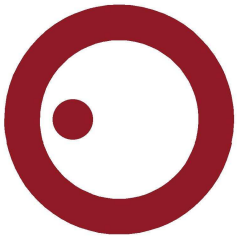


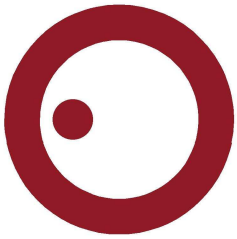
escola superior de tecnologia e gestão
instituto politécnico de leiria

Instituto Politécnico de Leiria
Escola Superior de Tecnologia e Gestão
Departamento de Engenharia Informática
Mestrado em Cibersegurança e Informática Forense

INTELIGÊNCIA ARTIFICIAL APLICADA À ANÁLISE DIGITAL FORENSE DE VÍDEOS

ESTUDANTE DANIEL ANTÓNIO SOUSINHA
Nº 2222940

Leiria, Setembro de 2024



escola superior de tecnologia e gestão
instituto politécnico de leiria

Instituto Politécnico de Leiria
Escola Superior de Tecnologia e Gestão
Departamento de Engenharia Informática
Mestrado em Cibersegurança e Informática Forense

INTELIGÊNCIA ARTIFICIAL APLICADA À ANÁLISE DIGITAL FORENSE DE VÍDEOS

ESTUDANTE DANIEL ANTÓNIO SOUSINHA
Nº 2222940

Dissertação realizada sob orientação do Professor Patrício Rodrigues Domingues, Professor Miguel Monteiro de Sousa Frade e Professor Miguel Cerdeira Marreiros Negrão.

Leiria, Setembro de 2024

AGRADECIMENTOS

Quero expressar a minha gratidão a todos os que contribuíram para que este projeto fosse concluído com êxito.

Os meus agradecimentos especiais vão para os meus orientadores Professor Patrício Rodrigues Domingues, Professor Miguel Monteiro de Sousa Frade e Professor Miguel Cerdeira Marreiros Negrão pelo apoio, orientação e a partilha de conhecimentos valiosos que sempre deram ao longo do desenvolvimento deste projeto.

Para terminar, quero agradecer à minha família e aos meus amigos, que de certa forma foram colaborando e apoiando ao longo deste percurso.

RESUMO

O número de dispositivos digitais, os respetivos volume de dados e a complexidade dos mesmos está a criar longas filas de espera nos laboratórios afetos à análise digital forense. Isso é particularmente evidente nas análises que envolvem conteúdos fotográficos e vídeos, pois a captura de imagens e vídeos tornou-se trivial e omnipresente com os *smartphones* que uma grande parte da população traz permanente consigo. Similarmente, a grande proliferação de câmaras de vídeo-vigilância, e o associado volume de vídeos que capturam tem contribuído para uma maior procura de recursos para análise e interpretação dos mesmos.

A IA tem sido apontada como um dos caminhos mais promissores para através da automação que a mesma permite, ser possível aliviar os analistas forenses de conteúdos digitais, delegando parte das tarefas para algoritmos de IA. Em particular, a análise computacional de conteúdos visuais apresenta já soluções interessantes, tanto ao nível da deteção e classificação de objetos, como no reconhecimento facial.

Este projeto aborda a integração da IA no processo da análise forense digital, com foco principal na área de análise de conteúdo de vídeo e os desafios que daí advém. Estes desafios podem prejudicar a exatidão e a eficiência de uma investigação, exigindo uma solução robusta. A solução proposta integra tecnologias de IA para melhorar a velocidade e a precisão da análise de vídeo, centrando-se no desenvolvimento de modelos para o YOLOv5 para deteção de objetos e o uso de reconhecimento facial para identificação de indivíduos. Para o efeito foi desenhada uma interface web assente no ambiente Vue.js, integrando as funcionalidades de deteção e classificação de classes de objetos disponibilizadas pelo YOLOv5, bem como o reconhecimento facial através das funcionalidades da biblioteca *Facial Recognition*, ela mesmo assente no *software* dlib. Procurou-se com a aplicação desenvolvida, minimizar a exposição do utilizador às minúcias das metodologias de IA subjacentes, disponibilizando ao utilizador um conjunto de funcionalidades aptas para que o mesmo se foque essencialmente no trabalho de análise e interpretação.

Outra atividade relevante deste projeto foi o treino de classificadores de classes de objetos específicos com vista à criação de modelos para o YOLOv5. De facto, uma perícia digital forense pode requerer a procura de classes de objetos não suportados pelo YOLOv5, pelo que se considerou pertinente integrar neste projeto

o estudo dos passos subjacentes ao desenvolvimento de classificadores de objetos à medida. O estudo apresentado neste relatório abrange desde a criação e recolha dos conjuntos de dados, as diligências necessárias ao treino e a avaliação dos resultados, nomeadamente da respetiva precisão e tempo de execução. Para o efeito foram criados quatro classes de objetos matriculas de veículos, espingarda, pistola e faca, descrevendo-se o processo e os resultados. Tal permitiu aquilatar dos desafios de tal atividade, sendo que as maiores dificuldades foram sentidas na criação dos datasets e nas restrições derivadas dos limitados recursos computacionais.

ABSTRACT

The number of digital devices, their volume of data and their complexity is creating long queues in digital forensic analysis laboratories. This is particularly evident in analyses involving photographic and video content, as capturing images and videos has become trivial and ubiquitous with the smartphones that a large part of the population carries around with them at all times. Similarly, the proliferation of video surveillance cameras and the associated volume of videos they capture has contributed to a greater demand for resources to analyze and interpret them.

Artificial intelligence (AI) has been touted as one of the most promising avenues for relieving forensic analysts of digital content, by delegating part of the tasks to AI algorithms. In particular, the computational analysis of visual content already presents interesting solutions, both in terms of the detection and classification of objects and in facial recognition.

This project addresses the integration of artificial intelligence in the process of digital forensic analysis, with a main focus on the area of video content analysis and the challenges that arise from this. These challenges can jeopardize the accuracy and efficiency of an investigation, requiring a robust solution. The proposed solution integrates AI technologies to improve the speed and accuracy of video analysis, focusing on the development of models for YOLOv5 for object detection and the use of facial recognition for identifying individuals. To this end, a web interface based on the Vue.js environment was designed, integrating the object class detection and classification functionalities provided by YOLOv5, as well as facial recognition through the functionalities of the Facial Recognition library, itself based on the dlib software. The aim of the application developed was to minimize the user's exposure to the minutiae of the underlying AI methodologies, providing the user with a set of functionalities that allow them to focus essentially on analysis and interpretation.

Another relevant activity of this project was the training of classifiers for specific object classes with a view to creating models for YOLOv5. In fact, digital forensics may require the search for object classes not supported by YOLOv5, which is why it was considered pertinent to include in this project the study of the steps underlying the development of tailor-made object classifiers. The study presented in this report covers everything from the creation and collection of data sets, the steps required

for training and the evaluation of the results, namely their accuracy and execution time. To this end, four classes of objects were created: rifle, pistol, knife, and the process and results are described. This made it possible to assess the challenges of this activity, with the greatest difficulties being experienced in the creation of datasets and the restrictions deriving from limited computing resources.

ÍNDICE

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

AGI	Artificial General Intelligence.
ANI	Artificial Narrow Intelligence.
API	Interface de Programação de Aplicações.
ASI	Super Inteligência Artificial.
CART	Equipa de Análise e Resposta a Computadores.
CCTV	Circuito Fechado de Televisão.
CNN	Redes Neurais Convolucionais.
CPUI	Camâras portáteis de uso individual.
CSS	Folhas de Estilo em Cascata.
FBI	Federal Bureau of Investigation.
FP	False Positives.
GPUs	Unidades de Processamento Gráfico.
HTML	Linguagem de Marcação de Hipertexto.
IA	Inteligência Artificial.
JSON	JavaScript Object Notation.
NIJ	National Institute of Justice.
NMS	Supressão Não-Máxima.
OCR	Reconhecimento Ótico de Caracteres.

OpenCV	Biblioteca de Visão Computacional de Código Aberto.
PNL	Processamento de Linguagem Natural.
RAM	Memória de Acesso Aleatório.
RGB	Red/Green/Blue.
RNN	Redes Neurais Recorrentes.
TP	True Positive.
VLC	VLC Media Player.
YOLO	You Only Look Once.

1

INTRODUÇÃO

A crescente dependência de dispositivos digitais na sociedade moderna resultou num aumento significativo do volume de dados gerados, particularmente no que diz respeito a conteúdo multimédia. Este aumento é impulsionado, em grande parte, pela proliferação de *smartphones*, que permitem a captura de vídeos de forma rápida e acessível, além do crescente número de câmaras de videovigilância espalhadas em áreas urbanas e comerciais. Conforme mencionado no Ericsson Mobility Report, o tráfego de dados móveis deverá atingir 466 exabytes até ao ano de 2029 *Mobile data traffic forecast* 2024. Além disso, a análise de dados multimédia apresenta desafios significativos em termos de volume, variedade e velocidade, exigindo soluções de big data e inteligência artificial para processar e interpretar esse enorme fluxo de dados *Big Multimedia Data and Applications* 2024. Perante esta realidade, torna-se evidente a necessidade de automatizar o processamento de vídeos, utilizando tecnologias avançadas, como a inteligência artificial, para garantir a eficiência e a exatidão nas investigações forenses.

1.1 MOTIVAÇÃO

A Inteligência Artificial (IA) tornou-se parte integrante das nossas vidas e devido a sua progressão a IA pode ser utilizada para auxiliar as pessoas nas mais diversas tarefas, nomeadamente para automatizar tarefas repetitivas e tediosas permitindo que estas se concentrem nos trabalhos mais importantes e criativos. Neste contexto, a IA pode ser vista como um potencial assistente humano que nos permite atingir os objetivos com mais eficiência. As técnicas de IA de ponta, incluindo principalmente o *machine learning* e a variante *deep learning*, poderão encontrar um lugar na análise forense nas áreas de tratamento de grandes quantidades de dados.

1.2 OBJETIVOS

Este trabalho tem como objetivo explorar a junção entre a análise forense digital e a inteligência artificial no que respeita as suas aplicações na análise de conteúdos de vídeo. Para o efeito, serão abordados neste trabalho a definição e os marcos da análise forense digital, a sua aplicação na análise de vídeo e os desafios que daí advêm. Segue-se uma revisão dos progressos da inteligência artificial e das suas caraterísticas, abrangendo noções básicas sobre *machine learning*, *deep learning* e redes neuronais. Após esta apresentação, será demonstrada uma prova de conceito utilizando ferramentas de IA para analisar vídeos, a fim de provar a praticabilidade da metodologia em questão. Em súmula pretende-se automatizar o processo de reconhecimento de pessoas e de objetos em vídeo, avaliando-se para o efeito varias metodologias.

1.3 CONTRIBUTOS

Este projeto contribui para a aplicação da inteligência artificial na análise de vídeo forense, centrando-se na automatização e na melhoria da precisão das investigações. Foi realizada uma revisão do estado da arte no processamento de vídeo baseado em IA, identificando as tendências atuais, os desafios e as oportunidades de melhoria.

Com base nisto, foi desenvolvida uma ferramenta para detetar pessoas e objetos em vídeos utilizando o modelo YOLOv5, acelerando significativamente o processo de investigação. Além disso, a tecnologia de reconhecimento facial, através do FaceRecognition, foi integrada para identificar eficazmente indivíduos em imagens de vídeo.

Para facilitar a utilização destas ferramentas, foi criada uma interface Web com Vue.js, oferecendo uma experiência de fácil utilização aos investigadores. Por fim, foram avaliadas as limitações, como as restrições de *hardware* e a necessidade de conjuntos de dados equilibrados, com recomendações para melhorias futuras.

