|  |  |
| --- | --- |
| CURSO DE CIÊNCIA DA COMPUTAÇÃO – TCC | |
| ( ) PRÉ-PROJETO     ( X ) PROJETO | ANO/SEMESTRE: 2022/2 |

Protótipo DE visão computacional para busca DE FOTOS utilizando algoritmos de reconhecimento facial

Jadiel dos Santos

Prof. Aurélio Faustino Hoppe - orientador

# Introdução

Nós, seres humanos, somos excelentes em reconhecer instantaneamente uma grande variedade de objetos de forma automática, mesmo sem perceber. Isso ocorre devido a capacidade de processamento do nosso cérebro em fazer operações complexas, tornando-se possível o reconhecimento de padrões (SANCHEZ-MORENO *et al*., 2021).

Neste sentido, segundo Almeida e Bento (2022), uma das características mais notáveis da visão humana é a capacidade de reconhecer faces. Esta capacidade, desenvolvida durante a infância, está na base das nossas relações sociais e tem desempenhado um papel de grande importância na nossa evolução enquanto espécie (ALMEIDA; BENTO, 2022).

Segundo Faria (2022), o reconhecimento facial é um dos processos de identificação mais utilizados pelos seres humanos, pois permite identificar rapidamente qualquer pessoa e assim definir o tipo apropriado de interação com ela. Para Zhang e Zhang (2010), o reconhecimento facial é uma parte essencial na comunicação entre pessoas e máquinas. Ainda de acordo com os autores, o reconhecimento facial também possibilita o uso de diversos algoritmos de análises faciais em ambientes reais, dos quais se encaixam: modelagem, detecção, reconhecimento e rastreamento de faces.

Para Yang, Kriegman e Ahuja (2002), o reconhecimento facial serve para determinar a existência de um rosto em uma imagem e, se existir, deve também retornar aonde ele está posicionado na imagem. Embora seja fácil para uma pessoa, para uma máquina é uma história totalmente diferente, quando levado em consideração as dificuldades apresentadas nesse processo por conta das diversas condições existentes que variam desde tamanho, posição, grau de rotação, luminosidade, oclusões até expressões da pessoa detectada.

Na literatura sobre reconhecimento facial, encontram-se diversas abordagens quanto a detecção, as quais, em sua maioria, se baseiam nas características do rosto, tais como a distância entre os olhos, comprimento do nariz, ossos laterais e cavidade orbital, entre outros. Já em relação as etapas realizadas para efetuar o reconhecimento, constam (i) a detecção das faces em uma imagem, (ii) extração de suas características, (iii) busca por imagens semelhantes em uma base de dados. Neste sentido, a grande problemática do reconhecimento facial encontra-se justamente em relação a detecção das faces na imagem. Oh *et al*. (2016) listam duas dimensões capazes de afetar a identificação de um indivíduo num sistema de reconhecimento facial: a quantidade de cabeças etiquetadas e o tipo de oclusão do rosto.

De acordo com He *et al.* (2019, p. 1), o reconhecimento facial, como um problema fundamental da área de visão computacional, tem sido bastante pesquisado nos últimos anos. O primeiro sistema semiautomatizado para o reconhecimento da face exigia que o seu administrador localizasse características nas fotografias (por exemplo, olhos, orelhas, nariz e boca) antes do sistema calcular distâncias e características para um ponto de referência comum. Atualmente, a tecnologia de reconhecimento facial está sendo utilizada para o combate à fraude, acesso a contas bancárias ou empresas, para a identificação de criminosos, seleção ou busca de conteúdos, entre muitos outros cenários (ALMEIDA; BENTO, 2022).

Ainda no contexto fotográfico, no qual atualmente a necessidade de organização se torna mais latente, pois os usuários capturam diariamente muitas imagens e, ao mesmo tempo, enfrentam dificuldades para recuperar fotos de uma determinada pessoa ou momento sem ter que percorrer ou visualizar todas a imagens. Isso ocorre, segundo Almeida e Bento (2022), porque ao serem transferidas ou armazenadas, as imagens apenas são identificadas por um código de números e/ou letras como título, não possuindo uma identificação pelo seu conteúdo.

Diante deste cenário, este trabalho levanta as seguintes questões de pesquisa: (i) algoritmos de reconhecimento facial, técnicas de visão computacional e aprendizado de máquina podem viabilizar um protótipo que emula a percepção visual humana em detectar e reconhecer uma mesma face em várias fotos? (ii) em qual circunstâncias/cenários isso se torna possível? (iii) fotógrafos e usuários comuns podem usufruir deste recurso para realizar a separação das fotos por pessoas que constam nas imagens capturadas?

## OBJETIVOS

O objetivo deste trabalho é disponibilizar um protótipo de visão computacional que utilize algoritmos de reconhecimento facial para localizar/agrupar fotos a partir de uma imagem de entrada.

Os objetivos específicos são:

1. investigar as técnicas de aprendizado de máquina mais utilizadas na detecção e reconhecimento de faces;
2. analisar a qualidade mínima das imagens para realizar a detecção e o reconhecimento facial;
3. disponibilizar uma aplicação *web* que facilite o gerenciamento e busca das imagens/fotos;
4. analisar e comparar a precisão dos algoritmos utilizados na seleção e agrupamento das imagens por indivíduo.

# trabalhos correlatos

Nesta seção são apresentados trabalhos com características semelhantes aos principais objetivos do estudo proposto. A seção 2.1 descreve o processo de detecção de faces por meio das regiões e atributos faciais: cabelo, olho, nariz, boca e barba (YANG *et al*., 2018b). A seção 2.2 discorre sobre o método de reconhecimento de faces implementado através de uma rede de Face Deformável (HE *et al*., 2019). Por fim, a seção 2.3 apresenta o método para reconhecimento facial em vídeos e fotos em tempo real (SANCHEZ-MORENO *et al.*, 2021).

