UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

BEATRIZ FRÉCCIA AMANTE

RASTREAMENTO DE PESSOAS POR TÉCNICAS DE APRENDIZAGEM DE MÁQUINA

GUARAPUAVA

2025

BEATRIZ FRÉCCIA AMANTE

RASTREAMENTO DE PESSOAS POR TÉCNICAS DE APRENDIZAGEM DE MÁQUINA

Person tracking using machine learning techniques

Proposta de Trabalho de Conclusão de Curso de Graduação apresentado como requisito para obtenção do título de Tecnólogo em Tecnologia em Sistemas para Internet do Curso Superior de Tecnologia em Sistemas para Internet da Universidade Tecnológica Federal do Paraná.

Orientador: Drª Kelly Lais Wiggers

GUARAPUAVA

2025

RESUMO

Este trabalho propõe o desenvolvimento de uma API *RESTful* acessível e personalizável para rastreamento e reidentificação de pessoas em vídeos, utilizando técnicas de aprendizado de máquina e visão computacional. A proposta busca democratizar o uso de sistemas baseados em inteligência artificial (IA) ao integrar modelos de detecção, rastreamento e extração de características visuais em uma interface funcional e intuitiva. Tais tecnologias, embora amplamente estudadas, ainda enfrentam barreiras de adoção devido à complexidade técnica e à escassez de ferramentas públicas e bem documentadas. A solução desenvolvida permitirá identificar indivíduos em transmissões ao vivo ou vídeos enviados, com suporte à adição de identificadores únicos e proteção da privacidade por meio de técnicas de criptografia. A iniciativa visa preencher lacunas existentes entre bibliotecas de pesquisa e aplicações práticas, oferecendo uma plataforma modular aplicável em contextos como segurança urbana, automação, controle de acesso e projetos sociais. O projeto utiliza uma biblioteca consolidada, YOLO aliada a frameworks *web* como *FastAPI*, para construir um *pipeline* de rastreamento end-to-end com foco em desempenho e acessibilidade por desenvolvedores e pesquisadores.

Palavras-chave: inteligencia artificial; deep learning; people tracking; re-identification; real-time video.

ABSTRACT

This project proposes the development of an accessible and customizable RESTful API for person tracking and re-identification in videos, using machine learning and computer vision techniques. The goal is to democratize the use of AI-based systems by integrating object detection, tracking, and feature extraction into a functional and user-friendly interface. Despite significant advances in this area, adoption remains limited due to technical complexity and the lack of public, well-documented tools. The proposed solution identifies individuals in both live video streams and uploaded recordings, supports unique identifiers, and includes data privacy measures such as encryption. The system bridges the gap between research libraries and practical applications, offering a modular platform suitable for urban security, automation, access control, and social projects. Built with an established library, YOLO, DeepSORT, and web frameworks like FastAPI, this end-to-end pipeline emphasizes performance and accessibility for developers and researchers.

Keywords: artificial intelligence; deep learning; people tracking; re-identification; real-time video.

SUMÁRIO

1 INTRODUÇÃO

Com o avanço da era da informação, a sociedade tem se adaptado de forma contínua às inovações tecnológicas, que permeiam desde aspectos cotidianos, como a substituição de linhas telefônicas por dispositivos móveis, até aplicações complexas de Inteligência Artificial (IA) voltadas à automação e otimização de processos em escala.

Assistentes virtuais, *chatbots* e sistemas inteligentes tornaram-se cada vez mais comuns no cotidiano da população, contribuindo para automatizar tarefas, reduzir a necessidade de intervenção humana e minimizar falhas operacionais. Tais sistemas são baseados em arquiteturas de redes neurais artificiais inspiradas no funcionamento do cérebro humano, compostas por múltiplas camadas ocultas (*hidden layers*) que simulam sinapses neurais e realizam cálculos matemáticos em paralelo para transformar entradas (como imagens ou textos) em saídas interpretáveis. O aprendizado ocorre a partir de grandes volumes de dados previamente rotulados, utilizando algoritmos de retro-propagação e otimização, de modo que o sistema passe a inferir padrões e tomar decisões com base em novas informações (GOODFELLOW; BENGIO; COURVILLE, 2016).

