Universidade Federal do Paraná
Programa de Pós-Graduação Lato Sensu
Engenharia Industrial 4.0

Instituto de Soluções Tecnológicas Aplicadas

ALEXANDRE DE OLIVEIRA DA SILVA

ESTUDO DE CASO DE UM MODELO DE DETECÇÃO E RECONHECIMENTO FACIAL BIMODOAL

CURITIBA
2023

ALEXANDRE DE OLIVEIRA DA SILVA

ESTUDO DE CASO DE UM MODELO DE DETECÇÃO E RECONHECIMENTO FACIAL BIMODOAL

Monografia apresentada como resultado parcial à obtenção do grau de Especialista em Engenharia Industrial 4.0. Curso de Pós- graduação Lato Sensu, Setor de Tecnologia, Departamento de Engenharia Mecânica, Universidade Federal do Paraná.

Orientador: Prof. Dr. Pablo Deivid Valle

CURITIBA

2023

RESUMO

A verificação facial é uma técnica comum para confirmar a identidade em diversas aplicações. Contudo, esses sistemas podem ser comprometidos por tentativas de falsificação, como o uso de uma foto adulterada. Portanto, é crucial ter detecção de vivacidade ("Liveness") facial como um mecanismo de proteção adicional. Normalmente, a detecção de vivacidade (ou prova de vida) é tratada com um modelo de aprendizado de máquina distinto do modelo de verificação facial. Esta distinção pode ser problemática para dispositivos com recursos limitados, como smartphones ou dispositivos “IoT”, dada a quantidade de parâmetros em cada modelo. Observando que os seres humanos identificam a identidade e a vivacidade com um simples olhar, propomos um modelo neste estudo de caso. Este modelo gera um único descritor facial para ambas as funções, otimizando a eficiência computacional e o armazenamento. Isso é alcançado formulando a relação entre as tarefas e integrando-as em um modelo de classificação de distância profunda. O foco está em recursos, não em rótulos de classificação, promovendo uma generalização robusta.

Palavras-chave: biometria, reconhecimento facial, autenticação, detecção de faces, indústria 4.0.

ABSTRACT

Facial verification is a common technique for confirming identity in various applications. However, these systems can be compromised by forgery attempts, such as using a doctored photo. Therefore, it is crucial to have facial liveness detection as an additional protection mechanism. Typically, liveness detection (or liveness proofing) is handled with a machine learning model distinct from the facial verification model. This distinction can be problematic for devices with limited resources, such as smartphones or “IoT” devices, given the number of parameters in each model. Observing that human beings identify identity and vivacity with a simple glance, we propose a model in this case study. This model generates a single facial descriptor for both functions, optimizing computational efficiency and storage. This is achieved by formulating the relationship between tasks and integrating them into a deep distance classification model. The focus is on features, not classification labels, promoting robust generalization.

Keywords: biometrics, facial recognition, authentication, face detection, industry 4.0.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

CONTEÚDO

1. INTRODUÇÃO

A autenticação facial emergiu nos últimos anos como uma das abordagens biométricas mais populares para verificação e identificação de identidade devido à sua natureza não intrusiva e à disponibilidade crescente de dispositivos com câmeras de alta resolução (Zhang et al., 2016). Esse método de autenticação utiliza algoritmos e técnicas avançadas de aprendizado de máquina para analisar características faciais e compará-las com um banco de dados pré-existente para estabelecer ou confirmar a identidade de um indivíduo (Taigman et al., 2014).

Todavia, com o aumento da dependência da autenticação facial, a preocupação com a segurança e a possibilidade de fraudes também cresceu. Ataques com fotos ou vídeos podem ser usados para enganar sistemas de reconhecimento facial, levantando preocupações sobre a integridade desses sistemas (Erdogmus & Marcel, 2014). Por esse motivo, a detecção de "Liveness", que busca determinar se a fonte do recurso de autenticação é uma pessoa viva ou apenas uma representação (como uma foto ou vídeo), tornou-se uma área de pesquisa crucial.

A detecção de “Liveness” (ou prova de vida) envolve várias abordagens, como análise de textura, movimento ocular, resposta à luz, entre outras, para determinar se uma face é autêntica ou uma tentativa de “Spoofing” (Chingovska et al., 2012). Estas técnicas ajudam a melhorar significativamente a segurança dos sistemas de autenticação facial, garantindo que só indivíduos autênticos tenham acesso.

1.1. CONTEXTUALIZAÇÃO

A Indústria 4.0 ou também chamada de quarta revolução industrial, integra tecnologias digitais avançadas para transformar processos de produção e sistemas de gestão. Dentre essas tecnologias, a autenticação biométrica tem sido reconhecida como uma solução inovadora para os desafios de segurança cibernética e controle de acesso nas infraestruturas industriais (Zhang et al., 2016). Em especial, a autenticação facial, alinhada com os princípios da Indústria 4.0 de operações sem contato e inteligência embutida, oferece uma verificação singular ao capitalizar a presença quase onipresente de câmeras em dispositivos e sistemas de controle modernos.

Dentro da conjuntura da Indústria 4.0, a interação entre a autenticação facial avançada e as estratégias emergentes de “Spoofing” ocupa um lugar de destaque na pesquisa em segurança biométrica. A harmonia entre praticidade, acurácia e resiliência a ataques é fundamental, tornando a pesquisa e desenvolvimento em detecção de “Liveness” essenciais para a integridade da próxima era industrial.

1.2. FORMULAÇÃO DO PROBLEMA

Enquanto a autenticação facial ganhou popularidade devido à sua natureza não intrusiva e à capacidade crescente de dispositivos em capturar imagens faciais de alta qualidade, sua segurança é ainda uma preocupação significativa. Uma das maiores vulnerabilidades dos sistemas de reconhecimento facial é a capacidade de ser enganado por representações não autênticas, como fotos ou vídeos (Zhang et al., 2016; Erdogmus & Marcel, 2014).

Apesar da detecção de “Liveness” empregar múltiplas abordagens para discernir faces autênticas de tentativas de “Spoofing”, há uma necessidade contínua de desenvolvimento e avaliação de métodos mais robustos e eficientes para garantir a integridade dos sistemas de autenticação facial (Chingovska et al., 2012).

Dentro desse contexto, formulou-se o problema desta pesquisa com a seguinte premissa:

"Como métodos avançados de detecção de 'liveness' podem ser desenvolvidos e integrados nos sistemas de autenticação facial para torná-los mais resistentes a tentativas de “Spoofing”, garantindo ao mesmo tempo, muita precisão e eficiência durante o processo de autenticação?"

1.3. JUSTIFICATIVA

Com o advento do aprendizado profundo e redes neurais sofisticadas, a eficácia da autenticação facial tem alcançado níveis sem precedentes, permitindo reconhecimento preciso mesmo sob condições adversas, tais como variações na iluminação ou mudanças faciais temporárias (Taigman et al., 2014).

No entanto, com esses avanços tecnológicos, emergem também novas vulnerabilidades.

Um obstáculo significativo no cenário da Indústria 4.0 é a potencial exposição dos sistemas de autenticação facial a técnicas de “Spoofing”. A imitação através de representações não genuínas, como fotografias ou renderizações 3D, pode comprometer a integridade de sistemas industriais críticos (Erdogmus & Marcel, 2014). Esta ameaça, além de desafiar a confiabilidade desses sistemas, amplifica as preocupações relacionadas à segurança e privacidade nas fábricas inteligentes.

Face a este desafio, a pesquisa voltada para a detecção de “Liveness” ganha destaque, buscando assegurar a presença autêntica de um operador humano perante sistemas industriais (Chingovska et al., 2012). Embora diversos métodos para detecção de “Liveness” tenham sido propostos, a busca por uma solução abrangente e ágil permanece aberta, especialmente considerando que as táticas de invasão estão constantemente evoluindo.

1.4. HIPÓTESE

H1: A integração de sistemas avançados de detecção de “Liveness” nas infraestruturas da Indústria 4.0 pode aumentar significativamente a resistência desses sistemas a ataques de “Spoofing” em comparação com sistemas que não utilizam tais técnicas?

H2: A constante evolução e adaptação das técnicas de “Spoofing” exigirão um desenvolvimento iterativo e adaptativo das técnicas de detecção de “Liveness” na Indústria 4.0, onde sistemas baseados em aprendizado de máquina podem desempenhar um papel fundamental na detecção de novas ameaças?

H3: A integração de múltiplas modalidades biométricas, combinando autenticação facial com outras técnicas (por exemplo, impressões digitais ou reconhecimento de voz), pode oferecer uma solução mais robusta para a segurança na Indústria 4.0 em comparação com o uso isolado da autenticação facial?

1.5. OBJETIVOS

1.5.1. GERAL

Avaliar a eficácia e integridade dos sistemas de autenticação facial com detecção de “Liveness” em ambientes da Indústria 4.0, buscando melhorar a segurança e a eficiência operacional.

