**Não-combinado**	**Combinado**
1	$3n$	-
2	$5n$	$5n$
3	$7n$	$8n$

(a) Não combinado



(b) Combinado

Figura 5 – Algumas das imagens resultantes dos dois métodos de combinação usando as operações Brilho, Contraste e Saturação.

4.1.3 Extração de características

Utilizar uma rede neural convolucional como extrator de características consiste em extrair apenas a informação de saída de uma camada escolhida como vetor de representação. Neste trabalho, a camada selecionada para representar as imagens é a penúltima camada completamente conectada (camada 34), que apresentou bons resultados como ferramenta extratora de caractetística em outros trabalhos (WAN et al., 2017).

Várias bases de dados foram geradas ao se combinarem as operações apresentadas na subseção anterior. Para cada uma das combinações, foram extraídas características da rede neural convolucional (CNN) VGG-Face (PARKHI; VEDALDI; ZISSERMAN, 2015a). A arquitetura VGG-Face foi escolhida para extrair as características das imagens no sistema porque é uma rede originalmente treinada com imagens de face e que apresentou bons resultados. Seus resultados foram comparáveis ao estado de a arte, exigindo menos dados (quando em comparação com os trabalhos (TAIGMAN et al., 2014; SCHROFF; KALENICHENKO; PHILBIN, 2015)) e usando uma arquitetura de rede mais simples.

A entrada da rede é imagem de tamanho 224×224 com a imagem média da base (calculada a partir do conjunto de treinamento) subtraída. As características são compostas da penúltima camada totalmente conectada. Esse processo gera um vetor de tamanho 4096 como representação de cada uma das imagens de face. O conjunto de vetores é usado para treinar os classificadores.

4.1.4 Treinamento dos classificadores

Cada uma das bases tem as características extraídas como descrito anteriormente. As características são usadas para treinar os três classificadores selecionados. Os três classificadores escolhidos são Máquina de Vetores de Suporte (Support Vector Machine - SVM) Linear (CORTES; VAPNIK, 1995), o K-Vizinhos mais próximos (K-Nearest Neighbor) (FIX; JR, 1951) e *Bagged Trees* (BREIMAN, 1996).

O SVM foi inicialmente proposto para classificação binária. No entanto, o problema multiclasse foi reduzido a um problema binário com a abordagem um-contra-todos (ALLWEIN; SCHAPIRE; SINGER, 2000). O *kernel* utilizado neste trabalho o SVM com o kernel linear. Para o classificador KNN o número de vizinhos considerado é $k = 1$. Na configuração do classificador *Bagged Trees* o número de árvores escolhido para o *ensemble* é 10, por ser o padrão da ferramenta *scikit-learn*, que foi utilizada.

Ao final da fase de treinamento, tem-se três modelos treinados para cada base de dados. Estes modelos são usados posterioremente no estágio de classificação.

4.2 Reconhecimento facial

Neste trabalho, é proposto um método de reconhecimento facial. Aqui são descritos os estágios que compõem a parte de reconhecimento do sistema. O sistema funciona é composto pelas etapas: detecção de pele, detecção de face, extração de características e classificação da face.

Primeiramente, a detecção de face acontece em cada imagem. Para cada face detectada na etapa anterior, são extraídas as características, que são submetidas aos classificadores para determinar a que classe a face pertence. A Figura 6. mostra um fluxograma resumindo esses processos. As subseções seguintes descrevem em detalhes essas etapas.

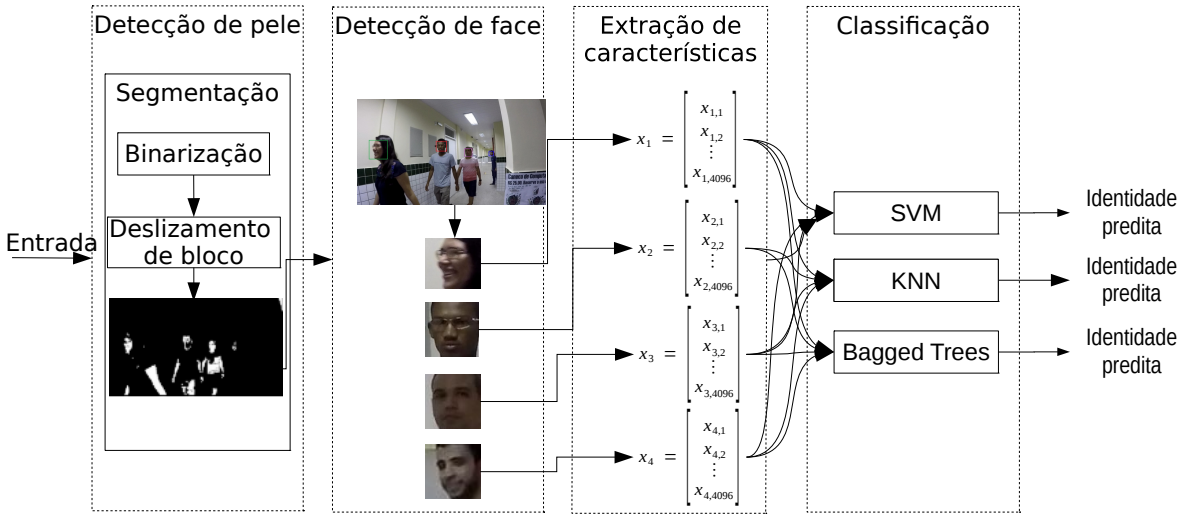


Figura 6 – Fluxograma do Reconhecimento facial.

4.2.1 Detecção de pele

Experimentos anteriores a este trabalho mostraram que o detector de face Viola-Jones (VIOLA; JONES, 2004) apresenta muitos falsos positivos (UCHÔA et al., 2016). Optou-se pela utilização de uma etapa de detecção de pele como uma solução para a redução de falsas detecções de face.