Entre os maiores desafios da IA está o campo da visão computacional, mais especificamente o reconhecimento e o rastreamento de objetos em vídeos. Essa subárea da IA busca capacitar máquinas a interpretarem o conteúdo visual de imagens e vídeos, com o objetivo de detectar e acompanhar, ao longo do tempo, objetos ou indivíduos em movimento. Suas aplicações abrangem áreas como segurança pública, monitoramento urbano, controle de acesso - Wei *et al.* (2020), varejo e marketing comportamental - Junior e Martini (2019), sistemas de transporte inteligentes e saúde pública - Silva (2022), por exemplo, para análise de fluxo em hospitais ou acompanhamento de pacientes com distúrbios de mobilidade (YADAV *et al.*, 2022).

Modelos modernos permitem, além da detecção de indivíduos, seu reconhecimento com base em características visuais, mesmo diante de ambientes com múltiplas câmeras, iluminação variável ou alta densidade populacional. Técnicas como detecção de objetos(por exemplo, o algoritmo *YOLO — You Only Look Once*, que divide imagens em grades e identifica objetos em tempo real) - Redmon *et al.* (2016), rastreamento multi-objetos (como o DeepSORT, que combina informações espaciais com *features* visuais para acompanhar indivíduos ao longo de sequências) - Wojke, Bewley e Paulus (2017) e re-identificação de pessoas (*person re-identification*), que utiliza vetores numéricos chamados *embeddings* para reconhecer a mesma pessoa em diferentes contextos, são frequentemente utilizadas em conjunto para garantir que uma mesma pessoa seja reconhecida ao longo de diferentes momentos e cenas (MCLAUGHLIN; RINCON; MILLER, 2016).

Apesar do avanço dessas soluções, sua adoção ainda é limitada pela complexidade técnica envolvida em sua implementação. Muitos dos *frameworks* existentes exigem conhecimentos especializados em aprendizado profundo (*deep learning*), engenharia de software e manipulação de vídeo em tempo real, o que restringe sua aplicação a empresas com equipes

altamente capacitadas. Além disso, o processo de treinamento de modelos próprios demanda poder computacional significativo e acesso a grandes bases de dados anotadas, o que inviabiliza o uso por desenvolvedores independentes ou instituições de pequeno porte.

Diante desse cenário, este projeto propõe o desenvolvimento de uma API acessível e personalizável que possibilite o rastreamento de pessoas em vídeos ao vivo (via *websocket*) ou por meio de arquivos enviados. O sistema utilizará modelos previamente treinados, com o intuito de facilitar sua aplicação em contextos diversos, como segurança urbana, empresas e ambientes domésticos. A proposta visa democratizar o acesso a essa tecnologia, unindo conceitos de processamento de vídeo, aprendizado de máquina e desenvolvimento web em uma solução funcional e intuitiva.

Adicionalmente, nota-se uma escassez de interfaces e APIs públicas que ofereçam suporte eficiente para reconhecimento de pessoas por imagem ou vídeo. Quando existentes, essas ferramentas costumam ser mal documentadas, desatualizadas ou de difícil integração com sistemas modernos. Este projeto busca preencher essa lacuna, utilizando modelos pré-treinados e *frameworks* de código aberto, como YOLOv8 -Redmon *et al.* (2016), DeepSORT - Ahmad (2023) e FastAPI - Ramírez (2023), para simplificar a adoção da tecnologia mesmo por usuários sem formação em IA.

Cabe destacar que, por lidar com dados sensíveis como imagens de pessoas, questões como desempenho, acurácia e segurança são cruciais. A aplicação proposta incluirá criptografia de identificadores (por meio de *hashes* gerados com funções seguras como *SHA-256* - Standards e Technology (2015)) a fim de proteger a privacidade dos usuários e mitigar riscos de ataques ou uso indevido das informações. A interface será concebida de forma clara e segura, promovendo facilidade de uso sem comprometer a proteção dos dados.

Ao final, espera-se obter um protótipo funcional de API que permita rastrear indivíduos em vídeos com o uso de IA, com possibilidade de extensão futura para funcionalidades como alertas em tempo real, reconhecimento corporal, múltiplos alvos e *dashboards* analíticos.

1.1 Objetivos

Aqui abrangeremos os objetivos que esse projeto visa lidar.

1.1.1 Objetivo geral

Desenvolver uma API funcional para rastreamento de pessoas em vídeos, com suporte tanto para transmissões em tempo real quanto para *uploads*, utilizando técnicas de aprendizado de máquina e visão computacional.