## Faceness-Net: Face Detection through Deep Facial Part Responses

Yang *et al*. (2018b) propuseram um método para melhorar o processo de detecção facial. Segundo os autores, embora as *bounding boxes* de face sejam utilizadas como padrão de entrada em detectores de faces, a utilidade desses atributos ainda é pouco explorada. Neste sentido, eles apresentam o conceito de que a supervisão de atributos faciais pode aumentar a capacidade de uma rede neural artificial em lidar com oclusões e posições diversas (YANG *et al*., 2018b).

Segundo Yang *et al*. (2018b), utilizando-se do conceito de supervisão dos atributos faciais, criou-se a Faceness-Net que possui dois estágios: (i) a rede neural convolucional (CNN) recebe uma imagem como *input* e gera um *partness map[[1]](#footnote-1),* indicando 5 partes da face (cabelos, olhos, nariz, boca e barba). A partir disso, se estabelece o ranqueamento de algumas janelas candidatas utilizando pontos faciais; (ii) as janelas candidatas são refinadas objetivando otimizar as classificações de face e as *bounding box regression* (YANG *et al.,* 2018b). A Figura 1 demonstra os principais elementos da arquitetura Faceness-Net.

Figura – Arquitetura da Faceness-Net

|  |
| --- |
|  |

Fonte: Yang *et al*. (2018b).

Yang *et al*. (2018b) também apontam que o primeiro estágio da linha base da Faceness-Net consiste de múltiplas redes de reconhecimento de atributos para gerar mapas de resposta de diferentes partes da imagem. Para isso são utilizadas 5 redes, uma para cada componente da face, as quais compartilham da mesma estrutura com 7 camadas de convolução e 2 de *max-pooling*. Ao fim da sétima convolução é gerado o *response map*, no qual aplica-se a normalização L2 e média ao longo dos canais, indicando a localização do componente facial. Posteriormente, realiza-se o *unpooling* para que seja obtido o *partness map* que será do mesmo tamanho da imagem inicial. Dessa forma, soma-se os *partness map* para que sejam obtidas janelas da *face proposals* (YANG *et al*., 2018b).

Segundo Yang *et al*. (2018b), a segunda etapa do Faceness-Net refina a janela de candidatos gerada a partir do primeiro estágio usando uma CNN multitarefa, onde a classificação de face e a *bounding box regression* são otimizadas em conjunto. Esta CNN multitarefa é pré-treinada nas camadas convolucionais de 1 a 4 do primeiro estágio. Na quarta camada, a CNN é dividida em 2 ramos, os quais consistem em 2 camadas convolucionais. Um destes ramos irá executar a classificação de face, e o outro fará a *bounding box regression* (YANG *et al*., 2018b).

Nos experimentos realizados, o método obteve bons resultados nos *datasets* *Annotated Faces in the Wild* (AFW) (ZHU; RAMANAN, 2012), *Pattern Analysis, Statistical Modelling and Computational Learning* (PASCAL) (NADA *et al.*, 2018), e WIDER FACE (YANG *et al.*, 2016a). Sua acurácia no subconjunto de nível fácil foi de, respectivamente, 98,05%, 92,11% e 71,7% (YANG *et al.*, 2018b).

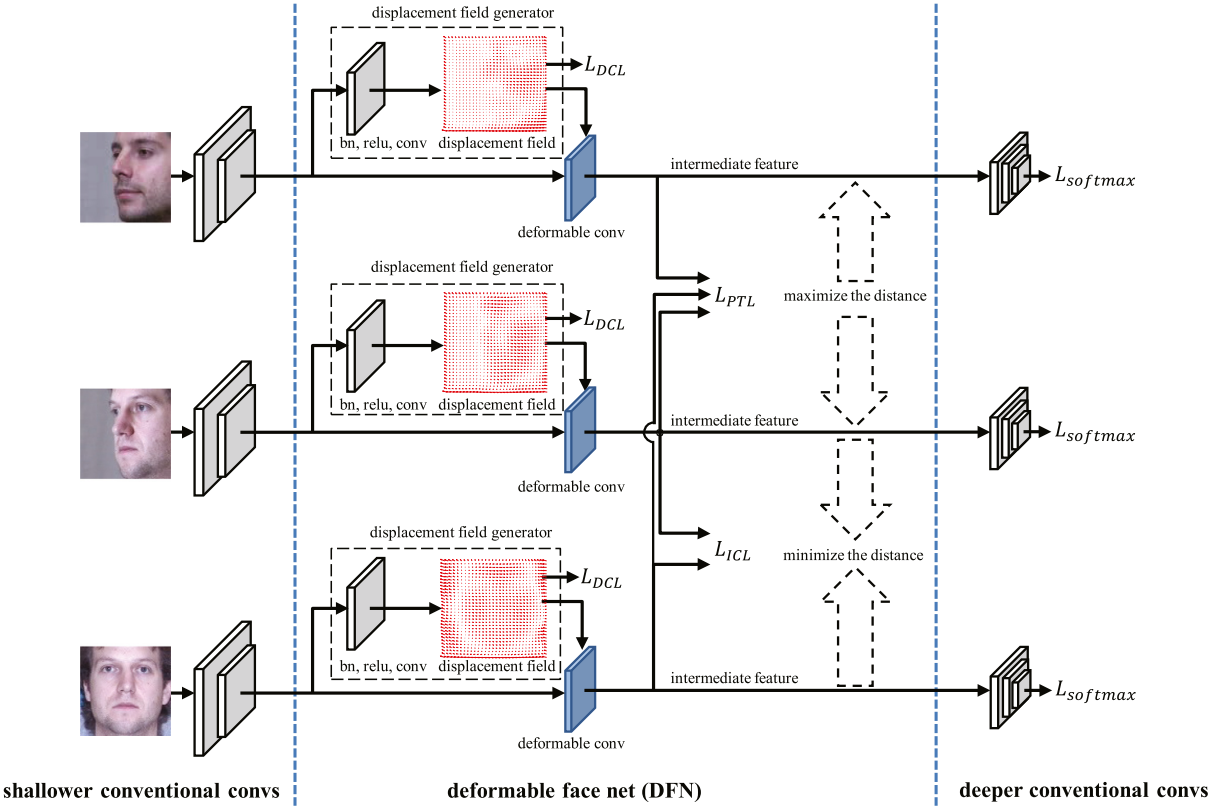
De acordo com Yang *et al*. (2018b), estas redes de reconhecimento de atributos podem compartilhar camadas de convolução iniciais, assim reduzindo em 83% os parâmetros, bem como as computações, além de melhorar a robustez para a detecção facial. Por fim, a eficácia do método também pode ser atingida a partir de atributos de imagens de faces não cortadas, sem supervisão explícita das partes.” (YANG *et al*., 2018b, p. 12, tradução nossa). Além disso, eles também destacam que obtiveram sucesso ao introduzir o conceito de *faceness score*, o qual pode ser usado para ranquear janelas candidatas em qualquer técnica de proposta de região, podendo gerar conjuntos de propostas faciais com alta taxa de recall (YANG *et al*., 2018b).

