Revista ARACÊ
ISSN: 2358-2472

ANÁLISE E AVALIAÇÃO DE MODELOS DE DETECÇÃO E RECONHECIMENTO FACIAL

https://doi.org/10.56238/arev7n3-003

Data de submissão: 03/02/2025 **Data de publicação:** 03/03/2025

Rhuan Lima Ruiz de Oliveira
Graduado em Engenharia de Computação
Departamento de Computação – Universidade Federal de Sergipe
E-mail: rhuan.ruiz@dcomp.ufs.br
ORCID: https://orcid.org/0009-0008-8864-6194

Sálvio Roberto Freitas Reis
Mestre em Ciência da Computação
Programa de Pós-Graduação em Ciência da Computação – Universidade Federal de Sergipe
E-mail: salvio.reis@dcomp.ufs.br
ORCID: https://orcid.org/0009-0002-2924-3084

Rafael Oliveira Vasconcelos
Doutor em Computação (DI/PUC-Rio)
Programa de Pós-Graduação em Ciência da Computação – Universidade Federal de Sergipe
E-mail: rafael @ dcomp.ufs.br
ORCID: https://orcid.org/0000-0001-7974-304X

RESUMO
Com tecnologias como Internet das Coisas em constante evolução, surge a oportunidade para que outras áreas, como reconhecimento facial, incorporem novos conceitos ao seu funcionamento e conquistem novas aplicações nas cidades inteligentes. Segurança pública é um dos aspectos que compõem as cidades, com maneiras criativas de aprimorar surgindo a todo momento, elas também desfrutam dos benefícios que a implementação de sistemas de reconhecimento facial trazem. Esses sistemas são capazes de captar imagens de indivíduos e detectar faces através de algoritmos, que são submetidos a modelos capazes de realizar o reconhecimento dessas pessoas comparando com uma base de dados. Esse trabalho realiza um estudo acerca do processo de reconhecimento facial, visando a análise e avaliação de modelos de detecção e reconhecimento facial no contexto de segurança pública.

Palavras-chave: Reconhecimento Facial. Detecção Facial. Segurança. Internet das Coisas.

Revista ARACÊ
ISSN: 2358-2472

1 INTRODUÇÃO

Com o avanço das tecnologias e o grande impacto da Internet das Coisas e das cidades inteligentes na vida cotidiana, um portfólio de possibilidades se apresenta nas mais diversas áreas que compõem a sociedade. De acordo com (SALAM, et al., 2019), a internet das coisas revolucionará como a sociedade funcionará, conectando dispositivos capazes de tomar decisões inteligentes com o mínimo de intervenção humana.

Internet das Coisas *(IoT)* refere-se a todo e qualquer objeto incorporado a outras tecnologias, como softwares e sensores, com o objetivo de promover comunicação entre estes e outros dispositivos. Desde itens cotidianos até ferramentas industriais, conforme observado por (ILYAS, 2021), esses objetos são capazes de monitorar e coletar um amplo leque de informação, que são então processados utilizando *edge computing* ou nuvem para a tomada de decisões. Segundo a (ORACLE, 2020), é esperado que o número de dispositivos *IoT* alcance 22 bilhões até 2025.

Nesse contexto, a Internet das Coisas surge como um componente crucial das cidades inteligentes. De acordo com (ILYAS, 2021), as cidades inteligentes desfrutam de todas as tecnologias disponíveis para aprimorar a qualidade de vida de seus habitantes, incentivando maior engajamento da comunidade, além de reduzir custos operacionais e otimizar o uso de recursos públicos. A comunicação promovida pelos dispositivos IoT beneficia diretamente as cidades inteligentes, sendo possível observar suas aplicações nas mais diversas áreas que a compõem, dentre elas, segurança pública.

Com o contínuo avanço das tecnologias, a todo momento surgem novas abordagens para aprimorar a segurança pública das cidades inteligentes. Os sistemas de Circuito Fechado de Televisão (CCTV) representam uma dessas abordagens, responsáveis por captar e transmitir imagens de câmeras, permitindo gravação e visualização destas imagens. Para (DOSHI et al., 2022), a implementação de câmeras é capaz de proteger tanto os bens quantos os indivíduos na região onde o sistema CCTV está instalado, oferecendo benefícios na prevenção de roubos e detecção de atividades suspeitas.

No entanto, conforme enfatizado por (ALSHAMMARI; RAWAT, 2019), a crescente quantidade de dados provenientes de monitoramento por vídeo trazem consigo aumento de erros humanos, tendo em vista o limite com o qual equipes são capazes de processar vasta quantidade de informação. Portanto, para um monitoramento inteligente, é necessário a implementação de softwares de reconhecimento facial. Essas ferramentas são capazes de identificar o rosto de indivíduos através da medição dos componentes faciais de uma imagem.

Revista ARACÊ
ISSN: 2358-2472

Assim, o uso de sistemas CCTV aliado a implementação de softwares de reconhecimento facial provê diversas oportunidades para o aprimoramento da segurança pública. (DOSHI et al., 2022) diz que reconhecimento facial é o responsável por tornar o sistema inteligente, facilitando a identificação de indivíduos suspeitos. Esses sistemas já estão presentes em grande volume em residências e estabelecimentos, tornando-se viável a sua utilização para identificação, envio de imagens e informações para os órgãos responsáveis pela segurança das cidades.

1.1 JUSTIFICATIVA

Diante da necessidade de aprimorar a segurança pública das cidades em um cenário onde métodos tradicionais mostram-se insuficientes, o uso de tecnologias surgem como alternativa para atingir tal feito. Hoje, indivíduos já são identificados de diversas maneiras no dia a dia, através de senhas, chaves eletrônicas, identificação por impressão digital, voz, dentre outros. Porém, muitos desses métodos podem apresentar falhas de segurança ou não ser tão eficazes na identificação dos cidadãos.

Nesse sentido, reconhecimento facial é um método em ampla ascensão, capaz de captar imagens das faces de indivíduos e atribuí-las a um indivíduo identificado. É uma tecnologia com grande versatilidade, possuindo aplicações em diversas áreas, tendo como maior benefício, em comparação a outros métodos, a capacidade de identificar indivíduos de maneira passiva, sem a necessidade de interação do usuário com o sistema. Essa característica mostra-se vantajosa na identificação de indivíduos que possuam históricos criminais e que tenham suas imagens submetidos a um sistema de reconhecimento facial.

Diante desse cenário, surge a oportunidade de realizar um estudo sobre todo o processo de reconhecimento facial e realizar uma análise e avaliação de alguns dos modelos de detecção e reconhecimento facial disponíveis, contextualizando para segurança pública.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

O objetivo deste trabalho é realizar uma análise dos modelos de detecção e reconhecimento facial disponíveis no contexto de segurança pública, visando utilizar métricas de avaliação para determinar os modelos mais eficientes.

1.2.2 Objetivos Específicos

- Realizar um estudo sobre reconhecimento facial;

Revista ARACÊ
ISSN: 2358-2472

- Analisar e discutir impactos do uso de reconhecimento facial na segurança pública e privacidade;
- Analisar bases dados disponíveis para os fins do trabalho;
- Discutir acerca dos modelos de detecção e reconhecimento facial;
- Analisar modelos de detecção e reconhecimento facial.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 RECONHECIMENTO FACIAL

A utilização de técnicas de reconhecimento facial está difundida nas mais diversas áreas da sociedade. No contexto de segurança pública, é utilizada para o controle de acesso em áreas restritas e no controle de multidões em eventos públicos ou privados, além de ser utilizado na identificação de atividades suspeitas. No entanto, diante da sensibilidade dos dados coletados, levanta-se uma série de questionamentos quanto a possíveis ferimentos à privacidade e integridade dos indivíduos que eventualmente possam ser monitorados.

No Brasil, o monitoramento com reconhecimento facial já é amplamente utilizado e mostra um futuro promissor, com diversos projetos na área em desenvolvimento. Segundo (ABDALA, 2023), quase 48 milhões de brasileiros estão potencialmente sob monitoramento com os sistemas de câmeras que utilizam reconhecimento facial, com maior concentração nas regiões sudeste e nordeste, respectivamente. Alguns estudiosos questionam a extensão desse uso, tendo em vista que o reconhecimento facial ainda não mostrou resultados práticos significativos.

O reconhecimento facial é uma técnica capaz de reconhecer indivíduos, por meio da análise das características físicas de suas faces, possibilitando sua identificação. É baseada nos princípios e técnicas da biometria, tendo padrões biométricos construídos a partir das características faciais únicas de um indivíduo, como por exemplo olhos, sobrancelhas, nariz, lábios e maxilar (BELUCO; FILHO, 2023). Esses padrões então são comparados com registros previamente armazenados para que a identificação possa ser feita.

Todo sistema de reconhecimento facial possui um modelo operacional, dividindo-se em etapas que irão compor todo o processo do reconhecimento das faces. De maneira geral, (CAMARA, 2021) diz que a primeira etapa consiste na captura geral da imagem, sendo seguido pela detecção facial. Nesta etapa, a face humana é destacada e segmentada do restante da imagem capturada, permitindo uma melhor análise das características faciais.

Revista ARACÊ
ISSN: 2358-2472

Em seguida, a imagem é submetida à extração de atributos, concentrando-se no auferimento das características geométricas da face. Posteriormente, a imagem facial procede para a etapa da análise em si, onde realmente é feita a detecção do rosto e a classificação da imagem, de modo que possa ser associada à uma identidade pré-cadastrada no sistema. A depender do sistema implementado, a imagem pode passar por etapas extras ao longo do processo de reconhecimento.

Figura 1 – Etapas do Reconhecimento Facial

Fonte: IDEC (2020)

2.1.1 Detecção Facial

A etapa de detecção facial é responsável por identificar as faces humanas existentes na imagem e tudo aquilo que possa ou não ser relevante para o processo de reconhecimento. Em outras palavras, conforme mencionado por (IDEC, 2020), trata-se do momento responsável por categorizar a imagem ou porções delas, segmentando as faces da imagem do restante do quadro e categorizando-as como faces ou não. Como porções da imagem podem conter elementos que não são do interesse da análise, como animais e objetos, a categorização é de suma importância para o processo.

