Reconhecimento facial automatizado de solicitantes de certificados digitais com redes neurais e imagens geradas artificialmente

João Vitor dos Santos Mendonça[1]*; Sandro Ricardo Fuzatto[2]

[1] Tech Lead. Rua Augusto dos Anjos, 225, Ap133 – Melville Empresarial II; 06485-370 Barueri, São Paulo, Brasil
[2] Engº Agrônomo especialista em Genética de Plantas. Rua Phenom, 35 – Portal das Araras; 79644-256, Três Lagoas, MS, Brasil
*autor correspondente: vitorjoaocomp1997@gmail.com

Reconhecimento facial automatizado de solicitantes de certificados digitais com redes neurais e imagens geradas artificialmente

Resumo

Nos últimos anos, com o avanço da digitalização, as transações online tornaram-se comuns em diversas áreas, como comércio, serviços bancários e governamentais. Para garantir a segurança dessas interações e validar assinaturas que antes exigiam presença física, os certificados digitais passaram a ser amplamente adotados, substituindo as assinaturas tradicionais por versões eletrônicas. Contudo, o aumento do uso desses certificados também trouxe um crescimento nas fraudes envolvendo sua emissão, comprometendo a confiança no sistema de certificação digital e criando vulnerabilidades para empresas e indivíduos. Diante desse cenário e inspirado por um desafio da Financiadora de Estudos e Projetos (FINEP), este projeto propos o desenvolvimento de um modelo de redes neurais, treinado com imagens geradas artificialmente e em quantidade limitada, para reconhecimento facial. A proposta consistiu em uma prova de conceito em um ambiente controlado, simulando a análise de dois repositórios de imagens: um representando as fotos dos solicitantes de novos certificados digitais, e outro, apelidado de "lista de bloqueio por fraude", contendo imagens de fraudadores já identificados. O objetivo foi automatizar o processo de análise das imagens dos solicitantes, comparando-as com a "lista de bloqueio por fraude", utilizando uma rede neural pré-treinada e um modelo de rede siamesa. Nesse ambiente controlado, são abordados os desafios relacionados à limitação de imagens e à capacidade de processamento.
Palavras-chave: ICP-Brasil, fraudes, rede siamesa, assinatura digital

Automated Facial Recognition of Digital Certificate Applicants Using Neural Networks and Artificially Generated Images

Abstract

In recent years, with the advancement of digitalization, online transactions have become common in various sectors such as commerce, banking, and government services. To ensure the security of these interactions and validate signatures that previously required physical presence, digital certificates have been widely adopted, replacing traditional signatures with electronic versions. However, the increased use of digital certificates has also led to a rise in fraud related to their issuance, undermining trust in the digital certification system and creating vulnerabilities for companies and individuals. In response to this scenario and inspired by a challenge from the Brazilian Funding Authority for Studies and Projects (FINEP), this project proposed the development of a neural network model trained with artificially generated and limited image datasets for facial recognition. The proposal consisted of a proof of concept in a controlled environment, simulating the analysis of two image repositories: one representing photos of applicants for new digital certificates and another, called the "fraud blocklist," containing images of known fraudsters. The objective is to automate the process of analyzing applicants' images by comparing them to the " fraud blocklist," using a pre-trained neural network and a siamese network model. In this controlled setting, the challenges related to image limitations and processing capacity were addressed.
Keywords: ICP-Brazil, fraud, siamese network, digital signature

Introdução

O uso da tecnologia tem sido uma das principais estratégias para otimizar e aumentar a eficiência na análise de dados. Nesse contexto, a Financiadora de Estudos e Projetos [FINEP] lançou uma série de desafios, identificando diversos problemas que podem ser solucionados com o auxílio de soluções tecnológicas. Entre os desafios propostos, destaca-se a “solução de IA para identificar possíveis fraudadores na cadeia de certificados”, a qual chamou especial atenção em função do aumento significativo de fraudes na emissão de certificados digitais nos últimos anos. Conforme reportado por diversos veículos de comunicação, o sistema de certificação digital apresenta vulnerabilidades na identificação dos reais usuários que assinam digitalmente, tornando-se alvo de fraudes recorrentes (Portal JUS Brasil, 2022).

Conforme Resende (2009), um certificado digital é um documento eletrônico que contém informações do titular, como nome, e-mail, CPF, além de duas chaves criptográficas, pública e privada e a assinatura digital da Autoridade Certificadora [AC] responsável por sua emissão. O uso de certificados digitais tem se mostrado crucial para garantir a segurança nas interações online, como compras, transações bancárias e o envio de e-mails, proporcionando ao usuário maior confiança e uma sensação reduzida de vulnerabilidade em suas atividades virtuais.

As fraudes envolvendo certificados digitais são diversas e podem impactar até mesmo órgãos públicos. Em 2023, um certificado digital foi utilizado para desviar R$ 4 milhões de reais da Justiça do Trabalho, utilizando o certificado de um juiz. Segundo ofício emitido por Edith Tourinho, desembargadora presidente do TRT-1, oito alvarás fraudulentos foram emitidos na 80ª Vara do Trabalho do Rio de Janeiro (Alecrim, 2023). Além disso, informações divulgadas pelo Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil [CERT.br] em 19 de abril de 2023, indicam um aumento significativo nas tentativas de fraude e ataques na internet. O portal VCert reporta um crescimento de 116,3% nas notificações de ataques no primeiro trimestre de 2023, e um aumento de 182,1% em comparação ao mesmo período do ano anterior (VCert, 2024).

A Infraestrutura de Chaves Públicas Brasileira [ICP-Brasil], regulamentada pelo Decreto nº 3.996 de 2001 e pela Lei nº 11.419 de 2006, é responsável por garantir a validade jurídica dos documentos digitais, desempenhando um papel fundamental na segurança das transações eletrônicas (Resende, 2023). Esse sistema hierárquico é gerido pelo Instituto Nacional de Tecnologia da Informação [ITI], vinculado à Casa Civil da Presidência da República, o qual coordena a emissão de certificações no Brasil (Tsukada, 2021).

De acordo com Baldissera et al. (2017), o Instituto Nacional de Tecnologia da Informação [ITI] tem a responsabilidade de revisar, atualizar e, quando necessário, ajustar os procedimentos e práticas estabelecidos para a Infraestrutura de Chaves Públicas Brasileira

[ICP-Brasil]. Além disso, é incumbido de garantir a compatibilidade tecnológica do sistema e promover a atualização de suas tecnologias, assegurando conformidade com as políticas de segurança. No exercício de suas funções, o [ITI] atua também na fiscalização, podendo aplicar sanções conforme previsto em lei. Outro papel relevante do [ITI] é fomentar o desenvolvimento de projetos tecnológicos e pesquisas científicas, com o objetivo de ampliar a cidadania e inclusão digital, além de popularizar o uso de certificados digitais. Sua atuação abrange ainda sistemas criptográficos, softwares livres e hardwares compatíveis com padrões abertos, entre outras atribuições.

Conforme descrito por Gomes et al. (2020), o processo de emissão de certificados digitais para pessoas físicas exige a presença do indivíduo em uma autoridade certificadora para verificação de documentos e coleta de dados biométricos. No entanto, a análise e revogação de certificados fraudulentos ainda ocorre de forma manual, com um tempo médio de processamento entre sete e dez dias, o que abre brechas para a atuação de fraudadores.

Diante desse cenário, este projeto foi desenvolvido com o objetivo de propor um modelo baseado em dois algoritmos de análise de dados: um utilizando a biblioteca “face_recognition” e outro utilizando a biblioteca “TensorFlow”, ambas utilizadas com o Python.