## Deformable face net for pose invariant face recognition

Segundo He *et al*. (2019), a acurácia do reconhecimento facial tem aumentado rapidamente com o uso de Rede Neural Convolucional (CNN). Entretanto, os autores ressaltam que sob circunstâncias não controladas, como poses, iluminação, resolução, oclusão e expressão, a performance do sistema de reconhecimento facial é afetada significativamente. Por conta disso, He *et al*. (2019) propuseram um método de alinhamento em nível de *feature* para lidar com variações de pose no reconhecimento de rosto.

He *et al*. (2019) ressaltam que o método inicialmente aplica uma *Deformable* Face Net (DFN), a qual busca aprender simultaneamente o alinhamento em nível de *feature* e a extração de *features* para reconhecimento de rosto por meio de convoluções deformáveis com um campo de deslocamento espacial. Dessa maneira, estes campos permitem que a DFN seja capaz de alinhar as *features*, independente de variações de pose. A Figura 2 apresenta a arquitetura da DFN.

Figura – Arquitetura da *Deformable* Face Net



Fonte: He *et al*. (2019).

A partir da Figura 2, pode-se observar que a DFN é composta de uma camada de convolução inicial com função de ativação *ReLU* e com a função de perda *Displacement Consistency Loss* (DCL) para que seja gerado um *displacement field*. Em seguida, passa por uma *deformable convolution*, a qual gera uma *feature* intermediária otimizada pelas funções de perda *Identity Consistency Loss* (ICL) e *Pose-Triplet Loss* (PTL). Por fim, as camadas mais profundas de convolução utilizam a função de perda *Softmax*.

Com o propósito de que os campos de deslocamento espacial sejam consistentes em sua orientação e amplitude, a DCL é utilizada para que obter um alinhamento melhor para reconhecer a face (HE *et al*., 2019). Neste sentido, as funções ICL e PTL, servem para minimizar a variação de *features* intraclasse causada por poses diferentes e maximizar a distância de *recursos* entre classes sob poses semelhantes. Dessa forma, He *et al*. (2019) afirmam que a DFN supera os métodos de última geração nos seguintes *datasets* de testes: MegaFace (KEMELMACHER-SHLIZERMAN *et al*., 2016), MultiPIE (GROSS *et al*., 2008) e CFP (SENGUPTA *et al*., 2016). No MegaFace, o método obteve 82,11% de precisão. No MultiPIE, obteve-se nas poses de ± 90◦ uma acurácia de 85,66%. Por fim, o *dataset* CFP, obteve-se uma acurácia de 94,01%. Com isso, He *et al*. (2019) afirmam que o método aumenta significativamente o reconhecimento de rosto em todas as poses.

## Efficient Face Recognition System for Operating in Unconstrained Environments

Sanchez-Moreno *et al*. (2021) desenvolveram um de sistema reconhecimento facial para ambientes irrestritos. Eles utilizaram o YOLO-Face modificado com um esquema de classificação que permite aumentar a acurácia da detecção com redução do tempo computacional. Os autores também propuseram uma arquitetura que combina uma CNN com um algoritmo de aprendizagem supervisionada, como por exemplo, o *Support Vector Machine* (SVM).

Segundo Sanchez-Moreno *et al*. (2021), subdividiu-se o algoritmo em 4 estágios: no primeiro estágio, realiza-se a detecção da face por meio de um *input* de *frame* de vídeo ou imagem. Para isso, eles utilizaram o YOLO-Face, ao qual detecta as faces em tempo real. Desse modo, os autores resolveram o problema de velocidade. A Figura 3 apresenta os estágios do algoritmo de reconhecimento facial.

Figura – Estágios do algoritmo de reconhecimento facial

|  |
| --- |
|  |

Fonte: Sanchez-Moreno *et al*. (2021).

No segundo estágio, realiza-se o redimensionamento, normalização L2 e conversão da imagem para escala de cinza, tendo como intuito a preparação para aplicar o *machine learning model* (SANCHEZ-MORENO *et al*. 2021, p. 4-7). No terceiro estágio ocorre a extração das *features* através dos métodos de classificação como Support Vector Machine (SVM), Random Forest (RF) e K-Nearest Neighbors (KNN). Por fim, no último estágio, as *features* faciais são comparadas com as *features* conhecidas de usuários cadastrados para identificação pessoal.