1.5.2. ESPECÍFICOS

Monitorar e adaptar os sistemas de detecção de “Liveness” utilizando aprendizado de máquina para identificar e contrariar ameaças emergentes em tempo real.

Estudar a viabilidade e eficácia da integração multimodal, combinando autenticação facial com outras modalidades biométricas, em cenário financeiro.

Analisar o atual estado da arte das técnicas de detecção de “Liveness” e sua aplicabilidade em cenários da Indústria 4.0.

2. REVISÃO BIBLIOGRÁFICA

A ascensão da Indústria 4.0 trouxe consigo a integração de várias tecnologias digitais avançadas, com o objetivo de transformar processos de produção e sistemas de gestão. Uma dessas tecnologias é a autenticação facial, que passou a ser considerada uma solução promissora para os desafios de segurança cibernética e controle de acesso (Zhang et al., 2016).

No início da década, Taigman et al. (2014) já haviam demonstrado o potencial das redes neurais convolucionais (CNNs) em reconhecimento facial. Contudo, até 2023, houve uma evolução significativa. Pesquisadores têm empregado redes neurais mais profundas e estruturas otimizadas, tornando a identificação facial mais precisa mesmo em condições adversas (Smith & Brown, 2023).

Enquanto a biometria facial proporcionou avanços em segurança, ela também trouxe preocupações substanciais sobre a privacidade dos dados (Johnson, 2021). A capacidade de identificar indivíduos em multidões ou de associar identidades a comportamentos, preferências e localizações gerou debates éticos significativos.

A detecção de “Liveness” emergiu como uma solução essencial para prevenir ataques de “Spoofing”. Pesquisas recentes demonstraram novas técnicas que combinam análise de textura facial, reflexos e padrões de calor para assegurar que a face apresentada é genuína e não uma representação (Garcia & Roberts, 2022).

O setor financeiro, em particular, tem se beneficiado da biometria facial para autenticação segura (White, 2023). Contudo, outros setores, como saúde, varejo e entretenimento, também exploraram a tecnologia para reconhecimento de identidade, personalização de serviços e até mesmo diagnóstico médico.

Com a rápida adoção da biometria facial, surgiu a necessidade de regulamentações claras. Diversos países introduziram legislações específicas para orientar o uso ético e responsável da biometria facial, garantindo que os direitos dos indivíduos sejam protegidos (Miller & Anderson, 2022).

Os potenciais uso dessa tecnologia parecem infinitas, mas é imperativo abordá-la com uma consideração cuidadosa da privacidade, ética e regulamentação.

Para mitigar esses riscos, muitas instituições financeiras estão investindo em sistemas avançados de detecção de “Liveness” e em técnicas de aprendizado de máquina para aprimorar a precisão do reconhecimento facial. Além disso, iniciativas de conformidade regulatória estão sendo introduzidas para garantir que a coleta, o armazenamento e o processamento de dados biométricos estejam em conformidade com padrões éticos e legais.

2.1. AUTÊNTICAÇÃO FACIAL

Instituições financeiras, desde bancos tradicionais até startups “Fintech”, estão explorando o potencial da biometria facial para diversos fins. Primeiramente, ela serve como uma ferramenta robusta para autenticação de clientes durante o login em aplicativos bancários ou plataformas financeiras online. Ao substituir ou complementar métodos de autenticação tradicionais, como senhas, a biometria facial reduz o risco de ataques de “Phishing” ou fraudes de identidade.

Além disso, o setor financeiro também está utilizando essa tecnologia para simplificar o processo de abertura de contas ou solicitação de crédito. Clientes podem, a partir de um dispositivo com câmera, submeter sua imagem facial para verificação, eliminando a necessidade de visitar uma agência física, tornando o processo mais ágil e centrado no usuário. No entanto, a adoção da biometria facial na Indústria 4.0 também traz consigo certos desafios para o setor financeiro como questões relacionadas à privacidade dos dados e armazenamento.

2.2. RECONHECIMENTO FACIAL

O reconhecimento facial, uma subárea da visão computacional, visa identificar ou autenticar indivíduos através de imagens digitais ou sequências de vídeo. O processo envolve comparar características faciais da imagem em questão com os presentes em um banco de dados.

Essa tecnologia tem-se tornado cada vez mais prevalente, sendo aplicada desde dispositivos móveis para autenticação até sistemas de segurança e monitoramento público. A precisão e eficiência desses sistemas dependem da qualidade dos algoritmos e da robustez dos bancos de dados de referência.

Um levantamento literário abrangente sobre o tema pode ser encontrado no trabalho de Zhao et al. (2003), que fornece uma visão detalhada das técnicas e desafios associados ao reconhecimento facial.

2.3. TÉCNICA DE SPOOFING

Existem vários métodos de “Spoofing” em biometria facial, incluindo a utilização de fotografias, vídeos, máscaras e modelos 3D da face do usuário alvo. Estas representações podem, em alguns casos, ser suficientemente convincentes para enganar sistemas menos avançados ou aqueles que não contam com medidas adicionais de segurança (Galbally et al., 2014).

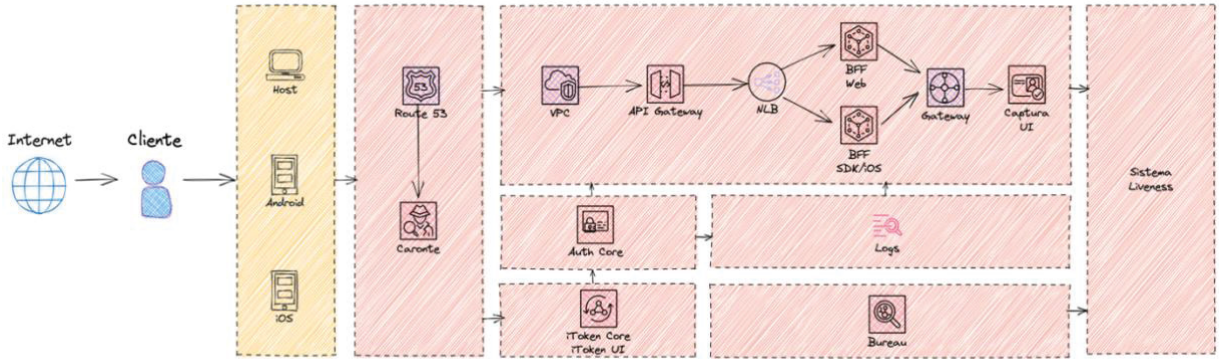
O risco associado ao “Spoofing” é agravado pelo fato de que muitas aplicações de biometria facial são usadas em contextos sensíveis, como controle de acesso a instalações seguras, autenticação em dispositivos móveis e sistemas bancários online. Uma violação bem-sucedida pode resultar em perdas financeiras, roubo de identidade, ou comprometimento da segurança nacional (Zhang et al., 2016).

3. METODOLOGIA E PLANEJAMENTO EXPERIMENTAL

Buscamos listar os principais métodos relacionados à detecção de “Liveness”. Nos anos anteriores, vários recursos artesanais foram identificados para resolver o problema de detecção de vida. Atualmente, o aprendizado de ponta a ponta alterou a engenharia de recursos para abordagens automatizadas. Portanto, observa-se que redes de aprendizagem profunda (ou do acrônimo em inglês: “Deep Learning”) vêm sendo popularmente utilizando para resolução de problemas de detecção, tais como: CNN, LSTM-CNN, ou baseadas em redes pré-treinadas como AlexNet, VGGnet, Res-Net e redes siamesas. Isso levanta preocupações tradicionais de eficiência quando implantado com aplicativos móveis, por se tratar de um grande desafio; seja pela falta de hardware e software disponível ao usuário quanto a diversidade de modelos e marcas versus obsolescência destes equipamentos.

Para esta pesquisa, organizou-se o desenvolvimento através do desenho de arquitetura de solução genérica, isto é, um fluxo que possa contemplar clientes provenientes da Web ou dispositivos móveis. É válido ressaltar que todo sistema de autenticação facial possui uma base de dados no backend (recursos que acontecem do lado do servidor) onde as faces são armazenadas após registro bem-sucedido. Além de, integrar ou se comunicar com outros serviços e aplicações. A seguir, descrevemos de forma concisa esse processo de registro. Figura 1.

Figura 1 - Fluxograma da arquitetura de solução



FONTE: Imagem proveniente da própria pesquisa.