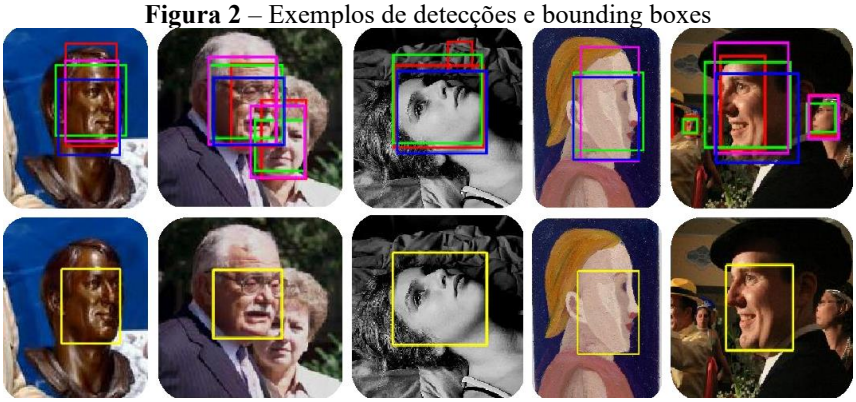
Após a detecção, algoritmos utilizam *bounding boxes* (ou caixas delimitadoras) para realizar a demarcação das faces. As caixas delimitadoras são definidas pelos pontos dos cantos superior e inferior, são amplamente utilizadas nas mais diversas tarefas de detecção, além de descreverem a posição e o tamanho de objetos na imagem. De acordo com (AYADATA, 2023), as caixas são utilizadas por algoritmos para que possam aprender sobre o conteúdo de uma imagem, de maneira que seja possível coletar características interessantes para o modelo e realizar a rotulagem das mesmas, facilitando o reconhecimento de objetos semelhantes.

Revista ARACÊ
ISSN: 2358-2472

Contudo, uma série de fatores podem afetar negativamente a detecção de uma ou mais faces na imagem. Conforme mencionado, as imagens podem conter elementos que não são interessantes para análise e que são descartados pelos algoritmos, mas (IDEC, 2020) mostra como elementos do ambiente também podem alterar a percepção da imagem. Aspectos como iluminação, rotação de faces também são alguns dos elementos que podem afetar a leitura das características de uma face.

Nesse sentido, sistemas também podem incorporar etapas intermediárias entre a detecção facial e a extração dos atributos. É possível minimizar alguns dos problemas mencionados anteriormente através do processo de normalização. Segundo (IDEC, 2020), durante esse processo, padrões como cor, rotação e iluminação são alterados na imagem por meio de um recorte padrão, visando a realização de uma análise mais consistente.

Figura 2 – Exemplos de detecções e bounding boxes



Fonte: (FENG et al., 2017)

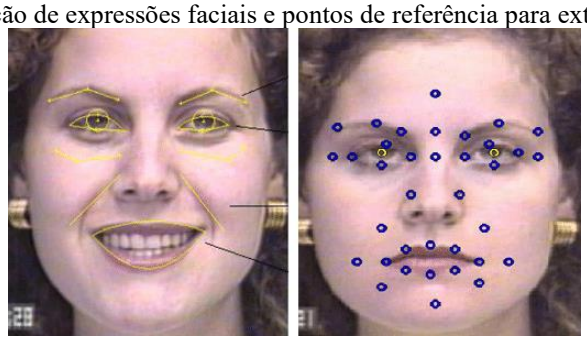
2.1.2 Extração de Atributos

Após a conclusão da etapa de detecção facial, a imagem prossegue para o processo de extração de atributos. Durante esta etapa, todos os pontos de referência faciais são analisados, com informações relevantes sobre a imagem sendo extraídas para etapas posteriores. Como mostra (IDEC, 2020), durante a fase de extração, dados acerca das características geométricas das faces como forma, localização e distância dos componentes faciais (boca, nariz, sobrancelhas etc.) são identificados e registrados para uso subsequente.

As características obtidas são então organizadas e utilizadas para formar um vetor de atributos. Propriedades e tamanho desse vetor podem variar a depender do algoritmo em uso. De acordo com (TIAN; KANADE; COHN, 2011), após a implementação da sequência, a face do indivíduo e a localização aproximada dos atributos faciais são detectados no frame inicial, com o restante dos atributos sendo identificado após a inicialização do algoritmo. A eficácia do algoritmo tem um impacto direto na precisão com os quais as faces de indivíduos sob monitoramento são identificadas.

Revista ARACÊ
ISSN: 2358-2472

Figura 3 – Extração de expressões faciais e pontos de referência para extração de atributos



Fonte: (TIAN; KANADE; COHN, 2011)

O papel desempenhado pela etapa de extração de atributos é crucial para o processo de reconhecimento facial, pois os atributos identificados pelos algoritmos impulsionam o restante das etapas. Com os dados obtidos, sistemas de reconhecimento facial terão em suas bases informações essenciais que contribuirão para a identificação e comparação das características individuais de cada usuário que seja submetido ao processo de reconhecimento.

2.1.3 Registro e Análise

Subsequente, dependendo das propriedades do sistema implementado, os dados podem ser descartados ou encaminhados para a etapa de registro. Nela, os dados são efetivamente registrados e capacitados para que possam ser comparados entre o que há armazenado e o que é obtido em tempo real. Dessa maneira, é possível realizar a identificação de um indivíduo cuja imagem tenha sido capturada pelo sistema.

A etapa de análise do reconhecimento facial passa pelas etapas de categorização, autenticação e identificação, de modo que possam ser empregadas de forma independente ou integrada, a depender do sistema implementado. Além da própria análise, também são comparados modelos biométricos e imagens armazenadas no banco de dados para efetuar a identificação.

Enfatizado por (CAMARA, 2021), durante a categorização ocorre a classificação da imagem biométrica em parâmetros como idade e gênero. Já na autenticação, é onde ocorre a verificação do indivíduo, comparando dois modelos biométricos e utilizando dados estatísticos para identificar a probabilidade do indivíduo ser o esperado pelo sistema. Por fim, na identificação a imagem é comparada com outras armazenadas no banco de dados para identificar de maneira mais abrangente o indivíduo.

Como ilustrado na Figura 4, no caso de uma verificação bem-sucedida, o indivíduo é reconhecido e classificado pelo sistema como identificado. Em contrapartida, caso a verificação seja mal-sucedida, o indivíduo é classificado como desconhecido.

Revista ARACÊ
ISSN: 2358-2472

2.2 REDES NEURAIS CONVOLUCIONAIS

Redes neurais convolucionais, também conhecidas como CNNs (Convolutional Neural Networks), são um tipo de modelo de aprendizado profundo projetados para processar dados, como imagens, estruturados em formato de grade. Segundo (YAMASHITA et al., 2018), os modelos são inspirados pelo córtex visual animal e designados para que possam automaticamente aprender, de forma adaptativa, hierarquias espaciais de características de padrões tanto de baixo quanto de alto nível. As CNNs são comumente utilizadas nas mais diversas tarefas relacionadas à visão computacional, em especial, são soluções bem sucedidas nos campos de reconhecimento facial e detecção de objetos.

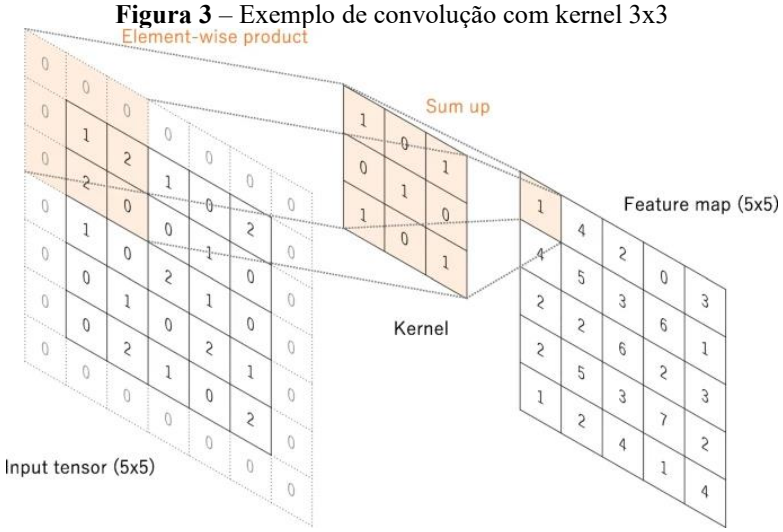
Além disso, as CNNs oferecem uma solução mais sofisticada em comparação a outras técnicas que ainda utilizam extração de atributos de maneira manual. De acordo com (ALBAWI; MOHAMMED; AL-ZAWI, 2017), as CNNs não utilizam extração manual de atributos e não exigem a segmentação de propriedades anatômicas por especialistas, ou seja, é uma solução computacionalmente mais cara, tendo em vista que requer um volume muito maior de dados para o aprendizado de seus parâmetros. Durante seu funcionamento, as CNNs devem ser capazes de localizar os atributos onde quer que estejam na matriz.

A camada de convolução é um dos componentes fundamentais da arquitetura das CNNs. Nela, são aplicados filtros na imagem, também conhecidos como kernels, que consistem em vetores de números que deslizam por toda a imagem, gerando o mapa de atributos ao fim do processo. Para (INDOLIA et al., 2018), o processo de convolução se dá pelo deslizamento do filtro tanto horizontalmente quanto verticalmente ao longo de toda a imagem, extraindo ao longo do processo N atributos em uma única camada, que representa atributos distintos. Portanto, ao fim do processo, existirão N filtros para N mapas de atributos.