O processo de detecção de pele se dá através de um bloco de $n \times n$ *pixels* que é deslocado por todo o quadro da imagem colorida, na região correspondente ao primeiro plano. Nos experimentos utilizamos $n = 15$. Cada pixel dentro do bloco tem seus valores nas três bandas do espaço de cores RGB analisados a fim de se determinar se o pixel pertence a uma região de pele. O deslocamento da janela tem comportamento semelhante aos filtros morfológicos básicos, no entanto, cada *pixel* é analisado somente uma vez. Essa

análise é feita na forma da Equação 4.1,

$$px = \begin{cases} 1, & \text{se } R > 20, G > 30, B < 100, \\ & R > G, R > B \\ 0, & \text{caso contrário} \end{cases}, \tag{4.1}$$

onde R, G e B são os valores dos *pixels* nas bandas vermelha, verde e azul, respectivamente. O valor na banda R deve ser maior que nas outras duas, pois tons de pele geralmente são cores quentes, onde o vermelho predomina. A imagem Im' resultante desse passo é uma imagem binária que é usada como entrada na avaliação através da janela deslizante. Se mais de cinquenta por cento dos *pixels* de Im' dentro da janela forem iguais a 1, os valores de toda a janela são copiados nos *pixels* correspondentes em Im. Caso contrário, as posições equivalentes em Im recebem 0. Esse último caso visa corrigir os valores de *pixels* possivelmente mal classificados. O processo se resume na Equação 4.2,

$$Im'(x,y) = \begin{cases} 0, \text{se} \sum_{\substack{i=a \\ j=b}}^{\substack{i \leq a+n-1 \\ j \leq b+n-1}} Im(i,j) < 0.50 * n^2 \\ Im(x',y'), \text{ caso contrário,} \end{cases}, \tag{4.2}$$

onde Im representa a imagem de entrada desta etapa, Im' é a imagem resultante, os pares (x, y) e (x', y') são pontos correspondentes nas duas imagens, n a dimensão da janela, e (a, b) as coordenadas do *pixel* superior esquerdo da janela.

A avaliação ocorre desse modo para evitar que pequenos objetos com cores semelhantes aos tons de pele sejam erroneamente classificados como pele. A imagem Im' resultante da fase de detecção de pele é uma imagem binária em que os *pixels* brancos indicam as regiões com presença de pele, onde o detector de face atuará. Em seguida, utilizamos o algoritmo de (VIOLA; JONES, 2004) para detectar as faces na imagem em escala de cinza correspondente à região de pele. Experimentos mostraram que o número de falsos positivos foi reduzido em média em 26%.

4.2.2 Detecção de face

Para detectar faces foi utilizado o método proposto por Viola e Jones (VIOLA; JONES, 2004). O método usa uma cascada de classificadores fracos para construir um classificador forte. As imagens que passam por toda a cascata e chegam ao estágio final são classificadas como face.

4.2.3 Extração de características

Quando a etapa de detecção de face acaba, as imagens de face cortadas são usadas como entrada na rede neural convolucional VGG-Face. As características das imagens

são extraídas da mesma forma que as características para teste conforme descrito na Seção 4.1.3. Portanto, cada face nas imagens corresponde a um vetor de tamanho 4096.

4.2.4 Classificação

Cada vetor de tamanho 4096 correspondente a uma imagem de face é submetido aos três modelos de classificadores treinados: Linear Support Vector Machine, K-Nearest Neighbor and Bagged Trees. Cada classificador emite como saída a identidade predita para uma dada imagem de face. Assim que todas as imagens de face provenientes do vídeo de teste são classificadas, pode-se calcular a acurácia de cada modelo treinado.

4.3 Reconhecimento facial com uma única amostra por pessoa

Um subproblema do reconhecimento facial é o problema de uma amostra por pessoa. Como o nome sugere, nesse tipo de problema, há somente uma amostra por pessoa disponível. Esse cenário é recorrente em problemas reais, por exemplo em sistemas de vigilância em que apenas a foto de documentos oficias está disponível. Pela importância do problema, neste trabalho também se busca um método que obtenha bons resultados em um cenário como este.

Foram realizados experimentos seguindo essa restrição. Para o primeiro caso, há somente uma amostra por pessoa. A base contendo uma imagem por classe passou pelo processo de aumento de dados e teve outras 11 bases derivadas da mesma. As 12 versões dessa base foram utilizadas para treinar os classificadores. À exceção do número de amostra por classe, o procedimento de classificação para esse subproblema é o mesmo relatado nas seções acima.

5 Resultados e Discussão

Esta seção expõe os resultados obtidos com os experimentos realizados para avaliar a eficácia do aumento de dados para reconhecimento facial. O sistema é proposto para a realização de reconhecimento facial em vídeo. Foram feitos experimentos com a base LFW e com a base proprietária UFD.

5.1 Experimentos com a base UFD

Para criar uma base de dados de imagens proprietárias, imagens de doze pessoas foram capturadas com sua colaboração. Oito fotos diferentes de cada pessoa foram tiradas. Portanto, a base de dados proprietária contém originalmente 96 imagens faciais. As transformações de aumento de dados foram aplicadas na base resultando em um total de 12 bases derivadas. Cada uma delas foi usada separadamente como conjunto de treinamento. Para a fase de teste, os classificadores treinados com essas bases foram testados em uma sequência de vídeo como o Capítulo 4 explica. Esses resultados foram obtidos a partir do vídeo que contém as pessoas do conjunto de dados caminhando arbitrariamente. Para cada imagem de face detectada no vídeo, verificamos se o rótulo previsto corresponde ao rótulo real.

A Tabela 3 mostra a precisão dos experimentos com o KNN, SVM e Bagged Trees para os diferentes conjuntos de treinamento aumentados da base UFD. Os valores de acurácia são a porcentagem exata de imagens faciais classificadas corretamente. Os resultados mostram que o melhor valor de acurácia foi alcançado com o classificador KNN para a base em sua versão Brilho, Contraste e Saturação Combinados com a precisão de

Tabela 3 – Acurácia para cada base derivada da base UFD.