O objetivo deste trabalho foi automatizar o processo de análise de imagens simuladas dos solicitantes de certificados digitais, comparando-as com uma lista de fraudadores simulados, permitindo a identificação automática de padrões suspeitos. Para isso, foram simulados diferentes volumes de dados para treinar o modelo de reconhecimento facial, buscando melhorar a eficiência na identificação automática de padrões suspeitos. Os resultados em ambiente controlado e utilizando um número limitado de imagens gerados artificialmente não representou resultados positivos quanto à identificação de rostos no repositório de fraudadores.

Material e Métodos

Para o desenvolvimento do projeto de análise de imagens voltadas para a detecção de possíveis fraudes na emissão de certificados digitais em ambiente controlado, foi utilizada de maneira adaptada a metodologia “Cross Industry Standard Process for Data Mining” [CRISP-DM]. Esta modelagem é uma abordagem consolidada que transforma dados do negócio em conhecimento e informações gerenciais. Com mais de vinte anos de existência, a [CRISP-DM] surgiu em resposta à necessidade dos profissionais da área de tecnologia da informação, que enfrentavam desafios significativos na conversão de dados em informações úteis, especialmente devido ao elevado volume de dados gerados. Assim, a metodologia

[CRISP-DM] foi desenvolvida para atender projetos que envolvem a análise de grandes volumes de dados, permitindo um processo mais organizado e eficaz na extração de conhecimento (Andrade et al., 2023). As etapas propostas pela metodologia são as seguintes:

1. Compreensão do negócio;
2. Entendimento dos dados;
3. Preparação dos dados;
4. Modelagem;
5. Avaliação;

Conforme descrito, a metodologia foi utilizada como guia para o desenvolvimento das etapas do projeto de maneira adaptável, focando na validação das propostas e não contemplando a aplicação do modelo em um ambiente de produção.

Compreensão do negócio

Conforme apontado em um vídeo disponibilizado no canal da Enap (2023), a abordagem atual do [ITI] é reativa. Isso significa que eles aguardam informações das autoridades certificadoras ou do poder judiciário que indiquem a existência de uma possível fraude em um certificado. A partir dos arquivos enviados pelas Autoridades Certificadoras [ACs], o [ITI] verifica e responde se o certificado é fraudulento ou não. Com a utilização de tecnologias, o objetivo do [ITI] é aumentar a rapidez e a confiabilidade na verificação dessas informações, agilizando o processo de resposta. Esta entrega de valor pela inteligência artificial transformaria a pesquisa de reativa para proativa. Assim, a implementação de um modelo que automatize a análise das imagens dos certificados enviados pelas [ACs], comparando-as com as imagens de fraudadores conhecidos, poderia representar uma contribuição significativa para um sistema mais amplo de automação nas análises realizadas pelo ITI. Com base nas necessidades do Instituto Nacional de Tecnologia da Informação (ITI), considerou-se relevante e oportuno o desenvolvimento de um modelo em ambiente controlado que, utilizando tecnologias de aprendizado de máquina, automatizasse o processo de análise de imagens.

Entendimento dos dados

Para a criação de dados fictícios que refletissem a realidade, foi utilizada a biblioteca “Faker” do “Laravel”, uma ferramenta amplamente adotada para a geração de dados

sintéticos, como nomes, endereços, e-mails, números de telefone e datas (Aramburu, 2024). A flexibilidade dessa biblioteca foi crucial para a construção de uma base de dados estruturada e coerente, permitindo não apenas a geração de informações pessoais como nome e CPF mas também dado, como endereços completos.

As diferentes tabelas criadas no banco de dados foram interligadas por meio de um identificador único, conhecido como [UUID] “Universal Unique Identifier”, denominado [PESID], atribuído a cada pessoa fictícia. Isso garantiu a integridade referencial e a consistência entre as informações, permitindo a organização eficiente dos dados e facilitando futuras consultas e análises. O uso de UUID foi fundamental para evitar colisões de identificadores e assegurar que cada registro fosse único no sistema, uma prática comum em sistemas de grande escala e alta performance.

As variáveis utilizadas no projeto estão na Tabela 1, de acordo com seus respectivos nomes e significados.

Tabela 1. Lista de variáveis qualitativas geradas com o apoio da biblioteca “Faker“ do “Laravel”.

Variável	Significado
pesid	Identificador único da pessoa
nome	Nome da pessoa
data_nascimento	Data nascimento da pessoa
sexo	Gênero da pessoa
endid	Identificador único do endereço
rua	Nome da rua do endereço
cidade	Nome da cidade do endereço
estado	Nome do estado do endereço
cep	Código postal do endereço
Imgid	Identificador único da imagem
telefone	Número de telefone da pessoa
endereco_id	Identificador único do endereço da agência
blid	Identificador único da lista de bloqueio por fraude
Motivo	Razão para inclusão na lista de bloqueio por fraude
data_inclusao	Data de inclusão na lista de bloqueio por fraude

Fonte: Dados originais da pesquisa

Para gerar imagens fictícias de pessoas, utilizou-se a plataforma "This Person Does Not Exist", que emprega Redes Gerativas Adversariais [GANs]. As [GANs] são uma técnica avançada de aprendizado profundo que visa aprender a partir de um conjunto de exemplos reais, gerando novas imagens que seguem a mesma distribuição dos dados originais (Mariz, 2018). A utilização desta tecnologia permitiu a criação de imagens realistas de rostos humanos inexistentes, de forma dinâmica e em tempo real, garantindo que as imagens utilizadas no projeto tivessem alta qualidade e fossem adequadas às necessidades do estudo.

Para validar o conceito proposto, foi criada uma base de dados contendo 300 imagens geradas, equilibradas entre rostos femininos e masculinos. Desse total, 166 eram rostos masculinos e 134 femininos, o que assegurou uma distribuição representativa de gênero. Adicionalmente, 110 dessas imagens foram rotuladas como pertencentes a fraudadores conhecidos, cujas informações foram inseridas em uma “lista de bloqueio por fraude”. Esses fraudadores foram associados às demais informações do banco de dados por meio de um identificador único, o [PESID], que é um “Universally Unique Identifier” [UUID]. Essa abordagem permitiu o vínculo direto entre as imagens e as informações cadastrais fictícias, garantindo a consistência e a rastreabilidade dos dados no sistema.

Na Figura 1, é apresentado o repositório que contém as imagens associadas aos usuários do gênero masculino. Já na Figura 2, observa-se o repositório destinado ao armazenamento das imagens correspondentes aos rostos do gênero feminino. Já na Figura 2, observa-se o repositório destinado ao armazenamento das imagens correspondentes aos rostos do gênero feminino.

Figura 1. Rostos Masculinos
Fonte: Dados originais da pesquisa.

Figura 2. Rostos Femininos
Fonte: Dados originais da pesquisa.

Preparação dos dados

Para a criação das models que representam as tabelas no banco de dados, foi utilizado o “framework Laravel”, no qual foi aplicada a biblioteca “Faker” para popular essas bases de dados com informações fictícias. O ambiente de desenvolvimento escolhido para a codificação foi o Visual Studio Code, proporcionando uma interface eficiente para a implementação do código.

A base de dados foi previamente modelada, de forma que não foi necessário realizar transformações adicionais nos dados antes de seu uso. Para o armazenamento das informações, optou-se pelo uso do “MySQL”, que garantiu a integridade e a escalabilidade necessárias ao projeto.