1.1.2 Objetivos específicos

- Investigar e selecionar bases de dados e um modelo de detecção, rastreamento e re-identificação de pessoas que ofereça bom desempenho em tempo real.
- Integrar o modelo selecionado em uma aplicação acessível via API.
- Permitir a inserção de identificadores personalizados para reconhecimento individual.
- Criar uma interface mínima para interação com o sistema, com foco em usabilidade e segurança.
- Avaliar o desempenho do sistema em diferentes contextos (qualidade de vídeo, iluminação, número de indivíduos, métricas estatísticas).
- Implementar mecanismos de criptografia para garantir a privacidade dos dados dos usuários.
- Documentar a API de forma clara, visando sua reutilização por outros pesquisadores e desenvolvedores.

1.2 Justificativa

O tema proposto é de relevância no cenário contemporâneo, especialmente no contexto da crescente urbanização, da digitalização de serviços e do aumento da demanda por soluções tecnológicas voltadas à segurança, automação e gestão inteligente de espaços. O uso de sistemas de rastreamento e reconhecimento de pessoas por vídeo tem potencial para impactar diretamente áreas críticas como segurança pública, cidades inteligentes (*smart cities*), controle de acesso, logística, varejo, saúde e ambientes corporativos (MELO; SERRA, 2022).

Em meio a esse panorama, destaca-se a dificuldade de acesso a ferramentas que possibilitem a implementação de tais sistemas de forma prática e acessível. As soluções atualmente disponíveis para rastreamento e identificação de pessoas por vídeo são, em sua maioria, restritas a grandes empresas ou instituições com equipes altamente especializadas, uma vez que exigem domínio técnico em aprendizado profundo (*deep learning*), redes neurais convolucionais (RNCs), visão computacional e manipulação de fluxos de vídeo em tempo real. Essa barreira tecnológica restringe a adoção mais ampla da IA, impedindo que pesquisadores independentes, pequenas empresas e até mesmo desenvolvedores experientes — mas não especializados em IA — consigam criar suas próprias soluções personalizadas (ALMASAWA; ELREFAEI; MORIA, 2019).

Além disso, falta no mercado um ecossistema sólido de APIs (termo refere-se à Interface de Programação de Aplicação, ou seja, um programa intermediário que se comunica com solicitações e respostas) públicas bem documentadas que permitam integração rápida com sistemas

já existentes, o que limita o uso prático de IA em vídeo em contextos fora do meio corporativo. Este projeto responde diretamente a essa lacuna, propondo uma API robusta, adaptável e fácil de usar, que permita a qualquer pessoa — com ou sem conhecimento profundo em IA — implementar sistemas de rastreamento com segurança e eficiência.

Ao desenvolver uma API acessível, customizável e de fácil integração, este trabalho visa democratizar o acesso à tecnologia de rastreamento por vídeo baseada em IA. A proposta é permitir que diferentes perfis de usuários — incluindo pesquisadores, desenvolvedores web e profissionais de TI — possam aplicar essas técnicas sem a necessidade de treinar modelos do zero ou compreender toda a complexidade dos algoritmos subjacentes. O projeto também se destaca por abordar um ponto sensível no desenvolvimento de sistemas de monitoramento: a proteção da privacidade e a segurança da informação. Por isso, o sistema proposto incorpora práticas como criptografia de identificadores (*hashes*) e uma interface clara e funcional, com foco em confiabilidade e proteção dos dados sensíveis dos usuários.

Adicionalmente, o projeto está alinhado aos princípios de viabilidade técnica e aplicabilidade prática. Existem bibliotecas e *frameworks* consolidados no ecossistema de código aberto — como OpenCV, YOLO, DeepSORT e Flask/FastAPI — que podem ser aproveitados para estruturar uma solução funcional dentro do tempo e dos recursos disponíveis no desenvolvimento de um trabalho de conclusão de curso.

Por fim, o projeto é também motivado por um interesse pessoal da autora nas áreas de aprendizado de máquina e desenvolvimento *web*. Trata-se, portanto, de uma oportunidade de aplicar conhecimentos adquiridos ao longo da formação acadêmica em um desafio real, que une teoria e prática e contribui tanto para o avanço pessoal quanto para o debate e a produção de conhecimento dentro da área de computação aplicada.

