UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS UNIVERSITÁRIO DE TUCURUÍ
FACULDADE DE ENGENHARIA DE COMPUTAÇÃO

ELTON SILVA BARBOSA

PROTÓTIPO DE SISTEMA DE VIGILÂNCIA BASEADO EM IA PARA PREVENÇÃO DE CRIMES COM ARMAS E DISFARCES

TUCURUÍ
2025

ELTON SILVA BARBOSA

PROTÓTIPO DE SISTEMA DE VIGILÂNCIA BASEADO EM IA PARA PREVENÇÃO DE CRIMES COM ARMAS E DISFARCES

Trabalho de Curso apresentado à Faculdade de Engenharia de Computação, do Campus Universitário de Tucuruí, da Universidade Federal do Pará, como requisito parcial para obtenção do título de Bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Iago Lins de Medeiros

TUCURUÍ
2025

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

B238p Barbosa, Elton.
Protótipo de Sistema de Segurança Baseado em IA para Prevenção de Crimes com Armas e Disfarces / Elton Barbosa. — 2025. xv,36 f. : il. color.
Orientador(a): Prof. Dr. Iago Lins de Medeiros
Trabalho de Conclusão (Graduação) - Universidade Federal do Pará, Campus Universitário de Tucuruí, Faculdade de Engenharia da Computação, Tucuruí, 2025.

1. Visão computacional. 2. Inteligência artificial. 3. Detecção de objetos. 4. YOLO. 5. Vigilância inteligente. I. Título.

CDD 006.37

ELTON SILVA BARBOSA

PROTÓTIPO DE SISTEMA DE VIGILÂNCIA BASEADO EM IA PARA PREVENÇÃO DE CRIMES COM ARMAS E DISFARCES

Trabalho de Curso apresentado à Faculdade de Engenharia de Computação, do Campus Universitário de Tucuruí, da Universidade Federal do Pará, como requisito parcial para obtenção do título de Bacharel em Engenharia de Computação.

Data da aprovação: 15/09/2025

Conceito:

BANCA EXAMINADORA

Prof. Dr. Iago Lins de Medeiros
UFPA

Eng.(a) Tatianna Trindade Amador de Aviz
UFPA

Eng. Esp. Vigner Vieira dos Santos
UFPA

AGRADECIMENTOS

A realização deste trabalho é fruto de uma jornada repleta de desafios, aprendizados e crescimento pessoal e acadêmico. Por isso, é com gratidão que dedico estas palavras a todos que, direta ou indiretamente, contribuíram para a construção deste TCC e para a minha formação ao longo do curso.

Em primeiro lugar, agradeço a Deus, fonte inesgotável de força, sabedoria e esperança. Em momentos de incerteza e cansaço, foi na fé que encontrei motivação para seguir em frente.

Agradeço profundamente à minha família, base sólida de tudo que sou. Aos meus pais, por serem exemplos de integridade, esforço e amor incondicional. Seu apoio emocional, incentivo constante e compreensão diante das ausências e noites mal dormidas foram fundamentais para que eu pudesse chegar até aqui. Aos meus irmãos (ou outros familiares, se houver), pelo carinho, companheirismo e palavras de encorajamento nos momentos difíceis.

Ao meu orientador Iago registro meu sincero agradecimento pela dedicação, paciência e orientação ao longo deste trabalho. Sua expertise, atenção aos detalhes e incentivo à pesquisa foram determinantes para o desenvolvimento deste projeto. Agradeço não apenas pela contribuição acadêmica, mas também pelo apoio humano que me ajudou a superar os desafios enfrentados durante esta caminhada.

Aos professores e professoras do curso, minha gratidão por compartilharem seu conhecimento com seriedade e paixão pelo ensino. Cada disciplina, cada aula e cada conversa foram parte importante na construção do meu saber.

Aos colegas de curso, agradeço pelas trocas de ideias, pelo companheirismo nas dificuldades e pela amizade construída ao longo dos semestres. Aprendi muito com cada um de vocês, dentro e fora da sala de aula.

Aos amigos que estiveram ao meu lado, mesmo nos momentos de maior pressão, muito obrigado pela compreensão, pelo acolhimento e pela força nos momentos em que o cansaço parecia maior que a vontade. Vocês fizeram toda a diferença.

Não posso deixar de mencionar todas as instituições, bibliotecas, laboratórios, colaboradores e fontes de conhecimento que me permitiram pesquisar, estudar e desenvolver este projeto com seriedade e profundidade.

Encerro este agradecimento com o coração cheio de gratidão. Este trabalho é, sem dúvida, um marco importante na minha trajetória, e ele carrega um pedaço de cada pessoa que me apoiou nesta conquista.

Muito obrigado.

“Tecnologia não é apenas uma ferramenta. Ela pode dar forma ao mundo e protegê-lo.

Tim O’Reilly

RESUMO

Este Trabalho de Conclusão de Curso apresenta o desenvolvimento e a avaliação do protótipo de sistema inteligente de vigilância, com capacidade de detectar em tempo real a presença de armas de fogo e o uso de disfarces, como capacetes fechados e bonés, frequentemente utilizados em ações criminosas. O protótipo foi construído com base em técnicas de visão computacional e aprendizado profundo, empregando Redes Neurais Convolucionais (CNNs) com foco na arquitetura YOLO (You Only Look Once), em suas versões YOLOv8, YOLOv9 e YOLOv10. A metodologia consistiu na criação e organização de uma base de dados anotada, no treinamento dos modelos com técnicas de aumento de dados e no uso de métricas padronizadas como precision, recall e média de precisão média (mAP) para análise comparativa. Os resultados evidenciaram a evolução dos modelos nas versões mais recentes, com destaque para o YOLOv8, que apresentou o melhor desempenho geral em termos de acurácia e eficiência. Além da análise técnica, o trabalho discute os aspectos éticos e legais envolvidos na aplicação de sistemas automatizados de vigilância, como privacidade e viés algorítmico. A pesquisa contribui para o avanço de soluções tecnológicas voltadas à segurança, propondo uma ferramenta viável, precisa e de baixo custo, aplicável em diversos contextos urbanos e institucionais.

Palavras-chave: Vigilância inteligente, YOLO, visão computacional, redes neurais, detecção de armas, privacidade.

ABSTRACT

This Undergraduate Thesis presents the development and evaluation of a prototype intelligent surveillance system, capable of detecting in real time the presence of firearms and the use of disguises, such as full-face helmets and caps, commonly used in criminal activities. The prototype was implemented using computer vision and deep learning techniques, employing Convolutional Neural Networks (CNNs) based on the YOLO (You Only Look Once) architecture, in its YOLOv8, YOLOv9, and YOLOv10 versions. The methodology included the creation and organization of an annotated dataset, model training with data augmentation techniques, and performance evaluation using standardized metrics such as *precision*, *recall*, and mean Average Precision (*mAP*) for comparative analysis. The results demonstrated the evolution of the models across recent versions, highlighting YOLOv8, which achieved the best overall performance in terms of accuracy and computational efficiency. In addition to the technical analysis, the study discusses ethical and legal aspects related to the application of automated surveillance systems, such as privacy and algorithmic bias. This research contributes to advancing technological solutions for security, proposing a viable, accurate, and low-cost tool applicable in various urban and institutional contexts.

Keywords: Intelligent surveillance, YOLO, computer vision, neural networks, firearm detection, privacy.

LISTA DE SIGLAS

IA	Inteligência Artificial
CNNs	Redes Neurais Convolucionais
YOLO	You Only Look Once
mAP	Mean Average Precision
FPS	Frames por Segundo

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

A crescente sensação de insegurança nos centros urbanos, impulsionada pela incidência de crimes violentos, representa um sério desafio para a segurança pública e privada. Ações praticadas com uso de armas de fogo e disfarces, como capacetes fechados e máscaras, tornaram-se frequentes, dificultando a identificação de suspeitos e reduzindo a eficácia de sistemas de vigilância convencionais (CARVALHO; GUERRA, 2025), que são geralmente passivos e baseados em análise posterior das imagens. Diante desse cenário, torna-se evidente a necessidade de soluções tecnológicas mais eficientes, capazes de atuar preventivamente e com respostas em tempo real.

Com os avanços na área de visão computacional, tornou-se possível identificar comportamentos suspeitos e objetos perigosos em tempo real. A aplicação de soluções baseadas em inteligência artificial (IA), especialmente algoritmos de aprendizado profundo (deep learning), surge como uma alternativa promissora para o monitoramento inteligente.

A proposta deste trabalho é o desenvolvimento de um sistema de vigilância automatizado capaz de detectar, com base em dados visuais, a presença de armas e o uso de disfarces. O projeto baseia-se em Redes Neurais Convolucionais (CNNs), com foco na arquitetura

You Only Look Once (YOLO), que oferece alta precisão e velocidade, características fundamentais para aplicações em segurança. Ao desenvolver uma ferramenta de apoio estratégico, este projeto contribui para o avanço das pesquisas aplicadas em segurança inteligente, integrando ciência de dados e responsabilidade social.

Adicionalmente, o trabalho busca preencher uma lacuna ao oferecer uma proposta viável e de baixo custo que trata de forma aplicada o uso de modelos como YOLO em sistemas de prevenção criminal no contexto brasileiro. A pesquisa também considera os aspectos éticos e legais envolvidos, como o direito à privacidade e o uso responsável da IA. Assim, este estudo pretende oferecer uma resposta inovadora e prática a um dos problemas mais urgentes da sociedade contemporânea: a segurança urbana.

1.1 Objetivos

1.1.1 Objetivo Geral

Desenvolver um protótipo de sistema de monitoramento inteligente, baseado em técnicas de visão computacional, capaz de detectar em tempo real a presença de armas de fogo

e o uso de disfarces (como capacete e boné), com o intuito de contribuir para a prevenção de crimes.

1.1.2 Objetivos Específicos

- Investigar e selecionar modelos de (CNNs) adequados para a tarefa de detecção em tempo real, com ênfase na arquitetura YOLO.
- Utilizar um banco de dados público com imagens e vídeos com ocorrências de armas e disfarces comuns utilizados em assaltos.
- Realizar o treinamento e validação de modelos de detecção de objetos com base nos dados.
- Avaliar o desempenho dos modelos por meio de métricas como precision, recall, mAP (média de precisão média).
- Implementar um protótipo funcional que integre o modelo treinado a um ambiente de monitoramento simulado.

1.2 Metodologia de pesquisa

A presente pesquisa caracteriza-se como uma *pesquisa aplicada*, pois visa desenvolver uma solução prática voltada à resolução de um problema real: a prevenção de crimes por meio da detecção automática de ameaças visuais em ambientes monitorados. Além disso, trata-se de uma *pesquisa experimental*, uma vez que envolve testes com modelos computacionais, validação por métricas técnicas e observação do comportamento dos algoritmos frente a diferentes conjuntos de dados.

Do ponto de vista da abordagem, adota-se uma metodologia *quantitativa*, com foco na análise de desempenho dos modelos por meio de métricas objetivas, como precision, recall e frames por segundo. A pesquisa também possui caráter *descritivo*, pois busca documentar o funcionamento, a eficácia e os limites das ferramentas utilizadas, bem como apresentar o comportamento do sistema em cenários simulados de risco.