Outrossim, o compartilhamento dos *kernels* em todos os pontos da imagem é um dos elementos fundamentais da camada de convolução. De acordo com (YAMASHITA et al., 2018), o compartilhamento permite que os atributos extraídos sejam invariantes à translação, à medida que os kernels detectam padrões aprendidos enquanto percorrem a imagem. Além disso, também é possível que aprendam hierarquias espaciais de padrões de atributos e aumentem a eficácia do modelo ao reduzir o número de parâmetros a serem aprendidos.

Revista ARACÊ
ISSN: 2358-2472

Figura 3 – Exemplo de convolução com kernel 3x3



Fonte: (YAMASHITA et al., 2018)

A camada convolução é seguida pela etapa de sub-amostragem, representando, de maneira geral, uma operação de redução de amostragem (*downsampling*). Essa camada é responsável pela redução das dimensões dos mapas de atributos formados através da diminuição do número de parâmetros que podem ser aprendidos, permitindo o aprendizado direto a partir dos dados. Para (INDOLIA et al., 2018), a maior vantagem da etapa de subamostragem é justamente a redução dos atributos que podem ser treinados, além de introduzir invariância à translação.

A capacidade de invariância à translação e distorções, além da precisão na extração dos atributos das imagens faciais com seu aprendizado hierárquico, são algumas das características das CNNs. Dessa maneira, as redes neurais convolucionais se mostram poderosas para os sistemas de reconhecimento facial, permitindo a extração de atributos discriminativos das faces e facilitando o reconhecimento dos indivíduos.

2.3 MODELOS DE DETECÇÃO FACIAL

2.3.1 Viola-Jones

O algoritmo Viola-Jones é um método popular para detecção de objetos ou faces em imagens, sendo reconhecido por sua velocidade na detecção. É capaz de lidar com as aversões do ambiente da imagem e expressões faciais, sendo comumente utilizado nos sistemas de reconhecimento facial. Apesar disso, é um algoritmo que não lida bem com extremas variações de iluminação e com as grandes diferenças causadas pela variação de objetos distintos (LU; YANG, 2019).

Seu funcionamento ocorre ao longo de algumas etapas. Inicialmente, ocorre a extração dos atributos utilizando filtros denominados Haar, onde realiza-se a identificação de padrões entre rostos, como posição dos olhos e nariz, por meio da aplicação do filtro na imagem que calcula as diferentes

Revista ARACÊ
ISSN: 2358-2472

intensidade entre seus pixels. Em seguida, através do algoritmo *Adaboost*, um classificador de cascata é treinado fazendo uma combinação de classificadores mais fracos que são dependentes de um único atributo com o intuito de formar um classificador mais forte (VIOLA; JONES, 2001). Em cada estágio da cascata, aplica-se um classificador com outro atributo mais específico até chegar ao ponto de detecção ou rejeição para o restante do fluxo.

2.3.2 Single Shot Multibox Detector (SSD)

Single Shot Multibox Detector (SSD) é uma rede neural projetada para realizar detecção de vários objetos de classes variadas em uma mesma etapa. Destaca-se pela precisão de detecção, atingindo um mAP (mean Average Precision) de 74,3% e 76,9% em entradas com resolução 300x300 e 512x512, respectivamente, em testes realizados com o dataset VOC2007 (LIU et al., 2021), realizando as tarefas de localização e classificação em uma única passada (single-shot) pela rede. Essas características fazem com que o SSD seja uma rede neural com mais precisão e acurácia em comparação a outros detectores single-shots, como o YOLO (You Only Look Once), ou detectores que utilizam técnicas como pooling e RPN (Region Proposal Network) (LIU et al., 2021).

A arquitetura do SSD é caracterizada por sua detecção multiescala, envolvendo camadas convolucionais sucessivas que são compostas por convoluções de diferentes escalas responsáveis por detectar objetos em diversas faixas de tamanho. O modelo divide a imagem de entrada em formato de grade, delimita bounding boxes em diversos tamanhos e realiza outras predições que incluem a localização do objeto e a probabilidade do mesmo estar presente em cada delimitação (LIU et al., 2021).

O SSD utiliza bases convolucionais (VGG-16) para extrair as características da imagem antes de ser submetido à etapa multiescala, além de utilizar bounding boxes pré-definidas semelhantes às anchor boxes utilizadas na RPN (LIU et al., 2021). Essa pré-definição evita que o modelo necessite fazer o pooling dos atributos faciais extraídos e, ao invés disso, atribui uma pontuação para cada objeto em cada uma das bounding boxes, realizando a detecção e classificação dos objetos simultaneamente.

Apesar de possuir foco na detecção geral de objetos, (YE et al., 2015) mostra o potencial do modelo para detecção facial em tempo real. Sua implementação se mostra favorável para casos em que velocidade é uma prioridade, como sistemas de navegação, autenticação facial em dispositivos móveis e principalmente para os sistemas de vigilância CCTV. Entretanto, o modelo perde desempenho em situações em que seja necessário manipular múltiplas faces.

Revista ARACÊ
ISSN: 2358-2472

2.3.3 Multi-task Cascaded Convolutional Networks (MTCNN)

Multi-Task Cascaded Convolutional Networks(MTCNN) é uma das redes neurais voltadas para detecção facial mais reconhecidas e amplamente utilizada, destacando-se com a correlação entre os métodos de detecção e alinhamento facial (RAJPUT, 2020). Proposto por Zhang et al. (2016), o modelo trabalha em cascata entre três redes neurais que são responsáveis pelo processo de detecção e refinamento das bounding boxes. O modelo atingiu uma taxa de falsos positivos de 0,9504 e um recall de 0,851 em testes realizados durante o seu desenvolvimento (ZHANG et al., 2016).

Sua arquitetura em cascata é composta pelas redes P-Net (Proposal Network), R-Net (Refine Network) e O-Net (Output Network). P-Net é a rede responsável por identificar as faces candidatas da imagem e suas bounding boxes iniciais através de um processo de supressão chamado NMS, onde bounding boxes redundantes são eliminadas e um vetor de regressão é formado. Em seguida, a rede R-Net faz um refinamento da entrada recebida, sendo responsável por reduzir o número de falsos positivos, calibrar o vetor de regressão das bounding boxes e realizar mais uma etapa de NMS. Por fim, a O-Net realiza os ajustes finais nas bounding boxes e descreve cinco posições de referências faciais (ZHANG et al., 2016).

Sistemas que tenham a necessidade de realizar ambas tarefas de detecção e reconhecimento facial simultaneamente podem se beneficiar com o uso do MTCNN devido a sua precisão em relação ao alinhamento das faces. Por conta disso, suas aplicações estão comumentes presentes em sistemas biométricos, sistemas de segurança em geral, além dos próprios sistemas de vigilância, CCTV ou móveis. De maneira geral, o bom alinhamento resultante do modelo promove um reconhecimento facial posterior mais eficaz.

Em (JOSE et al., 2019), o sistema é capaz de utilizar múltiplas câmeras, identificar e manter registro de suspeitos, sustentando durante o processo uma precisão de 97%. No entanto, em situações em que o modelo trabalhe com imagens em situações adversas pode resultar em uma cascata menos eficiente, consequentemente afetando o desempenho geral do modelo.

2.3.4 RetinaFace

RetinaFace é um modelo capaz de calcular bounding boxes e pontos chaves das faces em condições e tamanhos variados. O modelo destaca-se em imagens de alta qualidade, detectando faces em condições de tamanho diversos, realizando a detecção das faces de maneira hierárquica por meio de pirâmides de atributos (feature pyramid) e prevendo pontos chaves que facilitam a detecção e alinhamento para posteriormente reconhecimento (DENG et al., 2020).

Revista ARACÊ
ISSN: 2358-2472

Sua arquitetura utiliza a rede FPN (Feature Pyramid Network) e ResNet50 como outra rede backbone profunda. Juntas, fornecem vetores de atributos a partir da ResNet50 para a fase de detecção, contribuindo para a identificação de faces de diversas escalas, além do alinhamento dos pontos chaves das faces encontradas ao longo do processo (COCHARD, 2024). Essa metodologia torna o modelo extremamente eficaz na detecção de pequenas faces na imagem, além de garantir maior precisão nos casos em que há variação de poses, iluminação, entre outras condições adversas.

Diante disso, o modelo floresce nas situações em que a detecção de múltiplas faces é mandatória. É amplamente utilizado por sistemas de vigilância devido a suas características de possuir alta precisão em situações de ambiente não controlado ou cenários com número elevado de pessoas, superando modelos como MTCNN em testes diretos (DENG et al., 2020). No entanto, outros modelos ainda podem ser priorizados nestes casos devido ao custo computacional mais elevado que envolve a implementação do mesmo.

2.3.5 MediaPipe

MediaPipe é um conjunto de bibliotecas e ferramentas desenvolvido pela Google com aplicabilidade nos campos de inteligência artificial e machine learning. No campo de detecção facial, fornece soluções otimizadas especialmente para dispositivos móveis, por conta da sua capacidade de operar à um custo computacional mais inferior, sendo projetado para plataformas web, Android e iOS. O Face Detector compõe a biblioteca, sendo um modelo capaz de localizar faces, seus atributos e definir os pontos chaves de imagens únicas ou de um fluxo contínuo de imagens (GOOGLE, 2021).

A arquitetura do MediaPipe se desdobra através de um pipeline que utiliza um conjunto de algoritmos ao passo que avança durante as etapas de detecção. Utilizando redes neurais, é capaz de identificar os atributos faciais de imagens e vídeos delimitando as bounding boxes e pontos chaves da face. Essa abordagem contribui para a versatilidade do modelo, em especial em situações que fogem do padrão de ambientes com grandes volumes e movimentação de indivíduos.

O modelo atua através de gráficos de fluxo de dados, o que permite um processamento concorrente de várias etapas do processo de detecção, tal como uma maior velocidade de execução. Além disso, também é capaz de rastrear com precisão faces em movimento ou com mudanças de ângulo, permitindo que o modelo atue realizando uma detecção contínua. A combinação entre precisão e monitoramento adaptativo tornam o modelo eficiente e preciso, sendo amplamente utilizado em sistemas que necessitem desse tipo de tarefa (MANJHI; RAJAWAT; SRIVASTAVA, 2023).