2 CONTEXTUALIZAÇÃO

2.1 Contextualização Geral

Atualmente, um dos maiores desafios enfrentados por profissionais e pesquisadores que desejam utilizar inteligência artificial para rastreamento de pessoas em vídeo é a ausência de soluções completas e acessíveis que conduzam todo o processo de ponta a ponta (ALMASAWA; ELREFAEI; MORIA, 2019). Em outras palavras, falta no mercado e na comunidade científica uma ferramenta integrada que vá desde a captura do vídeo até a entrega dos resultados processados de forma compreensível e utilizável.

Um dos elementos centrais dessa lacuna está na geração automática de regiões de interesse (ROIs). Regiões de interesse são áreas específicas dentro de uma imagem ou vídeo onde se presume haver informação relevante — neste caso, a presença de uma pessoa. Para que sistemas automatizados funcionem corretamente, é necessário que consigam identificar, sem intervenção humana, onde essas pessoas estão no vídeo. Essa tarefa é realizada por modelos de detecção, como o YOLO (You Only Look Once) (REDMON *et al.*, 2016), que segmentam os quadros do vídeo em tempo real, indicando com precisão onde os objetos — pessoas, neste caso — estão localizados.

Após a detecção, o segundo passo necessário é o processamento com inteligência artificial para a extração de características únicas dessas pessoas detectadas. Essas características, chamadas de *features* ou *embeddings*, são vetores numéricos que codificam informações visuais relevantes de cada indivíduo — como tipo de roupa, cor, estrutura corporal, e outros padrões que, juntos, permitem distinguir uma pessoa de outra. Esses vetores são fundamentais para realizar o que se chama de "re-identificação de pessoas", ou seja, reconhecer o mesmo indivíduo em câmeras diferentes, em momentos distintos, mesmo que ele não esteja sempre na mesma posição ou iluminação.

Diversas abordagens têm sido desenvolvidas para gerar *embeddings* eficazes para tal objetivo, e dentro elas, incluem-se:

- Aprendizado Métrico Profundo (*Deep Metric learning*):
  - Utiliza redes neurais treinadas com funções de perda específicas, como a *triplet loss*, para mapear imagens de pessoas em um espaço vetorial onde indivíduos semelhantes estão próximos e diferentes estão distantes.
  - É eficaz na captura de relações de similaridades e diferenças entre indivíduos, (LI *et al.*, 2023)
- Aprendizado de Características Locais (*Local Feature Learning*):

  - Foca na extração de partes específicas do corpo (por exemplo, cabeça, ombro, joelho e pé) para gerar *embeddings* mais robustos e variações de pose e oclusões.
  - Modelos como o PCB(Part-based Convolutional Baseline) dividem a imagem em segmentos horizontais e extraem características de cada parte, (WU *et al.*, 2024).

- Redes Adversarias Generativas (GANs):
  - Treina duas redes neurais generativas para competirem entre si e gerar novos dados mais autênticos a partir de um determinado conjunto de dados de treinamento.
  - Utilizadas para gerar imagens sintéticas ou transformar imagens existentes, ajudando a modelar variações de aparência e melhorar a generalização dos *embeddings*, (Amazon Web Services, 2023).

- Aprendizado de Características Sequenciais (Sequence Feature Learning):
  - Explora informações temporais em sequências de vídeo, utiliza redes recorrentes ou mecanismos de atenção para capturar padrões dinâmicos de movimento e comportamento.
  - Essa abordagem é valiosa para re-identificação em vídeos, onde o movimento pode fornecer pistas adicionais, (AL-JABERY *et al.*, 2020).

Enquanto o YOLO (*You Only Look Once*) é amplamente utilizado para detecção de objetos em tempo real, incluindo pessoas, ele não é projetado para re-identificação. O YOLO divide a imagem em uma grade e prevê *bounding boxes* e classes para cada célula, permitindo a detecção rápida de objetos. No entanto, para re-identificar indivíduos, é necessário extrair *embeddings* que capturem características únicas além da simples detecção.

Portanto, após a detecção inicial com o YOLO, é comum empregar modelos especializados em re-identificação para extrair *embeddings* e realizar a correspondência entre indivíduos ao longo do tempo e em diferentes câmeras.