1.4 ESTRUTURA DO DOCUMENTO

O trabalho encontra-se organizado da seguinte forma:

No capítulo 1, apresenta-se a introdução, que inclui a motivação para o estudo, os objetivos propostos, os contributos esperados e a estrutura do documento.

No capítulo 2, aborda-se a análise forense digital. Inicialmente, discutem-se a definição, os marcos históricos e o processo da análise forense digital. Em seguida, dá-se especial atenção à análise de conteúdo de vídeo, destacando-se o papel dos vídeos em investigações e os desafios associados à sua análise.

No capítulo 3, exploram-se os conceitos de inteligência artificial. Este capítulo começa com uma definição e breve história da IA, seguida por uma explicação dos seus tipos (i.e IA fraca e IA forte). Também se detalham os diferentes tipos de *machine learning* (supervisionado, não supervisionado e de reforço) e de *deep learning*, com ênfase em redes neuronais convolucionais (CNN) e recorrentes (RNN). Além disso, discutem-se as diferenças entre essas redes de forma a introduzir a prova de conceito. O capítulo termina com uma análise do estado da arte na área da IA.

O capítulo 4 foca-se na análise forense digital assistida por IA, abordando-se as aplicações existentes, como a automação de análise de dados, o processamento de linguagem natural (PLN), a análise de imagens e vídeos, e o reconhecimento de padrões. Também se discutem os trabalhos relacionados, as possíveis vantagens da utilização da IA nesse contexto, os desafios potenciais e as previsões futuras para o campo.

No capítulo 5, apresenta-se uma prova de conceito sobre a integração de IA para a análise de vídeo. Descrevem-se as ferramentas utilizadas, a arquitetura, o fluxo do sistema e uma introdução a criação de modelos YOLOV5.

No capítulo 6, demonstram-se os resultados obtidos através dos testes efetuados recorrendo ao uso da prova de conceito, bem como serão tecidos alguns comentários acerca dos mesmos.

Por fim, no capítulo 7, apresentam-se as conclusões gerais, que incluem os resultados finais, as observações mais relevantes e as sugestões para trabalhos futuros, destacando-se a aplicabilidade da metodologia desenvolvida.

2

ANÁLISE FORENSE

A análise forense é o estudo sistemático de evidências, envolvendo métodos e princípios científicos para analisar casos civis/criminais com o objetivo de resolvê-los. Entre esses tipos de informações estão as derivadas de fontes digitais, como arquivos e comunicações, além de dados extraídos de amostras físicas ou quaisquer outras origens que revelem detalhes; esse termo pode ser aplicado em processos judiciais. De acordo com a definição dada pelo National Institute of Justice (NIJ), a ciência forense é um ramo da ciência aplicada, preocupado principalmente com o estabelecimento de provas legais, que pertencem a um caso civil ou criminal, através da aplicação de leis e princípios das ciências naturais. *Recent Developments in Forensic Trace Evidence Analysis* 2024

2.1 ANÁLISE FORENSE DIGITAL

2.1.1 *Definição*

O campo da análise forense digital centra-se no exame e extração de provas digitais de vários dispositivos eletrónicos, incluindo computadores, *smartphones*, *tablets* e câmaras. Estas provas valiosas desempenham um papel crucial na descoberta da verdade por detrás de atividades criminosas, fraudes, violações de segurança, disputas legais e quaisquer outros casos que envolvam dados digitais.

2.1.2 *Marcos*

Seguidamente, apresenta-se uma sucinta evolução temporal da análise forense.

- **1984**: O primeiro laboratório forense digital é geralmente creditado ao Federal Bureau of Investigation (FBI) nos Estados Unidos. Em 1984, o FBI criou o Programa de Meios Magnéticos, que mais tarde evoluiu para a Equipa de Análise e Resposta a Computadores (CART). Esta iniciativa marcou o início

da perícia digital formalizada, com o objetivo de lidar com o volume crescente de provas digitais resultantes de crimes informáticos. *Digital Forensics Used to Help Law Enforcement, Employers Defend Against Cybercrime* 2024

- **1990**: A década de 1990 assistiu a um progresso significativo no desenvolvimento de ferramentas e técnicas especializadas para examinar provas digitais, refletindo a necessidade crescente das capacidades forenses para fazer face à crescente prevalência de dispositivos digitais em atividades criminosas. (Casey, 2011)
- **2000**: Na década de 2000, a proliferação de dispositivos digitais e a adoção generalizada da Internet expandiram o âmbito e a complexidade das investigações forenses digitais. Durante este período, a quantidade e variedade de provas digitais disponíveis para análise forense cresceram exponencialmente. (Carrier, 2005)
- **Presente**: O campo da análise forense digital passou por mudanças drásticas nos últimos anos com a infusão de inteligência artificial e *machine learning*, procurando atingir uma melhor produtividade e um aumento do grau de confiabilidade das avaliações forenses. Ferramentas que façam o uso destas novas tecnologias podem permitir a automação e a deteção de padrões e irregularidades rapidamente, o que significa que um enorme conjunto de dados pode ser processado com facilidade e precisão (Garfinkel, 2010). Note-se, contudo que o número de casos envolvendo dispositivos digitais e o volume de dados, nomeadamente dados multimédia continua a crescer de forma exponencial.

2.1.3 *Processo*

A análise forense é uma abordagem precisa e ordenada, planeada para revelar, conservar e interpretar provas para fins legais e de investigação. Este procedimento reveste-se de uma importância fundamental na era da tecnologia, em que as provas assumem, em grande medida, a forma de informação digital Carrier, 2005.

O procedimento contém normalmente as seguintes fases: recolha, preservação, exame, análise e elaboração de relatórios. Cada passo é controlado por diretrizes rigorosas para manter a integridade das provas e garantir que os resultados são legalmente defensáveis. Abaixo será possível entender mais em detalhe o processo que ocorre em cada uma destas fases:

1. **Recolha**: O ato de recolha implica a acumulação de possíveis provas de diferentes sectores: aparelhos digitais, redes e até locais. A fase em que a cadeia de custódia é mantida durante este processo é vital, uma vez que garante que as provas são mantidas intactas, sem alterações, e podem ser utilizadas de forma fiável em tribunal. De acordo com Casey (2011), "a recolha de provas digitais deve ser efetuada de forma a manter a integridade das provas e da cadeia de custódia". Isto sublinha a sua importância, uma falha resultaria na utilização de provas não fiáveis, mesmo que estas estivessem disponíveis, podendo em certos casos perder a sua legalidade.

   Algumas das ferramentas que podem ser usadas para a recolha de evidências digitais são:

   - **Bloqueadores de Escrita**: Dispositivos que impedem a escrita de dados nos dispositivos de armazenamento, preservando a integridade dos dados originais. Podem ser de hardware ou de software.
   - ***Imaging Tools***: Ferramentas que podem produzir réplicas duplicadas de dispositivos de armazenamento, incluindo informações ocultas ou que tenham sido eliminadas ou corrompidas. Exemplos: FTK Imager *FTK Imager* 2024
   - **Ferramentas forenses móveis**: Software concebido para recuperar dados de dispositivos móveis é frequentemente utilizado para obter contactos, mensagens, registos de chamadas, fotografias e vídeos ou quaisquer outros ficheiros presentes nos dispositivos. Exemplo: XRY *XRY Mobile Data Forensic Phone Extraction Recovery* 2024

2. **Preservação**: Proteger as informações recolhidas contra a corrupção ou a deterioração. Métodos como a criação de imagens forenses a partir do armazenamento eletrónico e a utilização de abordagens de armazenamento seguro são adotados para manter as provas na sua forma original. Tal como salientado por Carrier, 2005, "Fazer uma réplica bit a bit do dispositivo de memória salvaguarda que a prova original seja conservada dando assim a oportunidade aos examinadores forenses de lidarem com a prova copiada".

   Algumas das ferramentas que podem ser usadas para a preservação de evidências digitais são:

   - ***Hashing Tools***: Software que têm a capacidade de criar valores distintos e inalteráveis com base nos dados, formando essencialmente uma impressão digital para os dados. Estes valores podem ser utilizados para

validar a genuinidade e fiabilidade dos dados, comparando os valores gerados antes e depois dos processos de recolha, salvaguarda o exame dos dados.

- **Ferramentas de Encriptação**: Ferramenta para proteger dados, encriptando-os e tornando-os acessíveis apenas a indivíduos que possuam a chave de desencriptação. Estes programas funcionam como uma medida de proteção contra o acesso não autorizado, garantindo a máxima confidencialidade e segurança dos dados. Exemplos: VeraCrypt *VeraCrypt* 2024

3. **Exame**: Fase em que os peritos se debruçam sobre os elementos de prova com uma série de ferramentas e técnicas sofisticadas. Implica desvendar dados que possam ter sido ocultados, apagados ou encriptados, bem como compreender os pormenores técnicos essenciais das provas. Uma afirmação de Garfinkel, 2010 sublinha a importância desta fase: "O exame forense envolve a utilização de ferramentas e técnicas avançadas para descobrir artefactos digitais que podem fornecer informações essenciais sobre o caso".

   Algumas das ferramentas que podem ser usadas para o exame de evidências digitais são:

   - **Ferramentas de análise de ficheiros**: Software que permite aos utilizadores aceder, modificar, recuperar e restaurar informações de ficheiros numa vasta gama de formatos, incluindo documentos, folhas de cálculo, apresentações, imagens, vídeos e ficheiros de áudio. Exemplo: Autopsy *Autopsy* 2024
   - **Ferramentas de análise de memória**: Software que permite examinar a informação contida na memória de acesso aleatório (RAM) de um dispositivo. A RAM é um tipo de memória que é temporária e pode ser apagada quando o dispositivo é desligado. Estes programas têm a capacidade de recuperar vários tipos de dados, incluindo processos, ligações, malware e muito mais. Exemplo: Volatility Volatility, 2024
   - **Ferramentas de análise de rede**: Software que permite aos utilizadores intercetar, classificar, avaliar e reconstruir as informações transmitidas através de várias redes de comunicação, incluindo a internet, a intranet, o Wi-Fi, o Bluetooth, entre outras. Estes programas têm a capacidade de recuperar dados como protocolos, pacotes, endereços, portas e outras informações relevantes. Exemplo: Wireshark *Wireshark* 2024

- **Ferramentas de análise de espaço não alocado**: Software que permite aos utilizadores analisar e recuperar dados de espaços não alocados no disco. O espaço não alocado refere-se à parte do armazenamento que não foi atribuída ou alocada a ficheiros e ou documentos existentes. Apesar de parecer vazio, pode conter informações valiosas para as investigações, exigindo ferramentas especializadas para uma pesquisa eficiente devido ao grande tamanho dos discos rígidos modernos. PhotoREC, 2023

4. **Análise**: A interpretação/análise dos dados consiste em encontrar o significado do que deles foi extraído. Esta fase exige profissionais forenses capazes de estabelecer uma relação entre o que encontraram e o contexto do caso, discernir padrões a partir das descobertas, e concluir sobre os cenários prováveis que conduziram aos acontecimentos em causa. De acordo com Casey 2011, "a fase de análise é a fase em que os peritos forenses reúnem as provas para formar uma narrativa coerente que possa apoiar conclusões de investigação e jurídicas".
5. **Elaboração de Relatórios**: Última fase em que todos os resultados são apresentados de forma clara e pormenorizada. O relatório deve ser capaz de dar a conhecer uma visão abrangente das conclusões, mas não deve ser tão técnico que as pessoas fora dessa área específica não o consigam compreender. Isto inclui as partes interessadas em questões jurídicas, como júris ou advogados. Esta fase pode também significar testemunhar em tribunal, onde os peritos forenses têm de defender os seus processos e resultados de forma articulada. Carrier, 2005 refere que os relatórios devem ser claros: "Os relatórios forenses devem ser pormenorizados e precisos, fornecendo uma descrição clara dos métodos utilizados e das conclusões obtidas, para garantir que as provas sejam apresentadas de forma eficaz em contextos jurídicos."