A fase prática do trabalho compreende o uso de modelos de visão computacional baseados em (CNNs), com ênfase no algoritmo (YOLO), amplamente reconhecido pela eficiência em detecção de objetos em tempo real. Os experimentos incluem o treinamento de modelos sobre um banco de dados contendo imagens e vídeos com armas de fogo, capacetes, máscaras e outros itens comumente usados em ações criminosas.

As etapas da metodologia envolvem:

- **Coleta e organização de dados visuais**, por meio de bases públicas, com anotações específicas para armas e disfarces.
- **Pré-processamento dos dados**, com redimensionamento de imagens, aumento de dados (data augmentation) e rotulação compatível com os modelos escolhidos.
- **Treinamento dos modelos YOLO**, variando hiper parâmetros e arquiteturas (como YOLOv8, YOLOv9 e YOLv10) para fins de comparação.
- **Avaliação de desempenho**, utilizando métricas técnicas para validar a eficácia dos modelos e determinar o mais adequado para uso prático.
- **Implementação de um protótipo funcional**, capaz de processar vídeo em tempo real e emitir alertas diante da detecção de objetos suspeitos.

Por fim, será realizada uma análise crítica dos resultados, considerando não apenas os números obtidos do treinamento, mas também os desafios enfrentados durante o teste do protótipo e as possíveis aplicações do sistema em cenários reais.

1.3 Estrutura do trabalho

Este trabalho está organizado em cinco capítulos, além dos elementos pré-textuais e pós-textuais, conforme descrito a seguir:

- **Capítulo 1 - Introdução**: Apresenta o contexto do problema, os objetivos da pesquisa, a justificativa para sua realização, a metodologia empregada e a estrutura geral do trabalho.
- **Capítulo 2 - Fundamentação Teórica**: Revisa a literatura e os conceitos essenciais relacionados à segurança eletrônica, (IA), visão computacional, (CNNs) e à arquitetura YOLO, além de abordar aspectos éticos e legais pertinentes ao uso dessas tecnologias.
- **Capítulo 3 - Trabalhos Relacionados:** Discussão de pesquisas e soluções existentes que se aproximam do tema estudado, destacando limitações e contribuições que embasam a proposta deste trabalho.
- **Capítulo 4 - Metodologia**: Descreve o tipo de pesquisa adotado, as ferramentas utilizadas, a construção do banco de dados, os procedimentos de treinamento e avaliação dos modelos, e a forma como o sistema foi implementado e testado.
- **Capítulo 5 - Resultados e Discussão**: Apresenta os resultados obtidos nos testes experimentais, analisa o desempenho dos modelos aplicados, discute os achados à luz da literatura aponta as limitações e contribuições do protótipo proposto.
- **Capítulo 6 - Conclusão**: Traz as considerações finais do estudo, destacando as contribuições do trabalho, suas aplicações práticas e sugestões para pesquisas futuras.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Segurança Pública E Tecnologia

A segurança pública constitui um dos pilares fundamentais para o bem-estar coletivo e para o desenvolvimento de qualquer sociedade. Compreende o conjunto de políticas, ações e instituições voltadas à manutenção da ordem, à proteção dos direitos e à prevenção da criminalidade. Nos últimos anos, diante do crescimento da violência e da complexificação das ameaças, esse campo tem enfrentado desafios cada vez mais sofisticados (SANTOS; LIMA, 2020).

Nesse cenário, as tecnologias da informação e comunicação surgem como instrumentos indispensáveis. Diversas pesquisas têm demonstrado que a adoção de soluções tecnológicas na gestão da segurança pública contribui para a ampliação da capacidade de monitoramento e para a maior eficiência das operações policiais (KUMAR; SINGH, 2022). A utilização de dados em tempo real, integrados a sistemas inteligentes de vigilância, possibilita não apenas uma resposta mais rápida das autoridades, mas também estratégias preventivas de maior alcance.

A incorporação de novas ferramentas busca, sobretudo, otimizar o uso dos recursos humanos e materiais disponíveis. Entre as inovações mais presentes, destacam-se o videomonitoramento, os drones para patrulhamento aéreo, a automação de processos e, em especial, os sistemas baseados em visão computacional**.** Esses últimos têm recebido atenção crescente na literatura, pois permitem identificar comportamentos suspeitos e detectar objetos de risco em tempo real, elevando o potencial de prevenção e intervenção (ZHAO; WANG, 2021; SHARMA et al., 2021).

Assim, percebe-se que a relação entre segurança pública e tecnologia transcende a simples modernização de equipamentos: trata-se de uma verdadeira transformação digital, na qual a Inteligência Artificial (IA) e a análise de grandes volumes de dados se consolidam como ferramentas estratégicas para enfrentar a criminalidade contemporânea.

2.2 Conceitos de Vigilância Inteligente

A vigilância inteligente refere-se à utilização de tecnologias avançadas, como a IA, a visão computacional e big data que é o conjunto de técnicas e tecnologias utilizadas para coletar, armazenar e analisar grandes volumes de dados que não podem ser processados por métodos tradicionais. Esses dados podem ser variados e de alta velocidade, permitindo identificar

padrões e gerar insights relevantes para a tomada de decisão (Mayer-Schönberger & Cukier, 2013). Os sistemas de monitoramento automatizado para melhorar a eficiência da vigilância e da segurança pública Lima & Agostinho (UTFPR, 2022). Ao contrário dos sistemas tradicionais de vigilância, que se limitam a registrar dados visuais ou audíveis de maneira passiva, os sistemas de vigilância inteligente incorporam algoritmos capazes de realizar análises em tempo real, identificar comportamentos suspeitos e até mesmo tomar decisões autônomas sobre quando acionar uma resposta de segurança. A Figura 1 ilustra o Centro de Operações e Inteligência de Taboão da Serra na aplicação de tecnologias avançadas de monitoramento em tempo real, integrando diversos sensores e câmeras em um único ambiente de supervisão.

Figura 1 – Centro de Operações e Inteligência de Taboão da Serra

Fonte: Tv Band Vale (2024).

Esses sistemas são projetados para automatizar e aprimorar o processo de monitoramento, utilizando recursos tecnológicos que permitem a detecção e interpretação de eventos, como a presença de objetos específicos (por exemplo, armas ou máscaras) ou comportamentos suspeitos, com um nível de precisão muito maior do que o proporcionado por observadores humanos.

2.2.1 Tecnologias e Vantagens da Vigilância Inteligente

As tecnologias de vigilância inteligente vêm ganhando espaço na segurança pública pela capacidade de processar grandes volumes de dados em tempo real. Um dos pilares desse avanço é a visão computacional, que permite que máquinas “enxerguem” e interpretem o ambiente visual. Dentro dela, destacam-se as Redes Neurais Convolucionais (CNNs), como os modelos

YOLO, Faster R-CNN e RetinaNet, que são amplamente utilizados para identificar e localizar objetos em tempo real com alta precisão.

Outro recurso essencial é a análise preditiva baseada em big data, que avalia grandes volumes de informações para antecipar possíveis ocorrências a partir de padrões históricos. A integração com a Internet das Coisas (IoT) também representa um avanço, permitindo que sensores e câmeras compartilhem dados e gerem alertas automáticos diante de anomalias.

A principal vantagem dessas tecnologias é a eficiência em tempo real. A automatização do monitoramento reduz a necessidade de observação contínua por operadores humanos e proporciona respostas mais ágeis em situações de risco. Sistemas inteligentes eliminam vulnerabilidades como distração e fadiga, reduzindo erros humanos e aumentando a confiabilidade na detecção de eventos críticos. Além disso, a IA possibilita a detecção de padrões complexos que seriam imperceptíveis ao olhar humano, como comportamentos sugestivos de preparação para atos ilícitos. Por fim, tais sistemas permitem a tomada de decisões autônomas, como o acionamento imediato de alarmes ou a notificação direta às autoridades competentes

A Figura 2 demonstra um exemplo de detecção automática realizada por um sistema de IA evidenciando a capacidade de identificar objetos de interesse e emitir alertas de forma imediata.

Figura 2 - Exemplo de Detecção utilizando IA

Fonte: Verzani & Sandrini (2024).

2.2.2 Visão Computacional

A visão computacional é uma subárea da Inteligência Artificial (IA) que busca capacitar máquinas a compreender e interpretar o mundo visual a partir de imagens ou vídeos, de maneira

semelhante à visão humana. No contexto de segurança, essa tecnologia permite automatizar a vigilância, identificar comportamentos suspeitos, reconhecer objetos ou pessoas e gerar respostas rápidas a situações potencialmente perigosas, sendo um dos pilares da vigilância inteligente e da segurança automatizada.

Para processar e analisar imagens e vídeos, a visão computacional utiliza algoritmos e modelos matemáticos, seguindo etapas fundamentais, como:

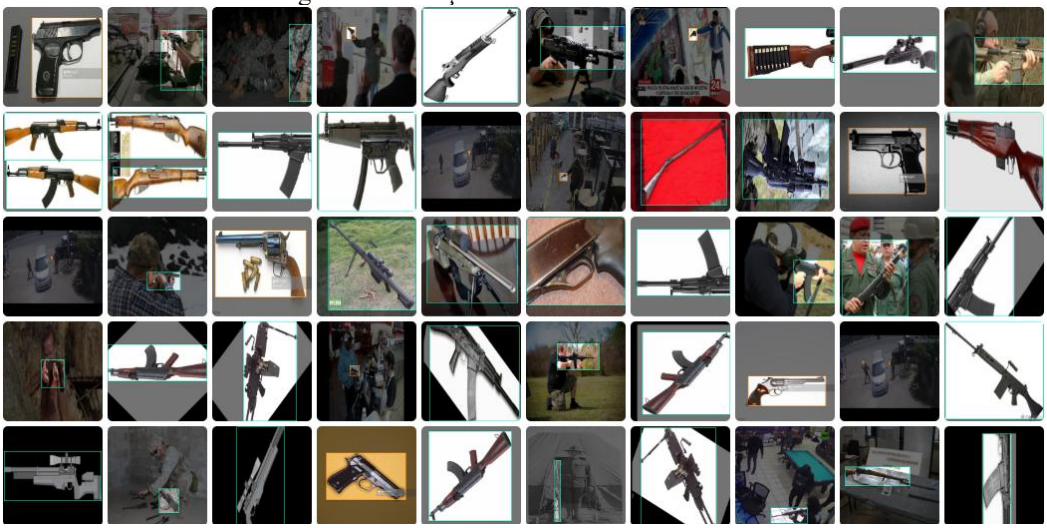
- **Aquisição de Imagens**: O processo de captura de imagens ou vídeos de câmeras de segurança, drones ou outros dispositivos. A aquisição de imagens, etapa inicial do processo de visão computacional, é exemplificada na Figura 3, que mostra a infraestrutura de um centro de monitoramento responsável por captar dados visuais em tempo real.

Figura 3 - Centro de monitoramento – Aquisição de Imagens

Fonte: Autor (2025).

- **Pré-processamento**: O tratamento das imagens para melhorar sua qualidade e destacar características relevantes, como a remoção de ruídos, ajuste de contraste ou transformação geométrica.
- **Segmentação de Imagens**: A divisão da imagem em partes ou regiões significativas, de modo que diferentes objetos ou características possam ser analisados separadamente.
- **Extração de Características**: A identificação de pontos ou regiões de interesse em uma imagem, como bordas, contornos ou texturas, que podem ser usados para distinguir diferentes objetos ou situações. Na Figura 4, observa-se o processo de extração de características, no qual pontos-chave das imagens são identificados para posterior análise pelo modelo de IA.