Para fins de avaliação do reconhecimento facial, Sanchez-Moreno *et al*. (2021) utilizaram os seguintes *datasets*: *Labeled Faces in the Wild* (LFW) (HUANG *et al*., 2007) e *YouTube Faces* (YTF) (WOLF; HASSNER; MAOZ, 2011). De acordo com os autores, no LFW, o método obteve uma acurácia de 99,1%. Já no YTF, a acurácia foi de 94,7%.

A partir desses resultados, Sanchez-Moreno *et al*. (2021) concluem que o sistema proposto melhora a precisão no banco de dados LFW, obtendo pontuação semelhante ou superior a diferentes estruturas de CNN utilizadas em reconhecimentos faciais. Assim, no que diz respeito ao reconhecimento facial, os autores resumiram que “[...] com os avanços dos algoritmos de deep learning aliados aos avanços tecnológicos, é possível desenvolver sistemas de reconhecimento facial em tempo real capazes de operar em ambientes de alta complexidade.” (SANCHEZ-MORENO *et al*., 2021).

# proposta

Esta seção apresenta a justificativa para o desenvolvimento desse trabalho, os principais requisitos e a metodologia de desenvolvimento que será utilizada. Também são relacionados os assuntos e as fontes bibliográficas que irão fundamentar o estudo proposto.

## JUSTIFICATIVA

No Quadro 1 é apresentado o comparativo entre os trabalhos correlatos. As linhas representam as características e as colunas os trabalhos.

Quadro – Comparativo dos trabalhos correlatos

|  |  |  |  |
| --- | --- | --- | --- |
| Trabalhos correlatos  Características | Yang *et al*.  (2018b) | He *et al*.  (2019) | Sanchez-Moreno *et al*. (2021) |
| Detecção de faces | Sim | Sim | Sim |
| Reconhecimento de faces | Não | Não | Sim |
| Uso das regiões faciais para aprendizado | Sim | Não | Não |
| Lida com ambientes irrestritos (oclusões, iluminação, dimensionamento e poses) | Sim | Sim | Sim |
| Arquitetura de rede neural | Faceness-Net | DFN | YOLO-faces |
| Datasets utilizados | AFW, PASCAL, FDDB e WIDER FACE | MegaFace, MultiPIE e CFP | LFW e YTF |
| Taxa de acurácia para os datasets | 98,05%, 92,11% e 71,7% | 82,11%, 85,66% e 94,01%. | 99,1% e 94,7% |

Fonte: elaborado pelo autor.

A partir do Quadro 1, pode-se observar que todos os estudos utilizam de redes neurais convolucionais (CNN) como método de aprendizado de máquina. No que diz respeito a detecção de faces, todos os estudos correlatos implementam diferenças métodos. Yang *et al*. (2018b) utilizaram atributos faciais para classificar e detectar faces. He *et al*. (2019) optaram pela rede *Deformable Face Net* (DFN), a qual simultaneamente alinha e extrai as *features*. Já Sanchez-Moreno *et al*. (2021) utilizaram o método YOLO-faces, que calcula a probabilidade em cada região, destacando-as através de *bounding boxes*.

No que diz respeito ao reconhecimento facial, nota-se que apenas Sanchez-Moreno *et al*. (2021), implementam tal funcionalidade. Já Yang *et al*. (2018b), com seu método Faceness-Net, utilizam regiões faciais, ou seja, cabelo, olhos, nariz, boca e barba para realizar a detecção facial. Já em relação aos ambientes irrestritos, isto é, levar em conta as oclusões, poses variadas, dimensões da imagem e iluminação, todos os estudos implementaram solução as quais conseguem obter bons resultados nestas condições adversas. Além disso, pode-se observar que a acurácia obtida pelo método de Yang *et al*. (2018b) nos *datasets* AFW, PASCAL, FDDB e WIDER FACE foi de 98,05%, 92,11% e 71,7%. Já as acurácias atingidas pelo método de He *et al*. (2019) nos *datasets* MegaFace, MultiPIE e CFP foram de 82,11%, 85,66% e 94,01%. Por fim, o método de Sanchez-Moreno *et al*. (2021) obteve acurácia de 99,1% e 94,7% nos *datasets* LFW e YTF.

A partir deste contexto, este trabalho implementará um protótipo de detecção e reconhecimento facial para que, dado um conjunto de imagens, realize a pré-seleção/localização/agrupamento automática das fotos, por indivíduo, a partir de uma imagem de entrada. O protótipo a ser desenvolvimento poderá ser utilizado por qualquer usuário, mas, neste trabalho, focará no auxílio a fotógrafos, aos quais após a realização de eventos como formaturas, casamentos e festas em geral, precisam realizar o processo de verificação de todas as imagens retiradas para encontrar e agrupar as fotos de acordo com quem aparece nas imagens. Ressalta-se que tal processo demanda muito tempo para ser feito manualmente, pois em um evento são tiradas milhares de fotos. Além disso, também se destaca a necessidade de estudos para definir qual é a qualidade mínima para o processamento e reconhecimento ocorrer com melhor precisão.

No cenário de fotografias profissionais, há alguns anos, o processo de envio das fotos tiradas pelo fotógrafo para que o cliente fizesse a seleção das mesmas era feito através de CDs ou DVDs, sendo posteriormente substituídos por *pendrives*. Deste modo, todos os fotógrafos que querem oferecer um serviço profissional e de qualidade para seu cliente, buscam formas de melhorar todos os processos que envolvem seu trabalho como um todo, e a tecnologia, técnicas de processamentos e aprendizado e máquina podem ajudá-los. Assim, este trabalho têm como essência a otimização do processo de seleção das fotografias, ao qual seus clientes apareçam nas imagens. Para isso, serão investigados e utilizados algoritmos de reconhecimento facial e redes neurais convolucionais que se adequem melhor aos ambientes aos quais as fotos são retiradas. Além disso, para facilitar a realização dos testes será desenvolvido uma aplicação *web* no intuito de facilitar o gerenciamento das fotos.

## REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

Esta seção mostrará os Requisitos Funcionais (RF) e Requisitos Não Funcionais (RNF) para a implementação do protótipo de seleção de fotos a partir de faces. Os requisitos do trabalho estão divididos em duas partes: requisitos da aplicação web e requisitos do módulo de reconhecimento.

A aplicação *web* deverá:

1. manter o cadastro de usuários (RF);
2. permitir o login de usuários já cadastrados (RF);
3. permitir ao usuário a criação de uma nova galeria de fotos (RF);
4. permitir ao usuário visualizar as fotos pré-selecionadas (RF);
5. utilizar o ambiente de desenvolvimento Visual Studio Code (RNF);
6. desenvolver a aplicação web utilizando NodeJS no backend e React no frontend (RNF);
7. utilizar o banco de dados PostgreSQL para persistir os dados (RNF).

O módulo de reconhecimento deverá:

1. utilizar técnicas de processamento de imagens para melhorar a qualidade das imagens obtidas (filtragem de ruído e normalização da iluminação) (RF);
2. realizar a normalização e a extração de características para estabelecer a qualidade mínima necessária da imagem para realizar a detecção e o reconhecimento facial (RF);
3. utilizar algoritmos de aprendizado de máquina para efetuar a detecção das faces e consecutivamente

a seleção das fotos a partir da face dos indivíduos (RF);

1. utilizar métricas de similaridade para comparar as faces de entrada com os presentes nas fotos (RF);
2. utilizar métricas de classificação: Principal Component Analysis (PCA), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN) ou identificar outro classificador que produza melhor resultado (RF);
3. ser desenvolvida na linguagem de programação Python (RNF);
4. utilizar as bibliotecas OpenCV, Keras e scikit-learn (RNF).

## METODOLOGIA

O trabalho será desenvolvido observando as seguintes etapas:

1. levantamento bibliográfico: pesquisar sobre reconhecimento facial, rede neural convolucional e trabalhos correlatos;
2. elicitação dos requisitos da aplicação web: baseando-se nas informações da etapa anterior, reavaliar os requisitos propostos para a aplicação, tanto funcionais como não funcionais;
3. especificação: utilizar a ferramenta de diagramação Enterprise Architect (EA) para elaborar os diagramas de caso de uso e de atividades de acordo com a Unified Modeling Language (UML);
4. implementação: a partir do item (c) implementar a aplicação web utilizando o NodeJS no *backend* e React no *frontend*;
5. testes da aplicação web: elaborar testes para validar a usabilidade da aplicação web;
6. coleta e montagem da base de dados: coletar fotos/vídeos ou utilizar bases de dados com faces obtidas/cedidas junto a projetos já existentes na comunidade acadêmica;
7. preparação das imagens: realizar o realce e melhoramento das imagens utilizando o OpenCV. Posteriormente subdividir as imagens em dois grupos: imagens de treinamento e de validação.
8. definição da arquitetura da rede neural convolucional: definir a arquitetura da rede neural convolucional mais aderente ao processo de reconhecimento facial;
9. implementação do reconhecimento facial: a partir do item (h) realizar a implementação do módulo de reconhecimento facial utilizando OpenCV para o processamento das imagens e a biblioteca Keras para aprendizado de máquina;
10. testes do reconhecimento facial: paralelamente à implementação do item (i), realizar os testes do módulo de reconhecimento facial, com base em um banco de imagens de familiares, amigos e eventos fotográficos.

As etapas serão realizadas nos períodos relacionados no Quadro 2.

Quadro – Cronograma de atividades a serem realizadas

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2023 | | | | | | | | | | | |
|  | jan. | | fev. | | mar. | | abr. | | maio | | jun. | |
| etapas / quinzenas | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| levantamento bibliográfico |  |  |  |  |  |  |  |  |  |  |  |  |
| elicitação de requisitos da aplicação web |  |  |  |  |  |  |  |  |  |  |  |  |
| especificação |  |  |  |  |  |  |  |  |  |  |  |  |
| implementação |  |  |  |  |  |  |  |  |  |  |  |  |
| testes da aplicação web |  |  |  |  |  |  |  |  |  |  |  |  |
| coleta e montagem da base de dados |  |  |  |  |  |  |  |  |  |  |  |  |
| preparação das imagens |  |  |  |  |  |  |  |  |  |  |  |  |
| definição da arquitetura da rede neural convolucional |  |  |  |  |  |  |  |  |  |  |  |  |
| implementação do reconhecimento facial |  |  |  |  |  |  |  |  |  |  |  |  |
| testes do reconhecimento facial |  |  |  |  |  |  |  |  |  |  |  |  |

Fonte: elaborado pelo autor.