2.2 ANÁLISE FORENSE DIGITAL: CONTEÚDO DE VÍDEO

2.2.1 *Papel do vídeo nas investigações modernas*

Um dos pilares das investigações forenses atuais são as provas de vídeo, que captam acontecimentos em tempo real e os preservam para a posteridade. Oferecem uma representação pictórica dos eventos que ocorreram, permitindo aos investigadores extrair pormenores das imagens. Com inúmeros dispositivos capazes de gravar imagens de vídeo, isto significa que a fonte de tais provas pode estar virtualmente

em qualquer lugar. Abaixo podemos observar os principais dispositivos utilizados para a recolha destas evidências:

1. **Circuito Fechado de Televisão (CCTV)**: Este tipo de câmaras podem ser encontradas em todo o lado, visto que captam continuamente imagens, estas podem desempenhar um papel importante para ajudar as forças policias a identificar suspeitos com base no local e na hora em que o crime ocorreu.
2. ***Camâras portáteis de uso individual (CPUI)***: As forças policiais estão a utilizar cada vez mais CPUI's, vulgo *bodycams* para documentar as diferentes interações e incidentes que ocorrem em serviço do ponto de vista dos agentes da polícia, para que possam assim ser capazes de proporcionar transparência e ajudar a proteger os agentes de falsas acusações.
3. **Gravações públicas e privadas**: Vídeos captados por dispositivos como *smartphones*, *dashcams* e outros aparelhos podem ser inestimáveis para algumas investigações forenses. Estes dispositivos frequentemente registam incidentes que, de outra forma, poderiam não ser documentados.

2.2.2 *Desafios na análise de vídeo*

Os principais desafios na análise forense de vídeo são:

- **Qualidade de Imagem**: A qualidade do vídeo pode ser gravemente comprometida por fatores como a má resolução, a iluminação inadequada e os ângulos de câmara insuficientes. Estas questões exigem a utilização de técnicas avançadas de melhoramento para extrair informações utilizáveis das filmagens (Casey, 2011).
- **Manipulação**: A integridade das provas de vídeo é fundamental, e a deteção de adulteração ou edição é um aspeto crítico da perícia de vídeo. A análise de metadados, a realização de inspeções fotograma a fotograma e a utilização de marcas de água digitais são técnicas essenciais utilizadas para verificar a autenticidade dos ficheiros de vídeo (Upadhyay e Singh, 2020). A criação de vídeo por IA tem evoluído significativamente, tornando ainda mais difícil distinguir o falso do verdadeiro, lançando assim novos desafios à sociedade em geral, e aos investigadores forenses em particular.
- **Armazenamento e processamento**: A manutenção da integridade dos dados durante as fases de armazenamento e análise é essencial para preservar o valor probatório das provas de vídeo. O armazenamento seguro e os métodos

de processamento fiáveis são fundamentais para garantir que o vídeo permanece inalterado e credível (Garfinkel, 2010).

Após a apresentação dos principais conceitos e marcos históricos da análise forense digital no capítulo anterior, com especial destaque para o papel dos vídeos como fonte de evidência, segue-se agora a introdução à inteligência artificial.

3

INTELIGÊNCIA ARTIFICIAL

Neste capítulo, será abordada a inteligência artificial, começando pela sua definição e evolução histórica. Em seguida, serão explorados os principais tipos de IA, com foco em *machine learning* e *deep learning*, e no papel das redes neuronais. Por fim, serão apresentadas algumas aplicações práticas, evidenciando os avanços recentes e suas implicações.

3.1 DEFINIÇÃO E HISTORIA

Inteligência Artificial (IA) refere-se ao uso de computadores e máquinas que pretendem imitar as capacidades humanas em termos de resolução de problemas e tomada de decisões. As origens da IA remontam à publicação de Alan Turing, publicada em 1950: “Computing Machinery And Intelligence”. Nesta, o autor colocou uma questão simples, mas profunda: “As máquinas podem pensar?” Turing, 1950, foi a partir desta questão que nasceu o famoso “Teste de Turing”, uma marca indelével na história da IA.

Uma análise mais abrangente sobre este tema foi posteriormente fornecida por Stuart Russell e Peter Norvig através do seu livro intitulado "Inteligência Artificial: Uma Abordagem Moderna"Russel e Norvig, 1996. Neste os mesmos estabeleceram quatro objetivos ou definições diferentes de IA que ajudam a distinguir os sistemas baseados na racionalidade e no pensamento daqueles baseados na ação:

1. **Sistemas que pensam como humanos**: Sistemas que tentam imitar o processo de pensamento humano. Russel e Norvig, 1996
2. **Sistemas que agem como humanos**: Sistemas que não só pensem, mas também ajam como humanos. Russel e Norvig, 1996
3. **Sistemas que pensam racionalmente**: Sistemas que são capazes de raciocinar e deduzir conclusões corretas com base em informações fornecidas, utilizando modelos matemáticos e lógicos de pensamento. Russel e Norvig, 1996

4. **Sistemas que agem racionalmente**: Sistemas que tomem ações racionais, ou seja, que ajam de forma a maximizar a probabilidade de alcançar metas desejadas com base em informações disponíveis. Russel e Norvig, 1996

Em suma, a definição de IA é o campo que combina a ciência da computação e conjuntos de dados robustos para possibilitar a solução de problemas, esta inclui subcampos como o *deep learning* Goodfellow et al., 2016, composto por algoritmos de IA que permitem o *machine learning* a partir de grandes volumes de dados.

3.2 TIPOS DE IA

3.2.1 *IA Fraca*

Também conhecida como *Narrow AI* ou *Artificial Narrow Intelligence (ANI)*. Este tipo de IA é desenvolvido e dedicado a tarefas específicas e constitui a maior parte da IA que nos é apresentada nos dias de hoje, sendo a mesma integrada em várias tecnologias e serviços. Desde assistentes virtuais como Siri e Alexa até sistemas de recomendação de conteúdo de plataformas de *streaming* e algoritmos de reconhecimento de voz de *smartphones*.

Apesar do seu elevado nível de especialização dirigido à eficiência nas tarefas, esta apresenta algumas limitações de generalização e adaptabilidade a novos contextos ou tarefas. A sua aplicabilidade está confinada ao âmbito específico para o qual foi concebido, significando que esta é incapaz de lidar com situações fora desse âmbito pré-determinado. *EDI Weekly: Engineered Design Insider* 2024

3.2.2 *IA Forte*

Embora esta ideia seja puramente hipotética, a mesma prevê um mundo onde as máquinas possuem inteligência equivalente ou superior à dos humanos, uma inteligência que é consciente de si mesma, capaz de resolver problemas, aprender e até prever o futuro. Esta IA em particular compreende Inteligência Geral Artificial (AGI) e Super Inteligência Artificial (ASI) dois componentes que buscam a consciência a nível humano. *EDI Weekly: Engineered Design Insider* 2024

3.3 MACHINE LEARNING

Machine Learning é um subcampo da IA que permite a um sistema aprender e melhorar autonomamente sem ser explicitamente programado. Estes algoritmos funcionam através do reconhecimento de padrões e dados fazendo previsões quando são introduzidos novos dados no sistema. Mitchell e Mitchell, 1997

Este subcampo pode ser dividido em três modelos, consoante a metodologia de treino adotada.

1. Supervisionado
2. Não Supervisionado
3. De reforço

De seguida, são analisados sucintamente cada um desses modelos:

3.3.1 *Supervisionado*

É um modelo de *machine learning* que usa, para efeitos de treino, dados rotulados por *labels* para mapear uma entrada numa saída. Simplificando, para treinar um modelo para reconhecer imagens de maçãs, é necessário alimentá-lo com imagens rotuladas como maçãs. Mitchell e Mitchell, 1997

Os tipos mais comuns deste tipo de aprendizagem são listados na tabela 1.

Tipo de Algoritmo	**Descrição**
Linear Regression	Algoritmo que assume uma relação linear entre a variável dependente e uma ou mais variáveis independentes.
Polynomial Regression	Uma extensão da regressão linear que usa uma equação polinomial para capturar relações não lineares entre as variáveis.
Naive Bayes	Baseado no Teorema de Bayes, este algoritmo assume que as características (features) são independentes entre si.
K-nearest Neighbors (KNN)	Um algoritmo que classifica ou faz previsões com base na proximidade dos exemplos de treinamento aos dados novos.
Decision Trees	Algoritmo que faz divisões sucessivas nos dados com base em atributos, criando uma estrutura em forma de árvore.

Tabela 1: Tipos de Algoritmos Supervisionados (Bishop, 2006)

3.3.2 *Não Supervisionado*

É um modelo de *machine learning* que utiliza dados não rotulados para aprender padrões. Ao contrário da aprendizagem supervisionada, o resultado não é conhecido antecipadamente, em vez disso o algoritmo aprende com os dados sem intervenção humana e categoriza-os em grupos com base em atributos. Por exemplo, se o algoritmo receber imagens de maças e bananas, ele trabalhará sozinho para categorizar qual das imagens é uma maça e qual é uma banana. Mitchell e Mitchell, 1997

Os tipos mais comuns deste tipo de aprendizagem encontram-se listados na tabela 2.

Tipo de Algoritmo	**Descrição**
Fuzzy C-Means	Algoritmo onde um ponto de dados pode pertencer a mais de um cluster com diferentes graus de associação.
K-means Clustering	Algoritmo de agrupamento que divide os dados em K clusters, onde cada ponto pertence ao cluster com o centro mais próximo.
Hierarchical Clustering	Algoritmo de agrupamento que cria uma árvore hierárquica de clusters, organizando-os em níveis de granularidade.
Principal Component Analysis (PCA)	Algoritmo de redução de dimensionalidade que transforma as variáveis originais em novas variáveis não correlacionadas, chamadas componentes principais.
Partial Least Squares (PLS)	Algoritmo de redução de dimensionalidade que encontra as direções de máxima variância que também explicam a correlação entre variáveis independentes e dependentes.

Tabela 2: Tipos de Algoritmos Não Supervisionados (Hastie et al., 2009)

3.3.3 *De Reforço*

É um modelo de *machine learning* que pode ser descrito através de uma série de experiências de tentativa erro. Um "agente"aprende a executar uma tarefa definida através de um ciclo de *feedback* até que o seu desempenho esteja dentro de um intervalo desejável. Este ciclo de *feedback* é feito através da devolução de um reforço positivo quando o agente executa bem a tarefa designada e um reforço negativo quando tem um mau desempenho. Mitchell e Mitchell, 1997

A tabela 3 apresenta os tipos mais comuns desta metodologia de aprendizagem.

Tipo de Algoritmo	Descrição
Baseados em Política	Aprendem diretamente uma política que mapeia estados para ações, otimizando a política para maximizar a recompensa.
Baseados em Valor	Aprendem uma função de valor que estima a recompensa esperada para cada estado ou ação, ajudando a escolher as melhores ações.
Ator-Crítico	Combinam uma política (ator) e uma função de valor (crítico) para otimizar a escolha de ações.
Baseados em Pesquisa	Usam simulações para explorar o espaço de decisões e identificar ações ótimas, comumente usados em jogos.
Reforço Profundo	Utilizam redes neuronais profundas para aproximar políticas e funções de valor em ambientes com grandes espaços de estados e ações.