As bases de dados construídas para este projeto abrangem tanto dados numéricos quanto categóricos, organizados de forma a permitir uma análise eficaz e representativa. Entre os dados numéricos, encontramos variáveis como o “CPF”, que, embora apresente um formato textual, podem ser utilizados para validação de identidade e categorização de usuários. A quantidade de imagens geradas, que totaliza 300, é um exemplo de dado numérico, permitindo quantificar a distribuição entre os gêneros, sendo 166 rostos masculinos e 134 femininos.

Em relação às imagens geradas pela plataforma "This Person Does Not Exist", foi necessário realizar uma triagem manual das imagens, uma vez que algumas apresentavam deformidades que poderiam comprometer a performance do modelo utilizado para reconhecimento facial.

Modelagem

O trabalho prático foi dividido em duas aplicações distintas: uma utilizando a biblioteca “face_recognition” e a outra utilizando a “TensorFlow”. A escolha por essa abordagem mista foi motivada por suas características diferentes. A biblioteca “face_recognition” já oferece um modelo pré-treinado que apenas realiza comparações de similaridade entre imagens faciais, o que elimina a necessidade de treinamento adicional. Essa facilidade permite que o sistema compare diretamente as imagens dos solicitantes com as imagens da “lista de bloqueio por fraude”, tornando o processo mais simples e rápido (Geitgey 2017).

Por outro lado, o “TensorFlow” permite maior flexibilidade e personalização ao permitir o treinamento de modelos a partir dos dados fornecidos. Isso é especialmente útil em cenários onde é necessário ajustar o modelo às especificidades do conjunto de dados em questão, oferecendo maior controle sobre o processo de aprendizado (Falcão, 2019).

Conforme apontado no estudo de Phillip (2012), o problema com a representação de imagens que temos é sua alta dimensionalidade. Imagens em escala de cinza bidimensionais [p×q] abrangem um espaço vetorial de dimensão [m=pq], então uma imagem com [100 × 100] pixels já se encontra em um espaço de imagem de 10.000 dimensões. Isso é excessivo para qualquer cálculo. Só podemos tomar uma decisão se houver alguma variância nos dados, então o que buscamos são os componentes que representam a maior parte das informações. A ideia é que um conjunto de dados de alta dimensão é frequentemente descrito por variáveis correlacionadas e, portanto, apenas algumas dimensões significativas representam a maior parte das informações.

Biblioteca face_recognition

A biblioteca “face_recognition”, portanto foi apropriada para se utilizar de dados fictícios. Este projeto permitiu a geração de apenas uma imagem aleatória por indivíduo, o que significa que as imagens comparadas para a validação do modelo teriam apenas uma representação por pessoa. Conforme descrito por Geitgey (2018), a biblioteca “face_recognition” oferece uma maneira simples de reconhecer e manipular rostos utilizando “Python” ou a linha de comando. Desenvolvida com o avançado modelo de reconhecimento facial da biblioteca [dlib], ela alcança uma impressionante precisão de “99,38%” no “benchmark Labeled Faces in the Wild”. Além disso, a biblioteca inclui uma ferramenta de linha de comando que facilita o reconhecimento facial em um diretório de imagens diretamente pelo terminal, tornando o processo mais acessível e eficiente.

O processo de reconhecimento facial usando a biblioteca “face_recognition” segue três etapas principais. Primeiro, os rostos nas imagens são detectados usando o método “Histogram of Oriented Gradients” [HOG], que converte a imagem em tons de cinza e analisa as mudanças de luminosidade (gradientes) para identificar padrões faciais. Em seguida, os rostos detectados são ajustados utilizando estimativas de pontos de referência faciais, como olhos e lábios, para alinhar os rostos e facilitar a comparação. Finalmente, a fase de codificação transforma os rostos em vetores numéricos “encodings” que podem ser comparados rapidamente para identificar correspondências com base em características faciais únicas, permitindo a classificação e o reconhecimento eficientes (Geitgey, A., 2016).

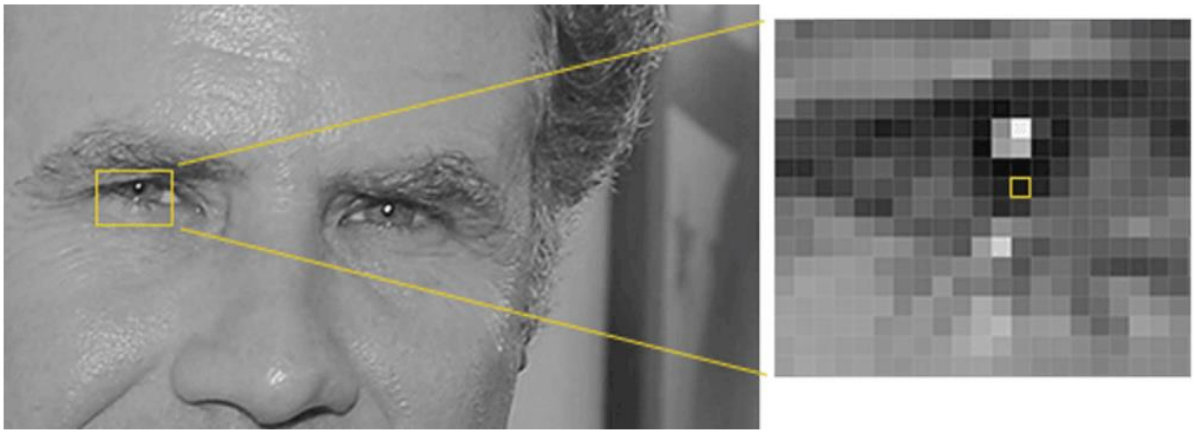

Figura 3. Observação de Pixels [HOG].
Fonte: Geitgey (2016).

Conforme apontado por Dalal et al (2005), o primeiro passo no HOG envolve calcular a intensidade e a direção dos gradientes da imagem. Isso é frequentemente feito usando filtros de Sobel para obter as derivadas nas direções x e y. Os gradientes são então combinados para formar uma imagem de gradiente, onde cada pixel contém a magnitude e a direção do gradiente, conforme a equação (1):

$$Gx = I * \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, Gy = I * \begin{bmatrix} 1 & 2 & 1 \\ 0- & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

$$G = \sqrt{G_x^2 + G_y^2}, \delta = tan^{-1}\begin{pmatrix} G_y \\ G_x \end{pmatrix} \quad (1)$$

em que (G_x) é representa o gradiente horizontal e (G_y) o gradiente vertical, (G) a magnitude de gradiente, a expressão (δ) é usada para calcular a orientação ou ângulo do gradiente m uma imagem, a imagem de gradiente é dividida em células (por exemplo, 6x6 pixels), e um

histograma de orientações é construído para cada célula. Cada célula acumula a magnitude dos gradientes em [bins] de orientação (por exemplo, 9 bins de 20 graus cada). Como a intensidade dos gradientes pode variar significativamente devido a condições de iluminação e contraste, a normalização local é essencial. Isso é feito agrupando células em blocos maiores (por exemplo, 2x2 células) e normalizando as respostas das células dentro de cada bloco. O vetor de descritor final é composto por todos os componentes normalizadas de todas as células em cada bloco.