Revista ARACÊ
ISSN: 2358-2472

Em linhas gerais, o MediaPipe é um modelo que se destaca pela sua otimização para dispositivos móveis e pela sua capacidade de detecção contínua, o que o torna um modelo eficiente. Essas características o tornam uma solução voltada para situações em que é necessário operar com recursos mais restritos, não alcançando o nível de robustez e precisão de outros modelos como o RetinFace e MTCNN. Sua área de atuação é eficiente em situações em que velocidade e fluidez são necessárias (GOOGLE, 2021).

2.4 MODELOS DE RECONHECIMENTO FACIAL

2.4.1 Eigenfaces

O modelo Eigenfaces oferece uma abordagem mais simplificada de se realizar reconhecimento facial, sendo amplamente considerado um dos modelos mais utilizados e servindo de base para outros algoritmos de reconhecimento facial. Para (ALAKKARI; COLLINS, 2013), Eigenfaces representa uma das primeiras tentativas de se criar uma espaço facial e comprimir os dados obtidos a partir das faces, sendo utilizado como um modelo para reconhecimento facial baseado na análise de componentes principais (PCA).

O funcionamento proposto por (TURK; PENTLAND, 1991) ocorre a partir das etapas de aprendizado e reconhecimento. Durante o aprendizado, o algoritmo recebe uma ou várias das chamadas imagens de treinamento, contendo as faces de indivíduos para que o algoritmo possa aprender sobre o modelo delas. Em seguida, na fase de reconhecimento, as imagens de treinamento são aplicadas à técnica PCA, transformando as faces em vetores, reduzindo a dimensionalidade da imagem e extraindo os principais atributos das faces.

Posteriormente, a imagem é projetada no subespaço e calcula-se as distâncias entre os vetores da imagem a ser reconhecida e das imagens de treinamento. Ao fim, o reconhecimento facial é feito com base na distância encontrada, ou seja, distâncias mais próximas representam um indivíduo identificado, enquanto que distâncias mais afastadas representam a face de um indivíduo que o algoritmo ainda precisa obter mais conhecimento (TURK; PENTLAND, 1991).

2.4.2 FisherFaces

Fisherface partilha semelhanças a outros modelos, em especial ao Eigenfaces, mas é considerado mais eficiente em relação a outros algoritmos por conta de sua separação de classes durante a etapa de treinamento. Segundo Reddy e Kumar (2021), semelhantemente ao modelo de Eigenfaces, Fisherface também utiliza a análise de componentes principais (PCA) para redução de dimensionalidade, mas apenas considera alguns atributos da imagem para realizar a detecção

Revista ARACÊ
ISSN: 2358-2472

utilizando a análise de discriminante linear (LDA). A sua característica de separação de classes faz com que o algoritmo lide bem com aversões comuns presentes nas imagens, como variação de expressões faciais e iluminação.

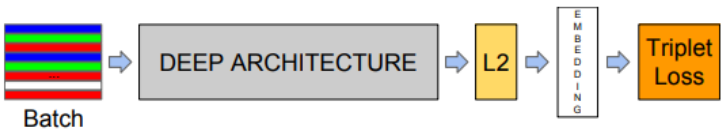
Com a aplicação do PCA e a diminuição da dimensionalidade, o LDA é aplicado nos atributos a fim de encontrar as melhores projeções que oferecem maior distinção entre as classes. Em seguida, as imagens de uma mesma classe são classificadas, distribuindo-as de maneira mais próxima caso o algoritmo enxergue-as como pertencentes a um indivíduo e de maneira mais distante caso pertençam a outro. Assim, com a variância entre as classes, o modelo é capaz de calcular a semelhança entre as imagens classificadas e determinar a identidade do indivíduo.

2.4.3 FaceNet

FaceNet é um modelo de aprendizado profundo, desenvolvido pela Google (2021), que trabalha no campo de reconhecimento facial utilizando as chamadas embeddings para o aprendizado das representações das faces. As embeddings formam vetores de atributos que podem ser utilizados na identificação e comparação das faces, se destacando pela sua eficácia nas medições de similaridade entre faces. As faces identificadas e atribuídas a um mesmo indivíduo são armazenadas em agropamentos chamados de face clusterings (SCHROFF; KALENICHENKO; PHILBIN, 2015).

O funcionamento do modelo ocorre através de redes neurais que trabalham diretamente em cima das embeddings, gerando vetores de atributos por face identificada (SCHROFF; KALENICHENKO; PHILBIN, 2015). Esses vetores contêm informações gerais acerca das faces, como formato dos olhos, estrutura facial, dentre outros. Através das tripletas, que define três embeddings: o âncora, que é a face de entrada; o positivo que é uma outra face mas também pertencente ao âncora; e o negativo pertencente a outro indivíduo. Dessa forma, o modelo é capaz de medir e comparar as faces entre si e determinar se pertencem ao mesmo indivíduo comparando a distância entre os vetores, de modo que quanto mais próximo os vetores do âncora e positivo, maior as chances de um verdadeiro positivo.

Figura 4 – Ilustração de funcionamento do modelo FaceNet



Fonte: (SCHROFF; KALENICHENKO; PHILBIN, 2015)

Revista ARACÊ
ISSN: 2358-2472

O FaceNet possui aplicações em sistemas biométricos, autenticação em dispositivos móveis e em segurança, nos sistemas de vigilância e monitoramento. Em testes realizados pelos autores, foram obtidas acurácias de 99,63% e 95,12% nas bases de dados do LFW (Labeled Faces in the Wild) e Youtube Faces DB (SCHROFF; KALENICHENKO; PHILBIN, 2015), o que evidencia a capacidade que o modelo possui na identificação e comparação de faces semelhantes. Tal eficiência vem com o custo mais elevado de pré-processamento, tendo em vista a qualidade de alinhamento necessária para que o modelo trabalhe com precisão.

2.4.4 OpenFace

O OpenFace é um modelo de reconhecimento facial de redes neurais com implementação baseada em Python e Torch. Ele é capaz de realizar detecção dos atributos faciais, realizar estimativas relacionadas a posição da cabeça do indivíduo e direção do olhar, dentre outros. É uma ferramenta flexível e capaz de ser desempenhada em tempo real, eliminando a necessidade de implementação de um ambiente com hardware especializado.

Desenvolvido por (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) e inspirado pelo modelo do FaceNet, utiliza redes neurais siamesas, que são projetadas para verificar a similaridade entre duas imagens, para aprender sobre os atributos das faces que são submetidas. Utiliza Torch, e seu treinamento foi feito utilizando o FaceNet para evitar o treinamento de grandes volumes de dados, já que possui foco em execução em tempo real, como dispositivos móveis. Seu treinamento gera 128 embeddings faciais representando atributos das faces, além de compactar os dados para tornar seu desempenho geral mais rápido.

2.4.5 VGG-Face

O VGG-Face é um modelo de aprendizado profundo baseado em redes neurais, também projetado para a tarefa de reconhecimento facial. Utilizando a arquitetura do grupo VGG (Visual Geometry Group), realiza pequenas convoluções de 3x3, extraindo atributos mais detalhados e discriminativos ao longo do processo de reconhecimento. Assim como o FaceNet e o OpenFace, esses atributos também são compactados nas chamadas embeddings faciais.

Dessa forma, as embeddings facilitam a comparação entre a face encontrada na imagem de entrada com uma face já conhecida no banco de dados. Ao calcular a distância entre os embeddings, é possível determinar a similaridade entre as faces, tornando o processo de reconhecimento facial mais ágil. Por conta disso, o VGG-Face também é mais adequado para aplicações voltadas para execução em tempo real, como segurança pública, autenticação e sistemas de monitoramento.

Revista ARACÊ
ISSN: 2358-2472

2.5 AMBIENTE DE EXECUÇÃO

Nos sistemas de reconhecimento facial, todo o processo pode ser feito completamente tanto localmente quanto pela nuvem. No primeiro caso, o sistema é capaz de realizar o reconhecimento diretamente das residências e estabelecimentos, oferecendo vantagens em termos de velocidade de detecção e proteção das imagens sensíveis dos usuários que eventualmente sejam submetidos ao sistema, tendo em vista que não há o compartilhamento de dados com a rede. No entanto, como a capacidade de detecção está diretamente ligada ao dispositivo local, isso representa uma limitação para a escalabilidade do sistema que pode possuir uma base de dados mais restrita em comparação com soluções em nuvem que possuem infraestruturas mais completas.

Por outro lado, um ambiente de execução em nuvem contrasta com os benefícios oferecidos por um ambiente com execução local. Utilizar os dados sensíveis de usuários sem a sua permissão pode trazer empecilhos jurídicos, de modo que abre a possibilidade do sistema infringir a lei geral de proteção de dados pessoais (LGPD) (DIáRIO OFICIAL DA UNIãO, 2018). Porém, como o sistema não depende apenas do hardware do dispositivo local, o sistema pode usufruir de recursos mais escaláveis que a nuvem pode oferecer, com maior base de dados e a possibilidade de utilizar diversos modelos de reconhecimento facial em conjunto.

O uso de ambientes com execução em nuvem em soluções voltadas para vigilância e segurança também promove mais versatilidade e poder de processamento. Tendo em vista a vasta quantidade de dados necessária para um funcionamento ideal, a nuvem permite o uso de modelos mais robustos em seu funcionamento, como MTCNN, FaceNet, dentre outros. Essas características simplificam o processo de manutenção e integração de novas features, tornando o sistema mais preciso, eficiente e a par dos avanços tecnológicos.