Tabela 3: Tipos de Algoritmos de Reforço (Sutton, 2018)

3.4 DEEP LEARNING

É um subconjunto de *Machine Learning* que recorre a utilização de redes neuronais para processar e analisar informações. Estas redes são compostas por nós computacionais que são colocados em camadas específicas de algoritmos de *Deep Learning*. Cada camada contém uma camada de entrada, uma camada de saída e uma camada oculta. A rede é alimentada com dados de treino que ajudam o algoritmo a aprender e a melhorar a sua precisão. Goodfellow et al., 2016

Estes algoritmos inspiram-se no funcionamento do cérebro humano e são utilizados para a análise de dados com uma estrutura lógica. Algoritmos como estes estão presentes em tarefas como o reconhecimento de imagens e de voz, deteção de objetos e o processamento de linguagem natural. Goodfellow et al., 2016

Seguidamente, são brevemente descritos algumas das variantes mais utilizadas no *Deep Learning*.

3.4.1 *FeedFoward*

O *Feedfoward* é um dos tipos mais simples de redes neuronais. Nesta rede, a informação move-se apenas numa direção, para a frente, a partir dos nós de entrada, através dos nós ocultos (se existirem) e para os nós de saída. Não há ciclos ou *loops*

na rede. As redes neuronais *feedforward* foram o primeiro tipo de rede neuronal artificial inventado e são mais simples do que as suas congéneres, como as redes neuronais recorrentes e as redes neuronais convolucionais. Goodfellow et al., 2016

A tabela 4 apresenta alguns dos principais casos de uso onde as redes *feedforward* podem ser utilizadas:

Aplicação	**Descrição**
Classificação de Dados Estruturados	Categorizar dados estruturados em classes específicas.
Regressão	Prever valores contínuos, como previsão de preços de imóveis ou demanda de produtos.
Reconhecimento de Dígitos Escritos à Mão	Tarefas simples de reconhecimento de padrões, como a classificação de dígitos da base de dados MNIST.
Sistemas de Recomendação	Recomendações de produtos ou conteúdo com base no histórico e preferências dos consumidores.
Modelos de Predição Financeira	Prever o comportamento de variáveis económicas, como preços de ações ou risco financeiro.

Tabela 4: Tipos de Usos Feedfoward (Goodfellow et al., 2016)

3.4.2 *Recorrentes (RNN)*

Este tipo de redes, cuja designação anglo-saxónica é Recurrent Neural Network (RNN) difere das redes *Feed Foward* na medida em que utilizam dados de séries temporais ou dados que envolvam sequências, tendo "memória"do que aconteceu na camada anterior como contingente para a saída da camada atual. Goodfellow et al., 2016

3.4.3 *Convolucionais (CNN)*

A designação rede convolucional inclui algumas das redes neuronais mais comuns na inteligência artificial moderna e utilizam várias camadas distintas que filtram diferentes partes de uma imagem antes de a voltarem a juntar. O nome destas redes deriva do facto de recorrer a operações de convolução. As Redes Neurais Convolucionais (CNN) apresentam resultados de elevada qualidade na deteção e identificação de aspetos em imagem/vídeo. Goodfellow et al., 2016

3.5 REDES NEURONAIS RECORRENTES & REDES NEURONAIS CONVOLUCIONAIS

3.5.1 *Redes Neuronais Convolucionais (CNN)*

Dado a relevância das CNN para este trabalho, seguidamente, são aprofundados os principais conceitos associados a CNN.

Conceito

Os computadores interpretam as imagens como conjunto de pixéis. Cada *pixel* representa um ponto da imagem, podendo ter, dependendo da codificação da imagem, um ou múltiplos valores. A imagem é ainda definida pela largura e altura.

As *Convolutional Neural Networks (CNN)* dão sentido a estes dados através de um mecanismo chamado filtros, ou seja, pequenas matrizes de pesos ajustadas para detetar características específicas numa imagem, como cores, arestas ou texturas. *What are Convolutional Neural Networks? | IBM* 2024

Funcionamento Geral

Nas primeiras camadas, também conhecidas como camadas convolucionais, um filtro é aplicado sobre a entrada, procurando correspondências entre a entrada e o padrão do filtro (ver Fig.1). Este processo resulta então numa nova matriz que indica as áreas onde a característica de interesse foi detetada, conhecida como mapa de características. Goodfellow et al., 2016

Na fase seguinte conhecida como a camada de *pooling*, estes mapas de características são reduzidos utilizando um filtro que identifica o valor máximo ou médio em várias regiões da imagem (ver Fig. 1). A redução dos mapas de características diminui consideravelmente o tamanho das representações de dados, tornando a rede neuronal muito mais rápida. Goodfellow et al., 2016

Por fim, as informações resultantes são inseridas na camada totalmente conectada da CNN (ver Fig.1), fazendo assim com que essa camada tenha em conta todas as características extraídas nas camadas anteriores permitindo assim que o modelo categorize novas imagens de entrada em várias classes. Goodfellow et al., 2016

Resumindo, a série de filtros constrói efetivamente uma rede que compreende cada vez mais a imagem a cada camada que passa, sendo que os filtros nas camadas iniciais detetam apenas características de baixo nível como arestas e consequentemente nas camadas mais profundas estes começam a reconhecer padrões mais complexos, como formas e texturas. Goodfellow et al., 2016

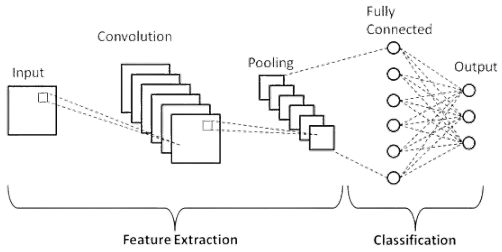


Figura 1: Arquitetura CNN Medium, 2022

Arquitetura

1. Camada de Entrada: A camada de entrada define as dimensões da imagem que a rede irá processar. Por exemplo, se estivermos a lidar com imagens a cores de 100x100 pixels e considerando três valores por codificação Red/Green/Blue (RGB), a camada de entrada seria configurada como 100x100x3, indicando uma largura de 100 pixels, uma altura de 100 pixels e três canais de cor (vermelho, verde e azul).

2. Camadas Convolucionais: As camadas convolucionais são responsáveis pela aprendizagem de padrões e características locais da imagem. Cada neurónio da camada convolucional está ligado apenas a uma pequena região local da entrada, chamada de campo recetivo. Durante o treino, a rede aprende filtros que são convoluídos com a entrada para detetar características como bordas, texturas e formas.

3. Camadas de Pooling (ou Subamostragem): As camadas de *pooling* são usadas para reduzir a dimensionalidade espacial da representação da imagem, mantendo as características-chave. O *pooling* é geralmente feito por meio de operações como o *max pooling*, onde apenas o valor máximo de uma região é mantido, ajudando a preservar as características mais proeminentes.

4. Camadas Totalmente Conectadas: Após as camadas convolucionais e de pooling, as camadas totalmente conectadas são responsáveis por combinar as características aprendidas numa representação mais global. Cada neurónio de uma camada totalmente conectada está ligado a todos os neurónios da camada anterior, preparando a rede para a tarefa específica, como classificação.

5. Camada de Saída: A camada de saída produz as previsões finais da rede. A escolha da função de ativação[1] depende do tipo de tarefa.

1 A função de ativação é um componente crucial em redes neuronais, responsável por introduzir não-linearidade ao sistema. (Exemplos: Sigmoid, ReLU, Tahn e Softmax)

3.5.2 *Redes Neuronais Recorrentes (RNN)*

Conceito

Como visto anteriormente, através da explicação do funcionamento das CNN's estas são apropriadas para reconhecer objetos, animais e pessoas, mas e se quisermos compreender o que está a acontecer numa imagem?

Se considerarmos uma imagem de uma bola no ar para determinar se a mesma está a subir ou a descer seria necessário mais contexto do que uma única imagem, ou seja, um vídeo cuja sequência pudesse esclarecer o que está a acontecer. Isto, por sua vez, exigiria que a rede neuronal se "lembrasse"das informações encontradas anteriormente e as considerasse em cálculos futuros.

As RNN's foram então concebidas para resolver exatamente o problema descrito acima, sendo que estas são capazes de processar dados sequenciais, como texto ou vídeo, utilizando ciclos que podem recordar e detetar padrões nessas sequências. As unidades que contêm esses ciclos de feedback são chamadas células recorrentes e permitem que a rede retenha informações ao longo do tempo. *What Is a Recurrent Neural Network (RNN)? | IBM* 2024

Funcionamento Geral

Quando é recebido um *input*, as células recorrentes combinam os novos dados com a informação recebida em passos anteriores. Utilizando esse *input* previamente recebido, estas atualizam os seus estados internos em resposta à nova entrada, permitindo a RNN identificar relações e padrões.

Arquitetura

A RNN recebe um vetor de entrada X e a rede gera um vetor de saída y, percorrendo os dados sequencialmente da esquerda para a direita, com cada passo temporal a atualizar o estado oculto e a produzir uma saída. Partilha os mesmos parâmetros em todos os passos temporais, conforme ilustrado na Figura 2. Isto significa que o mesmo conjunto de parâmetros, representados por U, V e W, é utilizado de forma consistente em toda a rede. U representa o parâmetro de peso que rege a ligação da camada de entrada X à camada oculta h, W representa o peso associado à ligação entre camadas ocultas e V a ligação da camada oculta h à camada de saída y. Esta partilha de parâmetros permite à RNN captar eficazmente as dependências temporais e processar dados sequenciais de forma mais eficiente, retendo a informação da entrada anterior no seu estado oculto atual.

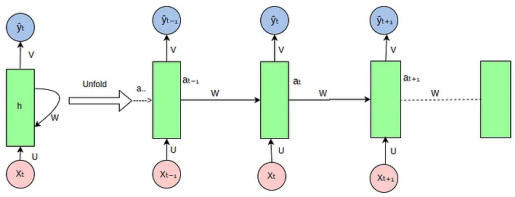


Figura 2: Arquitetura RNN Medium, 2023

3.5.3 *Comparações*

Seguidamente, apresenta-se uma breve comparação entre redes CNN e redes RNN.

- **Arquitetura**:
  - **Redes Neuronais Convolucionais (CNN)**: Utilizam camadas de convolução e *pooling* para processar e extrair características de dados espaciais, como imagens.
  - **Redes Neuronais Recorrentes (RNN)**: São redes que alimentam os resultados de uma iteração anterior de volta para a rede, permitindo a modelagem de dependências temporais em dados sequenciais, como texto ou vídeo.
- **Entrada/Saída**:
  - **Redes Neuronais Convolucionais (CNN)**: Tamanho de entrada e saída fixo, ou seja, recebem imagens de tamanho fixo e classificam-nas em categorias apropriadas com níveis de confiança.
  - **Redes Neuronais Recorrentes (RNN)**: Tamanho de entrada e saída variável, podendo receber diferentes comprimentos de texto e produzir traduções correspondentes, que podem ter mais ou menos palavras.
- **Uso Ideal**:
  - **Redes Neuronais Convolucionais (CNN)**: Ideal para dados espaciais, como imagens, onde a estrutura espacial é fundamental para a análise de reconhecimento de imagens e da classificação e deteção de objetos.
  - **Redes Neuronais Recorrentes (RNN)**: Mais adequadas para dados temporais ou sequenciais, como texto ou vídeo, onde a ordem e a dependência temporal são importantes, como na tradução de texto, na análise de sentimentos e na criação de texto
- **Cenário & Casos de Uso**:

- **Redes Neuronais Convolucionais (CNN)**: Útil em cenários onde a estrutura espacial é crucial, como no reconhecimento facial, análise médica, descoberta de drogas e análise de imagens em geral.
- **Redes Neuronais Recorrentes (RNN)**: Aplicável em cenários que envolvem dados sequenciais, como o processamento de linguagem natural, tradução de idiomas, análise de diálogos e análise de sentimentos em redes sociais.