Conjunto de dados	**KNN**	**SVM**	**Bagged Trees**
Sem aumento	0.8954	0.8299	0.6893
Brilho e Contraste não combinados	0.9287	0.8616	0.8123
Brilho	0.9290	0.8616	0.7696
Brilho e Saturação não combinados	0.9290	0.8616	0.7692
Brilho, Contraste e Saturação não combinados	0.9302	0.8622	0.8592
Contraste e Saturação Combinados	0.9320	0.8676	0.7640
Brilho e Contraste combinados	0.9326	0.8278	0.7176
Saturação	0.9369	0.8670	0.7180
Contraste e Saturação não combinados	0.9376	0.8560	0.7821
Contraste	0.9376	0.8560	0.7821
Brilho e Saturação combinados	0.9496	0.8594	0.8228
Brilho, Contraste e Saturação combinados	0.9541	0.8596	0.8084

95,41%. Podemos ver que o processo de aumento de dados foi capaz de aumentar a precisão em quase 6 pontos percentuais em comparação com a base sem aumento. De modo geral o classificador KNN apresentou melhores resultados que os outros dois classificadores.

Foram avaliadas quais transformações de aumento de dados têm mais influência no desempenho do sistema usando a base de dados UFD. A Tabela 4 mostra a contribuição de cada transformação de aumento de dados para as bases com resultados melhores que a média. Saturação é a transformação que aparece em 80% dos conjuntos com melhores acurácias. Isso mostra que, para o problema de reconhecimento de face, o uso dessa transformação de aumento de dados provavelmente vai melhorar os resultados. O mesmo aconteceu com o conjunto de dados LFW, já que o conjunto Saturação apresentou o melhor resultado dentre todos.

Tabela 4 – Participação em porcentagem de cada transformação de aumento de dados para as bases derivadas da base UFD para cada classificador.

Operação	**KNN**	**SVM**	**Bagged Trees**
Brilho	40	60	71,4
Contraste	60	60	71,4
Saturação	80	70	57,1

A Tabela 8 mostra a porcentagem do método de combinação de operações de aumento de dados nas bases com as melhores acurácias. Verifica-se que em todos os classificadores para a base proprietária os melhores resultados foram obtidos sem combinar as operações, apenas unindo as imagens de cada operação.

Tabela 5 – Porcentagem de participação de método de combinação nos melhores resultados para as bases derivadas da base proprietária para cada classificador.

Combinação	**KNN**	**SVM**	**Bagged Trees**
Combinados	40	40	28,6
Não-Combinados	60	60	71,4

Na Figura 7 encontra-se um gráfico relacionando o valor médio de acurácia com a quantidade de imagens nos conjuntos. Para o classificador *Bagged Trees* a acurácia média cresce conforme cresce o número de imagens no entanto para o maior número a média decresce. Para o KNN e SVM as médias vão oscilando quando se aumenta o número de imagens, não apresentam comportamento crescente ou descrescente.

5.2 Experimentos com a base LFW

Experimentos foram realizados na base de dados LFW. Essa base foi escolhida para fins de comparação, uma vez que muitos autores na literatura de reconhecimento facial apresentam seus resultados com ela.

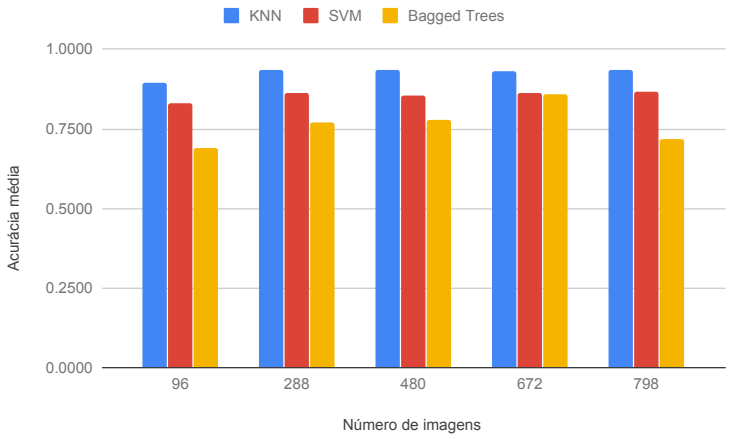


Figura 7 – Acurácia média por número de imagens para a base UFD.

Aplicaram-se as transformações de aumento de dados no LFW, resultando em 12 conjuntos de dados (1 não aumentado e 11 aumentados). Para cada um desses conjuntos de dados, o classificador foi treinado e validado com validação cruzada *k-fold*, onde $k = 10$. A Tabela 6 mostra a acurácia dos 12 conjuntos de dados derivados da LFW para o classificador KNN. Somente o KNN foi utilizado nos testes com essa base pois foi o classificador que apresentou melhores resultados para os experimentos anteriores a este.

As métricas na tabela são a acurácia média obtida da validação cruzada e o erro médio S_E. Os resultados aparecem ordenados classificados de pior para melhor precisão média. O conjunto de LFW sem aumento tem 61.01% de precisão. Podemos ver que o melhor resultado dentre todos os derivados da LFW foi o conjunto de dados aumentado com a operação Saturação com precisão de 98.43%, então essa transformação foi capaz de aumentar a precisão em 37.42 pontos percentuais em comparação com o conjunto não aumentado. Além disso, quase todos os conjuntos aumentados apresentaram melhores resultados que os não aumentados, exceto a versão Brilho.

Um dos propósitos era avaliar quais transformações são mais adequadas para a tarefa de reconhecimento facial. Para este propósito, calculou-se a acurácia média obtida pelas bases de dados derivadas do LFW. Aqueles com acurácia superior à média são considerados os melhores.

A Tabela 7 apresenta em qual porcentagem dos melhores resultados uma transformação de aumento de dados aparece. A tabela mostra que a transformação de Brilho aparece menos que as outras duas nos melhores resultados. Isso sugere que alterar o Brilho de imagens para criar novas amostras não é uma boa escolha para essa base de entrada. O fato de o conjunto de dados de Brilho ter resultados piores do que o LFW sem aumento ajuda a reforçar essa hipótese.

Também foi analisado qual método de combinação fornece melhores resultados. Para avaliar isso, considera-se a porcentagem de vezes que um método aparece entre os

Tabela 6 – Acurácia média de classificação û e erro padrão médio S_E para as bases derivadas da LFW.