Outro fator a se considerar é a capacidade de acesso à rede necessária para sua implementação. Levando em conta escalabilidade, execução local promove mais confiabilidade tendo em vista que o sistema não sofre com as instabilidades e ausência de rede que algumas áreas possuem. Sistemas em nuvem também podem sofrer com tempos de respostas mais elevados já que necessitam de boa comunicação com o servidor, o que pode impactar sua eficiência.

2.5.1 Dispositivos Embarcados

Sistemas embarcados (embedded systems) são aqueles em que o computador encontra-se completamente encapsulado ou dedicado a um dispositivo, possuindo um microprocessador para realizar o controle (SOUZA, 2018). São desenvolvidos para executar uma tarefa específica, sendo otimizado a fim de reduzir tamanho, custos computacionais e do produto, facilitando a conexão e uso

Revista ARACÊ
ISSN: 2358-2472

nos mais diversos tipos de projeto. Possui aplicação em praticamente todos os ramos tecnológicos, como aparelhos móveis, televisores, automóveis, dentre outros.

Os microcontroladores são circuitos semicondutores encapsulados reduzidos a tamanhos físicos muito pequenos que integram CPU, memória e dispositivos E/S, e que realizam conexões através de terminais, voltados para o controle de sistemas embarcados. Esse conjunto formado pela CPU, memória e periféricos é descrito como uma integração (embedded) em um mesmo chip, o que torna os microcontroladores ideais para aplicações em que o custo do produto e espaço ocupado são considerações mais relevantes do que poder computacional (MAZIDI; MCKINLAY; CAUSEY, 2008). O Arduino e o Raspberry Pi são algumas das placas disponíveis e acessíveis para testes.

A Figura 18 ilustra um Raspberry Pi, que é um microcomputador de baixo custo descrito como SBC (Single-Board Computer), com grande versatilidade e aplicações em aprendizagem, automação e robótica. Sua proposta é oferecer poder computacional limitado acessível, com desenvolvimento focado em tornar sistemas com excelente custo-benefício e eficiência (GUPTA et al., 2016). Devido a essas características e o vasto suporte oferecido, o Raspberry Pi é amplamente utilizado em aplicações voltadas à educação e prototipagem, o tornando poderoso nas tarefas de reconhecimento facial quando atrelado a ferramentas de visão computacional. Apesar da limitação de processamento que o Raspberry possui, é capaz de agir com eficiência em sistemas de segurança, monitoramento e vigilância. Modelos de detecção e reconhecimento facial menos complexos são capazes de atuar com bom desempenho, ideal para cenários com recursos limitados. Além disso, ao incorporar com outros sensores e periféricos, como GPUs, ao seu funcionamento, o potencial do sistema pode ser bastante aprimorado.

Figura 5 – Raspberry Pi 4

Fonte: (RASPBERRY PI TRADING, 2019)

Revista ARACÊ
ISSN: 2358-2472

2.6 DATASETS

Datasets são conjuntos dados utilizados em projetos de aprendizado profundo e machine learning para que algoritmos consigam aprender e realizar a função a qual são propostos. Geralmente dispostas em formato tabular, organizados em formato de linhas e colunas, são exportados em formatos que variam desde os arquivos de texto TXT até os arquivos CSV (Comma-Separated Values). Os dados são processados por bibliotecas como o Pandas do Python, que oferecem vasta gama de ferramentas para realizar o processamento e visualização. Além disso, a base normalmente é ramificada para que um conjunto de dados possa ser utilizado no treinamento dos algoritmos e um outro conjunto durante as etapas de teste, podendo até utilizar dados de outras fontes após a etapa de treinamento.

No tocante reconhecimento facial, os datasets são elementos fundamentais para o treinamento e avaliação dos modelos. Para um treinamento eficiente dos modelos é necessário datasets bem estruturados e que melhor se encaixem na tarefa proposta pelo sistema, devendo incluir quantidade de imagens suficientes, variações de poses, expressões, além de diversidade nas imagens em relação à idade, etnia e gênero. Algumas das bases de dados amplamente utilizada nas tarefas de detecção e reconhecimento facial incluem:

2.6.1 LFW (Labeled Faces in the Wild)

O LFW (HUANG et al., 2007) é uma base de imagens de faces voltada para a tarefa de reconhecimento facial. É composto por mais de 13.000 imagens de faces da web com mais de 5.000 identidades, contendo anotações para o nome das pessoas de cada imagem, além de garantir que ao menos 1.680 pessoas possuem uma ou mais fotos em toda a base. Além disso, possui imagens em situações não controladas, com variações de iluminação, pose e expressão facial, sendo uma boa escolha para o treinamento de sistemas voltados para vigilância. É um dataset referência na realização de avaliações dos modelos de reconhecimento facial, voltado para estudos e pesquisas que possam contribuir para avanço das técnicas de verificação facial.

2.6.2 IMDB-WIKI

É comum os datasets possuírem tamanhos pequenos a médio. Nesse âmbito, o IMDBWIKI (ROTHE; TIMOFTE; GOOL, 2018) tem a proposta de fugir desse padrão coletando imagens de mais de 100.000 celebridades dos sites do IMDb e Wikipedia agrupando os dados de nome, gênero, e data de nascimento de todas as imagens. O resultado disso é uma base composta por mais de 500.00 imagens com grande diversidade idade, gênero e etnia, com foco em especial para as tarefas de

Revista ARACÊ
ISSN: 2358-2472

predição de idade e análise demográfica. No entanto, a grande quantidade de dados e seus mais de dez atributos presentes nas anotações permitem que a base também possa ser utilizada nas tarefas de reconhecimento facial.

2.6.3 CelebA (CelebFaces Attributes Dataset)

O CelebA (LIU et al., 2015) contém mais de 200.00 imagens de celebridades de modo que cada imagem contém mais de 40 atributos de anotações. A base possui grande quantidade de imagens e anotações, assim como vasta diversidade de indivíduos com imagens contendo variações de pose, ambiente, expressões faciais, acessórios, sendo bastante utilizada para os estudos de atributos faciais devido a riqueza das anotações. A base pode ser implementada para os conjuntos de dados de treino e teste para diversas tarefas de visão computacional como detecção facial, reconhecimento de atributo facial, reconhecimento facial, identificação de pontos chaves, dentre outros.

2.6.4 VGGFace

O VGGFace (PARKHI; VEDALDI; ZISSERMAN, 2015) contém mais de 2,6 milhões de imagens de faces coletadas de mais de 2.622 identidades contendo anotações para localização de cada face e coordenadas das bounding boxes. Criado pela equipe da VGG (Visual Geometry Group), a base é voltada para o treinamento de redes neurais em reconhecimento facial, sendo fundamental no desenvolvimento e treinamento de diversos modelos, como o FaceNet e o ArcFace. O VGG trabalha sobre a suposição de que as identidades de celebridades encontradas são públicas, fornecendo uma grande quantidade de imagens e dados para fins de estudo, o tornando ideal para as tarefas de reconhecimento facial.

2.6.5 UTKFace

O UTKFace (ZHANG et al., 2017) é uma base que possui mais de 20.000 imagens de faces que variam de idades entre 0 e 116 anos, contendo anotações para a própria idade, assim como gênero etnia de cada uma das imagens. As imagens contém variação de pose, iluminação, resolução, oclusão, expressão facial, dentre outros, além das imagens alinhadas e cortadas com seus respectivos pontos chaves anotados. O dataset possui foco nas tarefas de detecção e reconhecimento facial voltado para predição de idade, além de análise e avaliação de precisão dos grupos populacionais.

Revista ARACÊ
ISSN: 2358-2472

2.6.6 Wider Face

O Wider Face (YANG et al., 2016) é uma base composta por imagens selecionadas de um outro dataset público, o Wider. A base contém 32.203 imagens com mais de 393.000 faces, variando em escala, pose, oclusão, maquiagem, iluminação, além das anotações das coordenadas de cada uma das faces presentes. O Wider Face é organizado em 61 classes de evento de modo que para cada classe são selecionados aleatoriamente 40% dos dados para serem utilizados no treinamento, 10% para validação e 50% para teste. Devido ao grande volume de faces presente nas imagens da base, ela se mostra excelente para a realização de testes e avaliação de modelos de detecção mais complexos, como é o caso do MTCNN e RetinaFace.

2.6.7 FDDB (Face Detection Dataset and Benchmark)

O FDDB (JAIN; LEARNED-MILLER, 2010) é uma ramificação do conjunto de dados do LFW, contendo 5.171 faces em um total de 2.845 imagens. É uma base voltada para o estudo de detecção facial em ambientes não controlados, com as faces anotadas em formato de elipses a fim de tornar a tarefa de detecção mais árdua para os modelos. A base também fornece imagens com variações de ângulo, oclusão e iluminação, sendo ideal para avaliação de desempenho de modelos utilizando as métricas da seção seção 2.7 como metodologia de avaliação.

2.7 AVALIAÇÃO E MÉTRICAS

As métricas são métodos de avaliação utilizados para garantir a eficiência de diversas tarefas de visão computacional, entre elas detecção e reconhecimento facial. Elas são utilizadas para obter parâmetros comparativos, além de medir a qualidade dos modelos, avaliando o poder de detecção das faces ou a capacidade que de um modelo identificar corretamente faces de indivíduos. Para isso, verdadeiros e falsos positivos são utilizados para o cálculo, atingindo resultados mais significativos a depender do modelo, tarefa, custo de diferentes classificações incorretos e se a base de dados está em equilíbrio com o objetivo do trabalho (GOOGLE, 2024). A seguir são apresentados os principais conceitos relativas à avaliação de modelos:

- Verdadeiro Positivo (VP): Indica uma predição correta que ocorre quando realmente há uma face real em uma detecção realizada pelo modelo;
- Verdadeiro Negativo (VN): Indica uma predição correta que ocorre quando o modelo identifica corretamente a inexistência de face em uma região ou por toda a imagem;
- Falso Positivo (FP): Indica uma predição incorreta que ocorre quando não há uma face real em uma detecção realizada pelo modelo;

Revista ARACÊ
ISSN: 2358-2472

- Falso Negativo (FN): Indica uma predição incorreta que ocorre quando o modelo falha e é incapaz de detectar uma face existente na imagem.