A tabela 5 apresenta alguns casos de uso de redes CNN e RNN.

Cenário	Casos de Uso
CNN	**Reconhecimento facial**: Utilizado em sistemas de autenticação facial, como o Face ID da Apple. Taigman, 2014
	Diagnóstico médico por imagem: Usado na análise de raios-x para detetar doenças como pneumonia ou cancro. Rajpurkar, 2017
	Descoberta de medicamentos: CNNs são usadas na previsão de interações proteína-ligante. Wallach et al., 2015
RNN	**Assistentes de voz**: Usado por assistentes como a Siri e Alexa para entender comandos de voz. Graves et al., 2013
	Tradução automática: RNNs alimentam sistemas de tradução, como o Google Translate. Sutskever et al., 2014
	Análise de sentimentos em redes sociais: Usado para classificar emoções em textos do Twitter ou Facebook. Tang e Liu, 2015

Tabela 5: Casos de uso redes neuronais

3.6 PROVA DE CONCEITO

3.6.1 *YOLOv5*

Conceito

YOLO, abreviatura de You Only Look Once, é uma abordagem inovadora à deteção de objetos utilizando uma única rede neuronal de ponta a ponta. Jiang et al., 2022 Esta rede neuronal prevê simultaneamente caixas delimitadoras e probabilidades de classe, o que a distingue dos métodos anteriores de deteção de objetos que se baseavam em classificadores reutilizados. A metodologia distinta do YOLO conduziu a resultados notáveis, ultrapassando outros algoritmos de deteção de objetos em tempo real por uma margem significativa. Redmon et al., 2016

Enquanto os algoritmos tradicionais, como o Faster RCNN, seguem um processo de várias etapas envolvendo redes de proposta de região e estágios de reconhecimento separados, o YOLO realiza todas as previsões através de uma única camada totalmente conectada. Ao contrário dos métodos que utilizam redes de proposta de região que requerem várias iterações na mesma imagem, o YOLO obtém os seus resultados numa única passagem, limitando dessa forma as necessidades computacionais. O nome You Only Look Once (YOLO) visa precisamente destacar que apenas é efetuada uma passagem por imagem.

Funcionamento

A ideia básica do YOLO é a divisão de uma imagem intitulada de **imagem de entrada** numa grelha de células e, para cada célula, prever a probabilidade da presença de um objeto e as coordenadas da caixa delimitadora do mesmo.

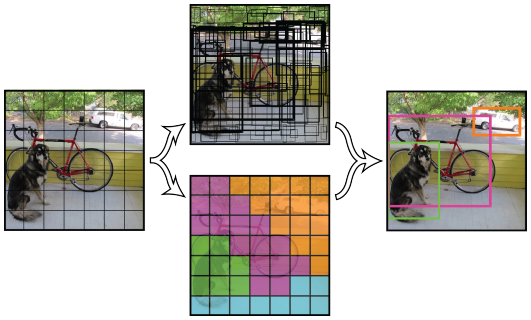


Figura 3: Caixa Delimitadora YOLOV5 Medium, 2023

Este processo explicitado acima e representado pela Figura 3 pode ser dividido nas seguintes etapas:

1. **Imagem de entrada** é passada através de uma CNN para proceder à extração das suas características.
2. As características extraídas são passadas por uma série de camadas totalmente ligadas, que preveem as probabilidades de classe e as coordenadas da caixa delimitadora.
3. Divisão da **imagem** em uma grelha de células sendo que cada célula será responsável pela previsão de um conjunto de caixas delimitadoras e probabilidades de classe, para possa ser obtido um conjunto de caixas delimitadoras e probabilidades de classe.
4. Filtração das caixas delimitadores usando um algoritmo de pós-processamento chamado supressão não-máxima (*NMS*) para remover as caixas sobrepostas e escolher a caixa com maior probabilidade.

5. Apresentação de um conjunto de caixas delimitadoras previstas e etiquetas de classe para cada objeto na imagem.

Caso De Uso

Uma aplicação prática do YOLOv5 pode passar pelo uso de sistemas de vigilância para detetar armas em vídeos captados por CCTV em tempo real. Esta aplicação poder ser utilizada em áreas de alto risco, como aeroportos, centros comerciais, eventos públicos e mesmo escolas, onde a identificação precoce de uma arma pode evitar ataques violentos. O YOLOv5 pode ser escolhido devido a sua eficiência, pois consegue processar rapidamente as imagens e identificar objetos com alta precisão, mesmo em ambientes complexos e com muitos detalhes.

Este caso de uso pode ser observado no seguinte estudo Al Amin e Paul, 2024 onde o YOLOv5 foi utilizado para identificar armas de fogo em tempo real em sistemas de CCTV. Este estudo concentrou-se na melhoria da precisão e a velocidade da deteção do modelo, podendo demonstrar uma aplicação fundamental para segurança pública em locais como escolas e aeroportos, onde a resposta rápida a possíveis ameaças é crucial.

3.6.2 *Reconhecimento Facial*

3.6.2.1 *Conceito*

Face Recognition é uma biblioteca Python que fornece ferramentas para deteção facial, reconhecimento e tarefas relacionadas. É construída sobre a dlib, que é uma biblioteca C++ conhecida pela sua eficiência em tarefas de visão computacional Boyko et al., 2018. A biblioteca Face Recognition facilita o trabalho com algoritmos de reconhecimento facial, fornecendo ainda uma interface Python de alto nível.

3.6.2.2 *Funcionamento*

A biblioteca dlib constitui a espinha dorsal do reconhecimento facial, esta possui funcionalidades que permitem a implementação de vários algoritmos baseados na aprendizagem automática. A biblioteca, desenvolvida em C++, utiliza modelos pré-treinados para efetuar a deteção de rostos, a deteção de marcas faciais e o reconhecimento de rostos. Estes modelos foram treinados em grandes conjuntos de dados para garantir taxas de precisão elevadas em todas as situações em que podem ser aplicados com algum nível de generalização em diferentes cenários.

A interface Python para o reconhecimento facial fornece uma API Python que oculta os aspetos técnicos de lidar com código C++ em níveis inferiores. Isto permite que os seus utilizadores encontrem uma forma fácil de infundir as suas aplicações Python com capacidades de reconhecimento facial.

A simplicidade, eficiência e facilidade de utilização são o foco da arquitetura do *Face Recognition* tornando-se assim uma das escolhas preferidas para uma simples integração de reconhecimento facial num projeto. No entanto, uma limitação identificada nos testes efetuados através desta biblioteca é que os modelos disponíveis podem apresentar desempenho inferior em indivíduos de diferentes raças e etnias, devido às semelhanças entre certos grupos, o que pode reduzir a precisão do reconhecimento facial em certos casos específicos.

3.7 ESTADO DE ARTE IA

Atualmente, a inteligência artificial está a fazer progressos significativos e está a transformar sectores como a saúde Davenport e Kalakota, 2019, os transportes, o entretenimento e media, entre outros.

Olhando para o setor da saúde como exemplo, podemos identificar os seguintes tipos de IA a operar *machine learning*, isto é, redes neuronais, deeplearning e Natural language processing.

1. **Machine Learning**: Aplicado na previsão do sucesso de um tratamento com base nos atributos do doente e no contexto do tratamento. Davenport e Kalakota, 2019
2. **Redes Neuronais & Deep Learning**: Aplicadas para tarefas como a previsão do aparecimento de doenças e a análise de imagens imagiológicas. Talaei Khoei et al., 2023
3. **Natural Language Processing**: Aplicado em tarefas como a análise de notas clínicas, preparação de relatórios, transcrição de interações com doentes e a condução de IA conversacional. Rajkomar et al., 2019

A IA apresenta um crescimento ágil, estando a acelerar o desenvolvimento de uma vasta gama de aplicações que serão capazes de revolucionar vários setores e atividades diárias. Contudo, por outro lado, é também importante reconhecer os seus desafios como os dilemas éticos para garantir a exploração desta tecnologia promovendo sempre o cumprimento dos valores morais básicos (Cath, 2018), e

os elevados recursos computacionais que muitos algoritmos requerem (Menghani, 2023).

4

ANÁLISE FORENSE DIGITAL ASSISTIDA POR IA

4.1 INTRODUÇÃO

Como visto nos capítulos anteriores, a análise forense digital envolve a recolha e análise meticulosa de provas digitais de vários dispositivos eletrónicos. Estas são cruciais para descobrir a verdade por detrás de atividades criminosas, fraudes, violações de segurança e disputas legais.

Do mesmo modo, explorámos os fundamentos da Inteligência Artificial (IA), que inclui o desenvolvimento de máquinas e algoritmos capazes de simular a capacidade humana de resolução de problemas e de tomada de decisões. Com base nestes fundamentos, este capítulo irá aprofundar a forma como as tecnologias de IA estão a transformar a análise forense digital.

A integração da IA na investigação forense digital marca um avanço significativo neste domínio aumentando a eficiência e a precisão das investigações forenses através da análise avançada de dados, do reconhecimento de padrões e da automatização de tarefas de trabalho intensivo. Este capítulo irá explorar as várias aplicações da IA na análise forense digital, os benefícios que oferece, os desafios que apresenta e as perspetivas futuras desta integração dinâmica.

4.2 APLICAÇÕES DE IA EXISTENTES

4.2.1 *Automação de Análise de Dados*

É possível utilizar a inteligência artificial para realizar o processamento eficiente e preciso de grandes quantidades de dados. O Nuix Neo[1], por exemplo, contextualiza os dados e os modelos linguísticos especificamente treinados para o seu caso de utilização com foco apenas no que precisa, melhorando a precisão e reduzindo a

1 Plataforma avançada da Nuix, projetada para lidar com grandes volumes de dados não estruturados.

quantidade de dados que precisa de rever e analisar. *Nuix Neo: uma plataforma para todos os seus desafios de dados complexos | Nuix* 2024

4.2.2 *Processamento de linguagem natural (PLN)*

O Processamento de Linguagem Natural (PNL) é fundamental na computação cognitiva, um domínio da IA que permite aos computadores recolher, analisar e interpretar dados. O PNL centra-se na compreensão e criação de linguagem escrita e falada, com base em tarefas como a criptoanálise e a tradução automática. Combina linguística computacional, aprendizagem automática e aprendizagem profunda para processar informação semelhante à cognição humana. *Nuix Neo: uma plataforma para todos os seus desafios de dados complexos | Nuix* 2024

No PNL, a linguagem é analisada em pensamentos ou ideias, ligados para estabelecer o contexto, ajudando a interpretar a intenção e o estado de espírito. Esta capacidade é valiosa na investigação forense digital para identificar rapidamente informações pertinentes. O PNL envolve o pré-processamento de dados e o desenvolvimento de algoritmos sendo o texto limpo e formatado para interpretação automática. Através do uso de várias técnicas, incluindo algoritmos alicerçados em regras, como analisadores sintáticos baseados em gramáticas (*Context-Free Grammar*) e expressões regulares, que utilizam estruturas da linguagem humana, é possível obter resultados precisos. Além disso, técnicas como o *stemming* e a lematização ajudam a processar o texto de forma mais eficiente, garantindo uma interpretação exata. Al et al., 2023

Um dos exemplos para este tipo de aplicações PLN é o API Cloud Natural Language desenvolvido pela Google que oferece aos desenvolvedores tecnologias de processamento de linguagem natural, como análise de sentimento, reconhecimento de entidades, análise de sentimento de entidades e outras anotações de texto. *API Cloud Natural Language | Cloud Natural Language API | Google Cloud* 2021