Conjunto de dados	**Acurácia média**	S_E
Brilho	0,61000	0,057497
Sem aumento	0,61008	0,574990
Brilho,Contraste e Saturação Combinados	0,93421	0,009054
Brilho e Contraste Combinados	0,93882	0,011564
Contraste e Saturação Combinados	0,96302	0,004807
Contraste	0,96945	0,00524
Brilho,Contraste Não combinados	0,97212	0,002991
Contrast e Saturação Não combinados	0,97494	0,0025
Brilho e Saturação Combinados	0,97542	0,002172
Brilho,Contraste e Saturação Não combinados	0,97629	0,001916
Brilho e Saturação Não combinados	0,97934	0,001769
Saturação	**0,98434**	**0,000836**

Tabela 7 – Participação em porcentagem de cada transformação de aumento de dados nas bases com melhores resultados para a base de entrada LFW

Transformação	Participação nos melhores resultados (%)
Brilho	60
Contraste	70
Saturação	70

melhores resultados (acurácia melhor que a média). Para os experimentos com a LFW, as bases com o método não combinado apresentaram melhores valores de acurácia.

Tabela 8 – Porcentagem do método de combinação de transformações de aumento de dados nos melhores resultados para as bases derivadas da LFW.

Método	**Participação nos melhores resultados(%)**
Combinados	40
Não combinados	60

Analisou-se a acurácia média em relação à quantidade de imagens no conjunto de dados. A Figura 8 ilustra essa análise. A maior média de acurácia não foi obtida com os conjuntos com o maior número de imagens.

A Tabela 9 mostra os resultados de várias técnicas que também foram testadas na base de dados LFW. Aqui são comparadas a abordagem proposta com outros trabalhos que se enquadram no protocolo *Unrestricted, Label-Free Outside Data*. Em outras palavras, não são coletados dados rotulados de outras fontes; usamos apenas imagens da base LFW. Conjuntos de dados marcados com um (*) são sistemas de reconhecimento comercial cujos algoritmos não foram publicados. O melhor resultado da abordagem proposta foi o classificador KNN treinado com o conjunto de dados Saturação derivado da LFW, com acurácia de 98,43 %. Isso significa que a abordagem superou os outros métodos,

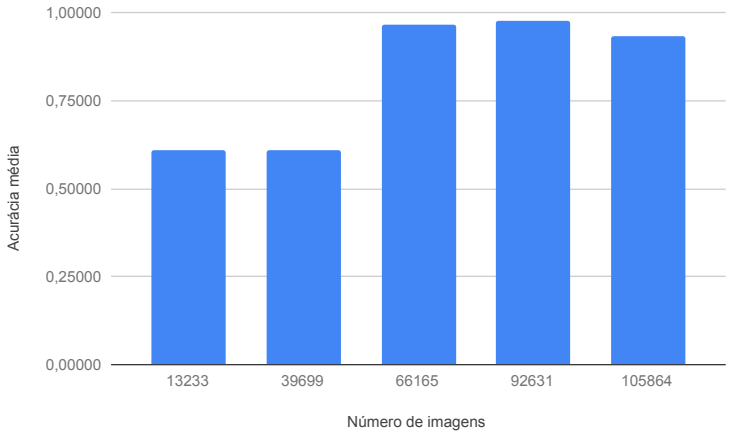


Figura 8 – Acurácia média por número de imagens para a base LFW.

demonstrando que o aumento de dados é eficiente na melhoria dos resultados do problema de reconhecimento facial.

Tabela 9 – Acurácia média û erro padrão da média S_E.

Método	**Acurácia média** $\pm S_E$
LBP multishot (TAIGMAN; WOLF; HASSNER, 2009)	0.8517 ± 0.0061
Attribute classifiers (KUMAR et al., 2009)	0.8525 ± 0.0060
LBP PLDA (PRINCE et al., 2012a)	0.8733 ± 0.0055
LDML-MkNN (GUILLAUMIN; VERBEEK; SCHMID, 2009)	0.8750 ± 0.0040
Combined smultishot (TAIGMAN; WOLF; HASSNER, 2009)	0.8950 ± 0.0051
SLBP* (HUANG; ZHU; YU, 2012)	0.9000 ± 0.0133
Combined PLDA (PRINCE et al., 2012b)	0.9007 ± 0.0051
Sub-SML (CAO; YING; LI, 2013)	0.9075 ± 0.0064
Combined Joint Bayesian (CHEN et al., 2012)	0.9090 ± 0.0148
CMD* (HUANG; ZHU; YU, 2012)	0.9170 ± 0.0110
ConvNet-RBM (SUN; WANG; TANG, 2016)	0.9175 ± 0.0048
VMRS (BARKAN et al., 2013)	0.9205 ± 0.0045
CMD+SLBP* (HUANG; ZHU; YU, 2012)	0.9258 ± 0.0136
VisionLabs ver. 1.0*	0.9290 ± 0.0031
MLBPH+MLPQH+MBSIFH (OUAMANE et al., 2014)	0.9303 ± 0.0082
Fisher vector faces (SIMONYAN et al., 2013)	0.9303 ± 0.0105
High-dim LBP (CHEN et al., 2013)	0.9318 ± 0.0107
Aurora* (SZEPTYCKI et al., 2014)	0.9324 ± 0.0044
HPEN + HD-LBP + JB (ZHU et al., 2015)	0.9487 ± 0.0038
HPEN + HD-Gabor + JB (ZHU et al., 2015)	0.9525 ± 0.0036
MDML-DCPs (DING et al., 2016)	0.9558 ± 0.0034
Método proposto	**0.9843 ± 0.0008**

5.3 Experimentos com uma amostra por classe

Nesta seção são discutidos os resultados obtidos com os experimentos da base com uma amostra por classe.

A Tabela 10 exibe os valores de acurácia resultantes de cada classificador para a base de uma amostra por classe e as bases aumentadas a partir dela. Nela vê-se que o classificador que produziu melhores resultados dentre todos os conjuntos de dados foi o SVM. O classificador *Bagged Trees* teve resultados bem abaixo dos apresentados pelo SVM e KNN, isso sugere que esse classificador é mais sensível à pequena quantidade de imagens utilizadas para treinar, mesmo com as operações de aumento de dados. O melhor resultado obtido dentre todas as combinações é com o conjunto de dados aumentado por Contraste e Saturação combinados com acurácia de 94.5% com o classificador SVM. Para o classificador KNN o conjunto Brilho e Contraste não combinados foi o com maior acurácia, cujo valor é 20% maior que a base sem aumento.