# REVISÃO BIBLIOGRÁFICA

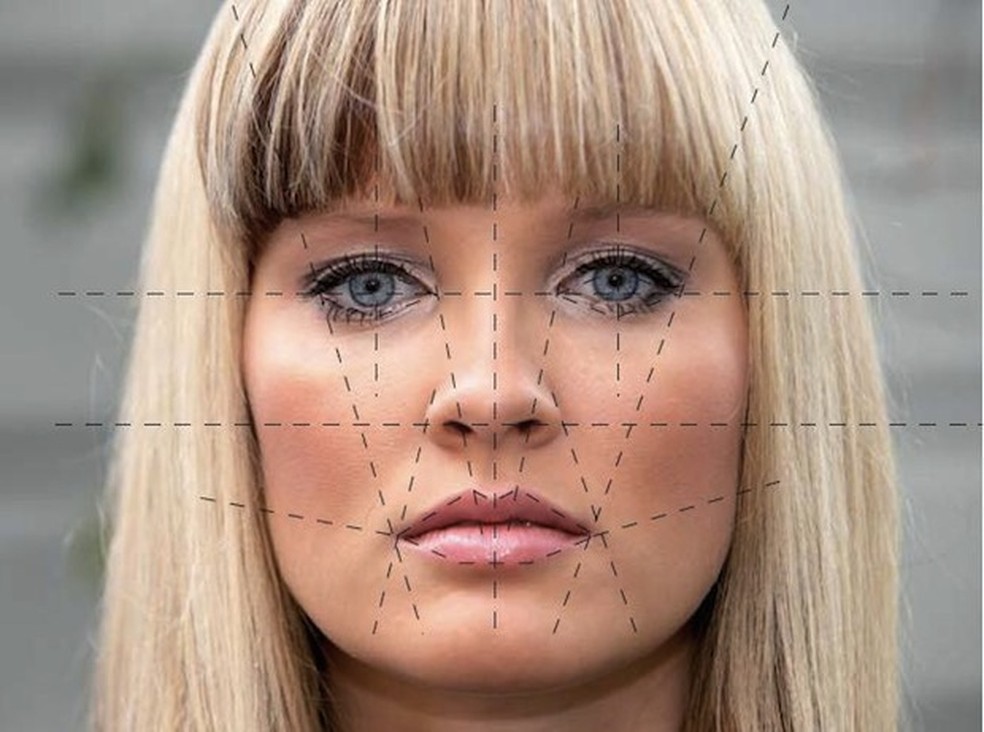
Esta seção tem como objetivo explorar os principais assuntos que fundamentarão o estudo a ser realizado. A seção 4.1 aborda reconhecimento facial. Por fim, a seção 4.2 discorre sobre rede neural convolucional.

## Reconhecimento facial

O reconhecimento facial é uma técnica de biometria que segundo Gogoni (2018) “[...] consiste na aplicação de métricas a atributos biológicos, para fins de aferição e identificação de um indivíduo.”, baseada nos traços do rosto das pessoas. É sabido que os seres humanos têm uma facilidade para reconhecer os rostos das pessoas com precisão. Mesmo que duas pessoas sejam parecidas, nós somos capazes de distingui-las. Essa técnica pode ser transmitida para os computadores, mas, para isso, precisa-se definir traços únicos que devem ser mapeados em códigos binários (PISA, 2022).

Segundo Pisa (2022), o rosto de uma pessoa é formado por diversas características, que são chamados pontos nodais. Segundo o autor, existem cerca de 80 pontos nodais na face humana. Alguns exemplos de pontos nodais são a distância entre os olhos, o comprimento do nariz, o tamanho do queixo e a linha da mandíbula. Neste cenário, como exemplificado na Figura 4, a medição e persistência destes pontos nodais formará a assinatura facial, completando assim a fase de extração de características.

Figura – Marcação de características da face humana



Fonte: Pisa (2022).

Pisa (2022) destaca que existem inúmeros algoritmos que permitem o reconhecimento facial. Em todos torna-se necessário seguir, pelo menos, três passos: (i) a detecção facial, (ii) a extração das características (*features*) e, então, (iii) a identificação dos rostos, ou seja, o reconhecimento facial. Para cada um dos passos, existem várias técnicas disponíveis. O autor ressalta que o grande desafio está em escolher a mais adequada a cada aplicação.

Ainda de acordo com Pisa (2022), os algoritmos de detecção mais populares são o Haar Feature-based Cascade e o Histogram of Oriented Gradient (HOG). O primeiro corresponde ao detector de faces de Viola e Jones (2004), ao qual possui quatro etapas: (a) seleção de características Haar-like retangulares, (b) criação da imagem integral, que permite o cálculo rápido das características, (c) treino de classificadores por um algoritmo de aprendizado de máquina baseado no AdaBoost e, por fim, (d) o uso de classificadores em cascata, que descarta as regiões de fundo para focar em áreas mais prováveis de conter uma face, conforme exemplifica a Figura 5.

Figura – Arquitetura do detector de faces Viola-Jones

|  |  |  |  |
| --- | --- | --- | --- |
| Uma imagem contendo Retângulo  Descrição gerada automaticamente |  |  |  |
| (a) Seleção características | (b) Criação imagem integral | (c) AdaBoost | (d) classificador em cascata |

Fonte: adaptado de Viola e Jones (2004).

O HOG é um extrator de característica que passou a ser popularmente utilizado na detecção de objetos após sua utilização por Dalal e Triggs (2005). O método HOG é dividido em 4 etapas: (a) cálculo do gradiente, (b) extração da direção do gradiente e magnitude, (c) geração dos histogramas e (d) normalização e concatenação e geração do vetor de características, conforme mostra a Figura 6.

Figura – Etapas do Histogram of Oriented Gradient (HOG)

|  |  |  |
| --- | --- | --- |
|  |  |  |
| (a) Cálculo gradiente e (b) direção da magnitude | (c) Geração dos Histogramas | (d) Vetor características |

Fonte: adaptado de Dalal e Triggs (2005).

Os métodos de extração mais conhecidos, Pisa (2022) enumera o Eigenfaces e o Fisherfaces. O método Eigenfaces usa o conceito de *Principal Component* (PC), que são partes consideradas mais relevantes dentre cada imagem treinada, descartando o resto. Nesse algoritmo, a iluminação é considerada uma *principal component*, tornando luz e sombra fatores importantes no reconhecimento. O método Fisherfaces é similiar ao Eigenfaces, diferenciando-se pelo fato de que as PC são extraídos de todas as imagens treinadas, como um todo. Assim, todos os componentes importantes para diferenciar o grupo de imagens são considerados *principal components*.