Biblioteca tensorflow

O “tensorflow” é uma biblioteca de software de código aberto para computação numérica usando grafos computacionais. Foi originalmente desenvolvido pela Google Brain Team na organização de pesquisa em “Machine Intelligence” do Google para aprendizado de máquina e pesquisa de redes neurais profundas “Deep Learning”, mas a biblioteca é geral o suficiente para ser aplicada em uma grande variedade de outros domínios também (DSAcademy, 2023). O “tensorflow” funciona combinando álgebra computacional e técnicas de otimização de compilação, o que facilita o cálculo de muitas expressões matemáticas. Outra vantagem fornecida pelo “tensorflow” é que os resultados do código do programa escrito serão mais fáceis de ler e entender por outros porque ele depende de bibliotecas para executar o Machine Learning (Basurah, 2023).

A aplicação do tensorflow no trabalho foi utilizada para a criação de um modelo de classificação de uma classe ("One-class classification") aplicando redes siamesas (“siamese networks”).

A classificação de uma classe é um tipo específico de classificação multi ou binária onde problema de classificação é resolvido examinando e analisando instâncias de apenas uma classe (Seliya et al, 2021).

Conforme descrito por Ji et al. (2022), a rede neural siamesa é uma arquitetura simples e eficiente. Inicialmente proposta para autenticação de assinaturas de cheques nos Estados Unidos, seu desenvolvimento foi limitado por restrições de hardware até 2010. A partir desse período, a rede neural siamesa passou a ser utilizada no campo do reconhecimento facial, combinando-se com redes neurais convulsionais para aprimorar sua aplicação em tarefas de verificação de identidade.

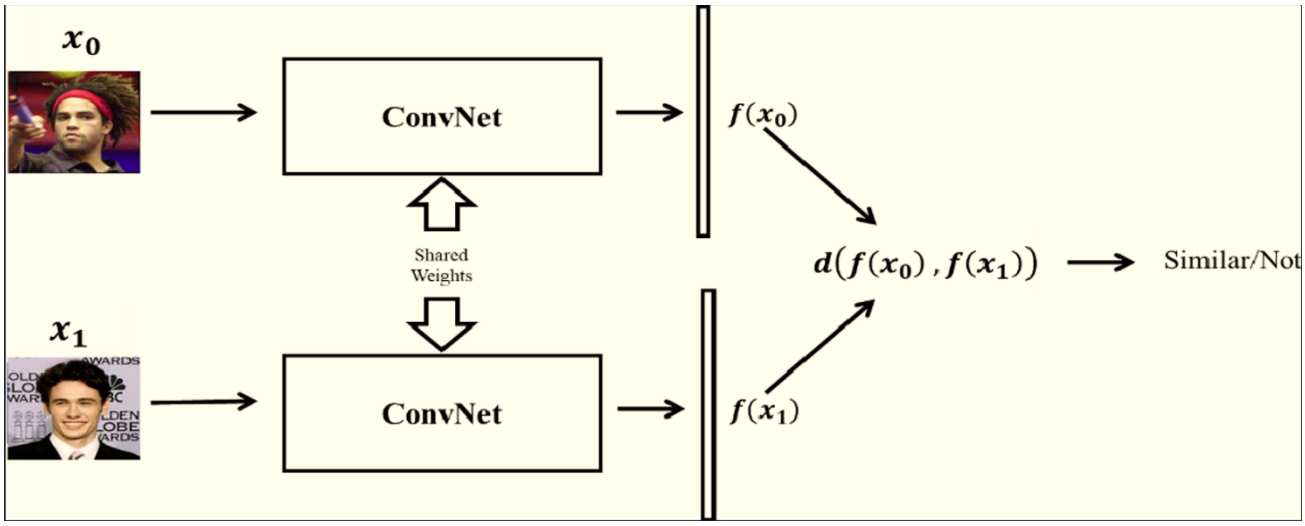

Figura 4. Aplicação do modelo de similaridade.
Fonte: Heidari et al, (2020).

A arquitetura de rede siamesa para reconhecimento facial utiliza (X0, X1) como um par de imagens de entrada, onde (f(X0), f(X1)) são os vetores de características extraídos para esse par de imagens por meio de uma rede neural convolucional. A função de similaridade, d(f(X0), f(X1)), calcula a distância entre os vetores de saída das duas redes (Heidari et al., 2020).

Avaliação

As imagens foram comparadas em um cenário de análise individual, ou seja, uma a uma. O diretório contendo as imagens dos solicitantes foi diretamente confrontado com o diretório “lista de bloqueio por fraude”. Esse processo visava identificar possíveis correspondências entre os solicitantes e as pessoas previamente listadas na lista de bloqueio por fraude.

Biblioteca face_recognition:

Na comparação realizada com a biblioteca face_recognition, cada imagem dos solicitantes foi comparada individualmente com as imagens presentes no diretório 'lista de bloqueio por fraude'. Sempre que uma similaridade era detectada, as informações, como o código do solicitante e o caminho da imagem correspondente, eram registradas na tabela solicitacao_suspeita do banco de dados.

Para a análise utilizando a biblioteca “face_recognition”, foi utilizada uma amostra de 100 imagens na pasta de solicitantes e 10 imagens na pasta “lista de bloqueio por fraude”,

visto que o modelo era pré-treinado. A métrica de avaliação adotada foi a quantidade de falsos positivos.

Biblioteca TensorFlow:

Na abordagem com “TensorFlow”, o treinamento do modelo foi realizado em 5 épocas, com um aumento gradual das amostras de imagens. No primeiro treinamento, foram utilizadas 100 imagens no diretório de solicitantes e 50 no diretório de 'lista de bloqueio por fraude'. Em seguida, foram ampliadas para 200 imagens de solicitantes e 80 de 'lista de bloqueio por fraude' no segundo treinamento. Por fim, no terceiro e último treinamento, foram trabalhadas 300 imagens de solicitantes e 110 no 'lista de bloqueio por fraude'.

Resultados e Discussão

O trabalho, realizado em um ambiente controlado utilizando duas abordagens, uma com um modelo pré-treinado da biblioteca "face_recognition" e outra com o treinamento de um modelo em um conjunto limitado de imagens fictícias teve como objetivo estudar uma solução rápida e eficiente para a proposta de um sistema que, futuramente, poderia evoluir para um modelo mais robusto de análise de imagens. Embora o estudo não tenha contradito essa possibilidade, a limitação dos dados disponíveis não permitiu obter resultados positivos, conforme descrito a seguir.

Resultado utilizando a biblioteca “face_recognition”

O script foi processado conforme as etapas descritas anteriormente, com um repositório representando os solicitantes e outro a “lista de bloqueio por fraude”. A biblioteca “face_recognition” foi utilizada para comparar as imagens de ambos os repositórios, realizando a análise facial e identificando possíveis correspondências. Os suspeitos identificados foram então armazenados na tabela de solicitações suspeitas.

Tabela 2. Tabela de armazenamento dos dados relacionados aos casos suspeitos.

Variável	Significado
Idsolicitacao_suspeita	Identificador único do registro de suspeita
pesid	Identificador único da pessoa
data	Data de inserção da análise suspeita
imagem	Caminho da imagem do suspeito no diretório

Fonte: Dados originais da pesquisa

O script que utiliza a biblioteca "face_recognition" apresentou uma margem alarmante de falsos positivos de 72,87%. Essa taxa é particularmente preocupante, uma vez que a comparação foi realizada com uma “lista de bloqueio por fraude” que contém apenas 10 imagens de fraudadores conhecidos. Ao analisar 100 novas imagens de solicitantes, o sistema identificou 37 casos suspeitos que foram destacados para investigação adicional. Essa situação levanta questões sobre a eficácia e a confiabilidade do modelo, sugerindo que a ferramenta pode não ser adequada para aplicações críticas em ambientes de produção, onde a precisão é fundamental para evitar enganos e alucinações em identificações de fraudes.