Tabela 10 – Acurácia para cada base derivada da base com uma amostra por classe para cada classificador.

Conjunto de dados	**KNN**	**SVM**	**Bagged Trees**
Brilho e Contraste não combinados	0,8468	0,8744	0,3783
Contraste e Saturação não combinados	0,6970	0,8803	0,3715
Contraste	0,8426	0,8859	0,4832
Saturação	0,8397	0,8930	0,3522
Brilho e Contraste Combinados	0,8349	0,9249	0,4558
Brilho	0,8310	0,9273	0,3041
Brilho e Saturação não combinados	0,8420	0,9282	0,3279
Brilho e Saturação Combinados	0,8466	0,9293	0,3616
Brilho, Contraste e Saturação não combinados	0,8190	0,9339	0,4361
Brilho, Contraste e Saturação Combinados	0,8426	0,9350	0,5817
Sem aumento	0,6673	0,9429	0,3179
Contraste e Saturação combinados	0,6764	**0,9450**	0,3457

Foi calculada a porcentagem de participação de cada operação de aumento nas bases aumentadas a partir da base de entrada com uma amostra por classe. Para os classificadores KNN e SVM a operação que apareceu na maior parte dos conjuntos com acurácia acima da média foi o brilho. A operação contraste apareceu em todos os conjuntos aumentados com melhores resultados para o classificador Bagged Trees.

Observou-se também a porcentagem de participação de cada método de combinação de operações de aumento de dados nas bases cujas acurácias estão acima da média. Essa análise está resumida na Tabela 12. Ao contrário do que foi observado para a LFW e a base UFD os melhores resultados foram obtidos por meio da combinação das operações.

A Figura 9 mostra um gráfico com o valor médio de acurácia de acordo com

Tabela 11 – Porcentagem de participação de operações de aumento nos melhores resultados para as bases derivadas da base com uma amostra por classe.

Operação	**KNN**	**SVM**	**Bagged Trees**
Brilho	77,7	75	75
Contraste	44,4	50	100
Saturação	55,5	37,5	50

Tabela 12 – Porcentagem de participação de método de combinação nos melhores resultados para as bases derivadas da base com uma amostra por classe.

Combinação	**KNN**	**SVM**	**Bagged Trees**
Combinados	75	62,5	62,5
Não-Combinados	25	37,5	37,5

a quantidade de imagens nos conjuntos de dados. Os classificadores SVM e KNN não atingiram as maiores médias com o maior número de imagens. Já o *Bagged Trees* apresentou um comportamento crescente do número de imagens em relação à acurácia média. Isso aliado ao fato de que esse classificador foi o que resultou em valores mais baixos pra a base de entrada com menor número de imagens mostra que é preciso de um grande volume de dados para que ele faça uma boa classificação.

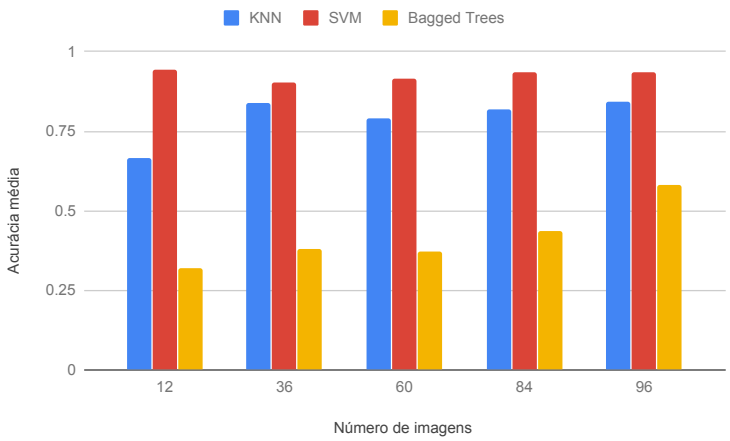


Figura 9 – Acurácia média por número de imagens para a base de uma imagem por classe.

5.4 Considerações Finais

Essa seção apresentou resultados dos experimentos de classificação de face treinados com as características extraídas da CNN VGG-Face. Três classificadores foram testados para os experimentos com a base UFD e a base com uma amostra por classe. Para a primeira os melhores resultados foram obtidos com o KNN enquanto o SVM apresentou maiores acurácias para o segundo.

Procurou-se avaliar qual operação de aumento de dados mais contribui para melhorar o valor de acurácia em relação às bases sem aumento. Para a base LFW e a base UFD a operação saturação fez parte da maioria dos conjuntos cuja acurácia é maior que a média. Para o experimento de uma amostra por classe a operação Brilho mais contribuiu com os melhores resultados para os três classificadores.

Além disso, também avaliou-se qual a melhor forma de combinar as imagens resultantes das operações de aumento de dados. Para a maior parte dos resultados acima da média dos experimentos com a base LFW e a base UFD as operações não foram combinadas, ou seja, as operações foram aplicadas somente nas imagens de entrada. Já para os experimentos com a base de uma amostra por classe combinar as operações, aplicando uma operação sobre o resultado de outra apresentou os melhores resultados.

Foi analisada a relação entre a quantidade de imagens e a acurácia média que os conjuntos com a mesma quantidade de imagens apresentaram. Foi visto que nem sempre um maior número de imagens implica em maior valor de acurácia.

6 Conclusões e continuidade da pesquisa

Neste trabalho, apresenta-se um método de reconhecimento facial com aprendizado de transferência de aprendizado de CNN. Foi apresentado um conjunto efetivo de operações de aumento de dados que abordam os obstáculos conhecidos do reconhecimento de faces, são elas Brilho, Contraste e Saturação.

O método é composto pelas etapas: preparação de dados, em que foram realizadas as operações de aumento de dados, extraíram-se características por meio da arquitetura de rede VGG-Face e essas características foram utilizadas para treinar os classificadores SVM, KNN e *Bagged Trees*.