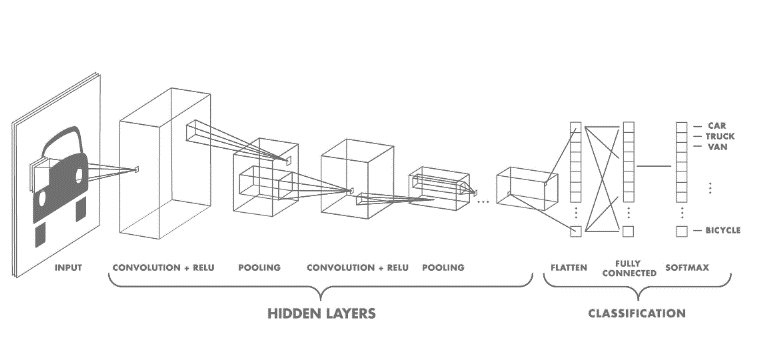
Por fim, de acordo com Sanchez-Moreno *et al*. (2021), a partir da extração das características que definem uma face ou qualquer outro objeto, pode-se utilizar os algoritmos ou métodos de aprendizado de máquina KNN, SVM e redes neurais tendo como intuito emula a capacidade do ser humano de reconhecer uma imensa variedade de objetos, animais, faces etc.

## Rede Neural Convolucional

Segundo Rosa (2018), a Rede Neural Convolucional, ou inglês Convolutional Neural Network (CNN), é uma classe da rede neural artificial que está sendo aplicada no processamento e análise de imagens digitais. A CNN é uma variante da rede *perceptrons* de múltiplas camadas, e foi inspirada no processo biológico de processamentos de dados visuais.

Segundo Vinayakumar, Soman e Poornachandran (2017), a CNN “é composta por uma camada convolution 1D, uma camada pooling 1D, uma camada fully connected e uma função não linear de ativação ReLU” e é uma extensão das tradicionais Feedforward Neural Networks (FNN). Na Figura 7 pode-se observar uma clássica arquitetura CNN.

Figura – Arquitetura da CNN



Fonte: Miyazaki (2017).

A partir da Figura 7 pode-se observar que uma CNN é composta pelas seguintes camadas: (a) input, corresponde à imagem de entrada; (b) convolution, é uma camada composta por uma série de filtros os quais são chamados de *kernel* e representam uma característica cada. Caso o filtro passe pela imagem e encontre algo que corresponda ao que está procurando, a operação convolucional gerará um grande número, ativando assim o filtro para tal característica. Caso contrário, o filtro não é ativado; (c) relu, que é uma camada de ativação a qual está presente após a camada convolucional (item b) e que ajuda a rede com a aprendizagem das funções difíceis e também com a redução do *overfitting*; (d) pooling, é um vetor de transformação escalar, o qual, assim como as convoluções, opera em cada região da imagem. Krinski e Todt (2019) apontam que nesta camada um filtro passa sobre o dado de entrada (item a) e as operações max (máximo), min (mínimo) e avg (média) de pooling são aplicadas; (e) flatten, essa camada costuma ser usada na divisão das duas partes da CNN: Hidden layer e Classification. Segundo Wang *et al*. (2020), é uma camada de nivelamento muitas vezes necessária antes da classificação para que a camada fully concected (item f) possa realizar as classificações; (f) fully connected é uma *Perceptron* multicamada composta por três outras camadas, sendo elas: (i) input: camada para qual a CNN manda as características geradas; (ii) hidden: é uma sequência de neurônios que aprenderão durante o treino da CNN. Pode haver mais de uma camada hidden em uma *perceptron* multicamadas; (iii) output: é também uma sequência de neurônios, porém com ativação diferente; (g) a função softmax costuma ser usada em um problema para gerar para cada categoria as suas probabilidades.

De acordo com Shrestha e Mahmood (2019), a implementação de uma rede neural é constituída por: (1) obter o conjunto de dados de treino e de teste; (2) treinar a rede; (3) fazer a predição com o dado de teste. Ainda de acordo com os autores, ao longo do tempo, foram criados diversos modelos com o intuito de alcançar objetivos mais específicos. Ceccon (2020) destaca que as CNNs podem ser separadas por funções: (i) reconhecimento de objetos em imagens, abrangendo as redes AlexNet, GoogleNet, VGG, ResNet e Inception, (ii) Detecção de objetos em imagens (You Only Look Once - Yolo e Fast R-CNN), (iii) segmentação semântica de objetos em imagens (UNet e SegNet) e (iv) Detecção com segmentação (Mask R-CNN). Já em termos de reconhecimento facial, o autor menciona que atualmente as redes de apresentam melhor performance e alta precisão são a Facebook DeepFace e Google FaceNet.

Referências

ALMEIDA, João M.; BENTO, Héder. **Biometria: Reconhecimento Facial**. Disponível em: <http://www.sinfic.pt/SinficWeb/displayconteudo.do2?numero=44661>. Acessado em: 20 nov. 2022.

CECCON, Denny. **Os tipos de redes neurais.** 2020. Disponível em: <https://iaexpert.academy/2020/06/08/os-tipos-de-redes-neurais/>. Acesso em: 07 dez. 2022.

DALAL, Navneet; TRIGGS, Bill. Histograms of Oriented Gradients for Human Detection. In: 2005 IEEE COMPUTER SOCIETY CONFERENCE VISION AND PATTERN RECOGNITION, 2005, [S.l.], **Proceedings...** [S.l.]: IEEE, 2005, p. 886-893.

FARIA, Alessandro de Oliveira. **Biometria: Reconhecimento Facial Livre**. 2022. Disponível em: <http://www.linhadecodigo.com.br/artigo/1813/biometria-reconhecimento-facial-livre.aspx>. Acessado em: 07 dez. 2022

GOGONI, Ronaldo. **O que é biometria? Os 6 tipos mais usados na tecnologia**. 2018. Disponível em: <https://tecnoblog.net/responde/o-que-e-biometria-tecnologia>. Acesso em: 29 nov. 2022.