O terceiro aspecto crítico é a entrega dos resultados em um sistema *web* funcional, que permita que esses dados processados sejam utilizados por outros sistemas ou por usuários humanos sem conhecimento técnico em IA. Isso envolve não apenas a construção de uma API pública (interface de programação de aplicações), que permita que sistemas externos enviem vídeos e recebam respostas com as identificações, mas também a criação de uma interface gráfica amigável, acessível por navegador, onde se possa testar e visualizar o funcionamento da solução. É nesse ponto que se torna clara a importância da arquitetura *end-to-end* — ou seja, que conecte todos os estágios do processo de forma integrada e fluida.

Tornar tecnologias avançadas de rastreamento de pessoas por vídeo acessíveis e reutilizáveis possui grande impacto social e tecnológico. A segurança pública e privada, por exemplo, frequentemente depende da análise de horas de gravações de câmeras, o que consome tempo e recursos humanos. Ferramentas automatizadas podem agilizar esse processo, reduzindo falhas humanas e otimizando respostas em tempo real. Em ambientes corporativos ou comerciais, como shoppings, supermercados ou escolas, o monitoramento de fluxo de pessoas permite, além da segurança, a análise de comportamento de clientes, otimização de rotas e melhoria da experiência do usuário. Em prédios públicos ou privados, o controle de acesso por meio de re-identificação permite sistemas mais inteligentes e adaptativos, promovendo automação predial e redução de custos com pessoal.

O projeto também se mostra essencial do ponto de vista da acessibilidade e inclusão digital, pois oferece uma alternativa gratuita, aberta e de baixo custo para uma tecnologia que atualmente é majoritariamente proprietária e cara. Isso possibilita que organizações de menor porte, iniciativas sociais e pesquisadores em início de carreira possam acessar e explorar o potencial da inteligência artificial aplicada ao vídeo, mesmo sem uma grande infraestrutura computacional.

2.1.1 Estudos Relacionados

Este trabalho se insere dentre múltiplas linhas de pesquisa consolidadas na área de ciência da computação e engenharia de software. Está diretamente relacionado à área de reconhecimento facial e re-identificação de pessoas, que estuda formas de identificar indivíduos por meio de características visuais, com ou sem uso de biometria direta. Assim também com o campo da vigilância inteligente, onde algoritmos analisam vídeos em tempo real para tomar decisões ou alertar operadores sobre eventos de interesse.

Um exemplo significativo é o trabalho de Chen *et al.* (2018), que aplicou Deep Metric Learning com a função de perda *triplet loss* para gerar *embeddings* discriminativos em tarefas de re-identificação. Os autores demonstraram que esse método pode atingir alta acurácia ao separar vetorialmente indivíduos distintos com eficácia, mesmo em grandes conjuntos de dados.

Já o modelo *PCB (Part-based Convolutional Baseline)*, proposto por Sun *et al.* (2021), adota uma abordagem de aprendizado local, segmentando a imagem em partes horizontais e extraindo características de cada uma delas. Essa estratégia mostrou-se particularmente eficaz em cenários com variações de pose e oclusão parcial, frequentemente encontradas em ambientes reais.

No contexto de melhoria de generalização e aumento da variabilidade visual, Karmakar e Mishra (2021) introduziram o uso de GANs para gerar amostras sintéticas de pessoas. Isso permitiu treinar modelos mais robustos mesmo com conjuntos de dados limitados, ampliando a capacidade dos *embeddings* em generalizar para novos indivíduos.

Quanto à re-identificação em vídeos, McLaughlin, Rincon e Miller (2016) propuseram uma rede recorrente convolucional (RCN) capaz de capturar padrões temporais para melhorar a identificação de indivíduos em sequências contínuas. O modelo demonstrou que o uso de informações temporais melhora significativamente o desempenho em comparação com abordagens baseadas apenas em imagens estáticas.

Além disso, conecta-se com pesquisas em rastreamento de multi-objetos (*multi-object tracking*), uma vertente da inteligência artificial que busca acompanhar simultaneamente várias entidades em movimento ao longo do tempo, como o estudo de Bergmann, Meinhardt e Leal-Taixe (2019) com o *Tracktor++*, que combina detecção com rastreamento baseado em *tracking by regression*. Outro elo importante está com a detecção de anomalias em vídeo, como ações suspeitas ou comportamentos fora do padrão, que utilizam muitos dos mesmos fundamentos técnicos aqui abordados.