Foi apresentada uma base de dados proprietária composta de voluntários de grupos de pesquisa da Universidade Federal do Piauí.

Foi demonstrado que simples transformações de imagem são capazes de aumentar a acurácia de classificador treinado com saídas de uma rede neural convolucional, um método mais simples do que realmente treinar uma rede complexa com milhões de parâmetros. E que este método atinge resultados comparáveis ao estado da arte.

Além disso, as transformações foram combinadas de várias maneiras, de modo a avaliar qual é a melhor maneira de usá-las para melhorar os resultados para determinada base. A acurácia dos classificadores treinados com cada base de dados foi testada com imagens de faces cortadas de uma sequência de vídeo de uma cena irrestrita. Também avaliaram-se os efeitos que o número de imagens nas bases de dados tem na precisão.

O conjunto de dados LFW aumentado com saturação foi capaz de melhorar os resultados em relação à base sem aumento em 37.42%, atingindo acurácia de 98.43% com o classificador KNN. Enquanto a maior acurácia encontrada para as bases derivadas da base proprietária UFD foi 95.41% com a versão aumentada com Brilho, Contraste e Saturação combinados também para o KNN. Já os experimentos da base com uma amostra por classe apresentaram melhores resultados com o classificador SVM, chegando a 94.5% de acurácia. No entanto, os resultados com o KNN ficaram bem próximos, indicando que esse classificador é uma boa escolha para o problema de reconhecimento facial com extração de características de CNN.

Os resultados, portanto, sugerem que para cada aplicação e base de dados deve-se avaliar que operações de aumento maximizam os valores de acurácia, já que as que apresentaram melhores resultados variaram de uma base para outra.

Além disso, a base LFW originalmente contém uma maior variedade de iluminação pois é composta de imagens provenientes de vários ambientes, horários e dispositivo de

captura diferentes. Enquanto na base UFD as imagens são mais uniformes por terem sido capturadas no mesmo dia sob as mesmas condições. Acredita-se que isso justifica o fato de que a melhor combinação de operações para a base UFD foi as três operações combinadas, uma vez que essa variação não ocorre naturalmente nessa base e que acontece na LFW.

Os experimentos mostram também que, ao contrário do que o senso comum leva a crer, uma maior quantidade de dados não necessariamente implica em maiores acurácias. Por isso, escolher cuidadosamente que operações usar para aumentar uma base de dados ajuda a garantir melhores resultados de classificação.

6.1 Seleção de características

A seleção de características é um fator muito relevante para decidir a dimensionalidade ideal a ser adotada em um problema de reconhecimento de padrões. Dois fatores que motivam a busca pela redução de dimensionalidade é o custo de predição em termos de tempo e memória e precisão do classificador ao se remover características ruidosas. Por isso há motivos para crer que reduzir a dimensionalidade da representação usada para as imagens acarrete em melhores resultados.

6.2 Arquiteturas de rede

A arquitetura de rede neural convolucional utilizada no método foi a VGG-Face, proposta e originalmente treinada para reconhecimento facial. No entanto, pode ser interessante analisar o uso de outras arquiteturas para a transferência de aprendizado, como a AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012b) e a Inception (SZEGEDY et al., 2015).

Referências

AJANKI, A. et al. An augmented reality interface to contextual information. *Virtual reality*, Springer, v. 15, n. 2-3, p. 161–173, 2011. Citado na página 2.

ALLWEIN, E. L.; SCHAPIRE, R. E.; SINGER, Y. Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of machine learning research*, v. 1, n. Dec, p. 113–141, 2000. Citado 2 vezes nas páginas 14 e 25.

BARKAN, O. et al. Fast high dimensional vector multiplication face recognition. In: *2013 IEEE International Conference on Computer Vision*. [S.l.: s.n.], 2013. p. 1960–1967. ISSN 1550-5499. Citado na página 33.

BARTLETT, M. S. et al. Real time face detection and facial expression recognition: Development and applications to human computer interaction. In: IEEE. *2003 Conference on computer vision and pattern recognition workshop*. [S.l.], 2003. v. 5, p. 53–53. Citado na página 2.

BELHUMEUR, P. N.; HESPANHA, J. P.; KRIEGMAN, D. J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 19, n. 7, p. 711–720, Jul 1997. ISSN 0162-8828. Citado na página 1.

Beveridge, J. R. et al. The challenge of face recognition from digital point-and-shoot cameras. In: *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. [S.l.: s.n.], 2013. p. 1–8. Citado na página 6.

BREIMAN, L. Bagging predictors. *Mach. Learn.*, Kluwer Academic Publishers, Hingham, MA, USA, v. 24, n. 2, p. 123–140, ago. 1996. ISSN 0885-6125. Disponível em: <http://dx.doi.org/10.1023/A:1018054314350>. Citado 3 vezes nas páginas 13, 15 e 25.

CAO, Q.; YING, Y.; LI, P. Similarity metric learning for face recognition. In: *2013 IEEE International Conference on Computer Vision*. [S.l.: s.n.], 2013. p. 2408–2415. ISSN 1550-5499. Citado na página 33.

CHEN, D. et al. *Bayesian Face Revisited: A Joint Formulation*. 2012. 566-579 p. Citado na página 33.

CHEN, D. et al. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In: *2013 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2013. p. 3025–3032. ISSN 1063-6919. Citado na página 33.

CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, v. 20, n. 3, p. 273–297, Sep 1995. ISSN 1573-0565. Disponível em: <https://doi.org/10.1023/A:1022627411411>. Citado 3 vezes nas páginas 8, 13 e 25.

CVC technical report. *AR Face dataset*. 1998. Citado na página 10.

DELAC, K.; GRGIC, M. A survey of biometric recognition methods. In: IEEE. *Proceedings. Elmar-2004. 46th International Symposium on Electronics in Marine*. [S.l.], 2004. p. 184–193. Citado na página 2.

DENG, J. et al. ImageNet: A Large-Scale Hierarchical Image Database. In: *CVPR09*. [S.l.: s.n.], 2009. Citado na página 7.

DING, C. et al. Multi-directional multi-level dual-cross patterns for robust face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 38, n. 3, p. 518–531, March 2016. ISSN 0162-8828. Citado na página 33.