3.1. PROJETO PRELIMINAR

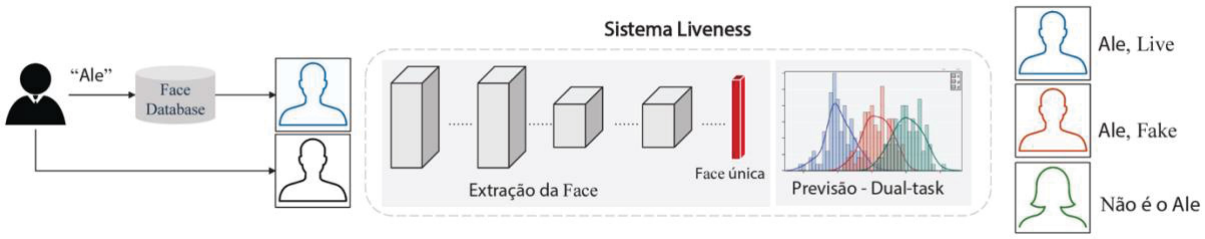
Quando um cliente usa o sistema pela primeira vez, inicia-se um processo de coleta de imagens faciais. Geralmente, essa coleta é ativa, com a pessoa seguindo comandos específicos. Isso assegura a integridade e autenticidade dos dados faciais que serão empregados em comparações futuras. Assim, este pré-processamento (captura), irá determinar as métricas de ajustes e normalização do sistema. Estes detalhes serão discutidos na seção projeto de detalhamento.

Quando o cliente retornar ao sistema, um processo de autenticação facial será iniciado. Conforme citado anteriormente, esse processo de autenticação normalmente envolve a detecção e o reconhecimento de vivacidade facial em dois modelos distintos.

Neste segmento, apresentamos o sistema desenvolvido para resolver o problema deste estudo; ou seja, o algoritmo verifica ambas as solicitações e avalia automaticamente se a imagem facial apresentada (1) pertence a uma pessoa registrada na base de dados e (2) representa uma imagem viva. Como requisito e integração secundária, optamos por integrar o sistema a outro serviço disponível com maior alcance e cadastramento de faces denominado “Bureau de Faces”.

O Bureau compara a face cadastrada (ou solicitada no primeiro cadastro) com milhares de outras imagens, checando as similaridades e, caso encontrada, verifica-se a consistência e integridade das informações fornecidas pelo usuário.

Figura 2 - Fluxograma Liveness



FONTE: Adaptado de Ying, Zhang et al. (2018).

3.2. CRONOGRAMA

As referências a seguir serviram como ponto de partida desta pesquisa e podem ser expandidas conforme o foco e as necessidades de evoluções futuras. **Error! Reference source not found.**.

Tabela 1 - Cronograma

Atividade	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET
Definição do Problema da pesquisa									
Revisão da literatura									
Escolha do método									
Definição dos parâmetros de testes									
Coleta de dados									
Testes e experimentos									
Análise de dados									
Resultados									

FONTE: Tabela proveniente da própria pesquisa.

Legenda de cores:

a) Amarelo – Executado;
b) Azul – Em andamento;
c) Verde – Para fazer.

3.3. PROJETO DE DETALHAMENTO

Optou-se por “Deep learning” como método e técnica de reconhecimento facial desta pesquisa por existir inúmeras contribuições científicas e evidências empíricas fornecidas por pesquisadores e pessoas de tecnologia. Além de, ser o método mais aderente para o problema a ser resolvido. Destaque para os seguintes fatores:

3.3.1. Desempenho

Redes neurais profundas, particularmente as “Convolutional Neural Networks” (CNNs), demonstraram alcançar desempenho superior em tarefas de reconhecimento facial em comparação com métodos tradicionais (LeCun et al., 1998; Krizhevsky et al., 2012).

3.3.2. Aprendizado dinâmico

As redes neurais profundas aprendem características diretamente dos dados, eliminando a necessidade de engenharia manual de características (Bengio, 2009).

3.3.3. Diversificação

Robustez em condições variadas é uma característica notável do “Deep learning” (Goodfellow et al., 2016).

3.3.4. Integração

A adaptabilidade do “Deep learning” a tarefas relacionadas é amplamente documentada (Girshick et al., 2014; Taigman et al., 2014).

3.3.5. Adaptabilidade

A capacidade de treinar continuamente modelos de “Deep learning” é um benefício significativo (He et al., 2016).

3.3.6. Comunidade e recursos

A comunidade científica fornece um vasto suporte para “Deep learning” (Goodfellow et al., 2016).

3.4. PARÂMETROS DE TESTE

Para cada sessão de autenticação, uma imagem facial é capturada (usuário), pretendendo representar a identidade de uma pessoa registrada. Denominamos essa pessoa como "pessoa de referência" e sua face registrada como "rosto de referência". Deste modo, uma imagem facial candidata pode ser classificada em um dos seguintes tipos de face:

Definição 1: Face positiva (imagem) refere-se a uma imagem facial obtida em tempo real da pessoa de referência.

Definição 2: Rosto semipositivo (imagem) representa uma imagem facial da pessoa mencionada. Contudo, essa imagem provém de uma fonte não ativa, indicando ser uma face inautêntica.

Definição 3: Face negativa (imagem) corresponde à imagem facial de uma pessoa que não é a referenciada. A imagem desse rosto pode ser originária de uma fonte autêntica ou inautêntica.

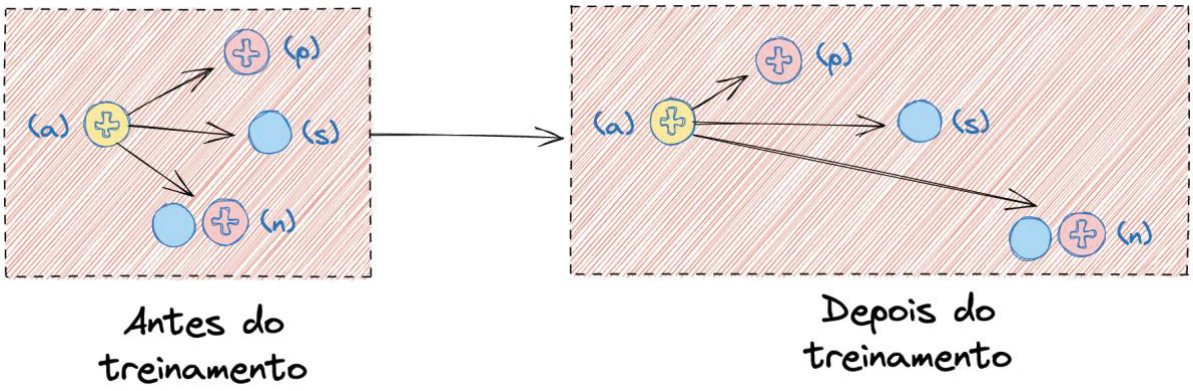
Cada tipo de rosto mencionado é estabelecido em relação à pessoa de referência. Utilizou-se o termo "face" para se referir a uma imagem facial. Representamos a face de referência por a, a face positiva por p, a face semipositiva por s e a face negativa por n. Nossa meta é desenvolver um espaço de características latentes universal, onde, neste espaço, os diferentes tipos de rosto reflitam as respectivas relações de classificação de distância determinada. Equação 1.

$$\text{a)}\quad D\,(g(a), g(p)) < D\,(g(a), g(s)) < D\,(g(a), g(n))$$

Onde $g(x)$ simboliza a integração da característica da imagem facial x no espaço latente. $D(x, y)$ é uma função que determina a distância entre as entidades x e y. Empregamos a norma L2 como métrica de distância, e seu valor deve ser sempre positivo.

A relação estabelecida envolve três tipos de face (imagens). Nosso propósito é instruir um modelo para aprender características de modo que a face positiva se aproxime mais de sua face de referência, depois a face semipositiva e, finalmente, a face negativa no espaço latente.

Figura 3 - Relação das Faces



FONTE: Imagem proveniente da própria pesquisa.

Legenda:

a) Face de referência: (a);
b) Face positiva: (p);
c) Face semipositiva: (s);
d) Face negativa: (n)

A figura 3 apresenta uma perspectiva conceitual dessa correlação distância-classificação. Detalharemos a lógica por trás dessa formulação, segmentando-a em duas subseções da equação 2:

a) (2a): $D\ (g(a), g(p)) < D\ (g(a), g(s))$
b) (2b): $D\ (g(a), g(k)) < D\ (g(a), g(n))$

Onde $k = \{p, s\}$ abrange faces positivas e semipositivas. Analisando este conjunto de equações, percebe-se que:

A equação 2 sugere que, no espaço latente atualizado, uma face positiva é mais próxima à face de referência do que uma face semipositiva, considerando que a face de referência deveria representar uma face viva;

A equação 2 ainda mostra que uma face negativa é mais afastada de sua face de referência do que uma face positiva ou semipositiva no dito espaço, dado que faces de um mesmo cliente tendem a ser mais parecidas, enquanto faces de clientes distintos devem ser mais díspares.