Por fim, a proposta também tangencia áreas mais aplicadas como engenharia de *software*, especialmente no que diz respeito à criação de infraestrutura moderna de APIs e interfaces *web* funcionais, que são essenciais para tornar as descobertas da inteligência artificial utilizáveis no mundo real, fora do ambiente restrito de laboratórios e grupos de pesquisa.

Entre as contribuições técnicas mais relevantes deste projeto, destaca-se o desenvolvimento de uma API RESTful completa, capaz de receber como entrada uma imagem de referência de uma pessoa e um vídeo, retornando como saída os trechos do vídeo onde esse indivíduo foi identificado. Esta API, baseada em modelos de detecção e re-identificação modernos, será documentada e acessível a desenvolvedores, pesquisadores ou profissionais sem expertise em inteligência artificial.

Complementarmente, será desenvolvida uma interface *web* intuitiva, que permitirá aos usuários visualizar e interagir com o sistema sem a necessidade de lidar com código. Isso facilitará a adoção da tecnologia em contextos diversos, como pequenas empresas, órgãos públicos e universidades.

Além da entrega técnica, será produzido um relatório técnico detalhado, documentando todo o *pipeline* desenvolvido — desde a detecção de pessoas nos quadros do vídeo, passando pela extração de *features* e pelo processo de re-identificação, até a entrega dos resultados por meio da API. Essa documentação tem valor acadêmico e prático, podendo ser reutilizada por outros projetos ou pesquisas que desejem evoluir a solução proposta.

Por fim, ao construir o sistema com uma arquitetura modular e adaptável, pretende-se que ele possa ser estendido ou adaptado para outras finalidades no futuro, como rastreamento de objetos específicos, monitoramento ambiental ou controle de acesso não biométrico, fomentando o avanço da pesquisa aplicada em inteligência artificial de forma ética, transparente e acessível.

3 PROPOSTA

Este Trabalho de Conclusão de Curso propõe o desenvolvimento de uma ferramenta *end-to-end* de rastreamento de pessoas baseada em inteligência artificial, com o objetivo de possibilitar a detecção e identificação de indivíduos em vídeos transmitidos em tempo real ou por meio de arquivos enviados. A solução combina técnicas modernas de visão computacional, como detecção de regiões de interesse (ROIs), extração de características únicas (*embeddings*) e comparação vetorial, de forma a reconhecer um indivíduo específico a partir de uma imagem de referência fornecida pelo usuário.

A ferramenta será estruturada como uma API acessível, com conexão via protocolo *web-socket* e suporte a requisições RESTful, além de oferecer um *dashboard* interativo para visualização dos resultados. Por meio dessa interface, o usuário poderá carregar uma imagem de uma pessoa de interesse e conectar uma fonte de vídeo — seja por *upload* direto ou por *streaming*. A API será responsável por identificar se, onde e quando o indivíduo-alvo aparece nos vídeos processados, retornando as informações em formato estruturado (JSON) e visualizando os dados em tempo real por meio do painel *web*.

Abaixo, apresenta-se o pipeline técnico proposto para a solução:

1. Definição de Base de Dados para Treinamento e Avaliação: Inicialmente, será selecionada uma base de dados pública apropriada para tarefas de detecção e re-identificação de pessoas. A base escolhida será utilizada tanto para treinar o modelo de re-identificação quanto para realizar testes controlados de desempenho, garantindo que a ferramenta desenvolvida seja eficaz em cenários variados, com múltiplas câmeras e mudanças de aparência.

2. Detecção de Regiões de Interesse (ROIs): Será empregado um modelo de detecção de objetos baseado em aprendizado profundo, como o YOLOv8, treinado para localizar e gerar caixas delimitadoras (*bounding boxes*) ao redor das pessoas em cada *frame* do vídeo. Esses recortes serão posteriormente processados para extração de características, constituindo o ponto de partida do *pipeline* de re-identificação.

3. Extração de *Embeddings*: Para cada pessoa detectada, será aplicado um modelo leve de re-identificação (Re-ID), como o OSNet (ZHOU *et al.*, 2019), que transforma a imagem recortada da pessoa em um vetor numérico (*embedding*). Esse vetor representa as características visuais únicas do indivíduo, como textura, proporção corporal, cor da roupa e outros padrões discriminativos. Como o sistema não parte de um banco de dados pré-existente, será possível fazer upload de um *embedding* de referência previamente extraído a partir de uma imagem fornecida pelo usuário final, por exemplo, uma captura de tela, uma imagem frontal da câmera ou uma foto da pessoa a ser rastreada.