2.7.1 Precisão

Precisão é a proporção de todas as classificações positivas do modelo que realmente são positivas, ou seja, a proporção de verdadeiros positivos (VP) entre todas as detecções realizadas pelo modelo, indicando quantos das faces detectadas são realmente verdadeiras. Matematicamente, é definido como:

$$Precisão = \frac{VP}{VP + FP}$$

2.7.2 Recall

É a proporção de positivos que foram classificados corretamente como positivos, ou seja, a proporção de verdadeiros positivos (VP) em relação ao total de casos positivos, indicando quantos das faces reais existentes na imagem foram de fato detectadas. O Recall é definido como:

$$Recall = \frac{VP}{VP + FN}$$

2.7.3 Acurácia

Acurácia é a proporção de todas as classificações corretas, verdadeiros positivos e verdadeiros negativos. Normalmente mede a capacidade do modelo de associar corretamente faces à verdadeira identidade, sendo a principal métrica para tarefas não genéricas ou não especificadas. Entretanto, é indicado utilizar também outras métricas para casos do mundo real, onde o conjunto de dados está desequilibrado. É definida como:

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

2.7.4 F1-Score

Proporciona equilíbrio entre as métricas de precisão e recall, calculando a média harmônica entre ambas as métricas. O F1-Score leva em consideração tanto falsos positivos (FP) quanto falsos negativos (FN) e é definido como:

Revista ARACÊ
ISSN: 2358-2472

$$F1 - Score = 2\,x\,\frac{Precisão\ x\ Recall}{Precisão + Recall}$$

3 DESENVOLVIMENTO

3.1 DATASET

Inicialmente, no escopo anterior do trabalho, foi pensado utilizar o VIRAT Video Dataset. O dataset consiste em vídeos gravados ao ar livre de ações corriqueiras do dia a dia e outras gravações feitas por aeronave não tripulada, possuindo mais de 20 tipos de eventos distribuídos por mais de 29 horas de vídeo (OH et al., 2011), projetado principalmente para a análise de atividades em vídeo. Apesar disso, não mostrou-se ideal para as tarefas de detecção e reconhecimento facial pois seu foco não está na captura de faces e sim no monitoramento e vigilância de atividade humana, com faces capturadas a longas distâncias, o que dificulta tanto sua detecção quanto reconhecimento. Diante disso, o VIRAT mostrou-se inadequado para o escopo proposto pelo trabalho.

Em seguida, o LFW (Labelled Faces in the Wild) foi considerado, dado sua reputação e eficiência nas tarefas de reconhecimento facial. No entanto, a ausência de anotações para as bounding boxes revelou-se um grande empecilho para as tarefas de detecção facial. Isso ocorre porque o LFW foi projetado principalmente para reconhecimento, não possuindo anotações para as coordenadas das faces das suas mais de 13.000 imagens, o que inviabiliza seu uso para as tarefas de detecção (HUANG et al., 2007). Tentou-se realizar a anotações manualmente, mas para o escopo dos testes optou-se por seguir com um outro dataset, tendo em vista que rotulá-las manualmente afetaria a consistência e qualidade dos resultados.

Diante disso, o dataset escolhido foi o IMDB-WIKI, contendo mais de 500.000 imagens com fotos de celebridades do IMDB e da Wikipedia. Porém, o grande volume de imagens mostrou-se desafiador, aumentando o tempo necessário para processar todas as tarefas, incluindo ajustes, pré-processamento e redimensionamento. Considerando as limitações computacionais, decidiu-se então utilizar apenas a parcela WIKI do dataset, que contém no total 62.328 imagens, reduzindo em quase 90% o número de imagens.

Durante o processo de análise da parcela WIKI, observou-se uma série de defeitos no dataset. As imagens não possuíam padronização de tamanho e outras estavam corrompidas ou completamente ausentes de conteúdo relevante para ambas as tarefas de detecção e reconhecimento. Além disso, verificou-se que muitas das anotações das bounding boxes eram imprecisas, resultando em detecções falhas e métricas inconsistentes durante os testes. Esse cenário revelou que, apesar de sua abundância,

Revista ARACÊ
ISSN: 2358-2472

o IMDB-WIKI não oferecia o necessário para testar os modelos de detecção e reconhecimento facial, sendo voltado para outras atividades de aprendizado profundo (ROTHE; TIMOFTE; GOOL, 2018).

Posteriormente, outros datasets foram contemplados, como o VGGFace, UTKFace e WiderFace. O VGGFace (PARKHI; VEDALDI; ZISSERMAN, 2015) também contém milhares de imagens de rostos com diversas anotações, mas enfrentou problemáticas similares às encontradas no IMDB-WIKI, como o grande volume de imagens e a variabilidade nas condições de captura. O UTKFace (ZHANG et al., 2017) não conta com informações dos nomes dos indivíduos ou anotações detalhadas de bounding boxes, o que o torna inadequado tanto para detecção quanto para reconhecimento. O WiderFace (YANG et al., 2016) também faltam informações essenciais como os nomes das pessoas, inviabilizando seu uso para a tarefa de reconhecimento facial.

Diante dessas dificuldades, chegou-se à conclusão de utilizar dois datasets distintos: um para detecção e outro para reconhecimento. Para detecção, o FDDB (Face Detection Data Set and Benchmark) foi selecionado devido à qualidade das anotações das bounding boxes assim como seu tamanho moderado, facilitando o processamento e atendendo às necessidade de um ambiente com recursos limitados. O dataset possui cerca de 5.171 faces em um total de 2.845 imagens, com diversidade de imagens em ângulos e condições desafiadoras (JAIN; LEARNED-MILLER, 2010).

Embora o LFW tenha se mostrado inadequado para a detecção facial, decidiu-se seguir com o mesmo para a tarefa de reconhecimento também pela qualidade das anotações, mas dessa vez voltadas para identificação dos indivíduos, como nome, mas também pelo seu tamanho moderado. A combinação desses dois datasets, evitam muitos dos problemas enfrentados e potencializa os resultados dos testes, tornando mais dinâmico e menos custoso computacionalmente.

3.2 PREPARAÇÃO DO AMBIENTE

Inicialmente, foi contemplada a ideia de conduzir os testes deste trabalho em um ambiente mais prático, utilizando Raspberry Pi. No entanto, visando a otimização de tempo, os experimentos foram realizados em um ambiente local configurado em um computador pessoal. A máquina possuía um poder computacional razoável, com 8 GB de memória RAMe um processador Ryzen 5 5600g com vídeo integrado.

A ausência de uma placa de vídeo dedicada e mais memória, entretanto, representou uma limitação para o trabalho. Muitos dos datasets abordados na seção 2.6 possuem grandes volumes de dados, exigindo maior poder de processamento, o que resultou em tempos de execução ainda mais longos. Além disso, os 8 GB de RAM disponíveis limitaram a capacidade de carregar e processar

Revista ARACÊ
ISSN: 2358-2472

imagens em lotes maiores, forçando a divisão dos dados em batches menores e aumentando o tempo total de execução.

Para a implementação dos modelos, foi utilizada a versão 3.10.9 do Python, juntamente com bibliotecas fundamentais como TensorFlow, PyTorch, NumPy, OpenCV, Matplotlib e Pandas. As bibliotecas foram utilizadas para a implementação dos modelos, assim como manipulação, análise e pré-processamento das imagens.

O Jupyter Notebook foi escolhido como ambiente de desenvolvimento devido à praticidade que oferece para ajustes no código e ao seu feedback mais visual, que foi útil durante o desenvolvimento e a análise inicial dos modelos. Contudo, para os testes em si e a coleta de métricas de desempenho, optou-se por executar os códigos diretamente no terminal para que o computador operasse com o mínimo possível de aplicações em execução, reduzindo o impacto de consumo de memória e processamento durante os experimentos.

3.3 PREPARAÇÃO DE MODELOS

Buscando garantir a uniformidade dos dados de entrada, a etapa inicial consistiu na preparação dos modelos realizando o pré-processamento das imagens de ambos os datasets. Todas as imagens foram redimensionadas para a dimensão padrão de 224x224 pixels buscando reduzir o custo computacional, além de padronizar a entrada de dados, potencializando o desempenho dos algoritmos.

Além disso, foi aplicado um filtro gaussiano às imagens visando suavizá-las e reduzir ruídos. Adicionalmente, visando eliminar informações desnecessárias durante a execução, imagens corrompidas (ou com anotações inconsistentes fornecidas pelos autores dos datasets) foram descartadas para evitar que comprometesse o desempenho dos modelos e que as métricas refletissem seu desempenho real.

A implementação dos modelos baseou-se em bibliotecas como TensorFlow e PyTorch. No entanto, durante esta etapa alguns desafios surgiram ao tentar implementar alguns modelos, como o SSD e ArcFace, devido a exigências destes com versões específicas de bibliotecas não compatíveis com o restante do ambiente, especificamente o TensorFlow, dificultando a execução de alguns testes planejados, uma vez que o ajuste de um modelo poderia inviabilizar outros. Para contornar essas dificuldades, optou-se por focar nos modelos que apresentaram maior viabilidade de implementação sem comprometer a integridade do ambiente configurado de modo que fosse possível otimizar a execução do trabalho durante o tempo proposto.

Revista ARACÊ
ISSN: 2358-2472

3.4 EXECUÇÃO DOS TESTES

Os modelos utilizados neste trabalho foram implementados já pré-treinados, pulando a etapa de treinamento. Levou-se em consideração esta estratégia por otimização de tempo, assim como o objetivo de testar o potencial dos modelos frente aos datasets escolhidos, colhendo resultados que refletissem diretamente seus desempenhos. Dessa maneira, foi possível colher as métricas e analisar de maneira mais objetiva os pontos em que cada modelo se destacam.