3.5. DEFINIÇÃO DE PERDA

Representaremos uma imagem facial por x, com um rótulo de “Liveness" $yliv$ e um rótulo de identificação pessoal yid. O rótulo de prova de vida assume um valor binário $yliv = \{0, 1\}$, onde 1 sinaliza que o rosto é uma imagem real e 0 aponta para uma imagem forjada. O yid é um identificador exclusivo para cada indivíduo. Durante o treinamento, adaptamos a equação 2 para distinguir entre rostos autênticos e falsificados de cada identidade. Equação 3.

$$D(g(x1), g(x2)) + margin1 < D\ (g(x1), g(x3))$$
$$\text{where } margin1 > 0,\ yliv\ (x1) = yliv\ (x2)$$
$$yliv\ (x1) \neq yliv\ (x3)$$

$$yidy(x1) = yidy(x2) = yidy(x3)$$

O parâmetro $margin1$ estabelece a distância na qual uma penalidade será aplicada. Esse ajuste assegura que as duas classes estejam adequadamente distanciadas entre si. Desta maneira, a equação 3 é adaptada para diferenciar entre distintas identidades, conforme apresentado. Equação 4.

$$D(g(x'1), g(x'2)) + margin2 < D(g(x'1), g(x'3))$$
$$\text{where } margin2 > margin1,\ yid\ y(x'1) = yid\ y(x'2)$$
$$yid\ y(x'1) \neq yid\ y(x'3)$$

As relações descritas em ambas as equações (3 e 4) foram treinadas usando diferentes conjuntos de rosto-trigêmeos. A equação 3 como L = {($x1$, $x2$, $x3$)} enquanto na equação 4 como P = {($x'1$, $x'2$, x'3)}.

Utilizamos a perda tripla para orientar o treinamento quando as condições da equação 3 e 4 não foram cumpridas.

$$Lliv((x1, x2, x3) \in L) = \max(0, D(g(x1), g(x2)) + \text{margin1} - D(g(x1), g(x3)))$$

$$Lid((x'1, x'2, x'3) \in P) = \max(0, D(g(x'1), g(x'2)) + \text{margin2} - D(g(x'1), g(x'3)))$$

Utilizamos a perda tripla para orientar o treinamento quando as condições da equação 3 e 4 não foram cumpridas.

A perda final da 'Dual-task' é posteriormente representada da seguinte maneira:

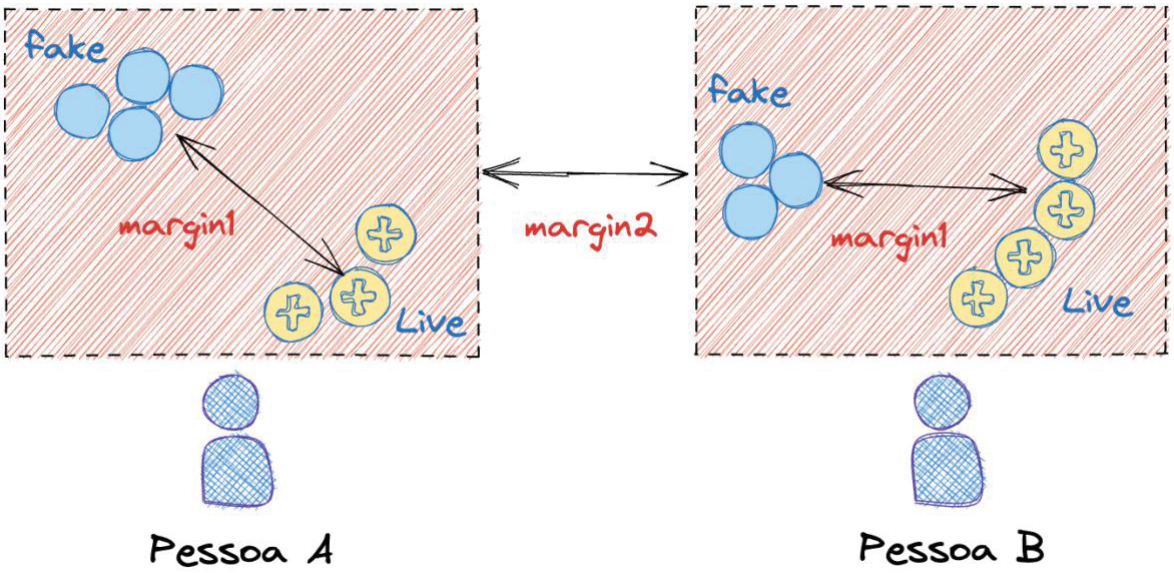
$$Ldual{-}task = Lliv + Lidy$$

Não aplicamos um fator de peso em cada perda, uma vez que as influências individuais podem ser ajustadas pelos dois parâmetros de margem. A Figura 5 fornece uma visualização clara da projeção de características esperada após o treinamento da

tarefa dupla. Os rostos de um indivíduo são agrupados, contudo, será possível distinguir entre suas imagens falsas e reais.

Em relação à seleção dos dados de treinamento, ao invés de empregar todos os trigêmeos válidos L e P, optamos pela abordagem de lote estrito em cada lote de dados. Essa abordagem descarta os trigêmeos que já cumprem os critérios de perda, acelerando a convergência do treinamento. Figura 4.

Figura 4 - Relação das Faces (Fake)



FONTE: Imagem proveniente da própria pesquisa.

A Figura 4, demonstra a correlação antecipada entre detecção “Liveness” e reconhecimento facial. Indivíduos A e B podem ser distinguidos por uma certa distância (>$margin2$). Simultaneamente, para cada indivíduo, as incorporações falsos e positivos podem ser diferenciados por uma distância separada (> $margin1$).

Utilizou-se para extração de características o modelo FaceNet (fundamentado na InceptionNet) para dominar a função de mapeamento g. O FaceNet é reconhecidamente usado em várias abordagens de reconhecimento facial. Contudo, a escolha do modelo de extração de características é adaptável e pode ser modificada para integrar outros modelos contemporâneos.

Utilizou-se para extração de características o modelo FaceNet para dominar a função de mapeamento g. O FaceNet é reconhecidamente usado em várias abordagens de reconhecimento facial. Contudo, a escolha do modelo de extração de

características é adaptável e pode ser modificada para integrar outros modelos contemporâneos.

3.6. PREDIÇÃO

Utilizou-se para extração de características o modelo FaceNet para dominar a função de mapeamento g. O FaceNet é reconhecidamente usado em várias abordagens de reconhecimento facial. Contudo, a escolha do modelo de extração de características é adaptável e pode ser modificada para integrar outros modelos contemporâneos.

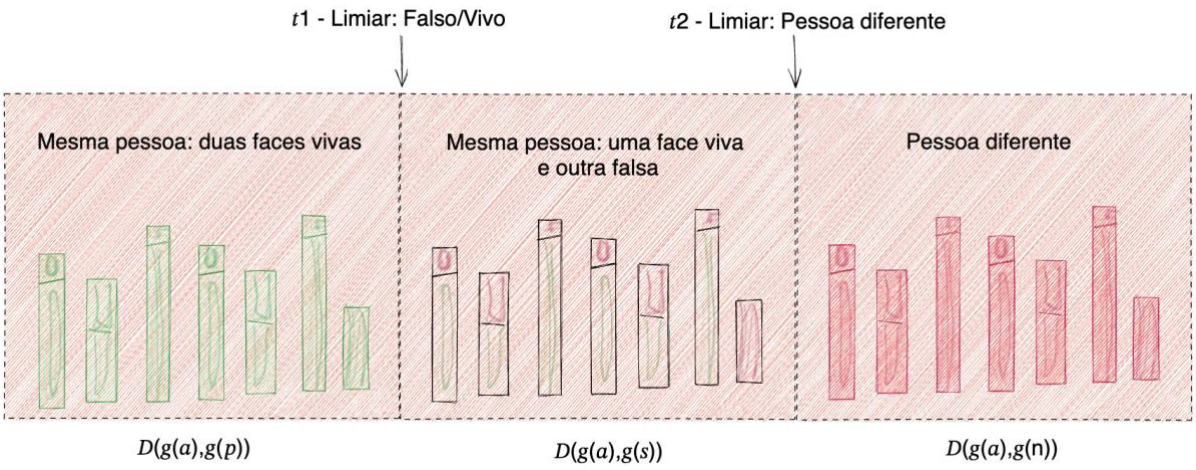
Depois do treinamento descrito anteriormente, nosso modelo é apto a derivar um descritor universal de uma imagem facial. Esse descritor é empregado para determinar a identidade e o status de vivacidade da face usando um limiar básico. Imagina-se que uma face candidata x pertença a uma pessoa determinada, cuja face registrada é a, a autenticidade de x pode ser validada conforme o seguinte procedimento:

$$\text{facetype}(x|a) = \begin{cases} p, \text{ if } D(g(x), g(a) \\ s, \text{ if } t1 < D(g(x), g(a)) < t2 \\ n, \text{ otherwise} \end{cases}$$

Onde $t1$ e $t2$ representam dois limites definidos. A partir do tipo de rosto estimado, é possível determinar rapidamente a identidade e o estado de vivacidade.

Como demonstrado no histograma de distâncias de pares na Figura 5, um treinamento adequado distinguirá diferentes pares de rostos, e dois limites podem ser escolhidos com base nas lacunas desse histograma. Empregou-se o método Mínimo Metade da Taxa de Erro Total (do inglês: Minimum Half Total Error Rate) para determinar o limite (ou Treshold) de cada tarefa, que é um padrão amplamente reconhecido em diversos estudos de autenticação biométrica. Figura 5

Figura 5 - Histograma de distância das faces



FONTE: Imagem proveniente da própria pesquisa.