4. Comparação Vetorial: O usuário da aplicação (*client*), será responsável por fornecer a imagem de refência do indivíduo que se deseja rastrear. Essa imagem será processada para gerar o *embedding* de referência que será mantido em memória durante a análise do vídeo. A cada detecção de pessoa no vídeo, o sistema irá calcular a distância entre o vetor de referência (utilizando métricas como distância euclidiana) e os vetores extraídos das pessoas detectadas no vídeo. As menores distâncias indicarão maior similaridade, permitindo identificar recorrências do mesmo indivíduo ao longo do tempo e de diferentes cenas.

5. Entrega via API: A comunicação com o sistema será viabilizada por meio de uma API desenvolvida com o *framework* FastAPI. Serão implementados *endpoints* específicos para as seguintes funções:

    - /upload-hash: Recebimento da imagem de referência, que será convertida em *embedding*.
    - /video-stream: Conexão com transmissões ao vivo, via protocolo *websocket*.
    - /upload-video: Envio de arquivos de vídeo para processamento.
    - /find/:personId: Requisições de busca por aparições da pessoa de interesse no vídeo.

6. Interface Web: Será desenvolvida uma interface gráfica minimalista para testes e demonstrações, que permitirá aos usuários visualizar o vídeo processado com marcações das detecções (*bounding boxes*), *timestamps* de ocorrência e identificações confirmadas. Essa interface visa facilitar a interpretação dos resultados, tornando a solução mais acessível para usuários não técnicos.

Com esse projeto, espera-se oferecer uma aplicação funcional e personalizável, que usa técnicas modernas de aprendizado de máquina, processamento de vídeo e desenvolvimento *web* em uma ferramenta prática, segura e intuitiva. A proposta busca não apenas resolver um problema técnico, mas também democratizar o acesso a tecnologias avançadas de rastreamento por IA, viabilizando seu uso em diferentes contextos como segurança, automação e análise comportamental.

4 CONSIDERAÇÕES FINAIS

Conclui-se que este projeto busca entregar uma solução funcional, acessível e escalável para o rastreamento de pessoas por vídeo com o uso de inteligência artificial, por meio do desenvolvimento de uma API *end-to-end*. A proposta se fundamenta na construção de uma ponte entre a complexidade técnica do reconhecimento visual baseado em redes neurais e a necessidade de interfaces simplificadas que viabilizem o uso por profissionais não especialistas em IA.

Especificamente, pretende-se disponibilizar uma API bem documentada, com *endpoints* REST organizados, exemplos de requisição e retorno em formato estruturado (JSON), além de uma interface visual intuitiva, que permita o teste e a validação das funcionalidades sem exigir conhecimento avançado em programação ou aprendizado de máquina.

O projeto procura superar dificuldades recorrentes na adoção de tecnologias de rastreamento automatizado, como a ausência de ferramentas acessíveis, a falta de documentação clara em muitos *frameworks* existentes e a complexidade na integração entre modelos de detecção, rastreamento e re-identificação. Ao focar em modularidade, privacidade e facilidade de uso, o sistema proposto visa democratizar o acesso a essas tecnologias, viabilizando sua aplicação em contextos diversos — da segurança pública ao uso doméstico.

Além disso, espera-se que o protótipo desenvolvido desperte o interesse da comunidade acadêmica e técnica para novas aplicações e extensões do sistema, como a inclusão de múltiplos alvos, alertas em tempo real, integração com câmeras IP, e *dashboards* analíticos para monitoramento contínuo. Em um cenário onde a privacidade dos dados é cada vez mais relevante, a proposta também levanta discussões importantes sobre ética e segurança, reforçando a necessidade de ferramentas tecnológicas aliadas a boas práticas de proteção da informação.

REFERÊNCIAS

AHMAD, S. Object tracking with deepsort |. 02 2023.

AL-JABERY, K. K. *et al.* 3 - clustering algorithms. *In*: AL-JABERY, K. K. *et al.* (Ed.). **Computational Learning Approaches to Data Analytics in Biomedical Applications**. Academic Press, 2020. p. 29–100. ISBN 978-0-12-814482-4. Disponível em: https://www.sciencedirect.com/science/article/pii/B9780128144824000036.

ALMASAWA, M.; ELREFAEI, L.; MORIA, K. A survey on deep learning based person re-identification systems. **IEEE Access**, PP, p. 1–1, 12 2019.