Diante disso, os modelos avaliados obtiveram os resultados dispostos nas Tabela 1 e Tabela 2, com variações voltadas a maneira como são implementados assim como o contexto em que estão sendo testados, no caso deste trabalho, os datasets do LFW e FDDB. Os resultados obtidos também destacam as capacidades dos modelos nas tarefas de detecção e reconhecimento facial, mas também é possível discutir sobre suas aplicações em cenários mais próximos do mundo real, como é o caso do contexto de segurança pública.

Tabela 1 - Resultados dos modelos de Detecção Facial

Modelo	Precisão	Recall	Acurácia	F1-Score
MTCNN	0,92701	0,97021	0,94355	0,98488
MediaPipe	0,99311	0,92032	0,89623	0,95851
RetinaFace	0,99474	0,65925	0,65922	0,79463

O MTCNN obteve o melhor resultado geral, atingindo valores superiores a 92% em todas as métricas, evidenciando que o mesmo foi capaz de gerar poucos falsos positivos, localizando boa parte das faces que constituem o dataset do LFW. O desempenho do MTCNN destaca o bom funcionamento de sua cascata de redes P-Net, R-Net e O-Net, passando através da identificação das faces candidatas, pela estruturação das bounding boxes até os refinamentos finais. Portanto, é possível afirmar que a estratégia proposta pelo modelo promove uma detecção mais refinada, não sendo tão afetada pelas variações que eventualmente ocorrerão em cenários cotidianos, como pose e iluminação.

Entretanto, observa-se que apesar das métricas altas a precisão foi a menor encontrada, em comparação com os outros algoritmos. É possível justificar tal fato afirmando que o MTCNN é mais dependente da qualidade das imagens de entrada, sendo necessário imagens com os atributos faciais mais nitídos para que a rede P-Net possa realizar uma identificação inicial das faces satisfatória, algo que possivelmente o LFW não ofereça em toda sua extensão do dataset.

Em contrapartida, os resultados do MediaPipe mostram que o modelo é um pouco mais conservador que o MTCNN. Durante a detecção, o modelo foi o segundo mais preciso, com poucos falsos positivos, porém, obteve recall com um valor 92%, o que sugere um cuidado maior para evitar previsões incorretas. Esse comportamento pode ser explicado por sua arquitetura que se desdobra

Revista ARACÊ
ISSN: 2358-2472

através de um pipeline utilizando redes neurais, sendo otimizado para contextos voltados a execução em tempo real.

Essa abordagem é vantajosa em aplicações para dispositivos embarcados e móveis, como mencionado na subseção 2.3.5, mas pode não ser ideal para detecção em contextos mais desafiadores, como monitoramento voltado à segurança pública. Nessas situações, é comum que as imagens de entrada se apresentem em uma vasta gama de variações tais como iluminação, nitidez, ângulos incomuns, dentre outros, que dificultariam a atuação do modelo. Portanto, o modelo mostrou-se confiável em contextos em que precisão e redução de falsos positivos sejam prioridades.

O RetinaFace obteve a precisão mais alta entre os modelos analisados com 99%, evidenciando sua habilidade em reduzir falsos positivos. Contudo, o recall foi consideravelmente inferior, sugerindo que uma parcela significativa de rostos nas imagens não foram identificados, também gerando valores de acurácia e F1-Score beirando os 65% e 79% respectivamente, indicando um equilíbrio entre acertos e erros na detecção. O desempenho inferior ao esperado pode ser justificado pelo fato de ser um modelo desenvolvido para capturar minuciosamente detalhes faciais e enfrentar desafios como iluminação desfavorável e diversas expressões faciais. Logo, essa especialização pode comprometer sua performance em situações com imagens de baixa qualidade ou com vários rostos pequenos, subestimando a existência de faces na imagem.

Uma outra possível explicação para os resultados obtidos pelo RetinaFace pode estar relacionada ao formato das anotações do dataset FDDB, que utiliza bounding boxes em formato elíptico. Apesar de as anotações terem sido convertidas para o formato esperado, essa transformação pode ter introduzido inconsistências. Assim como todos os modelos avaliados, o RetinaFace também é dependente da precisão das anotações para suas tarefas de aprendizado profundo, o que pode ter levado a interpretações incorretas das regiões faciais, impactando negativamente as métricas mencionadas.

3.4.1 Avaliação dos Modelos de Reconhecimento Facial

Tabela 2 – Resultados dos modelos de Reconhecimento Facial.

Modelo	Precisão	Recall	F1-Score	Acurácia	Tempo Médio (s)
VGG-Face	0,90327	0,78075	0,87684	0,78549	0,34730
FaceNet	0,85068	0,67107	0,80313	0,67895	0,35880
OpenFace	0,74015	0,22748	0,44302	0,23221	0,15730

O VGG-Face se sobressaiu com o desempenho mais sólido entre os modelos de reconhecimento facial analisados. Obteve precisão 90%, seguindo o padrão dos outros algoritmos

Revista ARACÊ
ISSN: 2358-2472

analisados, com sua arquitetura demonstrando eficácia no contexto empregado, com sua metodologia fundamentada em grandes e variados conjuntos de dados, o que muito provavelmente favoreceu sua habilidade de identificar mais atributos faciais. No entanto, o restante das métricas foram abaixo do ideal, um comportamento que foi observado em todos os modelos avaliados.

Uma das razões que pode explicar esse fenômeno reside nos obstáculos apresentados pelo dataset LFW, que possui imagens com grande variedade de expressões faciais, iluminação e ângulos, que podem oferecer resistência ao desempenho dos modelos. Ainda assim, é importante destacar o tempo médio de conclusão favorável, mesmo diante de um cenário desafiador, evidenciando a adaptabilidade do VGG-Face nos contextos em que for implementado.

O desempenho do FaceNet foi um pouco inferior ao do VGG-Face, mas ainda satisfatório, com o modelo obtendo 85% de precisão e F1-Score, enquanto que recall e acurácia foram de 67% e 80%, respectivamente. O FaceNet destacou-se por sua versatilidade, se mostrando um modelo consistente durante a avaliação, equilibrando custo computacional e bons resultados. Dessa maneira, considerando a necessidade de otimizar recursos, o FaceNet mostrou-se um modelo capaz de atuar nos sistemas de monitoramento voltado a segurança pública.

Entre os modelos avaliados, o OpenFace teve o desempenho mais modesto, com uma precisão de 74%, recall de 22% e F1-Score de 44%. Esses valores refletem que, embora o modelo tenha conseguido evitar muitos falsos positivos, ele enfrentou dificuldade considerável em identificar verdadeiros positivos. É provável que a causa disso tenha sido durante sua implementação, relacionado a bibliotecas necessárias para um bom funcionamento do modelo, indicando que com a realização de otimização e ajustes no código ou configurações de suas dependências o modelo possa desempenhar de maneira mais eficiente.

No entanto, no contexto geral da avaliação, e levando em conta o desempenho dos três modelos, é importante levar em consideração que os testes tiveram como foco o desempenho do modelo sobre unicamente a base utilizada, com ausência de uma etapa voltada ao treinamento. Além disso, as anotações presentes nos datasets podem divergir com os resultados obtidos pelos modelos, afetando o desempenho final durante a coleta das métricas.

4 CONCLUSÃO

O objetivo deste trabalho foi explorar e discutir conceitos que incorporam a tarefa de reconhecimento facial voltado para redes neurais convolucionais, assim como implementar e avaliar os desempenhos de alguns modelos. Para isso, foram analisadas algumas das abordagens mais populares, com foco em modelos pré-treinados, sem um treinamento específico no contexto deste

Revista ARACÊ
ISSN: 2358-2472

trabalho, a fim de avaliar o desempenho e eficácia das técnicas de detecção e reconhecimento facial em um ambiente local simulado, visando se aproximar de um contexto mais prático, similar ao encontrado por câmeras de segurança pública.

No escopo inicial do trabalho, a ideia era a construção de um sistema voltado para segurança pública, capaz de realizar a identificação de suspeitos relacionados à atividade criminosa, utilizando as câmeras de estabelecimentos residenciais e comerciais. Porém, essa abordagem mais ambiciosa deu espaço a outra com mais foco na discussão e avaliação de desempenho de modelos, visando uma análise das potencialidades, limitações e aplicabilidades destes algoritmos no contexto de segurança pública.

Para atingir tal feito, uma ampla quantidade de datasets foi testada e implementada. Foram utilizados propriamente os datasets do FDDB para as avaliações de detecção facial e LFW para reconhecimento facial. O dataset FDDB foi escolhido por sua qualidade e variedade para detectar faces em ambientes desordenados, enquanto o LFW forneceu uma boa base para avaliar o reconhecimento de faces.

Os modelos de detecção facial avaliados foram o MTCNN, MediaPipe e RetinaFace, que atingiram resultados variando de altos a mais modestos, dependendo das características e desafios dos respectivos modelos e dataset. O RetinaFace, em particular, foi mais sensível às variações nas anotações do dataset, enquanto o MediaPipe se destacou por sua precisão. Já o MTCNN obteve os melhores resultados, embora ainda tivesse algumas limitações na detecção de faces em ambientes mais complexos.

Já os testes de reconhecimento facial se deram sobre os modelos VGG-Face, FaceNet e OpenFace. O VGG-Face se destacou com boas métricas e melhor desempenho geral, enquanto que o FaceNet, mesmo com dificuldades na identificação de verdadeiros positivos, ainda foi preciso. O OpenFace, por outro lado, teve desempenho abaixo dos demais, com baixa acurácia e recall, possivelmente devido a conflitos entre bibliotecas ou inconsistências com as anotações do LFW.

Em conclusão, este trabalho mostrou o grande potencial que os modelos de detecção e reconhecimento facial avaliados possuem nas mais diversas aplicações. Esses algoritmos possuem qualidade para contribuir significativamente para sistemas de segurança pública, oferecendo uma plataforma eficaz e de alto desempenho e, em alguns casos, de baixo custo computacional. Atrelados a uma etapa de treinamento específica voltada para o contexto em que se é pensado sua implementação, os desempenhos aqui coletados neste trabalho podem ser ainda mais potencializados, contribuindo ainda mais para os sistemas de monitoramento voltados para segurança pública.