Figura 4 - Extração de Características



Fonte: Roboflow (2025).

Reconhecimento e Classificação: A identificação de objetos específicos, como pessoas, veículos, armas ou capacetes, por meio de técnicas como classificação de imagens ou detecção de objetos.

- **Pós-processamento**: A análise final dos dados extraídos para determinar o significado de um evento, como a detecção de um comportamento suspeito ou a identificação de uma ameaça.

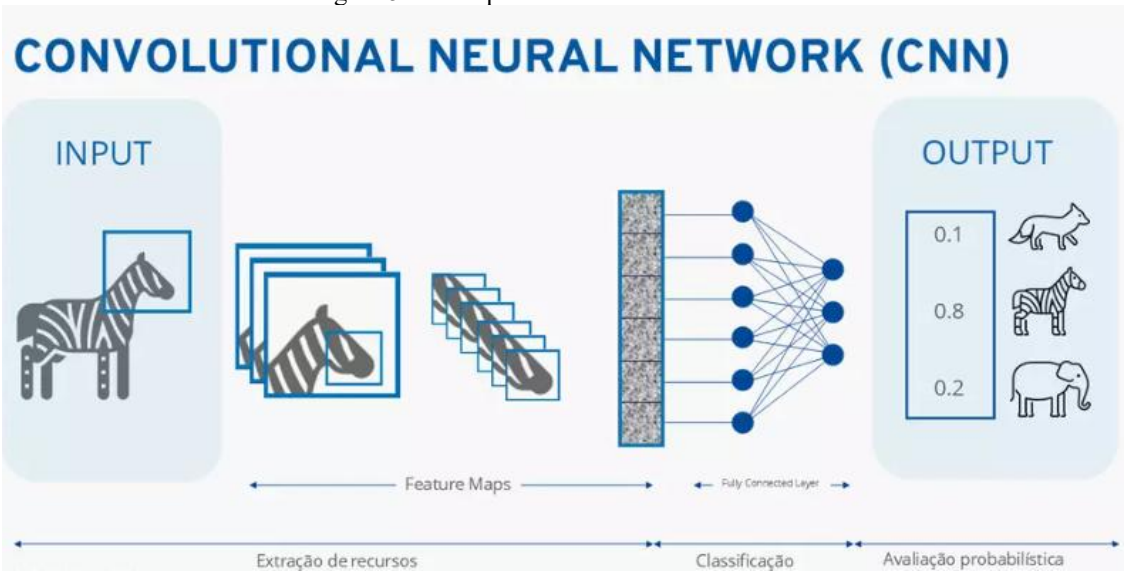
2.2.3 Redes Neurais Convolucionais (CNNs)

As CNNs são uma classe de redes neurais projetadas especificamente para processar dados em forma de matriz ou tensor, como imagens. A principal inovação das CNNs em relação às redes neurais tradicionais é a utilização de camadas convolucionais para detectar automaticamente características relevantes das imagens.

Arquitetura das CNNs

Uma arquitetura típica de CNN é composta por várias camadas, cada uma com um papel específico. A Figura 5 apresenta de forma esquemática como uma Rede Neural Convolucional processa as imagens, extraindo gradualmente padrões mais complexos a cada camada.

Figura 5 - Exemplo de funcionamento da CNN



Fonte: Ionos (2024).

1. **Camada Convolucional (Convolutional Layer)**: Esta é a camada principal das CNNs. Ela aplica **filtros** (ou kernels) a uma imagem para extrair características locais, como bordas, texturas e padrões. A convolução é um processo matemático que transforma a entrada da imagem em mapas de características.
2. **Camada de Pooling (Pooling Layer)**: Após a convolução, as CNNs frequentemente aplicam uma operação de pooling (como o **max pooling** ou **average pooling**) para reduzir a dimensionalidade dos dados e extrair as características mais importantes. Isso ajuda a reduzir o risco de overfitting e acelera o treinamento do modelo.
3. **Camada de Normalização (Normalization Layer)**: Camadas de normalização, como a **batch normalization**, ajudam a estabilizar e acelerar o treinamento ao normalizar as ativações de cada camada.
4. **Camada Totalmente Conectada (Fully Connected Layer)**: Nas camadas finais de uma CNN, as unidades neuronais são totalmente conectadas, permitindo que o modelo combine as características extraídas pelas camadas convolucionais e de pooling para fazer uma classificação ou previsão.
5. **Camada de Saída (Output Layer)**: A camada final da rede realiza a classificação ou detecção dos objetos. Por exemplo, no caso de um sistema de segurança, a saída pode indicar se há uma arma ou capacete presente em uma imagem, ou se a pessoa é um criminoso conhecido.

2.2.4 Como as (CNNs) Funcionam

Em resumo, quando uma imagem é processada pela CNN, as camadas convolucionais e de *pooling* extraem características progressivamente mais complexas, desde bordas e texturas até representações completas de objetos. Com o treinamento adequado em grandes volumes de dados, a rede aprende a identificar padrões relevantes e a generalizar esse conhecimento para detectar objetos em imagens nunca vistas anteriormente, tornando-se uma ferramenta poderosa para análise em tempo real.

2.3 Detecção de Objetos com YOLO

A detecção de objetos desempenha um papel fundamental em sistemas de monitoramento e vigilância inteligente, permitindo identificar comportamentos anômalos, objetos suspeitos e atividades que exigem atenção em tempo real. O (YOLO) é um dos algoritmos mais eficazes para detecção de objetos devido à sua velocidade e precisão. A evolução do YOLO ao longo das versões v8, v9 e v10 trouxe avanços significativos na arquitetura e na aplicação em contextos de segurança.

2.3.1 Arquitetura YOLO: Evolução e Versões

O YOLO passou por uma série de evoluções em sua arquitetura, mais recentes, como o YOLOv8, YOLOv9 e YOLOv10. A principal característica que diferencia essas versões é o aprimoramento contínuo da eficiência computacional, precisão e velocidade de inferência, além da capacidade de detectar objetos em tempo real, o que é crucial para o monitoramento de ambientes dinâmicos e com grande fluxo de informações, como áreas públicas e sistemas de segurança.

- YOLOv8: Esta versão trouxe avanços importantes em eficiência e precisão, com melhorias significativas na detecção de pequenos objetos e redução do tempo de inferência. O YOLOv8 introduziu aprimoramentos na estrutura de backbone e no uso de atenção espacial, além de técnicas avançadas de regularização para minimizar falsos positivos. Isso tornou o modelo mais adequado para cenários de segurança com alta variação de luz e densidade de objetos.
- YOLOv9: Focada em otimizar a detecção em tempo real, esta versão incluiu novos blocos de rede, como o C3K2 e o SPFF (Split Path Fast Fusion), que melhoram a eficiência na extração de características. O YOLOv9 também se beneficiou de algoritmos avançados de supressão de redundância e refinamento de bounding boxes, aumentando a precisão em ambientes com objetos móveis e complexos.

- YOLOv10: A principal inovação do YOLOv10 foi a introdução do mecanismo de detecção adaptativa com o uso de blocos CIB (Compact Inverted Blocks), que permitem ao modelo ajustar dinamicamente a granularidade dos detalhes extraídos. Esta versão também eliminou a necessidade de supressão não máxima (NMS) em alguns cenários, reduzindo a latência e melhorando a precisão em sistemas de vigilância urbana com alta densidade de objetos.

A evolução dessas versões reflete um contínuo esforço para tornar o YOLO cada vez mais adequado para ambientes de segurança pública, onde o tempo de resposta rápido, a precisão na detecção e a eficiência computacional são cruciais para a operação de sistemas de vigilância inteligente.

2.4 Técnicas para Prevenção de Crimes com Visão Computacional

2.4.1 Detecção de Armas de Fogo

A detecção de armas de fogo continua sendo uma das áreas mais críticas no uso de IA para prevenção de crimes. Armas de fogo são frequentemente usadas em crimes violentos como assaltos, tiroteios e furtos. O uso de IA na detecção de armas de fogo envolve principalmente análise de imagens e processamento de vídeo para identificar armas em ambientes públicos ou privados.

A IA pode ser aplicada em câmeras de segurança, drones ou outros dispositivos de vigilância equipados com sistemas de visão computacional e deep learning. Algoritmos como (YOLO) e SSD (Single Shot Multibox Detector) são eficientes para detectar pistolas, revólveres e fuzis em tempo real, permitindo que as forças de segurança intervenham rapidamente. Além disso, o reconhecimento acústico de disparos de arma de fogo, como o sistema ShotSpotter, utiliza IA para identificar e localizar a origem de disparos em grandes áreas urbanas.

2.4.2 Reconhecimento de Disfarces (capacete, boné e Pistola)

O uso de disfarces para ocultar a identidade ou intenções criminosas é uma tática comum em atividades ilícitas. A IA, particularmente por meio de visão computacional, tem se tornado essencial para a identificação de itens como capacetes, bonés e até mesmo pistolas escondidas por indivíduos em situações suspeitas.

Capacete: O uso de capacetes fechados, especialmente por motociclistas envolvidos em assaltos ou roubos, é uma prática comum. A inteligência artificial pode ser empregada para detectar a presença de capacetes em câmeras de vigilância. Modelos de deep learning, como YOLO ou RetinaNet, conseguem identificar capacetes de forma eficiente, mesmo em condições

adversas, como iluminação ruim ou ângulos de visão desafiadores (JIA et al., 2021; LI et al., 2024).

A IA pode analisar características específicas dos capacetes, como formato, cor e detalhes do visor, e determinar rapidamente se um indivíduo está usando um capacete de forma suspeita, sinalizando uma possível intenção criminosa. Além disso, algoritmos de detecção de objetos podem ser combinados com tecnologias de rastreamento para identificar a movimentação de indivíduos com capacetes em áreas públicas ou privadas, facilitando uma resposta rápida por parte das autoridades.

Bonés: O uso de bonés também é uma técnica comum de disfarce, frequentemente associada a roubos e assaltos, onde os criminosos tentam ocultar seu rosto. A IA, com base em modelos de reconhecimento de padrões e análise de imagens, pode identificar rapidamente capuzes, gorros ou outros tipos de vestimentas que cobrem a cabeça. Algoritmos de visão computacional podem ser treinados para reconhecer o padrão de um indivíduo com a cabeça coberta, alertando a segurança para a necessidade de uma verificação mais cuidadosa.

Pistola: O reconhecimento de pistolas em imagens e vídeos também se beneficia do uso de IA. Modelos de deep learning, como (CNNs), podem ser treinados para identificar as formas específicas de armas de fogo, mesmo quando parcialmente visíveis ou ocultas sob a roupa de um indivíduo. Quando integrados a sistemas de câmeras de segurança e câmeras portáteis (como as de agentes de segurança ou drones), esses modelos podem identificar pistolas com alta precisão e em tempo real.

Além disso, a detecção de armas também pode ser realizada em cenários dinâmicos, como interações de multidões, onde indivíduos podem esconder uma pistola sob roupas largas ou bolsas. Ao integrar algoritmos de detecção de objetos com rastreamento de movimento, os sistemas de IA podem identificar comportamentos suspeitos, como o manuseio ou posicionamento de uma pistola, e alertar automaticamente as autoridades.