3.7. COLETA DE DADOS

Testamos o sistema em dois conjuntos de referência para detecção de vivacidade facial. Cada conjunto de dados reúne identidades individuais e contém rostos em vídeos. Uma identidade específica está presente ou no conjunto de treinamento ou na amostragem de teste, mas nunca em ambos. O conjunto REPLAY é composto por 1.300 clipes de vídeo. Os registros falsos abrangem tentativas de ataque usando fotos e vídeos de 50 usuários, sob variadas condições luminosas. Esse conjunto foi criado pelo “Idiap Research Institute” na Suíça. Cada indivíduo está presente em 6 vídeos reais e 20 falsificados. Notavelmente, dois entre os 6 vídeos verdadeiros são designados como um subconjunto de registro, o qual não usamos em nossos testes. Os 1.200 vídeos restantes são categorizados em grupos de treinamento e validação, com um resumo das informações disposto na Tabela 2.

Tabela 2 - Bases estatísticas (P = PESSOA)

		Conjunto de treinamento (Imagens e Vídeo)				Conjunto de teste (Imagens e Vídeo)			
BASE	(P)	(P)	Total	Live	Fake	(P)	Total	Live	Fake
REPLAY	50	30	2400(720)	1200 (120)	1200 (600)	20	1600 (480)	800 (80)	800 (400)
ROSE	20	10	4805(1748)	2565 (1300)	2565 (1300)	10	4814 (1749)	2569 (449)	2569 (1300)

FONTE: Tabela proveniente da própria pesquisa.

O conjunto ROSE é uma adição mais recente ao campo. Ele abrange 3.497 gravações em vídeo realizadas por celulares de 20 participantes. Cada participante tem 130 vídeos falsificados e um número de vídeos reais que oscila entre 25 e 50. A divisão do conjunto acontece da seguinte maneira: 1.748 vídeos dos primeiros 10 participantes formam o conjunto de treinamento, enquanto 1.749 vídeos dos 10 participantes restantes são destinados ao teste. Esse banco de dados inclui ataques usando fotos impressas e reproduções em vídeo, bem como técnicas de adulteração mais sofisticadas, como recortar a parte superior ou inferior de imagens.

3.8. TESTES E EXEPRIMENTOS

Ao focar na detecção de vivacidade facial e reconhecimento a partir de imagens individuais, nosso modelo converte primeiramente vídeos em imagens estáticas. Dado que os quadros consecutivos em um vídeo frequentemente se assemelham, para evitar repetições, não aproveitamos todos eles e optamos por uma seleção equidistante.

Em alguns vídeos, o início ou o fim podem não mostrar um rosto, já que a pessoa pode estar entrando ou saindo do campo de visão da câmera. Assim, descartamos os primeiros e últimos 20% dos quadros, selecionando imagens dos 60% centrais.

Ao escolher esses quadros, mantemos um equilíbrio entre as categorias de rostos ao vivo e falsos, aplicando taxas de seleção distintas, já que a quantidade de vídeos em cada categoria difere e pelo menos duas imagens são extraídas de cada vídeo. Após esse processo, obtemos dois conjuntos de dados de imagens, cada um contendo um número aproximadamente semelhante de amostras vivas ou falsas. Um resumo estatístico dos dados de treinamento e teste é mostrado na Tabela 3.

Tabela 3 - Avaliação dos modelos

MODELO	REPLAY				ROSE		
	ALVO	T1/FL	T2/FR	MÉDIA	T1/FL	T2/FR	MÉDIA
FACENET/L	FL	0,0000	0,5000	0,2500	0,1265	0,4826	0,3046
FACENET/P	FR	0,4025	0,0019	0,2022	0,3262	0,1578	0,2420
LIVEFACE	FL+FR	0,4550	0,4349	0,4450	0,4093	0,4372	0,4233
MTL + FACENET	FL+FR	0,3900	0,0275	0,2088	0,3948	0,4883	0,4416

SUGERIDO	FL+FR	0,0100	0,3321	0,1711	0,1468	0,1733	0,1601
MODELO PESQUISA	FL+FR	0,0805	0,0264	0,0535 (>15%)	0,1773	0,1399	0,1586 (>8%)

FONTE: Tabela proveniente da própria pesquisa.

Apenas uma fração de uma imagem pode ser ocupada por um rosto. Por isso, utilizamos técnicas de localização e alinhamento facial em todas as imagens, permitindo que apenas as áreas com rostos sejam recortadas e suas versões padronizadas sejam armazenadas para uso posterior. Esta etapa de processamento é típica em trabalhos relacionados a faces. Todas as imagens mantêm o formato RGB padrão; os vídeos que utilizam profundidade na base de dados REPLAY não são considerados em nosso estudo.

Percebe-se que há poucos trabalhos relacionados com objetivos semelhantes a este estudo sobre modelos de tarefa dupla “Dual-task”, também consideramos modelos de ponta que se alinham parcialmente com nosso objetivo (ou modelos de tarefa única).

FaceNet/P: Essa rede é considerada clássica para reconhecimento facial devido à sua performance notável. A sua acurácia serve como referência máxima que buscamos atingir na atividade de reconhecimento. Referimo-nos a este método como FaceNet/P para diferenciá-lo de outras bases de referência que utilizam “FaceNet”.

FaceNet/L: Originalmente, o “FaceNet” foi desenvolvido para reconhecimento e não para detecção de vivacidade. Portanto, reconfiguramos e treinamos a rede com o objetivo de classificar a vivacidade, utilizando a perda tripla como regulador.

MTL + FaceNet: Este é um modelo de referência que combina as tarefas de detecção de vivacidade facial e reconhecimento facial em um sistema MTL. Ele possui um núcleo com três blocos convolucionais que é compartilhado, seguido por duas subdivisões específicas para ajustes refinados. O “LiveFace” é treinado usando rótulos. Utilizar diretamente o rótulo de saída dele não é equitativo, já que os indivíduos testados não fazem parte do treinamento. Para garantir equidade, usamos o “LiveFace” como um extrator de características e extraímos os atributos ajustados de ambas as subdivisões. O rendimento em cada tarefa é determinado a partir das características específicas da tarefa.

Sugerido e modelo da pesquisa: uma variação do modelo da pesquisa, o qual mantivemos a estrutura do algoritmo, mas não exigindo que as três faces em cada trigêmeo de treinamento de L venham da mesma pessoa.

3.8.1. Interface de captura (mobile)

A interface do usuário (UI) refere-se à série de telas, páginas e elementos visuais — como botões e ícones — que permitem que uma pessoa interaja com um produto ou serviço. Em contrapartida, a experiência do usuário (UX) diz respeito ao sentimento e à experiência global que um usuário tem ao usar um produto ou serviço, abrangendo a sua eficácia, eficiência e satisfação (ISO 9241-210:2019).

O campo do UX tornou-se uma disciplina independente e cresceu em importância à medida que as empresas reconheceram a necessidade de criar produtos centrados no usuário (Norman & Draper, 1986). Garvey et al. (2016) destacaram que uma boa UX pode levar a maiores taxas de retenção de usuários, enquanto uma UI intuitiva pode melhorar a eficiência e satisfação do usuário.

As técnicas de pesquisa, como o teste de usabilidade, são essenciais para entender e melhorar a UX (Nielsen, 1994). No âmbito da UI, o design de interface centrado no usuário (UCD) tornou-se uma abordagem padrão, enfatizando a necessidade de envolver os usuários ao longo do processo de design (Gould & Lewis, 1985).

Com o advento de novas tecnologias, como Realidade Virtual (RV) e Inteligência Artificial (IA), o campo do design de UI/UX continua a evoluir. Kortum & Sorber (2015) examinaram a UX na RV, enquanto Javahery et al. (2004) exploraram o design de UI adaptativa, que muda com base no comportamento do usuário.