Amazon Web Services. **O que é uma Rede Generativa Adversária (GAN)?** 2023. Acesso em: 30 abr. 2025. Disponível em: https://aws.amazon.com/pt/what-is/gan/.

BERGMANN, P.; MEINHARDT, T.; LEAL-TAIXE, L. Tracking without bells and whistles. *In*: **2019 IEEE/CVF International Conference on Computer Vision (ICCV)**. IEEE, 2019. p. 941–951. Disponível em: http://dx.doi.org/10.1109/ICCV.2019.00103.

CHEN, M. *et al.* Person re-identification by pose invariant deep metric learning with improved triplet loss. **IEEE Access**, PP, p. 1–1, 11 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, MA: MIT Press, 2016. ISBN 9780262035613. Disponível em: https://www.deeplearningbook.org.

JUNIOR, M. A. d. S.; MARTINI, J. S. C. **Utilização eficiente em larga escala de reconhecimento facial para análise preditiva de segurança em cidades inteligentes**. 2019. Dissertação (Mestrado) — Universidade de São Paulo, 2019.

KARMAKAR, A.; MISHRA, D. **Pose Invariant Person Re-Identification using Robust Pose-transformation GAN**. 2021. Disponível em: https://arxiv.org/abs/2105.00930.

LI, X. *et al.* Deep metric learning for few-shot image classification: A review of recent developments. **Pattern Recognition**, v. 138, p. 109381, 2023. ISSN 0031-3203. Disponível em: https://www.sciencedirect.com/science/article/pii/S0031320323000821.

MCLAUGHLIN, N.; RINCON, J. Martinez del; MILLER, P. Recurrent convolutional network for video-based person re-identification. *In*: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [*S.l.*: *s.n.*], 2016. p. 1325–1334.

MELO, P. V.; SERRA, P. Tecnologia de reconhecimento facial e segurança pública nas capitais brasileiras: Apontamentos e problematizações. **Comunicação e Sociedade**, n. 42, 2022. Postado online em 16 dez. 2022. Acesso em: 30 abr. 2025. Disponível em: http://journals.openedition.org/cs/8111.

RAMíREZ, S. **FastAPI: Modern, fast (high-performance), web framework for building APIs with Python 3.7+**. 2023. Acesso em: 30 abr. 2025. Disponível em: https://fastapi.tiangolo.com.

REDMON, J. *et al.* You only look once: Unified, real-time object detection. *In*: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [*S.l.*: *s.n.*], 2016.

SILVA, M. L. S. **As tecnologias de reconhecimento facial para Segurança Pública no Brasil: perspectivas regulatórias e a garantia de Direitos Fundamentais**. 2022. Monografia (Graduação em Direito) - Universidade Federal do Rio Grande do Norte, Natal, 87f.

STANDARDS, N. I. of; TECHNOLOGY. **Secure Hash Standard (SHS)**. [*S.l.*], 2015. Acesso em: 04 maio 2025. Disponível em: https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.180-4.pdf.

SUN, Y. *et al.* Learning part-based convolutional features for person re-identification. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 43, n. 3, p. 902–917, 2021.

WEI, Y. *et al.* Deep learning for retail product recognition: Challenges and techniques. **Computational Intelligence and Neuroscience**, v. 2020, p. 8875910, 2020.

WOJKE, N.; BEWLEY, A.; PAULUS, D. Simple online and realtime tracking with a deep association metric. *In*: **2017 IEEE International Conference on Image Processing (ICIP)**. IEEE Press, 2017. p. 3645–3649. Disponível em: https://doi.org/10.1109/ICIP.2017.8296962.

WU, J. *et al.* **Segment Anything Model is a Good Teacher for Local Feature Learning**. 2024. Disponível em: https://arxiv.org/abs/2309.16992.

YADAV, S. K. *et al.* Yognet: A two-stream network for realtime multiperson yoga action recognition and posture correction. **Knowledge-Based Systems**, v. 250, p. 109097, 2022. ISSN 0950-7051. Disponível em: https://www.sciencedirect.com/science/article/pii/S095070512200541X.

ZHOU, K. *et al.* Omni-scale feature learning for person re-identification. *In*: **Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)**. [*s.n.*], 2019. p. 3707–3716. Disponível em: https://arxiv.org/abs/1905.00953.