2.4.3 Impacto na Prevenção de Crimes

A detecção de capacete, boné e pistola pode ter um impacto significativo na prevenção de crimes. A IA permite a identificação precoce de indivíduos com intenções suspeitas ou criminosas, facilitando a intervenção das forças de segurança antes que o crime aconteça. Além disso, a utilização de IA nos sistemas de vigilância inteligente melhora a eficiência e a precisão na detecção de ameaças, minimizando a margem de erro humana e fornecendo uma resposta rápida e coordenada.

Essas tecnologias, combinadas com sistemas de monitoramento em tempo real, podem criar um ambiente mais seguro, tanto em áreas públicas quanto em espaços privados, e são ferramentas essenciais na luta contra crimes urbanos e violência generalizada.

2.5 Aspectos éticos e legais do uso de IA na vigilância

O uso de (IA) em sistemas de vigilância oferece benefícios consideráveis, como a detecção eficiente de crimes, a análise de grandes volumes de dados em tempo real e a otimização da segurança pública. No entanto, a implementação dessas tecnologias também levanta questões éticas e legais que precisam ser cuidadosamente analisadas e regulamentadas. A preocupação central reside no equilíbrio entre a segurança pública e os direitos individuais, especialmente no que diz respeito à privacidade, liberdade e não discriminação.

2.5.1 Privacidade e Monitoramento

Um dos principais desafios éticos e legais do uso de IA em sistemas de vigilância é a privacidade das pessoas. O monitoramento contínuo e em tempo real de indivíduos por meio de câmeras, drones e outros dispositivos pode violar a privacidade de cidadãos que não estão envolvidos em atividades criminosas. Além disso, a coleta de dados pessoais sem consentimento prévio pode gerar riscos de uso indevido e de vazamento de informações sensíveis (SOUZA; PEREIRA, 2023).

A Lei Geral de Proteção de Dados (LGPD — Lei nº 13.709/2018) estabelece diretrizes específicas para o tratamento de dados pessoais no Brasil, incluindo a necessidade de consentimento, a limitação da coleta de dados ao mínimo necessário e a responsabilidade das empresas e órgãos públicos em garantir a segurança das informações. De forma semelhante, o Regulamento Geral de Proteção de Dados da União Europeia (GDPR) reforça o direito à privacidade e à autodeterminação informativa dos cidadãos (EUROPEAN COMMISSION, 2024).

2.5.2 Discriminação Algorítmica e Viés

Outro desafio crítico é o viés algorítmico, que pode resultar em decisões injustas e discriminatórias. Sistemas de IA são treinados com grandes volumes de dados históricos que, muitas vezes, contêm preconceitos implícitos. Isso pode levar à identificação incorreta de certos grupos sociais, principalmente minorias raciais e étnicas, o que representa um risco significativo em aplicações de segurança pública (PINTO; ARAÚJO, 2024).

Pesquisas recentes apontam que sistemas de reconhecimento facial e vigilância inteligente tendem a apresentar maiores taxas de erro para indivíduos de pele mais escura e para mulheres, reforçando desigualdades estruturais (GARCEZ et al., 2025). A mitigação desses vieses exige curadoria criteriosa dos dados de treinamento e auditorias frequentes nos modelos utilizados.

2.5.3 Regulação e Legislação

A ausência de uma legislação específica para o uso de IA na vigilância ainda é um problema em muitos países. No Brasil, a LGPD fornece diretrizes gerais sobre proteção de dados, mas não aborda de forma detalhada as implicações do uso de IA para segurança pública. Projetos de lei em tramitação (BRASIL, 2024) buscam preencher essa lacuna, estabelecendo princípios de transparência, responsabilidade e supervisão humana obrigatória em sistemas automatizados.

Internacionalmente, a União Europeia tem avançado com o *AI Act*, aprovado em 2024, que classifica sistemas de IA de vigilância como de “alto risco” e exige avaliações de impacto ético e social antes de sua implantação (EUROPEAN PARLIAMENT, 2024).

2.5.4 Impacto Social e Psicossocial

Além das questões legais e éticas, o uso de inteligência artificial (IA) na vigilância pode ter um impacto social e psicossocial significativo. A sensação de estar constantemente sendo monitorado pode alterar o comportamento das pessoas, levando à autocensura ou até a alienação social. Esse fenômeno, conhecido como o efeito Panóptico, pode afetar a liberdade de expressão e o direito à privacidade de forma negativa.

O impacto psicológico desse tipo de vigilância é particularmente preocupante, pois a constante sensação de observação pode criar um ambiente de medo ou ansiedade entre os cidadãos, fazendo com que se sintam vigiados em todas as suas atividades diárias. Essa percepção contínua de monitoramento pode afetar negativamente o bem-estar psicológico e a liberdade pessoal dos indivíduos, sobretudo quando não há um equilíbrio adequado entre segurança e direitos civis (SOUZA; ALBUQUERQUE, 2024; RAMIRO; RAMIRO; TAMAOKI, 2024).

2.5.5 Responsabilidade dos Desenvolvedores e Usuários

A responsabilidade pelo uso ético da IA deve ser compartilhada entre desenvolvedores, instituições públicas e empresas privadas. É essencial que os sistemas sejam projetados com base em princípios de transparência, auditabilidade e prestação de contas. A implementação de comitês de ética e a realização de testes independentes antes da adoção em larga escala também são práticas recomendadas (OECD, 2024).

2.6 Definição de Termos Técnicos

A seguir são apresentadas definições resumidas dos principais termos técnicos utilizados neste trabalho, com o objetivo de facilitar a compreensão dos conceitos abordados:

- **Precisão (*precision*)** – Mede a proporção de previsões positivas corretas em relação ao total de previsões positivas realizadas pelo modelo. Indica o quão confiável é uma detecção quando ela ocorre.
- **Revocação (*recall*)** – Mede a proporção de previsões positivas corretas em relação ao total de instâncias positivas existentes. Avalia a capacidade do modelo de encontrar todos os elementos relevantes.
- **mAP (*mean Average Precision*)** – Métrica que representa a média da precisão obtida em diferentes níveis de *recall*. É amplamente utilizada para avaliar o desempenho geral de modelos de detecção de objetos.
- **FPS (Frames por Segundo)** – Quantidade de quadros processados por segundo pelo modelo. É uma medida de desempenho que indica a velocidade de processamento da rede.
- **Tempo de treinamento** – Intervalo total necessário para que o modelo complete seu processo de aprendizado com o conjunto de dados fornecido, desde a inicialização até o término da última época de treinamento.
- **Data augmentation** – Conjunto de técnicas utilizadas para aumentar artificialmente a diversidade do conjunto de dados de treinamento, aplicando transformações como rotações, espelhamentos, recortes e alterações de brilho, a fim de melhorar a capacidade de generalização do modelo.

3 TRABALHOS RELACIONADOS

A literatura e o mercado de segurança eletrônica têm avançado significativamente no desenvolvimento de sistemas de vigilância inteligentes capazes de detectar armas de fogo e comportamentos suspeitos em tempo real. Esses esforços combinam técnicas de visão computacional, Redes Neurais Convolucionais (CNNs) e modelos avançados de detecção de objetos, como a família You Only Look Once (YOLO), destacando a relevância e aplicabilidade do tema do presente trabalho. A seguir, são discutidos alguns dos principais trabalhos relacionados na academia, na indústria e em iniciativas internacionais de segurança pública.

3.1 Pesquisas Acadêmicas em Detecção de Armas

Diversos estudos científicos têm explorado o uso de algoritmos de deep learning para detecção automática de armas em vídeos de vigilância. Zhao e Wang (2021), por exemplo, implementaram o algoritmo YOLOv4 para identificação de armas de fogo em tempo real em ambientes monitorados por CFTV. Os autores demonstraram que o modelo atingiu alta acurácia e baixa taxa de falsos positivos, evidenciando o potencial dessas arquiteturas para uso em sistemas de segurança pública.

Outro trabalho relevante é o de Kumar e Singh (2022), que propuseram uma arquitetura integrada para cidades inteligentes, capaz de identificar não apenas armas, mas também comportamentos suspeitos em grandes multidões. Essa abordagem mostrou que a combinação de aprendizado profundo e análise de comportamento pode ampliar a eficácia das plataformas de monitoramento.

Esses resultados reforçam que a pesquisa acadêmica tem priorizado soluções em tempo real, explorando arquiteturas rápidas e eficientes, que podem ser facilmente adaptadas a diferentes contextos de vigilância.

3.2 Aplicações Comerciais na Indústria

No setor privado, empresas especializadas em Inteligência Artificial (IA) já disponibilizam soluções comerciais que incorporam tecnologias semelhantes às pesquisadas pela academia. Um exemplo notável é a Athena Security, empresa norte-americana que desenvolveu um sistema de detecção automática de armas de fogo integrado a câmeras de vigilância convencionais. A tecnologia é capaz de emitir alertas imediatos a autoridades locais, permitindo respostas rápidas em casos de risco iminente.

Esse tipo de iniciativa demonstra que as soluções baseadas em deep learning não se restringem a protótipos acadêmicos, mas já são aplicadas em cenários reais, como escolas,

aeroportos e espaços públicos. A existência dessas aplicações comerciais comprova a relevância prática do tema e abre espaço para melhorias, como redução de custos e maior acessibilidade em países em desenvolvimento.

3.3 Iniciativas Internacionais e Políticas Públicas

Além da academia e da indústria, diversos países têm investido em projetos de segurança pública que utilizam sistemas inteligentes de vigilância. Em particular, a Índia tem se destacado na implementação de plataformas de monitoramento em larga escala, combinando algoritmos de deep learning com redes de câmeras urbanas. O trabalho de Kumar e Singh (2022), já citado, reflete esse movimento, alinhando pesquisa acadêmica às necessidades de segurança pública.

Outros países, como Estados Unidos e Reino Unido, também têm discutido o uso ético e legal dessas tecnologias em ambientes públicos, buscando equilibrar eficiência de segurança com garantias de privacidade. Esses debates mostram que o desenvolvimento de sistemas de detecção de armas não deve se restringir ao aspecto tecnológico, mas considerar também implicações sociais, legais e culturais.

3.4 Considerações Finais

Os trabalhos relacionados demonstram que a detecção inteligente de armas é um campo em crescimento, com avanços relevantes na academia, aplicações práticas no setor privado e iniciativas em políticas públicas internacionais. A convergência dessas três dimensões reforça a importância do presente estudo, que busca contribuir para o desenvolvimento de soluções mais eficazes e adaptáveis ao contexto brasileiro, ampliando a segurança em espaços públicos e privados.

4 METODOLOGIA

A metodologia deste trabalho segue uma abordagem de *pesquisa aplicada e experimental*, com o intuito de desenvolver e validar uma solução prática para um problema real. O processo envolve testes com modelos computacionais, validação por métricas técnicas e observação do comportamento dos algoritmos, caracterizando-se como uma abordagem *quantitativa e descritiva*. O objetivo é documentar o funcionamento, a eficácia e os limites das ferramentas utilizadas em cenários simulados de risco.