Portanto, embora a “face_recognition” seja uma ferramenta útil para validação de conceitos e protótipos, ela apresenta várias limitações que a tornam inadequada para produção em larga escala.

Resultado utilizando a biblioteca “TensorFlow”

Primeiro Treinamento

A análise dos resultados obtidos nas três solicitações do projeto, onde foi aplicado uma modelo siamesa com Keras e ResNet50 para a análise de imagens de rostos humanos, revela informações importantes sobre a eficácia do modelo em diferentes cenários. Cada tabela reflete a performance do modelo em termos de loss, acurácia e tempo por etapa ao longo de cinco épocas de treinamento.

Tabela 3: Primeiro treinamento (Solicitantes: 100, Lista de bloqueio por fraude: 50, Épocas: 5)

Época	Tempo	Perda	Acurácia	Valor de Perda	Valor de Acurácia
1/5	143s	0.0577	0.9902	0.0652	0.9890
2/5	137s	0.0572	0.9902	0.0629	0.9890
3/5	136s	0.0551	0.9902	0.0606	0.9890
4/5	137s	0.0551	0.9902	0.0605	0.9890
5/5	138s	0.0556	0.9902	0.0613	0.9890

Fonte: Dados originais da pesquisa

A perda “loss” foi relativamente baixa e a acurácia manteve-se em 99% ao longo das cinco épocas. No entanto, a similaridade predita foi modesta, o que indica que, apesar da alta

acurácia, o modelo teve dificuldade em identificar correspondências relevantes entre as imagens.

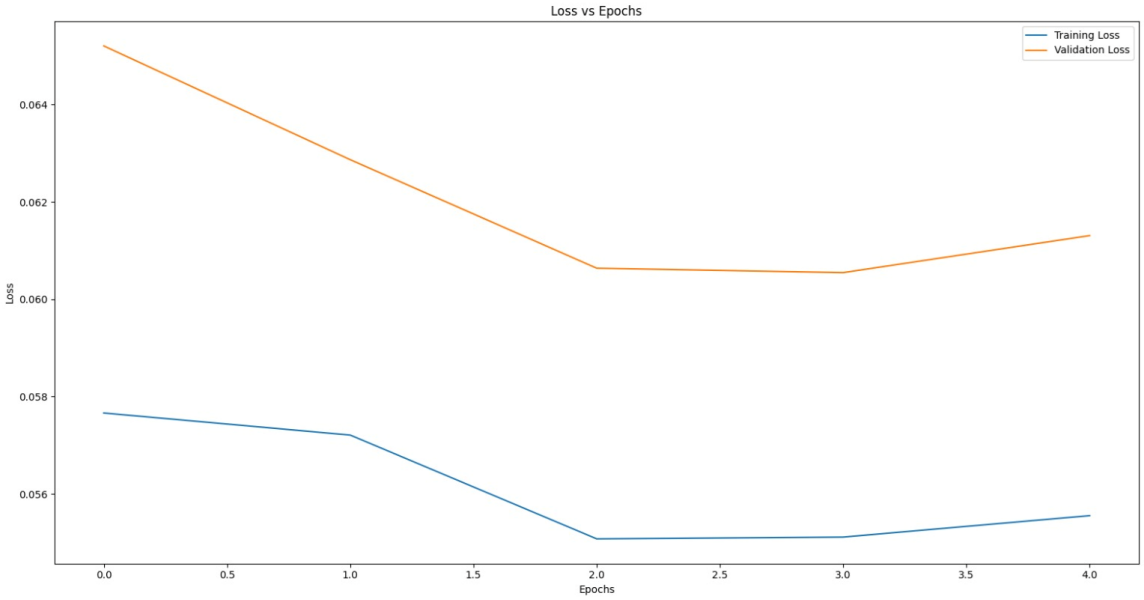

Figura 5. Gráfico de perda “loss” do primeiro treinamento.
Fonte: Dados originais da pesquisa.

Os resultados da primeira solicitação mostraram uma oscilação constante no plano cartesiano entre 0.0070 e 0.0090, o que sugere uma leve variação na similaridade entre os pares de imagens.

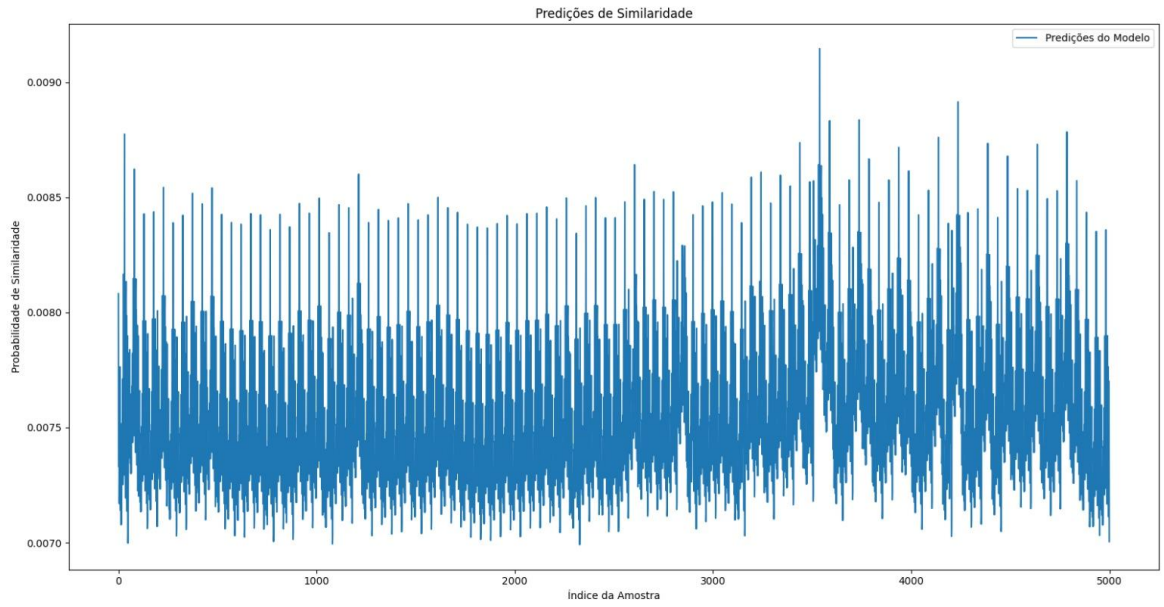

Figura 6. Gráfico de predição de similaridade do primeiro treinamento.
Fonte: Dados originais da pesquisa.

A performance do modelo, embora boa em termos de acurácia, não resultou em uma forte capacidade de discriminação entre as classes, possivelmente devido à quantidade limitada de solicitantes e imagens na “lista de bloqueio por fraude”.

Segundo Treinamento

Na segunda solicitação, o modelo mostrou uma notável capacidade de aprender com um maior número de amostras, refletindo uma melhor adaptação à tarefa proposta.

Tabela 4: Segundo treinamento (Solicitantes: 200, Lista de bloqueio por fraude: 80, Épocas: 5)

Época	Tempo	Perda	Acurácia	Valor de Perda	Valor de Acurácia
1/5	455s	0.5355	0.9829	0.4010	0.9947
2/5	465s	0.3143	0.9955	0.2453	0.9947
3/5	462s	0.1977	0.9955	0.1605	0.9947
4/5	459s	0.1323	0.9955	0.1115	0.9947
5/5	453s	0.0936	0.9955	0.0819	0.9947

Fonte: Dados originais da pesquisa

Os valores de “loss” apresentaram uma diminuição significativa ao longo das épocas, com a acurácia alcançando 99,55% no final do treinamento.