4.2.3 *Análise de imagem e vídeo*

Devido à popularidade dos dispositivos móveis inteligentes e ao baixo custo dos sistemas de vigilância, os dados visuais estão cada vez mais presentes no dia-a-dia e consequentemente utilizados nas investigações forenses digitais. Os vídeos digitais têm sido amplamente utilizados como fontes de provas fundamentais na identificação,

análise, apresentação e relatório de provas. É importante concentrarmos-nos no desenvolvimento de técnicas avançadas de análise de vídeo forense para ajudar a investigação forense. Al et al., 2023 Um exemplo proeminente desta tecnologia é a Clearview AI, uma plataforma revolucionária, que afirma possuir a maior base de dados conhecida com mais de 50 mil milhões de imagens faciais provenientes de fontes exclusivamente públicas da web e que pode ser usada pelas agências governamentais para ajudar a gerar pistas de investigação. Dul, 2022 Note-se, contudo, que o uso do reconhecimento facial tem levantado sérios problemas de ética e de precisão. Hill, 2021

4.2.4 *Reconhecimento de Padrões*

O reconhecimento de padrões é essencial para identificar e classificar tipos de dados em investigações. Através da utilização de classificadores é possível equilibrar a generalização e a especificidade para reconhecer com precisão tanto exemplos conhecidos como novos. O objetivo deste processo é minimizar os erros e, ao mesmo tempo, fazer corresponder eficazmente os padrões, muitas vezes com a ajuda de técnicas de aprendizagem automática. Al et al., 2023

Um dos exemplos destas aplicações é a DarkTrace, que utiliza inteligência artificial mais em específico *machine learning* para oferecer uma abordagem proativa à ciber-resiliência. Esta aplicação oferece uma plataforma de cibersegurança baseada em algoritmos de deteção de anomalias e análise comportamental para monitorar redes em tempo real *ActiveAI Security Platform | Darktrace* 2024. Com isso, a DarkTrace pode identificar atividades suspeitas e potenciais ameaças, mesmo aquelas que não foram previamente catalogadas. A resposta autónoma é possível através do seu sistema de 'Antigena', que reage automaticamente a ataques, neutralizando ameaças com base no contexto e gravidade sem intervenção humana, garantindo uma defesa eficaz contra ataques tanto conhecidos quanto desconhecidos *ActiveAI Security Platform | Darktrace* 2024.

4.3 TRABALHO RELACIONADO

Nesta secção, apresentam-se alguns dos trabalhos realizados por investigadores nos últimos anos, que se enquadram no âmbito do trabalho desenvolvido, nomeadamente a deteção de armas em vídeos e o reconhecimento facial. Primeiramente,

é apresentada uma tabela que enumera os trabalhos selecionados, seguida de um resumo dos estudos que constam na Tabela 6.

Trabalho	Tópico
Santos, 2024	Real-Time Weapon Detection in Surveillance Footage
Warsi et al., 2020	Automatic Handgun and Knife Detection Algorithms: A Review
Bhagyalakshmi. et al., 2019	Detection and Classification of Different Weapon Types Using Deep Learning
Khan et al., 2019	Face Detection and Recognition Using OpenCV

Tabela 6: Trabalho Relacionado

Santos, 2024 desenvolveu um sistema de deteção de armas em tempo real baseado na estrutura YOLOv7, que tinha como objetivo detetar a localização de uma arma de fogo em vídeos de vigilância. Esta abordagem recorreu a algoritmos genéticos para otimizar os hiperparâmetros, alcançando um melhor desempenho do que os modelos tradicionais, como o Faster R-CNN. Especialmente quando operando em condições adversas, como pouca luz ou armas de fogo pequenas. A abordagem proposta também pode ser integrado numa infraestrutura de vigilância já existente, maximizando a segurança do público através da deteção precoce e minimizando os custos operacionais.

O trabalho de Warsi et al., 2020 propõem um sistema multi-módulo para vigilância CCTV centrado na monitorização em tempo real e na deteção de eventos anormais. O sistema utiliza CNNs para deteção de objetos, com outros módulos para classificação de armas e disparo de alarmes. Com base no algoritmo de deteção de formas e em modelos pré-treinados como o ALEXNET Bangar, 2022, é alcançada uma elevada precisão de 89% mAP na identificação de armas de forma positiva tendo como objetivo a redução das taxas de criminalidade.

Bhagyalakshmi. et al., 2019 propõem um sistema de aprendizagem profunda para classificar sete classes de armas, incluindo espingardas de assalto e facas, utilizando a arquitetura VGGNet implementada com a biblioteca Keras em execução no TensorFlow. Atingiu uma taxa de precisão de 98,40%, superando os modelos de aprendizagem profunda mais avançados, como o VGG-16 e o ResNet-50. Um resultado deste tipo revela o potencial dos modelos adaptados em termos de melhoria da fiabilidade dos sistemas de deteção de armas.

Em conclusão, o trabalho Khan et al., 2019 enfatiza a importância da segurança biométrica através da criação de um sistema de reconhecimento facial em tempo real que emprega a Análise de Componentes Principais (PCA). Utilizando OpenCV e Python, o sistema a desenvolver tem como objetivo diminuir efetivamente a dimensionalidade dos dados para aumentar a precisão do reconhecimento facial,

proporcionando assim aplicações importantes na segurança para uma identificação célere.

Em conjunto, estes estudos demonstram avanços notáveis no desenvolvimento de tecnologias de vigilância automatizadas e sublinham a eficácia da aprendizagem profunda no reforço da segurança pública, melhorando a capacidade de deteção de eventos e de resposta eficiente.

4.4 POSSÍVEIS VANTAGENS

De acordo com algumas das ideias demonstradas a partir da apresentação de algumas das aplicações de IA no ramo da análise forense digital, de seguida serão apresentadas algumas possíveis vantagens que estas aplicações podem aportar:

1. **Capacidade de processamento**: A capacidade da IA para navegar em grandes volumes de dados pode melhorar a análise e fornecer uma melhor perceção desses dados. As capacidades analíticas da IA residem na identificação não só de padrões, anomalias e tendências previamente mencionados, mas também de correlações subtis e irregularidades matizadas que a análise manual tradicional poderia ignorar. Oxygen Forensics, 2024
2. **Automatização e aprendizagem**: A automatização e a aprendizagem da IA poderão simplificar tarefas morosas, como a recolha de dados, a análise e a criação de relatórios, podendo reduzir o tempo que os investigadores dedicam às tarefas de pesquisa. Ao utilizar a IA para realizar processos repetitivos e mundanos, os investigadores podem dedicar mais tempo ao pensamento crítico e aos aspetos de resolução de problemas de cada caso. Oxygen Forensics, 2024
3. **Simplificação de processos**: Os processos de pesquisa e análise simplificados encerram as investigações mais rapidamente, reduzindo os atrasos e os orçamentos. As capacidades automatizadas da IA podem ajudar as equipas de investigação a compensar os desafios de pessoal associados a escassez de recursos humanos ou as restrições orçamentais e de outros recursos. Oxygen Forensics, 2024

4.5 POTENCIAIS DESAFIOS

Como verificamos acima, esta integração aparenta ser muito promissora através das aplicações demonstradas e das potenciais vantagens identificadas. No entanto, é importante reconhecer que o desenvolvimento destas tecnologias dará origem a novas oportunidades e desafios que exigem uma análise cuidadosa. Neste contexto, identificam-se de seguida dois potenciais desafios que esta integração poderá trazer, com base num estudo realizado por Jarrett e Choo, 2021

1. **Fator Humano**: Presentemente a tecnologia baseada na IA só serve como ferramenta para facilitar as investigações que continuam a exigir a supervisão de investigadores humanos especializados. A exatidão do resultado forense depende, em certa medida, das capacidades do analista humano, uma vez que as ferramentas de AI ainda estão em desenvolvimento e podem nem sempre produzir informações exatas, completas ou robustas necessárias para os casos forenses Jarrett e Choo, 2021. Para ultrapassar este desafio, é necessário que o investigador receba uma formação específica e desenvolva as suas competências. Num cenário possível, os profissionais inexperientes agem com base em informações insuficientes devido à sua total dependência dos sistemas automatizados, aumentando a probabilidade de investigações falhadas James e Gladyshev, 2013. Além disso, a presença de investigadores inexperientes está frequentemente associada à falta de certificações e de controlos de qualidade do equipamento de laboratório envolvido na recolha de dados forenses James e Gladyshev, 2013. Adicionalmente, muitas investigações forenses digitais são adjudicadas a terceiros, que não são certificados nessa área James e Gladyshev, 2013. Outra dificuldade é a ausência de instituições de certificação e de organismos reguladores que garantam que apenas os investigadores forenses digitais certificados possam operar nesta área.

2. **Múltiplos Formatos**: Outro obstáculo com que se depara a investigação forense digital é a utilização de múltiplos e complicados formatos de suportes que podem revelar-se difíceis de adquirir ou analisar pelos atuais sistemas de IA. Por exemplo, podem ser utilizados formatos de esteganografia, encriptação e anti-forense. Os investigadores argumentam ainda que as diferentes fontes de provas digitais podem também revelar-se um obstáculo notável no caminho das investigações forenses digitais através da AI Oriwoh et al., 2013. A adoção e utilização generalizadas de dispositivos da Internet das Coisas (IoT) diversificam ainda mais os desafios deste paradigma específico, sendo

que alargam a disseminação e o fluxo de dados, conduzindo a uma premissa menos privada e menos segura para os utilizadores finais Oriwoh et al., 2013. O volume cada vez maior de dados e informações apreendidos e sujeitos a análise forense digital é um desafio que deve ser tratado por tecnologias de AI novas e melhoradas Jarrett e Choo, 2021. Talvez um lado bastante excitante, mas assustador, da aplicabilidade das tecnologias baseadas em AI à forense digital tenha sido explorado recentemente por Spencer Spencer, 2018. No seu estudo, foram discutidas as mentalidades complexas e imprevisíveis dos criminosos e formulada a teoria de que a automatização completa da perícia digital é potencialmente impossível Spencer, 2018. De facto, muitos casos criminais não obedecem a um padrão ou a tendências históricas Spencer, 2018. Além disso, a evolução das tecnologias e das técnicas abre caminho para que os criminosos adotem métodos novos e melhorados para cometer crimes, podendo eles mesmo recorrer também à IA para o efeito. Spencer, 2018.

4.6 PREVISÕES FUTURAS

O futuro da IA no ramo da análise forense digital, na minha perspetiva, demonstra ser um ponto de viragem, visto que oferecerá melhorias significativas em várias áreas fundamentais. Seguidamente, listam-se alguns desenvolvimentos que poderão vir a surgir no futuro baseados em alguns estudos feitos nesta área:

1. **Deteção avançada de *malware***: Os modelos de *deeplearning* podem detetar *malware* através da aprendizagem das características do código malicioso, tornando-os eficazes contra ameaças novas e desconhecidas Joshua Saxe, 2023
2. **Medidas reforçadas de cibersegurança**: Os sistemas de cibersegurança baseados em IA podem analisar o tráfego de rede e o comportamento dos utilizadores para detetar potenciais violações e anomalias, melhorando assim a postura de segurança e ajudando nas investigações forenses. Apruzzese et al., 2023
3. **Melhoria da análise de dados e do reconhecimento de padrões**: Os algoritmos de *machine learning* podem processar rapidamente grandes volumes de provas digitais, identificando padrões e correlações que poderiam passar despercebidos aos investigadores humanos. Este facto aumenta a capacidade da investigação forense digital para resolver casos complexos. Lillis et al., 2016

4. **Reconhecimento facial e análise de imagens**: Os recentes avanços na tecnologia de reconhecimento facial com recurso à IA são cruciais para a análise forense, permitindo a identificação de indivíduos em imagens e vídeos com elevada precisão aumentando assim a fidelidade dos resultados obtidos. Nguyen et al., 2019

A aplicação da IA no ramo da análise forense digital apresenta um elevado potencial de inovação, podendo reduzir consideravelmente todos os processos morosos envolvidos numa investigação. Porém, este tipo de aplicações não invalida a supervisão humano, sendo que pequenos detalhes podem ser desconsiderados por estes tipos de algoritmos, para além dos erros induzidos pela possibilidade de viés por estes tipos de algoritmos.