4.1 Ferramentas e Tecnologias Utilizadas

Para a implementação e análise do protótipo de monitoramento proposto, serão utilizadas diversas ferramentas tecnológicas avançadas. Abaixo estão as principais ferramentas e tecnologias que serão aplicadas na pesquisa.

4.1.1 Modelo de Visão Computacional - YOLO (You Only Look Once)

A tecnologia principal utilizada será o (YOLO) uma rede neural convolucional (CNN) usada para detecção de objetos em tempo real. O YOLO foi escolhido devido à sua alta eficiência e precisão na detecção de objetos em imagens e vídeos, além de seu desempenho em tempo real, o que é crucial para sistemas de monitoramento de segurança.

Serão testadas diferentes versões do YOLO (como o YOLOv8, YOLOv9 e YOLOv10) para comparar suas performances em cenários de detecção de armas de fogo, capacetes, bonés e disfarces. Cada versão será avaliada quanto à sua precision, recall, tempo de resposta e robustez ao lidar com diferentes condições de luz e ângulos de visão.

4.1.2 Ferramentas de Processamento de Imagens

Ferramentas de processamento de imagens serão utilizadas para preparar os dados de entrada, como filtros de aprimoramento para aumentar a qualidade das imagens, remoção de ruídos e normalização de dados. Além disso, serão aplicadas técnicas de aumento de dados (data augmentation) para gerar mais variações de imagens e treinar os modelos de IA em um conjunto de dados mais robusto.

Algumas das ferramentas específicas que serão utilizadas incluem:

- OpenCV: Biblioteca amplamente utilizada para processamento de imagens e vídeos em Python, que facilita a manipulação de imagens para realizar tarefas como detecção de bordas, segmentação de objetos e rastreamento de movimento.
- TensorFlow / PyTorch: Frameworks de deep learning usados para treinar e testar modelos de IA, como YOLO, com Redes Neurais Convolucionais (CNNs).

4.1.3 Métricas de Avaliação de Desempenho

Após o treinamento, será utilizado um sistema de avaliação para medir a eficiência dos modelos de IA. A avaliação será baseada em métricas como:

- Precision: A proporção de verdadeiros positivos em relação ao total de positivos previstos.
- Recall: A proporção de verdadeiros positivos em relação ao total de positivos reais.
- F1-Score: A média harmônica entre precision e recall, que fornece uma medida equilibrada da performance.

4.1.4 Infraestrutura de Teste

A pesquisa será conduzida utilizando uma infraestrutura composta por servidores de processamento do Google Colab, onde ocorrerá o treinamento dos modelos de Inteligência Artificial (IA), e por testes em vídeos executados no próprio computador, que permitirão a avaliação da aplicação em condições simuladas. Além disso, a implementação do protótipo do sistema incluirá cenas com objetos reais, com o objetivo de testar e validar a capacidade de detectar os objetos.

4.2 Base de Dados

4.2.1 Fontes de Imagens e Vídeos

A base de dados utilizada nesta pesquisa foi obtida através da plataforma Roboflow, conhecida por fornecer datasets prontos e compatíveis com frameworks de detecção de objetos como o YOLO. O conjunto de dados é composto por 7.593 imagens rotuladas, organizadas para atender às classes de interesse deste trabalho: pistola, casco (capacete fechado) e boné.

A distribuição por classes é a seguinte:

- **Pistola:** 4.239 instâncias
- **Casco (capacete fechado):** 1.449 instâncias
- **Boné:** 1.193 instâncias

As imagens foram coletadas de diferentes fontes visuais, incluindo câmeras de vigilância simuladas, fotos públicas de bancos de imagens, e registros de situações de monitoramento urbano. Todas as imagens foram previamente anotadas com bounding boxes, facilitando o processo de treinamento dos modelos de detecção.

4.2.2 Pré-processamento de Dados

O processo de preparação dos dados envolveu o redimensionamento das imagens para tamanhos padronizados (como 640×640 pixels), conforme exigido pela arquitetura YOLO, além da conversão de formatos, passando de JPEG ou PNG para arquivos compatíveis com a anotação YOLO, no formato TXT. Essas etapas garantiram que as imagens estivessem devidamente padronizadas, otimizando tanto a eficiência quanto a acurácia dos modelos no momento do treinamento.

4.3 Treinamento do Modelo

4.3.1 Configuração do Modelo YOLO

O modelo utilizado foi o (YOLO) em suas versões mais recentes: YOLOv8, YOLOv9 e YOLOv10, com arquitetura adaptada para tarefas de detecção em tempo real. A configuração foi realizada utilizando o ambiente PyTorch, com as seguintes definições reunidas na Tabela1. Cada versão foi treinada de forma independente, utilizando os mesmos parâmetros para fins de comparação.

Tabela 1 - Configurações de Treinamento dos Modelos YOLO

Parâmetro	Valor
Tamanho de imagem	640×640
Batch size	16
Número de épocas	100
Taxa de aprendizado inicial	0.001
Ancoragem adaptativa	habilitada
Transferência de aprendizado	os modelos foram iniciados a partir de pesos pré-treinados

Fonte: Autor (2025).

4.3.2 Técnicas de Aumento de Dados (Data Augmentation)

Para enriquecer o conjunto de treinamento e evitar overfitting, aplicaram-se técnicas de Data Augmentation. O Overfitting (sobre ajuste) ocorre quando um modelo de aprendizado de máquina se ajusta excessivamente aos dados de treinamento, aprendendo padrões e ruídos específicos desses dados em vez de generalizar para novos exemplos. Isso faz com que o modelo apresente alta precisão no conjunto de treinamento, mas baixo desempenho em dados inéditos, indicando baixa capacidade de generalização (ALPAYDIN, 2020; GOODFELLOW et al., 2016).

Para reduzir esse problema, foram aplicadas técnicas de *Data Augmentation*, que aumentam artificialmente a diversidade do conjunto de treinamento por meio de transformações como rotações, espelhamentos, alterações de brilho e recortes aleatórios, contribuindo para que o modelo aprenda padrões mais robustos e generalizáveis como mostra a Tabela 2.

Tabela 2 - Técnicas de Data Augmentation e Quantidade de Imagens Geradas

Data Augmentation	Estimativa de novas imagens (%)	Quantidade aproximada gerada
Giro e espelhamento (horizontal/vertical)	+40%	2519
Ajuste de brilho, contraste e saturação	+20%	1259
Ruídos aleatórios (Gaussian Noise)	+15%	944
Corte aleatório (random crop)	+15%	944
Zoom e rotação leve (affine transform)	+10%	629
Total de imagens geradas (apenas train)	+100%	6295

Fonte: Autor (2025).

A Tabela 3 apresenta a consolidação desse processo, evidenciando a distribuição final do dataset após a aplicação do *augmentation*. Observa-se que a quantidade de imagens em cada subconjunto (treinamento, validação e teste) foi proporcionalmente ampliada, passando de um total de 7.593 imagens originais para aproximadamente 13.888 imagens no dataset definitivo.

Tabela 3 - Distribuição Final do Dataset Após Augmentation

Pasta	Originais	Geradas (augmentation)	Total final
train/images	6295	6295	12590
valid/images	1147	0	1147
test/images	151	0	151
Total	7593	6295	13888

Fonte: Autor (2025).

4.4 Avaliação de Desempenho

4.4.1 Métricas Utilizadas

A avaliação dos modelos treinados foi realizada por meio de métricas padrões em tarefas de detecção de objetos: Precision, Recall, mAP@0.5 e mAP@0.5:0.95, FPS, F1-score e Confusion Matrix (Matriz de Confusão) para análise detalhada das classes mais confundidas. Essas métricas permitiram verificar tanto o desempenho em identificar corretamente os objetos quanto a agilidade do sistema em contexto de vigilância ao vivo.

4.4.2 Comparação entre Versões de Modelos

Foi realizada uma comparação entre as versões YOLOv8, YOLOv9 e YOLOv10, com base nos resultados das métricas citadas. A análise comparativa permitiu identificar a versão mais eficiente para o contexto de segurança pública e monitoramento.

Cada versão foi avaliada quanto à sua velocidade de detecção, capacidade de generalização e resistência a ruídos visuais (como baixa iluminação, ângulos desfavoráveis ou presença de múltiplos objetos). A escolha do modelo ideal levou em conta tanto o desempenho técnico quanto a viabilidade de implantação em ambientes reais de vigilância.

4.4.3 Pesquisa Descritiva e Aplicada

A pesquisa descritiva e aplicada será utilizada para descrever as características do sistema proposto e para aplicar as soluções em um cenário de vigilância simulado. A pesquisa descritiva envolve a coleta e análise de dados para descrever fenômenos ou situações específicas, enquanto a pesquisa aplicada tem como foco a implementação de soluções práticas para resolver problemas concretos.

Nesta pesquisa, o desenvolvimento do sistema de monitoramento será descrito detalhadamente, incluindo as tecnologias de visão computacional utilizadas, os algoritmos de detecção de objetos e reconhecimento de comportamentos e os resultados práticos da implementação dessas tecnologias em testes realizados no computador pessoal. O estudo

descreverá como as soluções propostas podem ser aplicadas em situações de vigilância simulada e seu potencial impacto na eficácia da segurança.

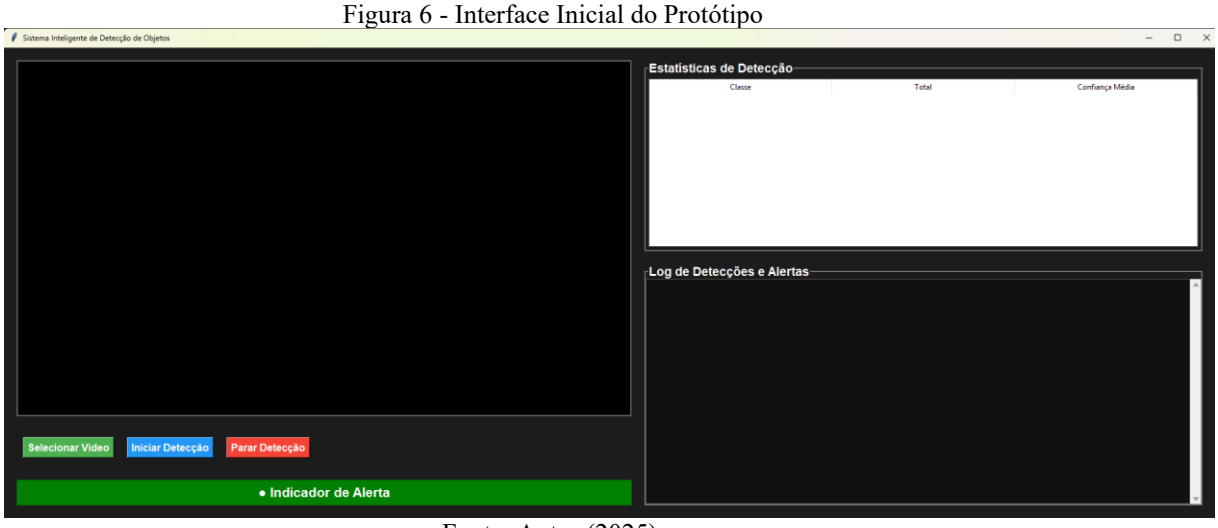
Além disso, a pesquisa será desenvolvida em etapas que contemplam desde a fundamentação teórica até a validação prática do sistema. Inicialmente, serão revisados trabalhos relacionados e tecnologias de detecção aplicadas à segurança pública. Em seguida, os modelos de visão computacional serão treinados e testados em ambientes controlados, com posterior aplicação em vídeos e cenários simulados no computador pessoal.