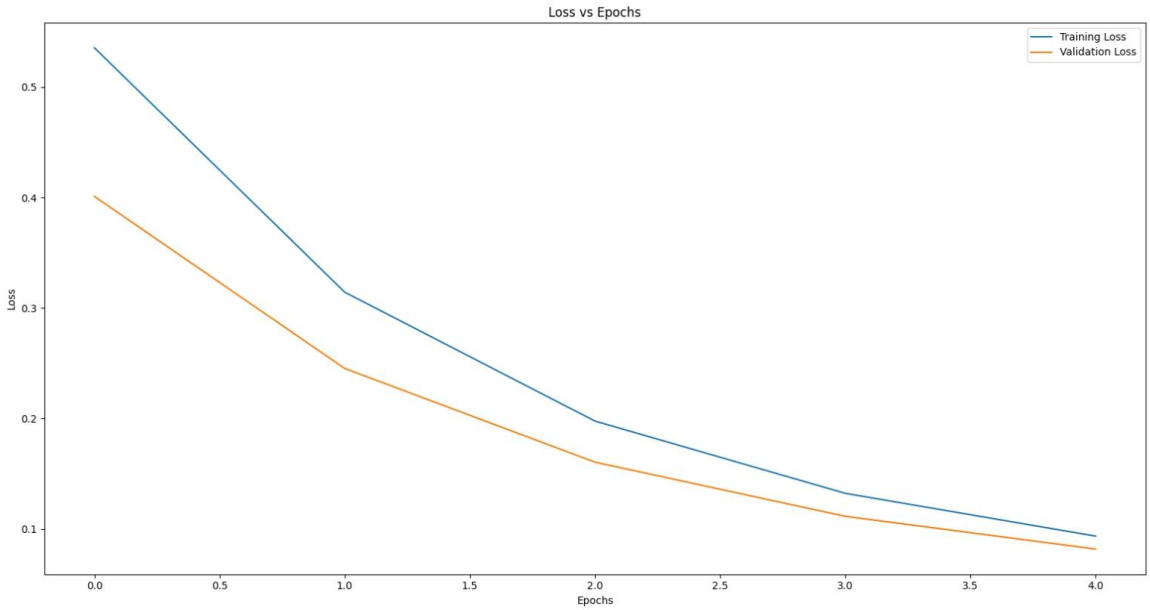

Figura 7. Gráfico de perda “loss” do segundo treinamento.
Fonte: Dados originais da pesquisa.

Na segunda solicitação, a predição de similaridade se estabilizou entre 0.065 e 0.066, indicando um aumento na identificação de padrões relevantes nas imagens.

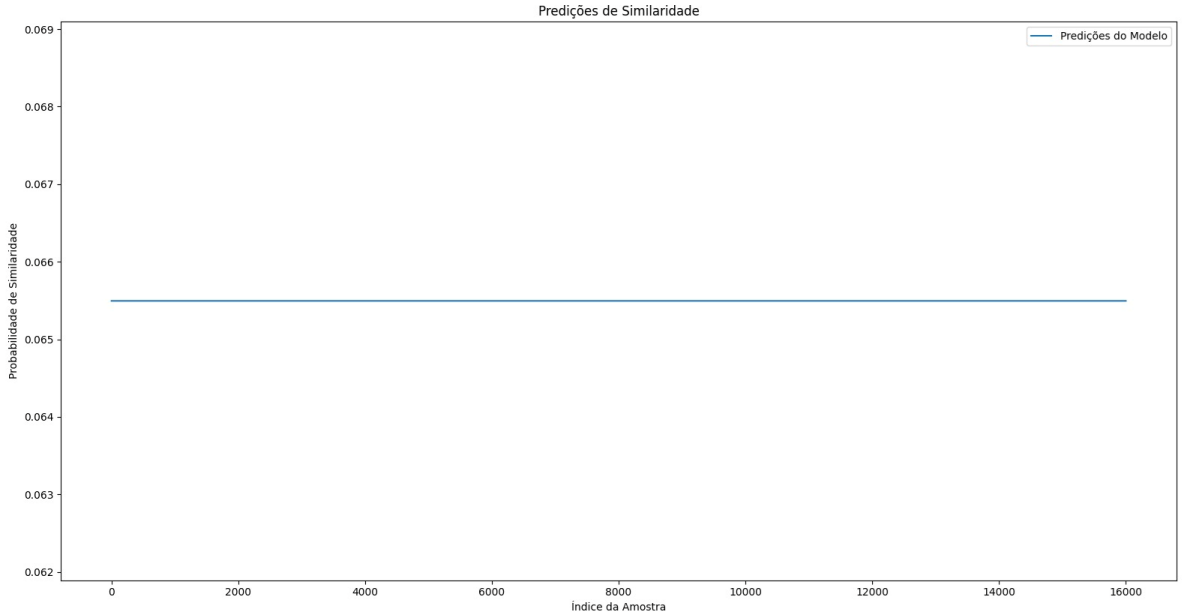

Figura 8. Gráfico de predição de similaridade do segundo treinamento.
Fonte: Dados originais da pesquisa.

O aumento no número de solicitantes e imagens na lista de bloqueio por fraude melhorou a capacidade do modelo de gerar predições de similaridade mais consistentes,

demonstrando que mais dados podem levar a um aprendizado mais robusto e a melhores resultados.

Terceiro Treinamento

Embora a acurácia tenha alcançado 99,65% e a perda tenha diminuído ao longo das épocas, a oscilação de similaridade foi consideravelmente mais baixa em comparação com as solicitações anteriores.

Tabela 5: Terceiro treinamento (Solicitantes: 300, Lista de bloqueio por fraude: 110, Épocas: 5)

Época	Tempo	Perda	Acurácia	Valor de Perda	Valor de Acurácia
1/5	1034s	0.4167	0.9916	0.2336	0.9974
2/5	1114s	0.1564	0.9965	0.0998	0.9974
3/5	1094s	0.0747	0.9965	0.0517	0.9974
4/5	1097s	0.0437	0.9965	0.0320	0.9974
5/5	1072s	0.0310	0.9965	0.0236	0.9974

Fonte: Dados originais da pesquisa

A performance do modelo continuou a melhorar em termos de “loss” e acurácia, mas a predição de similaridade não refletiu essa melhora.