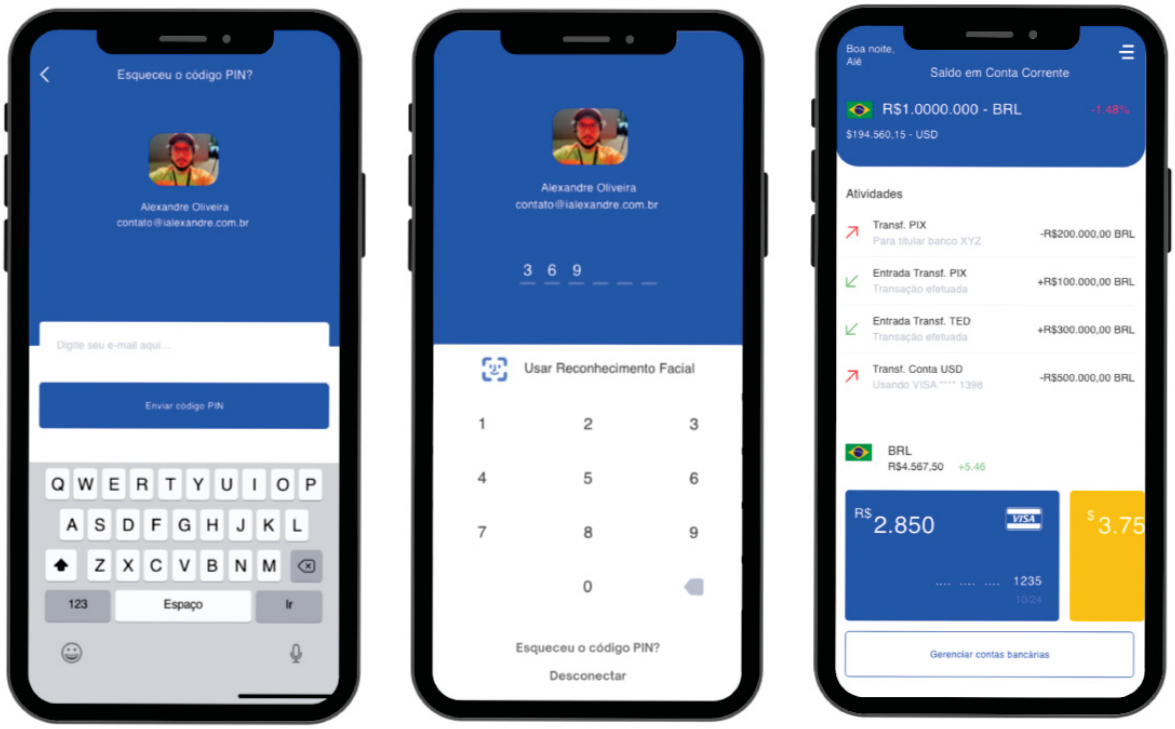
A tecnologia de captura facial tornou-se uma ferramenta essencial em muitas áreas, desde segurança até entretenimento. Com a evolução do aprendizado profundo e redes neurais convolucionais (CNNs), os aplicativos de captura facial agora possuem precisão e eficácia sem precedentes (Taigman et al., 2014).

4. RESULTADOS E DISCUSSÃO

Um aplicativo de captura facial geralmente funciona capturando imagens ou vídeos do rosto de um indivíduo e processando-os para identificar características únicas, que são então comparadas ou analisadas de acordo com o objetivo do aplicativo (Sagonas et al., 2016). Além das aplicações tradicionais em segurança e vigilância, a captura facial agora é amplamente utilizada em áreas como autenticação biométrica em smartphones, realidade aumentada, e até análise de sentimentos, avaliando emoções humanas através das expressões faciais (Kosti et al., 2017).

O objetivo deste projeto é se tornar uma proposta de negócio futuro, com foco no mercado financeiro e afins.

Figura 6 - Interface do usuário (mobile)



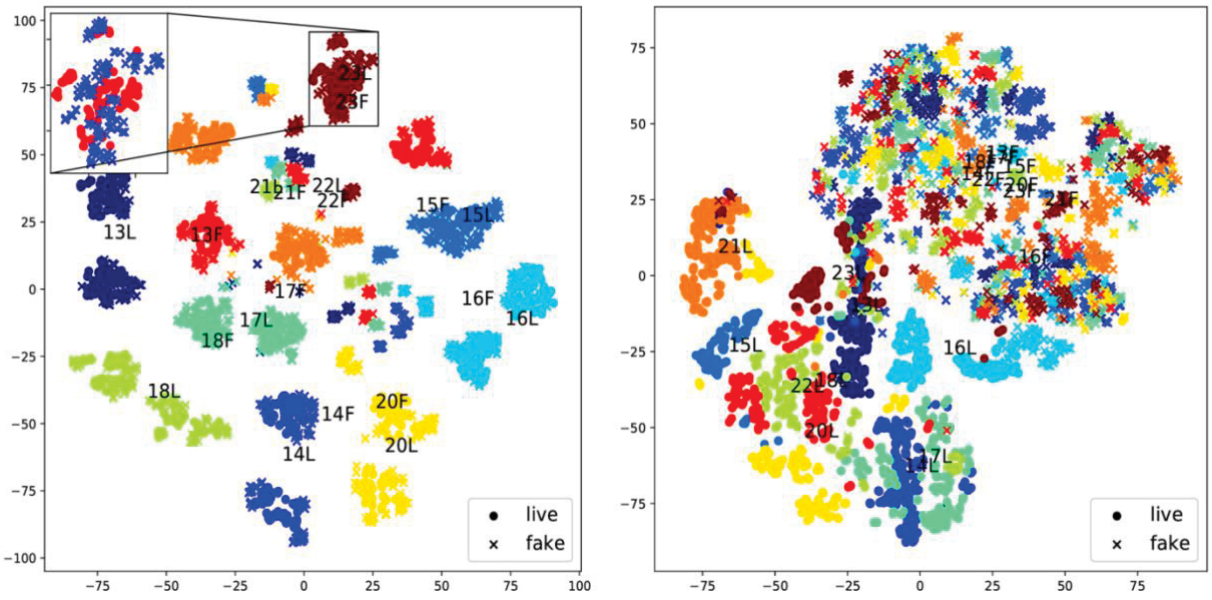
FONTE: Imagem proveniente da própria pesquisa.

4.1.1. Análise de dados

Observa-se que a nossa métrica HTER apresenta melhorias de 15% e 8% em comparação com os métodos FaceNet/P, FaceNet/L, LiveFace e MTL+FaceNet nos conjuntos REPLAY e ROSE, respectivamente. Algumas conclusões podem ser destacadas desta análise dos dados:

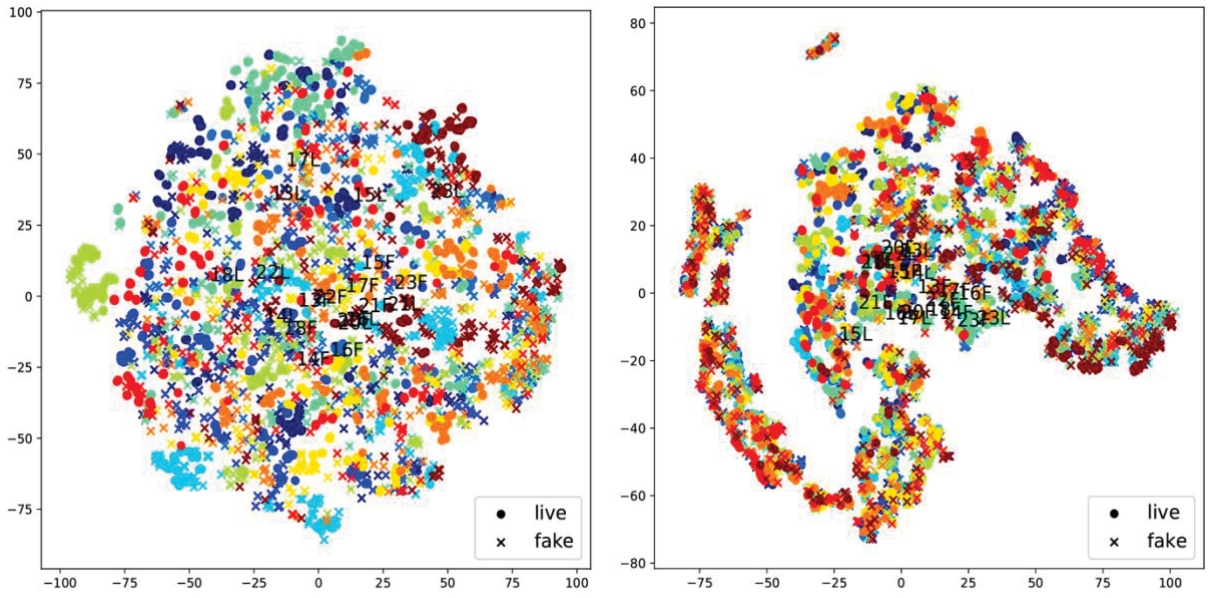
a) Modelos de tarefa única se sobressaem principalmente na tarefa para a qual foram projetados. Por exemplo, o FaceNet/P é eficiente em reconhecimento, mas não em detecção de vivacidade, particularmente no conjunto ROSE.
b) O modelo proposto pela pesquisa exibe boa performance em ambas as tarefas, aproximando-se dos melhores resultados obtidos pelos modelos de tarefa única.
c) Abordagens multitarefa, como LiveFace e MTL+FaceNet, não conseguem equilibrar eficientemente as duas tarefas. A perda em reconhecimento tende a convergir rapidamente, enquanto a detecção de vivacidade não apresenta melhorias significativas. Isso resulta em desempenho desequilibrado em conjuntos como REPLAY e ROSE. Assim, sugerimos a necessidade de uma abordagem (MTL) mais refinada para tarefas de variadas complexidades. Vide Figuras, 7, 8 e 9.