Essa abordagem permitirá analisar não apenas o desempenho técnico dos algoritmos, mas também sua viabilidade prática, considerando fatores como tempo de resposta, taxa de acertos e limitações observadas durante a execução. Espera-se que os resultados contribuam para a consolidação de um sistema de vigilância inteligente capaz de ampliar a eficiência da segurança, servindo como base para futuras pesquisas e para o desenvolvimento de soluções escaláveis em diferentes contextos sociais e institucionais.

4.5 Desenvolvimento do Protótipo

O protótipo desenvolvido consistiu em um sistema inteligente de detecção de objetos utilizando visão computacional baseada no modelo YOLOv8, com interface gráfica implementada em Tkinter. A Figura 6 mostra a interface inicial do protótipo desenvolvido, onde é possível carregar vídeos para análise, visualizar as detecções em tempo real e acompanhar estatísticas de confiança.

Figura 6 - Interface Inicial do Protótipo



Fonte: Autor (2025).

O sistema permite a análise de vídeos pré-gravados, destacando objetos de interesse, como bonés e capacetes, através de caixas delimitadoras e legendas, ao mesmo tempo em que registra logs e estatísticas de confiança média por classe. Além disso, o protótipo envia alertas automáticos via WhatsApp com o frame correspondente à detecção, permitindo resposta rápida a eventos críticos.

4.5.1 Funcionamento e Aplicação

O protótipo processa cada frame do vídeo em tempo quase real, realizando a detecção de objetos com confiança acima de um limiar previamente definido. Quando um objeto é identificado, o sistema atualiza as estatísticas, destaca visualmente o objeto e registra o evento nos logs. A aplicação em vídeos simulados demonstrou que a aplicação consegue identificar objetos relevantes mesmo em cenários com movimento e variações de iluminação, proporcionando alertas quase instantâneos e permitindo análise contínua por meio da interface.

5 RESULTADOS E DISCUSSÃO

5.1 Visão Geral dos Resultados

Para avaliar o desempenho dos modelos na detecção de pistolas, capacetes e bonés, foram adotadas métricas como *precision, recall, mAP@0.5, mAP@0.5:0.95*, perdas (*losses*) e tempo de treinamento. A Tabela 4 apresenta a comparação dos modelos.

Tabela 4 - Comparativo de desempenho entre YOLOv8, YOLOv9 e YOLOv10

Modelo	Precisio n	Recall	mAP@0. 5	mAP@0.5: 0.95	Box Loss	Cls Loss	DFL Loss	Tempo (s)
YOLOv8	0.93392	0.88151	0.92415	0.60830	1.38658	0.72794	1.32325	3634.66
YOLOv9	0.93862	0.86687	0.91235	0.60437	1.36098	0.72622	1.42129	10182.7 0
YOLOv10	0.90920	0.82460	0.88403	0.57154	2.92982	1.62651	2.67895	4168.52

Fonte: Autor (2025).

A avaliação comparativa teve como objetivo identificar qual versão da arquitetura YOLO apresenta o melhor equilíbrio entre desempenho e eficiência para a detecção de objetos críticos em sistemas de vigilância. Conforme os dados da Tabela 4, o YOLOv8 se destacou por apresentar um balanço superior entre acurácia, recall e velocidade, sendo o mais rápido para treinar. O YOLOv9, apesar da alta precisão, demonstrou um custo computacional muito elevado, enquanto o YOLOv10 apresentou um desempenho geral inferior às outras versões. Uma análise mais detalhada desses resultados é apresentada nas seções a seguir.

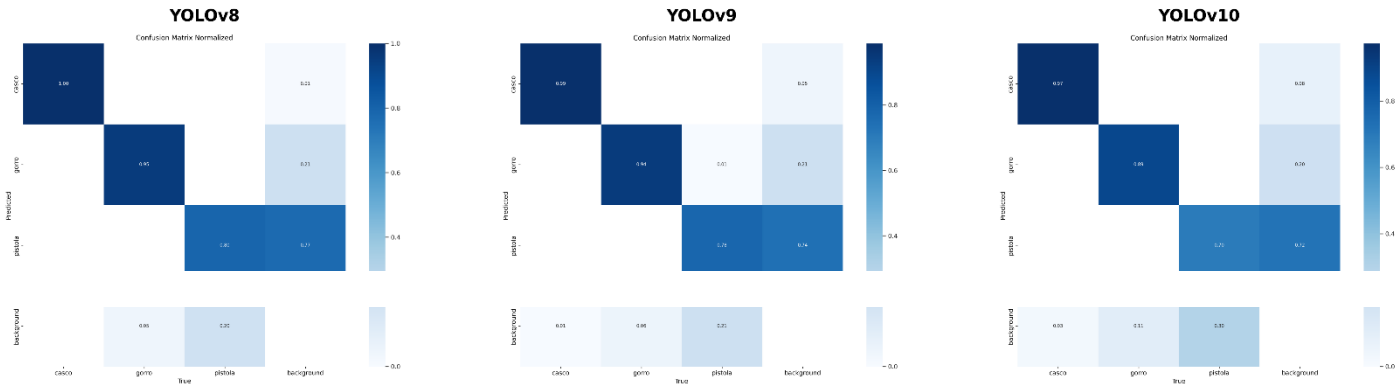
5.2 Análise Aprofundada do Modelo

A análise conjunta dos modelos YOLOv8, YOLOv9 e YOLOv10 permite compreender não apenas seus desempenhos individuais, mas também a evolução entre versões e as implicações práticas para sistemas de monitoramento em segurança pública. Foram avaliadas métricas como *precision, recall, F1-score, mAP@0.5, matrizes de confusão, curvas precisão-recall e recall-confiança*, que possibilitam uma visão mais ampla do comportamento dos modelos.

5.2.1 Desempenho geral das classes

As *matrizes de confusão normalizadas* revelam padrões consistentes entre os três modelos. A classe *capacete* foi a mais facilmente identificada, apresentando taxas de acerto acima de 95% em todas as versões. Já a classe *boné* mostrou confusão recorrente com capacete, principalmente no YOLOv8 e YOLOv9. A classe mais crítica foi a *pistola*, que apresentou baixos índices de recall nos modelos YOLOv8 e YOLOv9, com leve melhora no YOLOv10, mas ainda insuficiente para garantir confiabilidade em contextos de risco. A Figura 7 exibe as matrizes de confusão normalizadas que revelam padrões consistentes entre os três modelos.

Figura 7 - Matrizes de confusão comparativas dos modelos YOLOv8, YOLOv9 e YOLOv10.

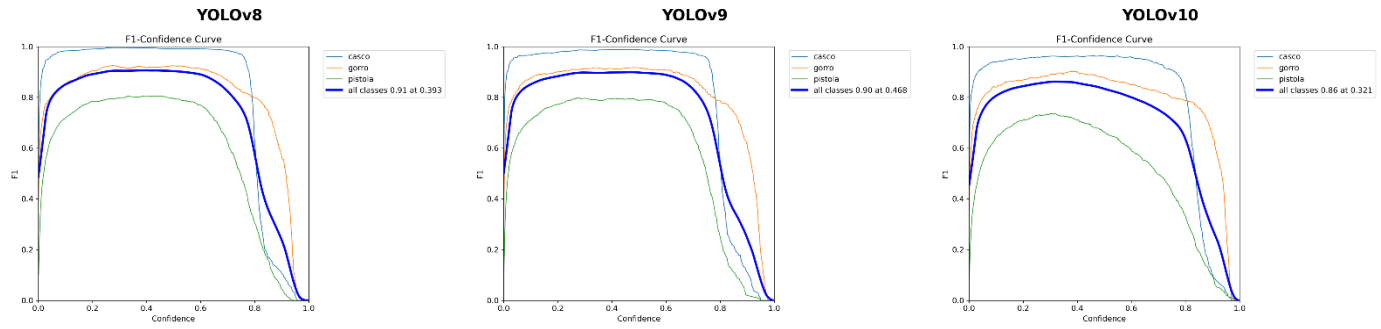


Fonte: Autor (2025).

5.2.2 Comparação das métricas quantitativas

A Figura 8 exibe as curvas *F1-Confidence*, onde o YOLOv8 apresentou o melhor equilíbrio entre precision e recall, alcançando mAP@0.5 de 0.924. O YOLOv9 destacou-se pela maior precisão absoluta (0.938), mas com menor recall, deixando passar muitas ocorrências reais. O YOLOv10 trouxe avanços estruturais, porém manteve dificuldades relevantes na classe pistola, com valor de AP abaixo de 0.75.

Figura 8 - Curvas F1-Confidence dos modelos YOLOv8, YOLOv9 e YOLOv10.



Fonte: Autor (2025).

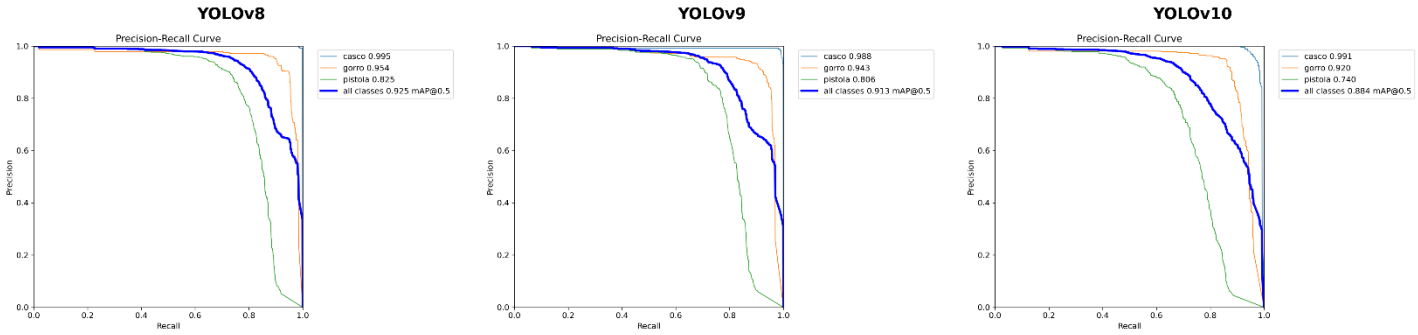
As *curvas precision-recall* reforçam esse cenário: o YOLOv8 apresentou desempenho consistente em todas as classes, o YOLOv9 manteve ótimo resultado para capacetes, mas falhou

em armas, e o YOLOv10 obteve desempenho intermediário, com queda acentuada na classe pistola.

5.2.3 Impacto do desbalanceamento dos dados

A *distribuição de classes e bounding boxes* evidencia o desbalanceamento: pistolas aparecem em menor número em relação a bonés e capacetes. Esse fator impactou diretamente no aprendizado, reduzindo recall e F1-score para armas. Esse problema é amplamente discutido na literatura e reforça a necessidade de estratégias como data augmentation direcionada e oversampling para mitigar o viés do dataset. A Figura 9 mostra as curvas precision-recall que reforçam o cenário do Yolov8 com o melhor desempenho.

Figura 9 - Curvas Precision-Recall comparativas dos modelos YOLOv8, YOLOv9 e YOLOv10.