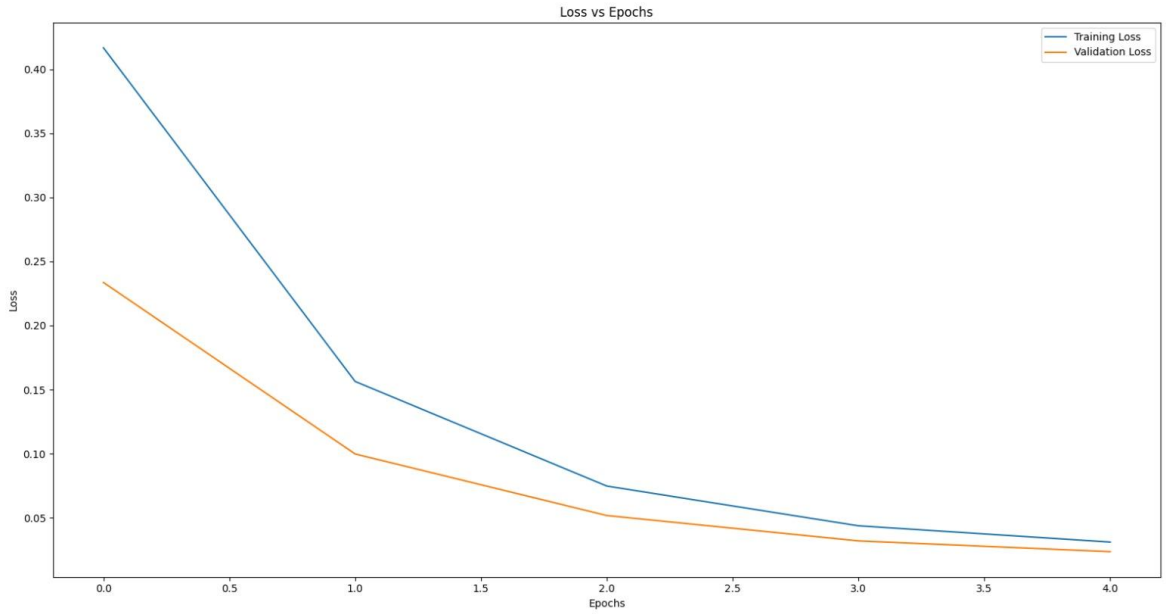

Figura 9. Gráfico de perda “loss” do terceiro treinamento.
Fonte: Dados originais da pesquisa.

A terceira solicitação apresentou uma predição de similaridade constante em 0.0121.

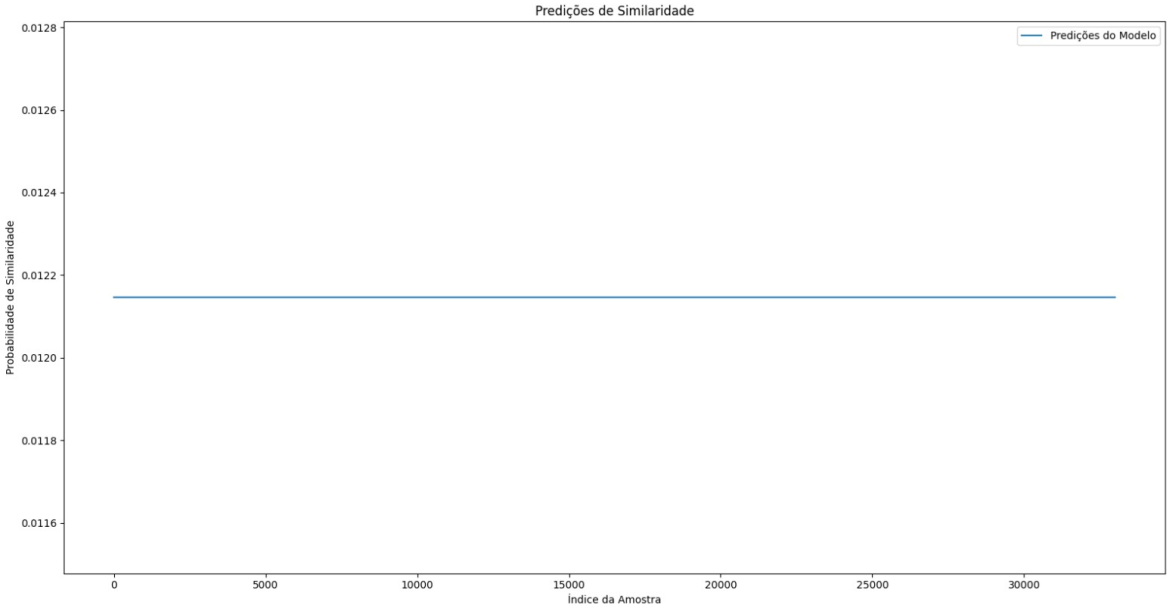

Figura 10. Gráfico de predição de similaridade do terceiro treinamento.
Fonte: Dados originais da pesquisa.

Apesar de um maior número de amostras, a predição de similaridade baixa sugere que o modelo pode estar sofrendo de uma forma de “overfitting”, onde se adapta bem aos dados de treinamento, mas não consegue generalizar adequadamente para dados novos. Isso é particularmente preocupante em cenários onde a quantidade de imagens correspondentes é limitada.

No trabalho de Koch et al. (2015), intitulado "Siamese Neural Networks for One-Shot Image Recognition", os autores alcançaram resultados expressivos ao utilizar redes siamesas para identificar similaridade entre pares de imagens. O treinamento foi conduzido com três tamanhos distintos de conjuntos de dados, contendo 30.000, 90.000 e 150.000 exemplos, gerados por meio de amostragem aleatória de pares iguais e diferentes. Eles garantiram uma representação uniforme de exemplos por alfabeto, assegurando que cada alfabeto tivesse o mesmo número de exemplos durante a otimização, ainda que essa uniformidade não fosse garantida para as classes de caracteres individuais dentro de cada alfabeto.

No caso do modelo do presente trabalho, o pequeno número de imagens e a constante oscilação nas predições de similaridade sugerem que o modelo pode não estar suficientemente robusto para capturar as sutilezas entre diferentes indivíduos ou as variações da mesma pessoa. Isso indica a necessidade de aumentar a diversidade dos dados e explorar

ajustes arquiteturais que possam melhorar o desempenho em cenários de comparação com uma única imagem por classe.

O modelo enfrentou dificuldades para capturar nuances sutis entre diferentes indivíduos ou variações do mesmo indivíduo, especialmente devido à limitação dos dados. Diferentemente do trabalho de Koch et al. (2015), que obteve resultados significativos com um grande volume de exemplos e diversidade de classes, o presente estudo demonstrou que dados limitados restringem a capacidade do modelo em generalizar e identificar similaridades com precisão em cenários reais.

Considerações Finais

O estudo mostrou que o aumento no volume de dados tende a melhorar a capacidade de aprendizado e a identificação de padrões visuais. Contudo, as oscilações nas predições de similaridade revelaram que a abordagem atual, com um número limitado de imagens e um cenário de comparação um a um, ainda não se mostra suficientemente robusta. O modelo teve dificuldades em capturar nuances sutis entre diferentes indivíduos ou variações do mesmo indivíduo, principalmente devido à limitação dos dados. Enquanto outros estudos com maior volume e diversidade de classes obtiveram resultados significativos, o presente trabalho demonstrou que dados restritos dificultam a generalização e a identificação precisa de similaridades em cenários reais. Para trabalhos futuros, sugere-se aumentar o volume e a diversidade de imagens, ajustar os hiperparâmetros e experimentar diferentes arquiteturas e técnicas de pré-processamento. Essas melhorias podem tornar o modelo mais eficiente e eficaz na identificação de padrões faciais, criando uma base para soluções mais robustas no reconhecimento de imagens faciais

Referências

Alecrim, E. 2023. Certificado digital é usado para desviar R$ 4 milhões da Justiça do Trabalho. Disponível em: <https://tecnoblog.net/noticias/certificado-digital-e-usado-para-desviar-r-4-milhoes-da-justica-do-trabalho/>. Acessado em: 20 jun. 2024.

Andrade, R., Pinho, F., A. 2023. Desenvolvimento de uma ferramenta de auxílio à tomada de decisão com base na CRISP-DM. Disponível em: <https://periodicos.unifei.edu.br/index.php/rtic/article/view/111>. Acessado em: 02 jun. 2024.