5

CASO DE ESTUDO: A APLICAÇÃO CYBERSYNC

Neste capítulo, será demonstrado um protótipo que integra tecnologias de IA para apoiar a análise de dados de vídeo de forma a produzir uma amostra do que poderá ser desenvolvido para auxiliar os investigadores forenses. A solução utiliza uma interface baseada na Web construída com a *framework* Vue.js que interage com dois *scripts* Python que alojam modelos de IA: o YOLOv5, para deteção de objetos, e o *Face Recognition* para reconhecimento facial de indivíduos. A demonstração ilustra a sinergia destes modelos de IA na decifração de dados visuais de vídeos e deteção de objetos ou pessoas que aparecem nas filmagens, para que seja possível obter de forma mais rápida e assertiva alguns dados que podem ser vitais para a resolução de algumas investigações forenses.

5.1 FERRAMENTAS

Seguidamente, são descritas as ferramentas e tecnologias utilizadas na construção do protótipo para análise de vídeo com integração de IA ao qual foi dado o nome de CyberSync. A escolha dessas ferramentas foi baseada em critérios de eficiência, compatibilidade, facilidade de uso e robustez para atender às necessidades específicas do projeto, para além de estarem disponíveis sob licença de código aberto.

1. **VueJS**: Framework baseada em JavaScript usada para construir interfaces de utilizador. É construído sobre HTML, CSS e JavaScript padrão e fornece um modelo de programação declarativo e baseado em componentes que o ajuda a desenvolver eficientemente interfaces de utilizador de qualquer complexidade. *Vue.js* 2024
2. **Flask**: Framework web que pode ser usada como um módulo Python que permite desenvolver facilmente aplicações web. *What is Flask Python - Python Tutorial* 2024
3. **YOLOv5**: Modelo popular de deteção de objetos e segmentação de imagens. Ultralytics, 2024a

4. **Face Recognition**: Módulo Python alicerçado no dlib que fornece uma interface de utilizador simples para tarefas de deteção e reconhecimento de rostos. *Face Recognition in Python - Javatpoint* 2024
5. **OpenCV**: A OpenCV (Open Source Computer Vision Library) é uma biblioteca de software disponibilizada em licença de código aberto para visão computacional e *machine learning*. A OpenCV foi criada para fornecer uma infraestrutura comum para aplicações de visão computacional e para acelerar a utilização da perceção automática nos produtos comerciais. *About* 2020
6. **Express.js**: Framework baseada em Node.js que se destaca por ser mínima e flexível. Fornece um conjunto de funcionalidades para aplicações web e móveis. *Express - Node.js web application framework* 2024

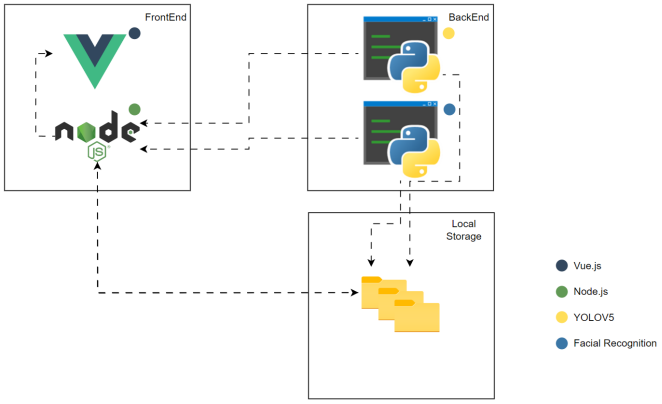
5.2 ARQUITETURA & FLUXO

5.2.1 *Arquitetura*

Esta secção apresenta a arquitetura usada para a incorporação de tecnologias de IA no apoio à análise de dados de vídeo, dando um exemplo do que pode ser utilizado para ajudar os profissionais forenses. Esta solução utiliza a interface Web desenvolvida em Vue.js e interage depois com dois scripts Python que alojam modelos de IA, YOLOv5 para deteção de objetos e FaceRecognition para reconhecimento de indivíduos.

A arquitetura da solução proposta consiste em muitas partes interligadas, sendo cada uma delas importante na captura e no processamento de informações visuais. A comunicação entre a interface Web e os scripts Python é mediada pelo servidor Web escrito Express.js baseado em Node.js.

Figura 4: Arquitetura



1. **Interface Web (Vue.JS)**:
    - Responsável pela interação com o utilizador, permitindo a criação de investigações e a visualização dos resultados obtidos nas mesmas.
2. **Servidor Web (Node.js com Express.js)**:
    - Gere os pedidos da interface web e coordena a execução dos scripts Python.
    - Fornece *endpoints* para enumerar ficheiros, disponibilizar imagens e vídeos, ler ficheiros JSON, e controlar a reprodução de vídeos no VLC media player.
    - Implementa a "segurança básica"para assegurar que apenas arquivos permitidos sejam acessados e processados.
3. **Script de Deteção de Objetos (YOLOv5)**:
    - Recorrendo à linguagem Python, utiliza o modelo YOLOv5 para identificar e classificar objetos em fotogramas de vídeo.
    - Retorna a localização e a identificação dos objetos detetados e guarda estes resultados através de um ficheiro JSON, para que depois possam ser processados pela interface web de forma a que o utilizador obtenha uma visão simplificada dos dados.
4. **Script de Reconhecimento Facial (FaceRecognition)**:
    - Faz uso da biblioteca *Face Recognition* para identificar indivíduos em fotogramas de vídeo.

- Compara os rostos detetados com uma base de dados desenvolvido, retornando a identificação e o grau de confiança das correspondências.
- Suporta o aprimoramento dos fotogramas para melhorar a precisão do reconhecimento.

Através da combinação de YOLOv5 para deteção de objetos e FaceRecognition para reconhecimento facial, a solução pretende oferecer uma análise mais rápida e precisa, auxiliando os investigadores forenses na obtenção de dados vitais para a resolução de investigações.

5.2.2 *Fluxo*

Procede-se agora à descrição detalhada do fluxo do sistema desenvolvido, a partir da descrição das etapas envolvidas no processamento de dados de vídeo utilizando as tecnologias de IA integradas. O objetivo é ilustrar como o sistema funciona desde o momento em que o utilizador cria a investigação até à obtenção de resultados.

1. **Criação de uma investigação**: O fluxo começa com a introdução de parâmetros, na interface Web. Antes do início da investigação, o utilizador deve fornecer parâmetros específicos, tais como o caminho para a pasta onde se localizam os vídeos, os critérios de seleção de fotogramas e os limiares de confiabilidade. Uma vez definidos estes parâmetros, inicia-se a fase de pré-processamento. A Figura 5 e a Figura 6 ilustram, respetivamente, a interface para a ativação em modo de reconhecimento facial, e modo de deteção de objetos.

Figura 5: Criação Investigação Facial

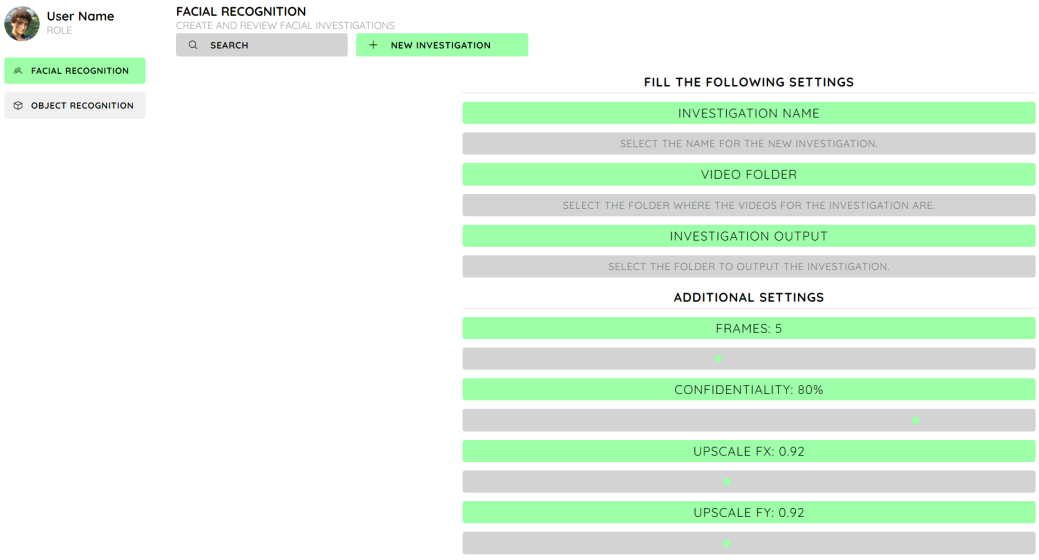
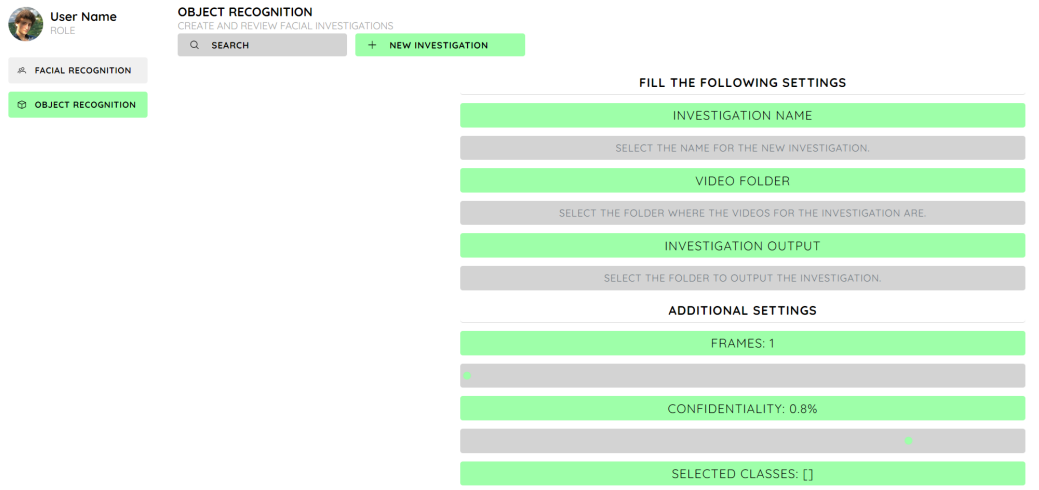


Figura 6: Criação Investigação Objetos



2. **Pré-Processamento**: O processo começa com a segmentação do vídeo, onde o vídeo é dividido em fotogramas consoante o intervalo de fotogramas pretendido. Após a segmentação, cada *frame* obtido no intervalo determinado é submetido a um processo de aprimoramento projetado para melhorar a visibilidade dos objetos ou rostos a detetar. Este melhoramento inclui o redimensionamento dos fotogramas, a aplicação da equalização e o aumento da nitidez das imagens. Estes ajustes garantem que os fotogramas estão nas condições ideais para as

tarefas de deteção e reconhecimento subsequentes, de forma a aumentar a precisão e a eficácia da análise.

3. **Reconhecimento**: Uma vez concluído o pré-processamento e dependendo da seleção do tipo de investigação, os fotogramas podem passar pelo modelo YOLOv5 para deteção de objetos ou pelo modelo do FaceRecognition para deteção de rostos.

    - **Reconhecimento de objetos**: Identificação de fotogramas para reconhecimento de objetos utilizando o modelo do YOLOV5.