DING, C.; TAO, D. Trunk-branch ensemble convolutional neural networks for video-based face recognition. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 40, n. 4, p. 1002–1014, 2018. Citado na página 5.

FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861 – 874, 2006. ISSN 0167-8655. ROC Analysis in Pattern Recognition. Disponível em: <http://www.sciencedirect.com/science/article/pii/S016786550500303X>. Citado na página 18.

Fei-Fei, L.; Perona, P. A bayesian hierarchical model for learning natural scene categories. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. [S.l.: s.n.], 2005. v. 2, p. 524–531 vol. 2. Citado na página 10.

FIX, E.; JR, J. L. H. *Discriminatory analysis-nonparametric discrimination: consistency properties*. [S.l.], 1951. Citado 4 vezes nas páginas 9, 13, 14 e 25.

GALBALLY, J.; MARCEL, S.; FIERREZ, J. Biometric antispoofing methods: A survey in face recognition. *IEEE Access*, IEEE, v. 2, p. 1530–1552, 2014. Citado na página 2.

GEORGHIADES, A.; BELHUMEUR, P.; KRIEGMAN, D. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, v. 23, n. 6, p. 643–660, 2001. Citado na página 9.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press, 2016. Citado na página 11.

GROSS, R. et al. Multi-pie. *Image and Vision Computing*, Elsevier, v. 28, n. 5, p. 807–813, 2010. Citado na página 9.

GUILLAUMIN, M.; VERBEEK, J.; SCHMID, C. Is that you? metric learning approaches for face identification. In: *2009 IEEE 12th International Conference on Computer Vision*. [S.l.: s.n.], 2009. p. 498–505. ISSN 1550-5499. Citado na página 33.

HERRMANN, C.; WILLERSINN, D.; BEYERER, J. Low-quality video face recognition with deep networks and polygonal chain distance. In: *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. [S.l.: s.n.], 2016. p. 1–7. Citado 2 vezes nas páginas 5 e 8.

HOWARD, A. G. Some improvements on deep convolutional neural network based image classification. *CoRR*, abs/1312.5402, 2013. Disponível em: <http://arxiv.org/abs/1312.5402>. Citado 3 vezes nas páginas 1, 6 e 7.

HU, G. et al. Frankenstein: Learning deep face representations using small data. *arXiv preprint arXiv:1603.06470*, 2016. Citado na página 8.

HUANG, C.; ZHU, S.; YU, K. Large scale strongly supervised ensemble metric learning, with applications to face verification and retrieval. *CoRR*, abs/1212.6094, 2012. Disponível em: <http://arxiv.org/abs/1212.6094>. Citado na página 33.

HUANG, G. B. et al. *Labeled faces in the wild: A database for studying face recognition in unconstrained environments*. [S.l.], 2007. Citado 2 vezes nas páginas 8 e 16.

HUANG, Z. et al. A benchmark and comparative study of video-based face recognition on cox face database. *IEEE Transactions on Image Processing*, v. 24, n. 12, p. 5967–5981, Dec 2015. Citado na página 6.

JAIN, A.; CHANDRASEKARAN, B. 39 dimensionality and sample size considerations in pattern recognition practice. In: *Classification Pattern Recognition and Reduction of Dimensionality*. Elsevier, 1982, (Handbook of Statistics, v. 2). p. 835 – 855. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0169716182020422>. Citado na página 17.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F. et al. (Ed.). *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012. p. 1097–1105. Disponível em: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>. Citado na página 7.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F. et al. (Ed.). *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012. p. 1097–1105. Disponível em: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>. Citado na página 38.

KUMAR, N. et al. Attribute and simile classifiers for face verification. In: *2009 IEEE 12th International Conference on Computer Vision*. [S.l.: s.n.], 2009. p. 365–372. ISSN 1550-5499. Citado na página 33.

LENG, B.; YU, K.; QIN, J. Data augmentation for unbalanced face recognition training sets. *Neurocomputing*, v. 235, n. Supplement C, p. 10 – 14, 2017. ISSN 0925-2312. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0925231216314886>. Citado na página 8.

LLOYD, S. P. Least squares quantization in pcm. *IEEE Trans. Information Theory*, v. 28, p. 129–136, 1982. Citado na página 10.

MASI, I. et al. Do we really need to collect millions of faces for effective face recognition? *CoRR*, abs/1603.07057, 2016. Disponível em: <http://arxiv.org/abs/1603.07057>. Citado na página 8.

Min, R.; Xu, S.; Cui, Z. Single-sample face recognition based on feature expansion. *IEEE Access*, v. 7, p. 45219–45229, 2019. ISSN 2169-3536. Citado na página 9.

ODIL, A.; OBILEN, M. M. A survey on comparison of face recognition algorithms. In: *2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT)*. [S.l.: s.n.], 2014. p. 1–3. Citado na página 1.

OUAMANE, A. et al. Multi scale multi descriptor local binary features and exponential discriminant analysis for robust face authentication. In: *2014 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2014. p. 313–317. ISSN 1522-4880. Citado na página 33.

PARKHI, O. M.; VEDALDI, A.; ZISSERMAN, A. Deep face recognition. In: *BMVC*. [S.l.: s.n.], 2015. v. 1, n. 3, p. 6. Citado 3 vezes nas páginas 1, 5 e 25.

PARKHI, O. M.; VEDALDI, A.; ZISSERMAN, A. Deep face recognition. In: *British Machine Vision Conference*. [S.l.: s.n.], 2015. Citado 2 vezes nas páginas 5 e 12.

PEARSON, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, v. 2, p. 559–572., 1901. Citado na página 7.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado na página 13.

PEREZ, L.; WANG, J. The effectiveness of data augmentation in image classification using deep learning. *CoRR*, abs/1712.04621, 2017. Disponível em: <http://arxiv.org/abs/1712.04621>. Citado na página 6.

PHILLIPS, P. J. et al. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 10, p. 1090–1104, Oct 2000. ISSN 0162-8828. Citado na página 9.

PRINCE, S. et al. Probabilistic models for inference about identity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 34, n. 1, p. 144–157, Jan 2012. ISSN 0162-8828. Citado na página 33.