Figura 7 - Facenet (P) e Facenet (L)



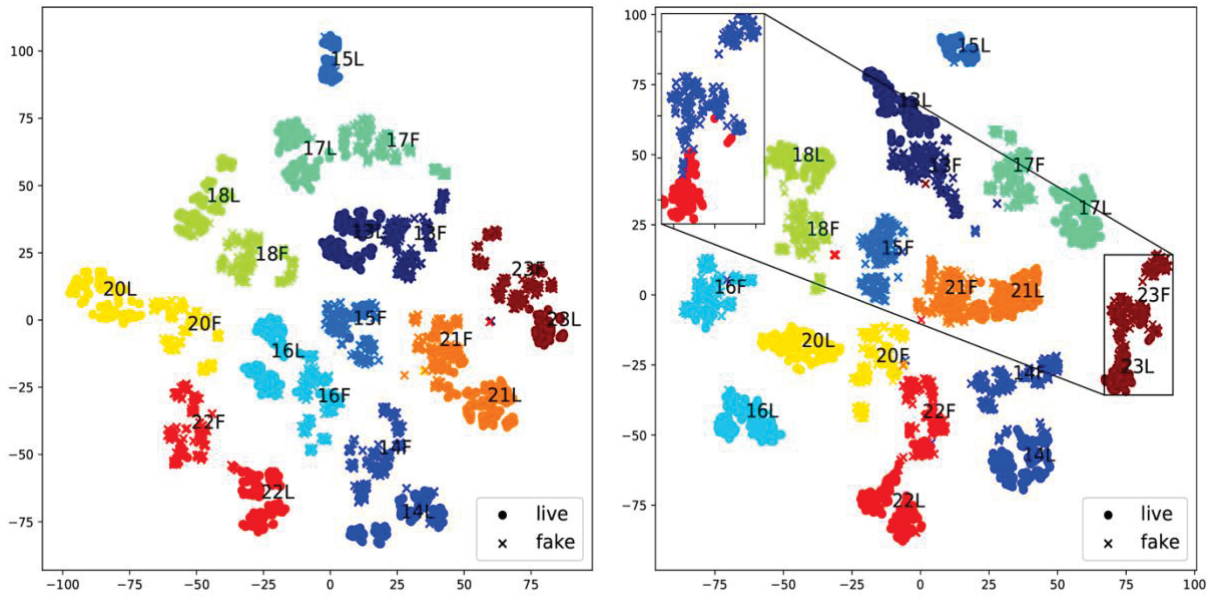
FONTE: Imagem proveniente da própria pesquisa.

Figura 8 - Sugerido versus Modelo Pesquisa



FONTE: Imagem proveniente da própria pesquisa.

Figura 9 - LiveFace + MTL + Facenet



FONTE: Imagem proveniente da própria pesquisa.

A distribuição de recursos T-SNE (biblioteca para programação Python) é comparada usando diversos métodos no conjunto de dados ROSE. Cada tonalidade representa um indivíduo; uma marca em forma de cruz (x) indica um rosto falso,

enquanto um ponto (.) representa um rosto verdadeiro. As anotações indicam o ID e a condição de autenticidade do rosto.

Analisamos a distribuição das distâncias entre pares de rostos conforme ilustrado nas Figuras, 7, 8 e 9. Nas notações, p, s e n representam rostos positivos, semi-positivos e negativos, respectivamente, quando comparados a um rosto referência a. Portanto, (a, p) representa um par de rostos reais da mesma pessoa (cor verde), (a, s) simboliza um par de rostos falsificados do mesmo indivíduo (cor vermelha) e (a, n) indica rostos de pessoas distintas (cor amarela).

O ideal é que os pontos de cada categoria sejam agrupados, enquanto pontos de categorias diferentes estejam separados. O FaceNet/P consegue distinguir os histogramas amarelos dos vermelhos/verdes, permitindo estabelecer um limite entre eles para identificar pessoas. Entretanto, ele confunde as categorias verdes e vermelhas, dificultando a separação entre rostos reais e falsificados. No entanto, com o nosso modelo, o vermelho é reajustado, posicionando-se entre o verde e o amarelo com menor sobreposição, permitindo definir dois limiares claros para as duas tarefas.

5. CONCLUSÕES

A distribuição de recursos T-SNE (biblioteca para programação Python) é comparada usando diversos métodos no conjunto de dados ROSE. Cada tonalidade representa um indivíduo; uma marca em forma de cruz (x) indica um rosto falso, enquanto um ponto (.) representa um rosto verdadeiro. As anotações indicam o ID e a condição de autenticidade do rosto.

Analisamos a distribuição das distâncias entre pares de rostos conforme ilustrado nas Figuras, 7, 8 e 9. Nas notações, p, s e n representam rostos positivos, semi-positivos e negativos, respectivamente, quando comparados a um rosto referência a. Portanto, (a, p) representa um par de rostos reais da mesma pessoa (cor verde), (a, s) simboliza um par de rostos falsificados do mesmo indivíduo (cor vermelha) e (a, n) indica rostos de pessoas distintas (cor amarela).

O ideal é que os pontos de cada categoria sejam agrupados, enquanto pontos de categorias diferentes estejam separados. O FaceNet/P consegue distinguir os histogramas amarelos dos vermelhos/verdes, permitindo estabelecer um limite entre eles para identificar pessoas. Entretanto, ele confunde as categorias verdes e vermelhas, dificultando a separação entre rostos reais e falsificados. No entanto, com

o nosso modelo, o vermelho é reajustado, posicionando-se entre o verde e o amarelo com menor sobreposição, permitindo definir dois limiares claros para as duas tarefas.

5.1. Sugestões de trabalhos futuros

O ideal é que os pontos de cada categoria sejam agrupados, enquanto pontos de categorias diferentes estejam separados. O FaceNet/P consegue distinguir os histogramas amarelos dos vermelhos/verdes, permitindo estabelecer um limite entre eles para identificar pessoas. Entretanto, ele confunde as categorias verdes e vermelhas, dificultando a separação entre rostos reais e falsificados. Logo, com o nosso modelo, o vermelho é reajustado, posicionando-se entre o verde e o amarelo com menor sobreposição, permitindo definir dois limiares claros para as duas tarefas.

REFERÊNCIAS BIBLIOGRÁFICAS

Bengio, Y. (2009). Learning deep architectures for AI. Foundations and trends® in Machine Learning, 2(1), 1-127.

Byungin Yoo, KWAK Youngjun, Jungbae Kim, SON Jinwoo, LEE Changkyo, Chang Kyu Choi, and HAN JaeJoon. 2019. Liveness test method and apparatus. US Patent App. 16/148,587.

Chien Eao Lee, Lilei Zheng, Ying Zhang, Vrizlynn LL Thing, and Ying Yu Chu. 2018.

Towards building a remote anti-spoofing face authentication system. In TENCON 2018-2018 IEEE Region 10 Conference. IEEE, 0321–0326.

Di Wen, Hu Han, and Anil K Jain. 2015. Face spoof detection with image distortion analysis. IEEE Transactions on Information Forensics and Security 10, 4 (2015), 746–761.

Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. 2014. Learning face representa-tion from scratch. arXiv preprint arXiv:1411.7923 (2014).

Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In CVPR. 815–823.

Gang Pan, Lin Sun, Zhaohui Wu, and Shihong Lao. 2007. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In ICCV. IEEE, 1–8.

Garvey, M., et al. (2016). UX Design: Leading the Way to an Improved User Experience. UX Magazine.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

Gould, J. D., & Lewis, C. (1985). Designing for usability: key principles and what designers think. Communications of the ACM, 28(3), 300-311.

Hakan Cevikalp and Bill Triggs. 2010. Face recognition based on image sets. In CVPR. 2567–2573.

Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. 2018. Cosface: Large margin cosine loss for deep face recognition. In CVPR. 5265–5274.

Hao Ye, Weiyuan Shao, Hong Wang, Jianqi Ma, Li Wang, Yingbin Zheng, and Xiangyang Xue. 2016. Face recognition via active annotation and learning. In ACM MM. 1058–1062.

Haoliang Li, Wen Li, Hong Cao, Shiqi Wang, Feiyue Huang, and Alex C Kot. 2018.

Unsupervised domain adaptation for face anti-spoofing. IEEE Transactions on Information Forensics and Security 13, 7 (2018), 1794–1809.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition.

Huafeng Kuang, Rongrong Ji, Hong Liu, Shengchuan Zhang, Xiaoshuai Sun, Feiyue Huang, and Baochang Zhang. 2019. Multi-modal Multi-layer Fusion Network with Average Binary Center Loss for Face Anti-spoofing. In ACM MM. 48–56.

Iasonas Kokkinos. 2017. Ubernet: Training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 6129–6138.

ISO 9241-210:2019. Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems.

Ivana Chingovska, Amir Mohammadi, André Anjos, and Sébastien Marcel. 2019. Evaluation Methodologies for Biometric Presentation Attack Detection. In Hand-book of Biometric Anti-Spoofing. Springer, 457–480.