Revista ARACÊ
ISSN: 2358-2472

REFERÊNCIAS

ABDALA, V. Over 47 mi may be under facial recognition surveillance in brazil. Rio de Janeiro, Brasil, 2023. Disponível em: . Acesso em: 19 mar 2024.

ALAKKARI, S.; COLLINS, J. J. Eigenfaces for face detection: A novel study. 2013. Disponível em: . Acesso em: 22 mar 2024.

ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. 2017. Disponível em: . Acesso em: 21 mar 2024.

ALSHAMMARI, A.; RAWAT, D. B. Intelligent multi-camera video surveillance system for smart city applications. 2019. Disponível em: <https://ieeexplore.ieee.org/document/8666579>. Acesso em: 12 mar 2024.

AMOS, B.; LUDWICZUK, B.; SATYANARAYANAN, M. Openface: A generalpurpose face recognition library with mobile applications. 2016. Disponível em: . Citado na página 28.

AYADATA. Bounding boxes in computer vision: Uses, best practices for labeling, and more. 2023. Disponível em: . Acesso em: 19 mar 2024.

BELUCO, D. C.; FILHO, J. L. F. Reconhecimento facial aplicado para registro de ponto. 2023. Disponível em: . Acesso em: 19 mar 2024.

CAMARA, G. Do reconhecimento facial: Estudo exploratório e análise comparativa entre brasil e portugal. 2021. Disponível em: . Acesso em: 19 mar 2024.

COCHARD, D. Retinaface: A face detection model for high resolution images. 2024.

DENG, J. et al. Retinaface: Single-shot multi-level face localisation in the wild. 2020.

DIáRIO OFICIAL DA UNIãO. Lei nº 13.709. Lei Geral de Proteção de Dados Pessoas (LGPD). Brasília, DF, 2018. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>.

DOSHI, M. et al. Smart surveillance system using face detection for residentials. 2022. Disponível em: <https://ieeexplore.ieee.org/document/9825346>. Acesso em: 12 mar 2024.

FENG, Z. et al. Face detection, bounding box aggregation and pose estimation for robust facial landmark localisation in the wild. 2017. Disponível em: <https://www.researchgate.net/publication/316788799_Face_Detection_Bounding_Box_Aggregation_and_Pose_Estimation_for_Robust_Facial_Landmark_Localisation_in_the_Wild>. Acesso em: 19 mar 2024.

GEEKSFORGEEKS. Ml | face recognition using eigenfaces (pca algorithm). 2021. Disponível em: <https://www.geeksforgeeks.org/ml-face-recognition-using-eigenfaces-pca-algorithm/>. Acesso em: 22 mar 2024.

Revista ARACÊ
ISSN: 2358-2472

GOOGLE. Guia de soluções do mediapipe. 2021. Disponível em: <https://ai.google.dev/edge/mediapipe/solutions/guide?hl=pt-br>.

GOOGLE. Classificação: precisão, recall, precisão e métricas relacionadas. 2024. Disponível em: <https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?hl=pt-br>.

GUPTA, I. et al. Face detection and recognition using raspberry pi. 2016. Disponível em:<https://ieeexplore.ieee.org/document/8009092>.

HUANG, G. B. et al. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. [S.l.], 2007. Disponível em: <https://vis-www.cs.umass.edu/lfw>.

IDEC. Reconhecimento facial e o setor privado: guia para adoção de boas práticas. 2020. Disponível em: <https://idec.org.br/sites/default/files/reconhecimento_facial_diagramacao_digital_2.pdf>. Acesso em: 19 mar 2024.

ILYAS, M. Iot applications in smart cities. 2021. Disponível em: . Acesso em: 11 mar 2024.

INDOLIA, S. et al. Conceptual understanding of convolutional neural network- a deep learning approach. 2018. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050918308019>. Acesso em: 21 mar 2024.

JAIN, V.; LEARNED-MILLER, E. FDDB: A Benchmark for Face Detection in Unconstrained Settings. [S.l.], 2010. Disponível em: <https://vis-www.cs.umass.edu/fddb/>.

JAN, A. et al. Artificial intelligent system for automatic depression level analysis through visual and vocal expressions. 2017. Disponível em: <https://www.researchgate.net/publication/318798243_Artificial_Intelligent_System_for_Automatic_Depression_Level_Analysis_Through_Visual_and_Vocal_Expressions>.

JOSE, E. et al. Face recognition based surveillance system using facenet and mtcnn on jetson tx2. 2019. Disponível em: <https://ieeexplore.ieee.org/document/8728466>.

LIU, W. et al. Ssd: Single shot multibox detector. 2021. Disponível em: <https://arxiv.org/abs/1512.02325>.

LIU, Z. et al. Deep learning face attributes in the wild. In: Proceedings of International Conference on Computer Vision (ICCV). [s.n.], 2015. Disponível em: <https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>.

LU, W. yao; YANG, M. Face detection based on viola-jones algorithm applying composite features. 2019. Disponível em: <https://ieeexplore.ieee.org/document/8806572>. Acesso em: 21 mar 2024

MANJHI, A. K.; RAJAWAT, A. S.; SRIVASTAVA, A. Design and analysis of programmed face monitoring system. 2023. Disponível em: <https://ieeexplore.ieee.org/document/10370072>.

MAZIDI, M. A.; MCKINLAY, R. D.; CAUSEY, D. PIC Microcontroller and Embedded Systems. [S.l.: s.n.], 2008. 24 p.

Revista ARACÊ
ISSN: 2358-2472

MULTICOMP LAB. Openface 2.2.0: a facial behavior analysis toolkit. 2018. Disponível em: <https://github.com/TadasBaltrusaitis/OpenFace>.

OH, S. et al. Avss 2011 demo session: A large-scale benchmark dataset for event recognition in surveillance video. 2011. Disponível em: <https://ieeexplore.ieee.org/document/6027400>. Acesso em: 26 mar 2024.

ORACLE. O que é iot? 2020. Disponível em: <https://www.oracle.com/br/internet-of-things/what-is-iot/>. Acesso em: 11 mar 2024.

PARKHI, O. M.; VEDALDI, A.; ZISSERMAN, A. Vgg face dataset. 2015. Disponível em: <https://www.robots.ox.ac.uk/~vgg/data/vgg_face/>.

RAJPUT, S. Face detection using mtcnn. 2020. Disponível em: <https://medium.com/@saranshrajput/face-detection-using-mtcnn-f3948e5d1acb>.

RASPBERRY PI TRADING. Raspberry pi 4. 2019. Disponível em: <https://www.raspberrypi.com/products/raspberry-pi-4-model-b/>. Acesso em: 26 mar 2024.

REDDY, N. V. M. C.; KUMAR, K. Comparison of hog and fisherfaces based face recognition system using matlab. 2021. Disponível em: <https://ieeexplore.ieee.org/document/9456366>. Acesso em: 22 mar 2024.

RODRIGUES, S. R. dos S. Desenvolvimento de um sistema de reconhecimento facial. 2020. Disponível em: <https://recipp.ipp.pt/handle/10400.22/17594>. Acesso em: 21 mar 2024.

ROTHE, R.; TIMOFTE, R.; GOOL, L. V. Deep expectation of real and apparent age from a single image without facial landmarks. International Journal of Computer Vision, Springer, v. 126, n. 2-4, p. 144–157, 2018. Disponível em: <https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/>.

SAINI, Y. Face recognition using fisherfaces. 2020. Disponível em: <https://iq.opengenus.org/face-recognition-using-fisherfaces/#google_vignette>. Acesso em: 22 mar 2024.

SALAM, A. et al. The future of emerging iot paradigms: Architectures and technologies. 2019. Disponível em: . Acesso em: 11 mar 2024.

SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. 2015. Disponível em: https://arxiv.org/abs/1503.03832

SILVA, A. L.; CINTRA, M. E. Reconhecimento de padrões faciais: Um estudo. 2015. Disponível em: https://www.researchgate.net/publication/341625381_Reconhecimento_de_padroes_faciais_Um_estudo>. Acesso em: 20 mar 2024.

SOUZA, M. Sistemas operacionais de sistemas embarcados. 2018. Disponível em: <https://medium.com/@matheussouza_42815/sistemas-operacionais-de-sistemas-embarcados-892edfc37cd>.

TIAN, Y.; KANADE, T.; COHN, J. F. Facial expression recognition. 2011. Disponível em: <https://www.researchgate.net/publication/227031714_Facial_Expression_Recognit

Revista ARACÊ
ISSN: 2358-2472

ion>. Acesso em: 20 mar 2024.

TURK, M.; PENTLAND, A. Eigenfaces for recognition. 1991. Disponível em: <https://ieeexplore.ieee.org/document/6793549>. Acesso em: 22 mar 2024.

VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. 2001. Disponível em: <https://ieeexplore.ieee.org/document/990517>. Acesso em: 21 mar 2024.

YAMASHITA, R. et al. Convolutional neural networks: an overview and application in radiology. 2018. Disponível em: https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9>. Acesso em: 21 mar 2024.

YANG, S. et al. Wider face: A face detection benchmark. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [s.n.], 2016. Disponível em: <http://shuoyang1213.me/WIDERFACE/>.

YE, B. et al. Face ssd: A real-time face detector based on ssd. 2015. Disponível em: <https://ieeexplore.ieee.org/document/9550294>.

ZHANG et al. Age progression/regression by conditional adversarial autoencoder. In: IEEE. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. Disponível em:<https://susanqq.github.io/UTKFace/>.

ZHANG, K. et al. Joint face detection and alignment using multitask cascaded convolutional networks. 2016. Disponível em: <https://ieeexplore.ieee.org/document/7553523>.