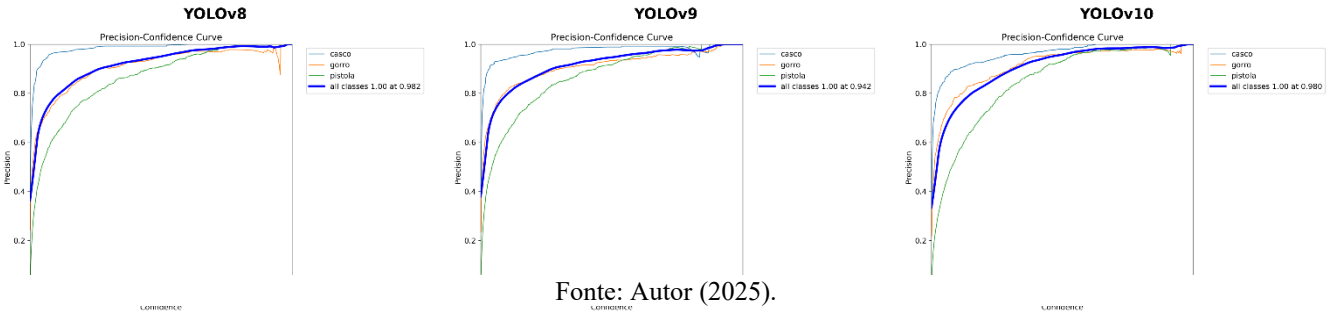


Fonte: Autor (2025).

5.2.4 Considerações práticas para a segurança pública

As curvas de *precision e recall em função da confiança* mostram que o YOLOv8 mantém bom desempenho mesmo em limiares baixos, sendo mais confiável em aplicações de tempo real. O YOLOv9 exige limiares elevados para garantir precisão, o que reduz recall. Já o YOLOv10 apresentou maior estabilidade para classes de proteção, mas desempenho insuficiente para pistolas como mostra a Figura 10.

Figura 10 - Curvas de Precision e Recall vs. Confiança dos modelos YOLOv8, YOLOv9 e YOLOv10.



Fonte: Autor (2025).

5.3 Limitações Observadas

Durante o desenvolvimento e avaliação dos modelos YOLO, foram identificadas algumas limitações que podem ter impactado os resultados e comprometido, em certa medida, a generalização dos modelos.

Uma das principais limitações está relacionada ao desequilíbrio de dados entre as classes presentes no conjunto de treinamento. Essa distribuição desigual pode favorecer o aprendizado de classes mais frequentes em detrimento das menos representadas, resultando em baixa taxa de detecção ou maior número de erros para essas categorias minoritárias. Como consequência, o modelo pode apresentar desempenho satisfatório em termos globais, mas com prejuízos significativos na acurácia por classe.

Além disso, foram observadas situações de falsa detecção (falsos positivos) e omissão de objetos (falsos negativos) em alguns cenários. Esses erros são críticos, especialmente em aplicações sensíveis, e indicam que, mesmo com métricas agregadas favoráveis, os modelos ainda podem falhar na identificação precisa de objetos em casos específicos.

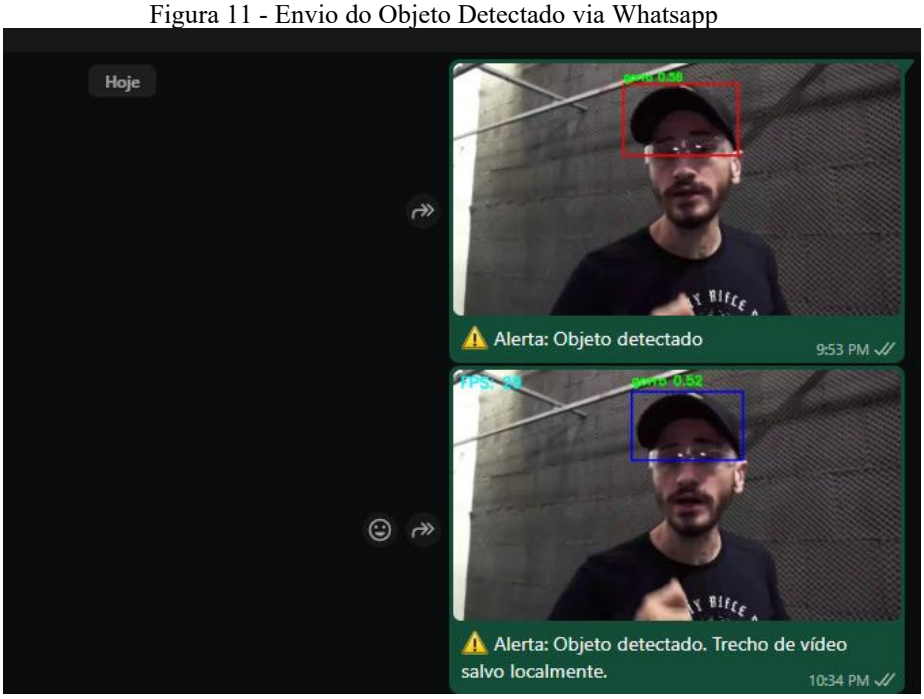
Outro fator relevante é a influência de condições externas, como variações de iluminação, ângulo de visão e resolução das imagens. Essas variáveis impactam diretamente na qualidade da detecção, podendo reduzir a confiança das predições ou dificultar a localização correta dos objetos. Como os modelos foram treinados e avaliados em um ambiente relativamente controlado, a presença dessas condições adversas em ambientes reais pode degradar substancialmente o desempenho observado.

Dessa forma, essas limitações evidenciam a necessidade de abordagens complementares, como balanceamento do conjunto de dados, uso de estratégias avançadas de data augmentation, calibração de confiança nas predições e testes em condições mais variadas, a fim de tornar os modelos mais robustos e aplicáveis em contextos reais.

5.3.1 Teste Simulado e Observações

A Figura 11 exemplifica o envio automático, via WhatsApp, do frame no qual o objeto foi detectado, permitindo resposta imediata das equipes de segurança. Para validar o protótipo, foi realizado um teste simulado com vídeos representativos de monitoramento urbano. Durante os testes, o sistema apresentou capacidade consistente de detectar objetos de interesse e enviar alertas via WhatsApp.

Figura 11 - Envio do Objeto Detectado via Whatsapp



Fonte: Autor (2025).

Na Figura 12, observam-se trechos de vídeo salvos automaticamente pelo sistema ao detectar objetos de interesse, possibilitando análise posterior dos eventos. Observou-se que, apesar da eficácia, pequenas falhas ocorreram em situações de baixa luminosidade. As observações indicam que o protótipo é funcional como sistema de alerta inteligente, fornecendo informações rápidas sobre eventos críticos. Outro ponto está relacionado a trecho da detecção salvos localmente.

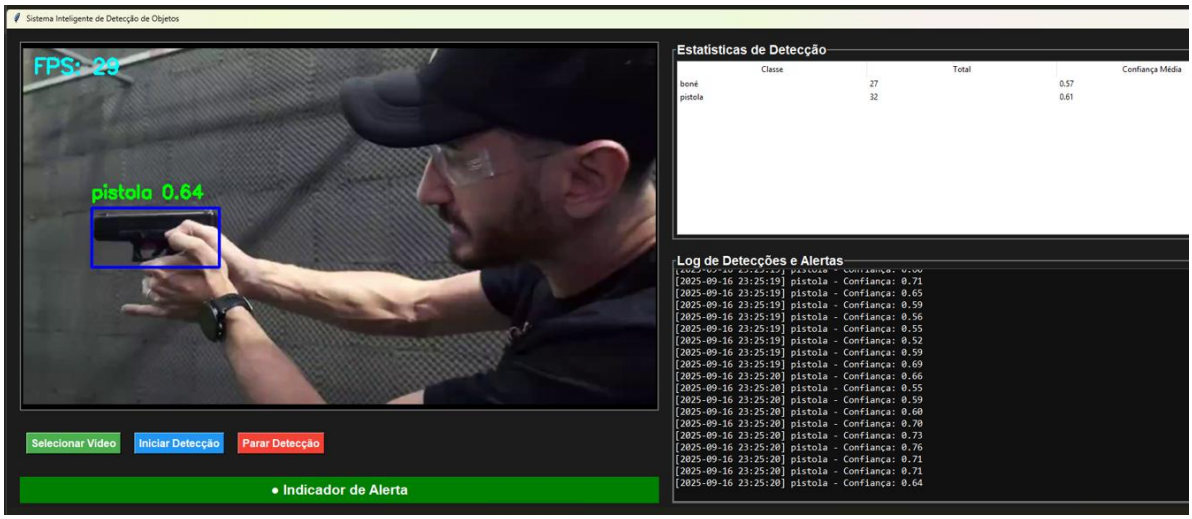
Figura 12 - Trechos de Detecção Salvos Localmente

Fonte: Autor (2025).

A Figura 13 mostra o protótipo em operação, realizando a detecção de múltiplos objetos em tempo quase real e destacando-os com caixas delimitadoras na tela. Os resultados mostraram que o YOLOv8 apresentou alta precisão nas detecções com uma taxa de confiabilidade de 60%,

mesmo em situações desafiadoras, como iluminação variável, movimento rápido ou objetos parcialmente ocultos. Algumas limitações foram observadas, como dificuldade em identificar objetos muito próximos ou parcialmente ocluídos, indicando oportunidades de ajustes no modelo e aprimoramentos nos dados de treinamento.

Figura 13 - Protótipo em Funcionamento



Fonte: Autor (2025).

5.4 Considerações para Aplicações Reais e Discussão Comparativa

A partir dos resultados obtidos na avaliação dos modelos da família YOLO, observa-se um potencial significativo para aplicação em contextos reais, especialmente em sistemas de vigilância urbana inteligente. A capacidade desses modelos de detectar objetos com alta precisão em tempo quase real os torna adequados para tarefas como monitoramento de vias públicas, identificação de veículos e reconhecimento de situações de risco. A eficiência computacional, especialmente observada no YOLOv8, reforça sua viabilidade para uso em sistemas de vídeo em tempo real.

Entretanto, a implantação prática de soluções baseadas em visão computacional envolve diversos desafios. Do ponto de vista técnico, a exigência de hardware compatível, como placas gráficas de alto desempenho, pode representar uma barreira, sobretudo em aplicações descentralizadas ou em larga escala. Além disso, questões legais e éticas, como o uso de imagens de pessoas em ambientes públicos, levantam preocupações quanto à privacidade, armazenamento de dados sensíveis e conformidade com legislações locais, como a Lei Geral de Proteção de Dados (LGPD) no Brasil ou o Regulamento Geral sobre a Proteção de Dados (GDPR) na União Europeia.

No contexto comparativo entre versões do modelo, o YOLOv8 se destacou como a versão mais adequada para o objetivo desta pesquisa, apresentando equilíbrio entre desempenho em classes críticas e eficiência computacional. Ajustes no dataset continuam sendo essenciais para aprimorar a acurácia, reduzir falsos negativos e mitigar falhas na detecção de objetos de risco, garantindo maior confiabilidade do sistema em situações reais de vigilância.