PRINCE, S. et al. Probabilistic models for inference about identity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 34, n. 1, p. 144–157, Jan 2012. ISSN 0162-8828. Citado na página 33.

RAUTARAY, S. S.; AGRAWAL, A. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial intelligence review*, Springer, v. 43, n. 1, p. 1–54, 2015. Citado na página 2.

SAMARIA, F. S.; HARTER, A. C. Parameterisation of a stochastic model for human face identification. In: *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*. [S.l.: s.n.], 1994. p. 138–142. Citado 2 vezes nas páginas 9 e 10.

SCHAPIRE, R. E.; SINGER, Y. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, v. 37, n. 3, p. 297–336, Dec 1999. ISSN 1573-0565. Citado na página 16.

SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832, 2015. Disponível em: <http://arxiv.org/abs/1503.03832>. Citado na página 25.

SIMONYAN, K. et al. *Fisher Vector Faces in the Wild*. 2013. 8.1-8.11 p. Citado na página 33.

SUN, Y.; WANG, X.; TANG, X. Deep learning face representation from predicting 10,000 classes. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2014. p. 1891–1898. Citado na página 5.

SUN, Y.; WANG, X.; TANG, X. Hybrid deep learning for face verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 38, n. 10, p. 1997–2009, Oct 2016. ISSN 0162-8828. Citado na página 33.

SZEGEDY, C. et al. Going deeper with convolutions. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. Citado na página 38.

SZEPTYCKI, T. H. adn P. et al. *Aurora Face Recognition Technical Report: Evaluation of Algorithm "Aurora-c-2014-1" on Labeled Faces in the Wild*. [S.l.], 2014. Citado na página 33.

TAIGMAN, Y.; WOLF, L.; HASSNER, T. Multiple one-shots for utilizing class label information. In: *The British Machine Vision Conference (BMVC)*. [s.n.], 2009. Disponível em: <https://www.openu.ac.il/home/hassner/projects/multishot>. Citado na página 33.

TAIGMAN, Y. et al. Deepface: Closing the gap to human-level performance in face verification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2014. p. 1701–1708. Citado 2 vezes nas páginas 6 e 25.

TAN, X. et al. Face recognition from a single image per person: A survey. *Pattern Recognition*, v. 39, n. 9, p. 1725 – 1745, 2006. ISSN 0031-3203. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0031320306001270>. Citado na página 17.

TURK, M. A.; PENTLAND, A. P. Face recognition using eigenfaces. In: *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 1991. p. 586–591. ISSN 1063-6919. Citado na página 1.

UCHÔA, V. de S. et al. Estimativa de movimento e detecção de pele aplicadas ao problema de detecção de face em vídeo no espaço de cores rgb. In: *2016 XXI Congresso Brasileiro de Automática (CBA)*. [S.l.: s.n.], 2016. Citado na página 26.

VEL, O. de; AEBERHARD, S. Line-based face recognition under varying pose. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 21, n. 10, p. 1081–1088, out. 1999. ISSN 0162-8828. Disponível em: <http://dx.doi.org/10.1109/34.799912>. Citado na página 1.

VIOLA, P.; JONES, M. J. Robust real-time face detection. *Int. J. Comput. Vision*, Kluwer Academic Publishers, Hingham, MA, USA, v. 57, n. 2, p. 137–154, maio 2004. ISSN 0920-5691. Citado 3 vezes nas páginas 15, 26 e 27.

WAN, L. et al. Face recognition with convolutional neural networks and subspace learning. In: *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*. [S.l.: s.n.], 2017. p. 228–233. Citado na página 25.

WANG, X. Intelligent multi-camera video surveillance: A review. *Pattern recognition letters*, Elsevier, v. 34, n. 1, p. 3–19, 2013. Citado na página 2.

WANG, Y. et al. Face recognition in real-world surveillance videos with deep learning method. In: *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*. [S.l.: s.n.], 2017. p. 239–243. Citado na página 5.

Wheeler, F. W.; Weiss, R. L.; Tu, P. H. Face recognition at a distance system for surveillance applications. In: *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. [S.l.: s.n.], 2010. p. 1–8. Citado na página 2.

Whitelam, C. et al. Iarpa janus benchmark-b face dataset. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. [S.l.: s.n.], 2017. p. 592–600. Citado na página 8.

Wolf, L.; Hassner, T.; Maoz, I. Face recognition in unconstrained videos with matched background similarity. In: *CVPR 2011*. [S.l.: s.n.], 2011. p. 529–534. Citado na página 6.

XU, Y. et al. Approximately symmetrical face images for image preprocessing in face recognition and sparse representation based classification. *Pattern Recogn.*, Elsevier Science Inc., New York, NY, USA, v. 54, n. C, p. 68–82, jun. 2016. ISSN 0031-3203. Disponível em: <http://dx.doi.org/10.1016/j.patcog.2015.12.017>. Citado na página 9.

YI, D. et al. Learning face representation from scratch. *CoRR*, abs/1411.7923, 2014. Disponível em: <http://arxiv.org/abs/1411.7923>. Citado na página 10.

YOSINSKI, J. et al. How transferable are features in deep neural networks? In: *Advances in neural information processing systems*. [S.l.: s.n.], 2014. p. 3320–3328. Citado 2 vezes nas páginas 1 e 6.

ZHANG, X. et al. Finding celebrities in billions of web images. *IEEE Transactions on Multimedia*, v. 14, n. 4, p. 995–1007, Aug 2012. ISSN 1520-9210. Citado na página 8.

ZHAO, W. et al. Face recognition: A literature survey. *ACM Comput. Surv.*, ACM, New York, NY, USA, v. 35, n. 4, p. 399–458, dez. 2003. ISSN 0360-0300. Disponível em: <http://doi.acm.org/10.1145/954339.954342>. Citado na página 1.

ZHU, X. et al. High-fidelity pose and expression normalization for face recognition in the wild. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. p. 787–796. ISSN 1063-6919. Citado na página 33.

Zhuang, L. et al. Single-sample face recognition with image corruption and misalignment via sparse illumination transfer. In: *2013 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2013. p. 3546–3553. ISSN 1063-6919. Citado na página 9.