Aramburu, R. 2024. Faker: Gerando dados para testes. Disponível em: <https://www.botecodigital.dev.br/php/faker-gerando-dados-para-testes/#google_vignette>. Acessado em: 05 jun. 2024.

Baldissera, J., Silveira, S, R. 2017. Proposta de um modelo para detecção de fraudes na emissão de certificados digitais. Disponível em: <https://core.ac.uk/download/pdf/141516019.pdf>. Acessado em: 04 abr. 2024.

Basurah, M; Swastika, W; Kelana, H, O; 2023. Implementation of face recognition and liveness detection system using tensorflow.js. Disponívem em <https://jurnal.polinema.ac.id/index.php/jip/article/view/3977/2759>. Acessado em 05 abr. 2024.

Blog DSAcademy, 2023. O que é o tensorflow machine intelligence platform? Disponível em <https://blog.dsacademy.com.br/o-que-e-o-tensorflow-machine-intelligence-platform/>. Acessado em 04 abr. 2024.

Blog VCert. 2024. Cresce número de fraudes e ataques na WEB. Disponível em: <https://validcertificadora.com.br/blogs/noticias/cresce-numero-de-fraudes-e-ataques-na-web?srsltid=AfmBOoojH-N3vXDhyLMfmSOMM_QX5gRMDiyLS6yo84kyM-_GmmcMdO2b>. Acessado em: 20 jun. 2024.

Canal do youtube Enap. 2023. Desafio: Solução de IA para identificar possíveis fraudadores na cadeia de certificados | ITI. Disponível em <https://www.youtube.com/watch?v=hOwcIXYoJvA&list=LL&index=11&t=987s>. Acessado em 22 jun. 2024.

Coutinho, T. 2022. Os 4 passos para emitir certificados digitais por meio da ICP Brasil. Disponível em: <https://voitto.com.br/blog/artigo/icp>. Acessado em: 06 abr. 2024.

Dalal, N., Triggs, B. 2005. Histograms of Oriented Gradients for Human Detection. Disponível em <https://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf>. Acessado em 20 jul. 2024.

Falcão, J, V, R; Moreira, V, A; Santos, F, A, O; Ramos, C, A. 2019. Disponível em <https://revistas.unifenas.br/index.php/RE3C/article/view/232>. Acessado 04 abr. 2024.

Galindo, S., W., M. 2019. Automação de controle de acesso por reconhecimento facial desenvolvido em linguagem Python. Disponível em: <https://repositorio.ufpe.br/handle/123456789/47186>. Acessado em: 07 mai. 2024.

Geitgey, A., 2016. Machine learning is Fun! part 4: modern face recognition with deep learning. Disponível em <https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78>. Acessado em 15 jul. 2024.

Geitgey A., 2017. Face Recognition. Disponível em < https://face-recognition.readthedocs.io/en/latest/readme.html>. Acessado em 01 out. 2024

Geitgey A., 2018. Face Recognition Documentation. Disponível em <https://readthedocs.org/projects/face-recognition/downloads/pdf/stable/>. Acessado em 01 out. 2024

Gomes, O. F.; Agostinho, M. B.; Baldisera, J.; Silveira, S. R.; Martina, E. J. 2020. Detecção de Fraudes na Emissão de Certificados Digitais dentro da Infraestrutura de Chaves Públicas Brasileira, Universidade Federal de Santa Catarina, Santa Catarina. Disponível em: <https://sol.sbc.org.br/index.php/sbseg/article/view/19239>. Acessado em: 04 abr. 2024.

Heidari, M; Fouladi-Ghaleh, K. 2020. Using siamese networks with transfer learning for face recognition on small-samples datasets. Disponível em <https://ieeexplore.ieee.org/document/9116915>. Acessado em 15 jun. 2024.

Instituto Nacional de Tecnologia da Informação. 2017. Infraestrutura de Chaves Públicas Brasileira (ICP-Brasil). Disponível em: <https://www.gov.br/iti/pt-br/assuntos/icp-brasil>. Acessado em: 10 abr. 2024.

Ji, S; Song, C. 2022. Face recognition method based on siamese networks under non-restricted conditions. Disponível em <https://ieeexplore.ieee.org/abstract/document/9756547>. Acessando em 15 jun. 2024.

Junior, C. M., F. L. 2011. Reconhecimento facial utilizando redes neurais. Disponível em: <https://aberto.univem.edu.br/bitstream/handle/11077/360/Reconhecimento%20Facial%20Utilizando%20Redes%20Neurais.pdf?sequence=1&isAllowed=y>. Acessado em: 14 mai. 2024.

Junior, P. D. 2022. Análise de algoritmos de reconhecimento facial: HOG (Histogram of oriented gradients) e YOLO (you Only look once). Disponível em: <https://www.academia.edu/102237413/Analise_De_Algoritmos_De_Reconhecimento_Facial_Hog_Histogram_of_Oriented_Gradients_e_Yolo_You_Only_Look_Once_>. Acessado em: 02 mai. 2024.

Koch, G; Zemel, R; Salakhutdinov, R. 2015. Siamese neural networks for one-shot image recognition. Disponível em <https://www.cs.utoronto.ca/~rsalakhu/papers/oneshot1.pdf>. Acessado em 02 out. 2024.

Mariz, F., L. 2018. Redes geradoras adversárias em geração de imagens. Disponível em: <https://bdm.unb.br/bitstream/10483/28451/1/2018_LucasDeFreitasMariz_tcc.pdf>. Acessado em: 16 jun. 2024.

Portal ID Brasil Digital. Porque o Certificado Digital é indispensável para a segurança das transações online. Disponível em: <https://idbrasildigital.com.br/blog/por-que-o-certificado-digital-e-indispensavel-para-a-seguranca-das-transacoes-online/>. Acessado em: 06 abr. 2024.

Portal JUS Brasil. 2022. Falsificação de certificados digitais e responsabilidade da certificadora. Disponível em: <https://www.jusbrasil.com.br/artigos/falsificacao-de-certificados-digitais-e-responsabilidade-da-certificadora/1666521031>. Acessado em: 06 abr. 2024.

Philipp, W. 2012. Face Recognition with Python. Disponível em <https://online.datasport.pl/logos/reg4791.pdf>. Acessado em 28 jun. 2024.

Resende, A. D. 2009. Certificação Digital. Disponível em: <https://egov.ufsc.br/portal/sites/default/files/artigo09_0.pdf>. Acessado em: 04 abr. 2024.

Resende, F. 2023. ICP-Brasil: entenda o que é e como ela funciona na prática. Disponível em: <https://blog.1doc.com.br/icp-brasil/#:~:text=ICP%2DBrasil%20%C3%A9%20um%20sistema,por%20v%C3%A1rios%20%C3%B3rg%C3%A3os%20e%20recursos>. Acessado em: 12 abr. 2024.

Seliya, N; Abdollah, Z, A; Khoshgoftaar, T,M. 2021. A literature review on one-class classification and its potential applications in big data. Disponível em <https://doi.org/10.1186/s40537-021-00514-x>. Acessado em 15 jun. 2024.

Tsukada, J. 2021. ICP-Brasil: o que é e qual sua relação com a assinatura digital. Disponível em: <https://assinei.digital/icp-brasil/>. Acessado em: 12 abr. 2024.

Ying. X; 2022. An iverview of overfitting and its solutions. Disponível em <https://iopscience.iop.org/article/10.1088/1742-6596/1168/2/022022/meta>. Acesso em 03 ago. 2024.