GROSS, Ralph *et al*. Multi-PIE. In: IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE & GESTURE RECOGNITION, 8., 2008, Amsterdam. **Proceedings…** Amsterdam, 2008. p. 1-8.

HE, Mingjie *et al*. Deformable Face Net: Learning Pose Invariant Feature with Pose Aware Feature Alignment for Face Recognition. In: IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE & GESTURE RECOGNITION, 14., 2019, Lille. **Proceedings…** Lille: Lille grand Palais, 2019. p. 1-8.

HUANG, Gary B. *et al*. **Labeled Faces in the Wild:** A Database for Studying Face Recognition in Unconstrained Environments. 2007. 11 f. Technical Report 07-49 - University of Massachusetts, Amherst.

KEMELMACHER-SHLIZERMAN, Ira *et al*. **The MegaFace Benchmark:** 1 Million Faces for Recognition at Scale. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 29., 2016, Las Vegas. **Proceedings…** Las Vegas: Caesars Palace, 2016. p. 4873-4882.

KRINSKI, Bruno Alexandre; TODT, Eduardo. **Convolutional neural network**. [S.l.], 30 nov. 2019. Disponível em: http://www.inf.ufpr.br/todt/IAaplicada/CNN\_Presentation.pdf. Acesso em: 26 nov. 2022.

MIYAZAKI, Caio Kioshi. **Redes neurais convolucionais para aprendizagem e reconhecimento de objetos 3D**. 2017. 56f, Monografia (Graduação em engenharia elétrica com ênfase em sistemas de Energia e Automação) – Escola de Engenharia de São Carlos da Universidade de São Paulo, 2017.

NADA, Hajime *et al*. **Pushing the Limits of Unconstrained Face Detection:** A Challenge Dataset and Baseline Results. 2018. Disponível em: <https://arxiv.org/abs/1804.10275>. Acesso em: 27 set. 2022

OH, Seong Joon *et al*. **Faceless person recognition:** Privacy implications in social media. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 14., 2016, Amsterdam. **Proceedings…** Amsterdam: Koninklijk Theater Carré, 2016. p. 19-35.

PISA, Pedro. **Como funciona o reconhecimento facial**. Disponível em: <https://www.techtudo.com.br/artigos/noticia/2012/04/como-funciona-o-reconhecimento-facial.html>.   
Acessado em: 13 set. 2022.

ROSA, Renan P. **Método de classificação de pragas por meio de rede neural convolucional profunda**. 2018. 101 f. Dissertação (Mestrado em Computação Aplicada) - Universidade Estadual de Ponta Grossa, Ponta Grossa, 2018.

SANCHEZ-MORENO, Alejandra S. *et al*. 2021. Efficient Face Recognition System for Operating in Unconstrained Environments. **Journal of Imaging**, v. 7, n. 9, p. 161, ago. 2021.

SENGUPTA, Soumyadip *et al*. Frontal to profile face verification in the wild. In: IEEE WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION, 16., 2016, Lake Placid. **Proceedings…** Lake Placid: Crowne Plaza, 2016. p. 1-9.

SHRESTHA, Ajay; MAHMOOD, Ausif. Review of Deep Learning Algorithms and Architectures. **Ieee Access**, [S.l.], v. 7, p. 53040-53065, abr. 2019

VINAYAKUMAR, Ravi; SOMAN, Kotti. P.; POORNACHANDRAN, Prabaharan. Applying convolutional neural network for network intrusion detection. In: 2017 INTERNATIONAL CONFERENCE ON ADVANCES IN COMPUTING, COMMUNICATIONS AND INFORMATICS (ICACCI), 2017, [S.l.]. **Proceedings Of The Ieee**.[S.l.]:IEEE, 2017, p. 1222-1228

VIOLA, Paul; JONES, Michael J. Robust Real-Time Face Detection. **International Journal Of Computer Vision**, [S.l.], v. 2, p.137-154, maio 2004.

WANG, Zijie J *et al*. CNN EXPLAINER: learning convolutional neural networks with interactive visualization. In: IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, 2020, [S.l.]. **Proceedings Of The Ieee.**[S.l.]:IEEE, 2020, p. 1-11.

WOLF, Lior; HASSNER, Tal; MAOZ, Itay. Face Recognition in Unconstrained Videos with Matched Background Similarity. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 24., 2011, Colorado Springs. **Proceedings…** Colorado Springs: Crowne Plaza, 2011. p. 529-534.

YANG, Ming-hsuan; KRIEGMAN, D.j.; AHUJA, N. **Detecting faces in images**: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, [S.l.], v. 24, n. 1, p.34-58, 7 ago. 2002. Institute of Electrical and Electronics Engineers (IEEE).

YANG, Shuo *et al.* **Wider Face:** A Face Detection Benchmark. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 29., 2016, Las Vegas. **Proceedings…** Las Vegas: Caesars Palace, 2016. p. 5525-5533.

\_\_\_\_\_. Faceness-Net: Face Detection through Deep Facial Part Responses. **IEEE Transactions on Pattern Analysis and** Machine Intelligence, v. 40, n. 8, p. 1845-1859, ago. 2018.

ZHANG, Cha; ZHANG, Zhengyou. **A survey of recent advances in face detection**. jun. 2010. Disponível em: <https://www.microsoft.com/en-us/research/publication/a-survey-of-recent-advances-in-face-detection/>. Acesso em: 30 nov. 2022

ZHU, Xiangxin; RAMANAN, Deva. Face detection, Pose Estimation, and Landmark Localization in the Wild. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 25., 2012, Providence. **Proceedings…** Providence: Rhode Island Convention Center, 2012. p. 2879-2886.

1. Partness map: São os mapas de resposta que destacam 5 partes diferentes da face (YANG *et al*., 2018b, p. 1-2, tradução nossa). [↑](#footnote-ref-1)