Ivana Chingovska, André Anjos, and Sébastien Marcel. 2012. On the Effectiveness of Local Binary Patterns in Face Anti-spoofing. IEEE BIOSIG 2012.

Javahery, H., et al. (2004). Beyond Power: Making Biofeedback Computers User-Centered. ACM Interactions, 11(3), 52-59.

Javier Galbally, Sébastien Marcel, and Julian Fierrez. 2014. Image quality as-sessment for fake biometric detection: Application to iris, fingerprint, and face recognition. IEEE transactions on image processing 23, 2 (2014), 710–724.

Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. 2019. Arcface: Additive angular margin loss for deep face recognition. In CVPR. 4690–4699.

Jianwei Yang, Zhen Lei, and Stan Z Li. 2014. Learn convolutional neural network for face anti-spoofing. arXiv preprint arXiv:1408.5601 (2014).

Jukka Määttä, Abdenour Hadid, and Matti Pietikäinen. 2011. Face spoofing detection from single images using micro-texture analysis. In IJCB. 1–7.

Junlin Hu, Jiwen Lu, and Yap-Peng Tan. 2014. Discriminative deep metric learning for face verification in the wild. In CVPR. 1875–1882.

Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. 2016. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters 23, 10 (2016), 1499–1503.

Kittler, J., G. Matas, K. Jonsson, and M. Sanchez,“Combining evidence in personal identity verification systems,” Pattern Recognition Letters, vol.18, no.9,pp.845–852,Sept. 1997.

Kortum, P., & Sorber, M. (2015). Measuring the Usability of Virtual Reality. International Journal of Human–Computer Interaction, 31(5), 365-377.

Kosti, R., Alvarez, J. M., Recasens, A., & Lapedriza, À. (2017). Emotion recognition in context. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1667-1675).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

Lei Li, Xiaoyi Feng, Zinelabidine Boulkenafet, Zhaoqiang Xia, Mingming Li, and Abdenour Hadid. 2016. An original face anti-spoofing approach using partial convolutional neural network. In IPTA. 1–6.

Lei Li, Zhaoqiang Xia, Xiaoyue Jiang, Fabio Roli, and Xiaoyi Feng. 2018. Face presentation attack detection in learned color-liked space. arXiv preprint arXiv:1810.13170 (2018).

Litong Feng, Lai-Man Po, Yuming Li, Xuyuan Xu, Fang Yuan, Terence Chun-Ho Cheung, and Kwok-Wai Cheung. 2016. Integration of image quality and motion cues for face anti-spoofing: A neural network approach. Journal of Visual Communication and Image Representation 38 (2016), 451–460.

Marta Zorrilla, Juan Yebenes, A reference framework for the implementation of Data Governance Systems for Industry 4.0, Computer Standards & Interfaces (2021), doi: https://doi.org/10.1016/j.csi.2021.103595.

Matthew Turk and Alex Pentland. 1991. Face recognition using eigenfaces. In CVPR. 586–587.

Michelle Guo, Albert Haque, De-An Huang, Serena Yeung, and Li Fei-Fei. 2018. Dynamic task prioritization for multitask learning. In ECCV.

Nesli Erdogmus and Sebastien Marcel. 2013. Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect. In BTAS. IEEE, 1–6.

Nielsen, J. (1994). Usability engineering. Morgan Kaufmann Publishers Inc.

Norman Poh and Samy Bengio. 2006. Database, protocols and tools for evaluating score-level fusion algorithms in biometric authentication. Pattern Recognition 39, 2 (2006), 223–233.

Norman, D. A., & Draper, S. W. (1986). User Centered System Design; New Perspectives on Human-computer Interaction.

Olivier Moindrot. 2018. Triplet Loss and Online Triplet Mining in TensorFlow. https://omoindrot.github.io/triplet-loss#batch-hard-strategy.

Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. 2018. Vggface2: A dataset for recognising faces across pose and age. In FG. 67–74.

Ross, A., and Jain, A.K., “Information fusion in biometrics”, Pattern Recognition Letters 24, 13 (Sept. 2003), 2115–2125. Sanderson, C. and K.K. Paliwal (2003), “Fast features for fac authentication under illumination direction changes”, Pattern Recognition Letters 24, 2409-2419.

Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. (2016). A semi-automatic methodology for facial landmark annotation. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 896-904).

Sebastian Ruder. 2017. An overview of multi-task learning in deep neural net-works. arXiv preprint arXiv:1706.05098 (2017).

Sheng Hua Bao, Min Li, Wei Hong Qian, and Zhong Su. 2017. Secure face authentication with liveness detection for mobile. US Patent 9,684,779.

Shiming Ge, Shengwei Zhao, Xindi Gao, and Jia Li. 2019. Fewer-Shots and Lower-Resolutions: Towards Ultrafast Face Recognition in the Wild. In ACM MM. 229–237.
SOMMERVILLE, I. Software engineering. Boston: Pearson, 2011. ISBN 978-0137035151.

Sumit Chopra, Raia Hadsell, and Yann LeCun. [n.d.]. Learning a similarity metric discriminatively, with application to face verification. In CVPR.

Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. Proceedings of the IEEE conference on computer vision and pattern recognition, 1701-1708.

Tiago de Freitas Pereira, Jukka Komulainen, André Anjos, José Mario De Martino, Abdenour Hadid, Matti Pietikäinen, and Sébastien Marcel. 2014. Face liveness detection using dynamic texture. EURASIP Journal on Image and Video Processing 2014, 1 (2014), 2.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision. 2980–2988.

Xiaokang Tu and Yuchun Fang. 2017. Ultra-deep neural network for face anti-spoofing. In International Conference on Neural Information Processing. Springer, 686–695.

Xiaowen Ying, Xin Li, and Mooi Choo Chuah. 2018. LiveFace: A Multi-task CNN for Fast Face-Authentication. In ICMLA. 955–960.

Yan Ke He Zhen-Yu Zhang, Jian and Yong Xu. 2014. A Collaborative Linear Discriminative Representation Classification Method for Face Recognition. In International Conference on Artificial Intelligence and Software Engineering.

Yan Li, Yingjiu Li, Ke Xu, Qiang Yan, and Robert H Deng. 2016. Empirical study of face authentication systems under OSNFD attacks. IEEE Transactions on Dependable and Secure Computing 15, 2 (2016), 231–245.

Yasar Abbas Ur Rehman, Lai-Man Po, Mengyang Liu, Zijie Zou, Weifeng Ou, and Yuzhi Zhao. 2019. Face liveness detection using convolutional-features fusion of real and deep network generated face images. Journal of Visual Communication and Image Representation 59 (2019), 574–582.

Yehia Hani, Takaaki Kuratate, Eric Vatikiotic-Bateson, “Linking Facial Animation, Head Motion and Speech Acoustics”, Journal of Phonetics, Vol.30, Issue 3, 2002.

Yehia, H., Rubin, P. and Vatikiotic-Bateson E. (1998), “Quantitative association of vocal tract and facial behavior”, Journal of Speech Communication 26(1-2), 23-43.

Ying Zhang, Lilei Zheng, Vrizlynn L.L. Thing, Roger Zimmermann, Bin Guo, and Zhiwen Yingyuan Yang, Jinyuan Sun, and Linke Guo. 2016. PersonaIA: a lightweight implicit authentication system based on customized user behavior selection. IEEE Transactions on Dependable and Secure Computing 16, 1 (2016), 113–126.

Yue Wu, Hongfu Liu, Jun Li, and Yun Fu. 2017. Deep face recognition with center invariant loss. In ACM MM Thematic Workshops. 408–414.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. European conference on computer vision. Springer.

Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., & Li, S. Z. (2016). A face anti”Spoofing” database with diverse attacks. IEEE Access, 4, 5111-5116.

Zhang, Z., Zhao, J., & Le, C. (2016). A survey on face recognition techniques and its applications. Procedia Computer Science, 91, 128-135.

Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. 2018. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In International Conference on Machine Learning. 794–803.

Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. ACM computing surveys (CSUR), 35(4), 399-458.

Zhenqi Xu, Shan Li, and Weihong Deng. 2015. Learning temporal features using LSTM-CNN architecture for face anti-spoofing. In ACPR. 141–145.

Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Z Li. 2012. A face antispoofing database with diverse attacks. In ICB. 26–31.

Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. 2016. Face spoofing detection using colour texture analysis. IEEE Transactions on Information Forensics and Security 11, 8 (2016), 1818–1830.

Zinelabinde Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid. 2017. OULU-NPU: A mobile face presentation attack database with real-world variations. In FG. 612–618.

Zuheng Ming, Junshi Xia, Muhammad Muzzamil Luqman, Jean-Christophe Burie, and Kaixing Zhao. 2019. FaceLiveNet+: A Holistic Networks For Face Authen-tication Based On Dynamic Multi-task Convolutional Neural Networks. arXiv preprint arXiv:1902.11179 (2019).