A análise comparativa entre YOLOv8, YOLOv9 e YOLOv10 evidenciou diferenças relevantes:

- **YOLOv8:** precisão de 93,39%, recall de 88,15%, mAP@0.5 de 92,42% e mAP@0.5:0.95 de 60,83%, com tempo de treinamento de 3.634,66 segundos. As perdas de validação mantiveram-se mais baixas, indicando melhor convergência e capacidade de generalização.
- **YOLOv9:** obteve precisão ligeiramente superior (93,86%), mas desempenho inferior em mAP e tempo de treinamento elevado (10.182,70 segundos), limitando sua aplicabilidade prática.
- **YOLOv10:** registrou os piores resultados, com precisão de 90,92%, recall de 82,46%, mAP@0.5 de 88,40% e mAP@0.5:0.95 de 57,15%, além de perdas de validação mais altas, inviabilizando sua recomendação, apesar do tempo de treinamento intermediário (4.168,52 segundos).

Dessa forma, conclui-se que o YOLOv8 representa a melhor alternativa entre os modelos testados, oferecendo o equilíbrio ideal entre precisão, eficiência computacional e capacidade de generalização. Sua adoção em cenários reais requer uma abordagem multidisciplinar que considere aspectos técnicos, éticos e legais, garantindo aplicação eficaz, segura e responsável da tecnologia.

6 CONCLUSÃO

A análise comparativa dos modelos YOLOv8, YOLOv9 e YOLOv10 demonstrou que o YOLOv8 se destaca como a alternativa mais equilibrada entre precisão e eficiência computacional para aplicações de detecção de objetos em tempo quase real. Além de apresentar as melhores métricas de desempenho (precision, recall e mAP), o modelo também se mostrou mais eficiente em termos de tempo de treinamento, oferecendo maior confiabilidade para cenários operacionais.

Embora o YOLOv9 tenha apresentado alta precisão, seu custo computacional elevado comprometeu sua viabilidade prática, enquanto o YOLOv10 apresentou desempenho geral inferior. Esses resultados evidenciam a necessidade de avaliar não apenas a acurácia isolada, mas todo o conjunto de fatores que influenciam a aplicação de modelos de visão computacional em contextos reais.

O desenvolvimento do protótipo revelou limitações típicas de sistemas baseados em IA, como desequilíbrio entre classes, ocorrência de falsos positivos e negativos, e sensibilidade a condições externas, como iluminação e ângulos de captura. Tais limitações reforçam a importância de adaptações específicas para o ambiente de uso, mitigação de erros críticos e consideração de aspectos éticos e legais, incluindo a conformidade com a Lei Geral de Proteção de Dados (LGPD) e a preservação da privacidade de indivíduos monitorados.

Como contribuição prática, esta pesquisa fornece uma análise detalhada das versões mais recentes do YOLO, oferecendo subsídios para profissionais e pesquisadores que buscam implementar soluções de visão computacional de maneira eficiente e responsável. Para trabalhos futuros, recomenda-se ampliar o conjunto de dados, aplicar técnicas avançadas de data augmentation, realizar ajustes finos (fine-tuning) para ambientes específicos e avaliar a performance em dispositivos embarcados. A integração com sistemas de análise comportamental ou predição de eventos também se apresenta como uma oportunidade promissora para aumentar a inteligência e autonomia do sistema.

Em síntese, este trabalho demonstra o potencial da arquitetura YOLO para aplicações reais, evidenciando tanto suas capacidades quanto os cuidados necessários para sua adoção segura, ética e eficaz, em conformidade com legislações de proteção de dados, como a LGPD, em ambientes operacionais.

REFERÊNCIAS

ALPAYDIN, E. Introduction to Machine Learning. 4. ed. Cambridge: MIT Press, 2020.

AMODEI, D. et al. Concrete Problems in AI Safety. arXiv:1606.06565, 2016. Disponível em: https://arxiv.org/abs/1606.06565. Acesso em: 15 out. 2025.

ATHENA SECURITY. AI Gun Detection System. 2020. Disponível em: https://athena-security.com/. Acesso em: 29 ago. 2025.

BOJARSKI, M. et al. End to End Learning for Self-Driving Cars. NVIDIA, 2016. Disponível em: https://arxiv.org/abs/1604.07316. Acesso em: 10 set. 2025.

BRASIL. Lei nº 13.709 (Lei Geral de Proteção de Dados - LGPD). Diário Oficial da União, Brasília, 14 ago. 2018.

BRASIL. Projeto de Lei n.º 2.338/2023. Dispõe sobre o uso da Inteligência Artificial no Brasil. Câmara dos Deputados, Brasília, 2024. Disponível em: https://www.camara.leg.br. Acesso em: 15 set. 2025.

BROWN, T. et al. Language Models are Few-Shot Learners. OpenAI, 2020. Disponível em: https://arxiv.org/abs/2005.14165. Acesso em: 5 nov. 2025.

CHEN, C. et al. MMDetection: Open MMLab Detection Toolbox. Journal of Machine Learning Research, v. 21, p. 1-7, 2020.

EUROPEAN COMMISSION. General Data Protection Regulation (GDPR). Brussels, 2024. Disponível em: https://gdpr.eu/. Acesso em: 15 set. 2025.

EUROPEAN PARLIAMENT. Artificial Intelligence Act (AI Act). Brussels, 2024. Disponível em: https://artificialintelligenceact.eu/. Acesso em: 15 set. 2025.

EUROPEAN UNION. General Data Protection Regulation (GDPR). Official Journal of the EU, 2016/679, 2016.

GARCEZ, R.; MENDES, T.; OLIVEIRA, L. Bias and fairness in facial recognition systems: A systematic review. Journal of Artificial Intelligence Ethics, v. 2, n. 1, p. 33-49, 2025.

GOODFELLOW, I. et al. Deep Learning. Cambridge: MIT Press, 2016.

HOWARD, A. et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. Google Research, 2017. Disponível em: https://arxiv.org/abs/1704.04861. Acesso em: 20 ago. 2025.

JIA, W.; ZHANG, Y.; WANG, H.; LI, F. Real-time automatic helmet detection of motorcyclists in urban traffic. IET Image Processing, v. 15, n. 3, p. 605–613, 2021. Disponível em: https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.12295. Acesso em: 16 set. 2025.

JOCHEM, T. et al. Real-Time Object Detection with YOLO. Journal of Artificial Intelligence Research, v. 68, p. 123-145, 2020.

KUMAR, S.; SINGH, P. Smart surveillance framework for public safety using deep learning. IEEE Access, v. 10, p. 11245–11257, 2022. DOI: 10.1109/ACCESS.2022.3145678.
LECUN, Y. et al. Deep Learning. Nature, v. 521, n. 7553, p. 436-444, 2015.

LI, H.; ZHANG, X.; WANG, J. YOLO-PL: Helmet wearing detection algorithm based on YOLOv4. Computers in Industry, v. 137, p. 103589, 2024. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S1051200423003780. Acesso em: 16 set. 2025.

MOURA, D.; LOPES, C. O efeito panóptico da vigilância inteligente: impactos psicossociais do monitoramento em massa. Revista Brasileira de Psicologia e Sociedade, v. 37, n. 2, p. 221–240, 2023.

OECD – Organisation for Economic Co-operation and Development. OECD Framework for the Responsible Use of Artificial Intelligence. Paris, 2024. Disponível em: https://oecd.ai. Acesso em: 15 set. 2025.

PINTO, A.; ARAÚJO, F. Viés algorítmico e discriminação em sistemas de segurança pública. Revista de Direito Digital e Sociedade, v. 11, n. 3, p. 77–95, 2024.

RAMIRO, H. L. L.; RAMIRO, M. G. N.; TAMAOKI, C. C. A sociedade pós-moderna e o panóptico digital: o livre desenvolvimento da personalidade em risco. Revista Direito & Liberdade, v. 12, p. 1-12, 2024. Disponível em: https://revistas.fucamp.edu.br/index.php/direito-realidade/article/view/3169/2019. Acesso em: 16 set. 2025.

REDMON, J. et al. YOLOv3: An Incremental Improvement. arXiv:1804.02767, 2018. Disponível em: https://arxiv.org/abs/1804.02767. Acesso em: 12 set. 2025.
RUSSELL, S.; NORVIG, P. Artificial Intelligence: A Modern Approach. 4. ed. Harlow: Pearson, 2021.

SANTOS, D. F.; LIMA, J. P. O impacto das tecnologias digitais na segurança pública no Brasil. Revista Brasileira de Segurança Pública, v. 14, n. 2, p. 72–89, 2020.

SHARMA, R.; GUPTA, M.; VERMA, S. AI-driven public safety: Emerging trends in computer vision for crime prevention. International Journal of Computer Applications, v. 183, n. 45, p. 25–33, 2021.

SILVA, E. B. Ética e IA: Dilemas na Vigilância Inteligente. São Paulo: Editora Tecnológica, 2023.

SOUZA, R.; PEREIRA, L. Proteção de dados e inteligência artificial: desafios da LGPD no contexto da vigilância. Revista Brasileira de Direito e Tecnologia, v. 6, n. 1, p. 15–34, 2023.

SOUZA, R. M.; ALBUQUERQUE, E. S. Panoptismo: reflexões sobre os efeitos da vigilância constante e o papel das câmeras na sociedade contemporânea. Revista Contradição, v. 5, n. 1, p. 1–14, jan./jun. 2024. Disponível em: https://www.researchgate.net/publication/381492325_O_panoptismo_atual_os_efeitos_da_vigilancia_constante_e_o_papel_das_cameras_na_sociedade_contemporanea. Acesso em: 16 set. 2025.

TORRALBA, A. et al. MIT Places Dataset. MIT CSAIL, 2020. Disponível em: http://places.csail.mit.edu. Acesso em: 3 nov. 2025.

UN. Report on AI and Human Rights. United Nations, 2022. Disponível em: https://www.ohchr.org. Acesso em: 8 out. 2025.

WANG, C. Y. et al. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art. arXiv:2207.02696, 2022. Disponível em: https://arxiv.org/abs/2207.02696. Acesso em: 25 ago. 2025.

ZHANG, H. et al. ResNeSt: Split-Attention Networks. arXiv:2004.08955, 2020. Disponível em: https://arxiv.org/abs/2004.08955. Acesso em: 17 set. 2025.

ZHAO, Z.; WANG, L. Real-time firearm detection in surveillance video using YOLOv4. Journal of Visual Communication and Image Representation, v. 78, p. 103145, 2021. DOI: 10.1016/j.jvcir.2021.103145.

ZHU, X. et al. Deformable DETR: Deformable Transformers for Object Detection. arXiv:2010.04159, 2020. Disponível em: https://arxiv.org/abs/2010.04159. Acesso em: 30 out. 2025.

ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. University of Oxford, 2014. Disponível em: https://arxiv.org/abs/1409.1556. Acesso em: 1 nov. 2025.

ANEXO A – REPOSITÓRIO DO PROJETO NO GITHUB

Título do Projeto: Protótipo de Sistema de Vigilância Baseado em IA para Prevenção de Crimes com Armas e Disfarces

Link de Acesso: https://github.com/eltonbarbosaa/tcc-yolo

Descrição:

Este repositório contém os códigos-fonte, datasets utilizados e documentação do projeto desenvolvido. Inclui:

- Implementações dos modelos YOLOv8, YOLOv9 e YOLOv10
- Scripts para pré-processamento de dados e treinamento
- Manual de instalação e requisitos do sistema