        – **Processo de deteção**: O YOLOv5 analisa cada *frame* para identificar e rotular objetos de interesse, tais como veículos, armas e outros objetos relevantes. O modelo devolve as coordenadas destes objetos dentro do *frame*.

        – **Anotação**: Os objetos detetados são anotados num *array* que depois será posteriormente exportado em ficheiro com o formato JSON.

    - **Reconhecimento Facial**: Identificação de fotogramas para reconhecimento de rostos utilizando o modelo FaceRecognition.

        – **Deteção de rostos**: O sistema deteta rostos em cada fotograma. Estes rostos são então comparados com uma base de dados local de indivíduos conhecidos. O analista forense deve providenciar pelo menos uma foto de rosto do(s) indíviduo(s) que se pretende reconhecer no material em análise.

        – **Reconhecimento e pontuação de confiança**: Para cada rosto detetado, o sistema tenta compará-lo com rostos conhecidos, e é gerada uma pontuação de confiança, indicando a probabilidade de o rosto detetado corresponder a um indivíduo conhecido.

4. **Armazenamento de dados**: Após o reconhecimento facial ou de objetos, o sistema organiza e armazena os resultados.

    - **Reconhecimento Facial**: Para proceder ao armazenamento dos resultados do reconhecimento facial, é criada uma pasta dentro do caminho identificado que contém para cada deteção facial um ficheiro em formato PNG que explicita o limiar de confiança bem como a *bounding box* gerada pelo modelo do *facial recognition*.

- **Reconhecimento de Objetos**: É gerado um ficheiro JSON de resumo, que detalha os vídeos processados individualmente e todos os objetos detetados durante a investigação.

5. **Visualização de resultados e interação com o utilizador**: Através da identificação de um caminho para a pasta criada para o processo de investigação, os resultados são então apresentados ao utilizador através da interface Web.

    - **Visualização de resultados**: Os utilizadores podem ver os fotogramas anotados ou os detalhes dos objetos detetados. Esta interface tem por objetivo permitir ao utilizador uma visualização simplificada dos resultados gerados pelos *scripts*, conforme ilustrado pela Figura 7.
    - **Funcionalidades Adicionais**: O sistema oferece ferramentas como a navegação baseada em *timestamps* de data/hora e a capacidade de abrir fotogramas específicos em software externo como o VLC Media Player, entre outras funcionalidades (Figura 8).

Figura 7: Menu Principal - Reconhecimento Facial

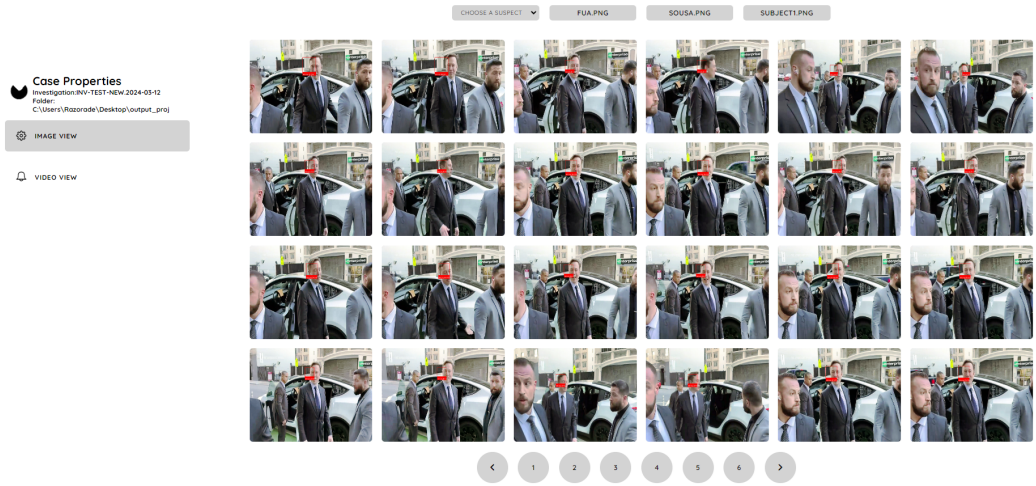


Figura 8: Vista Vídeo - Reconhecimento Facial

As Figuras 9, 10 e 11 são capturas de ecrã referentes à funcionalidade de deteção de objetos. Note-se que o módulo de estatísticas que indica o número de ocorrência de cada objeto por intervalo de tempo no vídeo. Tal permite ao utente da aplicação focar em secções do vídeo na qual apareçam objetos relevantes para o caso em análise.

Figura 9: Menu Principal - Reconhecimento de Objetos

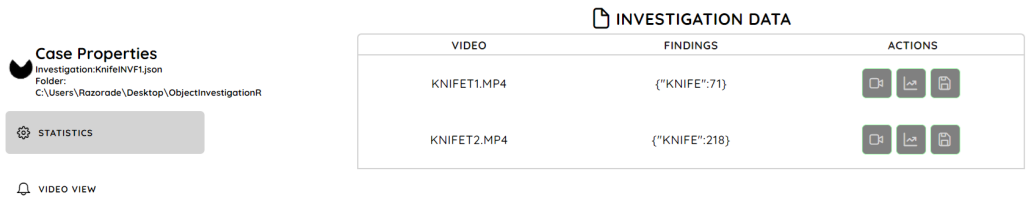


Figura 10: Vista Vídeo - Reconhecimento de Objetos

Figura 11: Vista Estatísticas - Reconhecimento de Objetos

Este fluxo de trabalho tenta demonstrar a natureza abrangente de um sistema que combina ferramentas de IA com interfaces de fácil utilização, de forma a agilizar a análise eficiente de dados de vídeo em investigações forenses. A abordagem adotada em cada etapa visa garantir que o sistema possa apoiar eficazmente os investigadores, através de informações precisas, facilmente acessíveis.

5.3 TREINO DE MODELOS PARA DETEÇÃO DE OBJETOS

Para que fosse possível proceder ao uso do YOLOV5 para a identificação de objetos na aplicação CyberSync foram delineados ao longo do tempo alguns processos de criação de *datasets*, bem como criados alguns modelos através dos mesmos para proceder à execução dos testes. Estes processos são descritos no anexo b onde será possível observar as seguintes secções:

1. **Criação de conjuntos de dados**: Na secção B.1 do anexo são discutidas as várias abordagens desenvolvidas para a criação dos conjuntos de dados necessários para o treino dos modelos. Como forma de resumo abaixo serão identificadas as abordagens utilizadas:
   a) Utilização de *datasets* disponíveis em sites como o Kaggle ou Roboflow
   b) Criação manual de um *dataset* e consequente anotação precisa dos objetos de interesse para a criação de classes específicas.
   c) Implementação de um script Python desenvolvido para proceder à junção de vários *datasets* existentes, de forma a que fosse possível a abrangência de um maior número de classes.
2. **Análise dos resultados do treino Yolov5** Na secção B.2 é possível observar a avaliação dos modelos criados, bem como a evolução das classes que foram delineadas ao longo do projeto. A precisão, a recuperação e a pontuação F1 foram as principais métricas investigadas para avaliar o desempenho do modelo em termos de eficácia da deteção. Em resumo, acabou-se por se proceder a uma redução do número de classes devido a ser necessário a construção de um *dataset* manual específico que pudesse superar os obstáculos mencionados nos desafios especificados no final do capítulo seguinte.

A consulta das secções do anexo é fundamental para apreciar as decisões que foram tomadas no desenvolvimento desta aplicação. É aqui que se descrevem as metodologias aplicadas e uma visão crítica dos resultados obtidos, lançando as bases

para as recomendações e ajustamentos necessários para afinar a aplicação de forma a que os modelos atuais possam ser estendidos para incorporarem um maior número de classes.

O próximo capítulo apresenta os resultados obtidos nos testes da aplicação CyberSync e descreve os desafios encontrados durante a criação da aplicação ou a realização de testes e possíveis melhorias que podem ser implementadas.

6

RESULTADOS EXPRIMENTAIS

Neste capítulo são discutidos os resultados obtidos, destacando as implicações, os desafios encontrados e as possíveis melhorias para uma versão futura de um sistema semelhante ao conceito desenvolvido. Durante os testes, tanto para o reconhecimento facial, como para a deteção de classe de objetos, foram avaliados dois exemplos representativos que são aqui apresentados. Note-se que foram ainda realizados mais testes para garantir a validade dos resultados.

6.1 AMBIENTE COMPUTACIONAL

Para a realização dos testes e consequente treino dos modelos, foi utilizado o ambiente computacional descrito na Tabela 7 e na Tabela 8. É importante salientar que a GPU empregue GeForce RTX 2060 é uma GPU de baixa/média gama com 1920 CUDA cores e 6 GiB de memória DDR6, e um preço aproximado de 250 euros no momento de redação desde documento. Tal limita necessariamente os resultados ao nível dos tempos de execução. O uso de uma GPU com maior capacidade computacional resultaria num melhor desempenho.

Componente	Especificação
Processador	Intel Core i5-9600KF 3.7 GHz (6 Nucleos)
GPU	EVGA GeForce RTX 2060 KO Ultra Gaming 6GB GDDR6
RAM	Kingston FURY Beast DDR4 3200 MHz 16GB 2x8GB CL16
Disco	SSD 2.5"Samsung 870 EVO 500GB

Tabela 7: Hardware

Software	Versão
Dlib	v19.24.4
Face Recognition	v1.3.0
Flask	v3.0.3
Node.js	v18.17.0
Python	v3.11.9
Tensorflow	v2.17.0
Torch	v2.4.0
Ultralytics	v5.10.0
Vue.js	v3.3.10
Windows 11 Pro	v10.0.22631
YOLO	v5

Tabela 8: Software

6.2 PARÂMETROS E MÉTRICAS

Para efeito de avaliação, foram definidos os parâmetros seguidamente definidos.

1. **Parâmetros de investigação**

   - **Intervalos de fotogramas**: A análise do vídeo é feita em intervalos de fotogramas, variando de 1 a 9, conforme selecionado. Isso significa que, por exemplo, se o valor de intervalo for 3, o sistema processa apenas um em cada três fotogramas (1,4,7,10,...).
   - **Fatores de aumento de escala (FX, FY)**: O sistema aplica ajustes de escala para melhorar a resolução das imagens. Esses valores variam de 1x1, 1.5X1.5 e 2x2, onde o fotograma original pode ser ampliado até duas vezes em altura e largura, usando interpolação cúbica para garantir maior clareza visual sem perder muita qualidade.
   - **Limiar de confiança**: O reconhecimento de faces só é considerada válido se a confiança na correspondência facial for superior a 70%. Esse limiar visa equilibrar precisão e evitar falsos positivo, de forma a proporcionar uma identificação mais confiável (reduzindo a probabilidade de erros) sem comprometer a usabilidade do sistema ou causar muitos falsos negativos.

- **Objetivo**: Avaliar o desempenho do sistema de reconhecimento facial em termos de deteções de verdadeiros positivos (TP) e falsos positivos (FP) sob diferentes configurações.

2. **Métricas** Para efeitos de avaliação de resultados são consideradas as seguintes métricas.

   a) **Resultados verdadeiros positivos**: O número de identificações positivas corretas de rostos.

   b) **Resultados falsos positivos**: O número de identificações positivas incorretas de rostos.

   c) **Pontuações de confiança**: Os níveis de confiança associados a cada deteção, variando de 70% (mínimo) até 100%.

6.3 TESTES DE RECONHECIMENTO FACIAL

Esta secção apresenta os resultados obtidos nos testes do modelo de reconhecimento facial, considerando as configurações e métricas especificadas. Procede-se ainda à análise desses mesmos resultados. Por fim, são tecidas algumas considerações sobre os principais resultados registados durante o processo. A Figura 12 apresenta os rostos empregues nas experiências de reconhecimento facial.

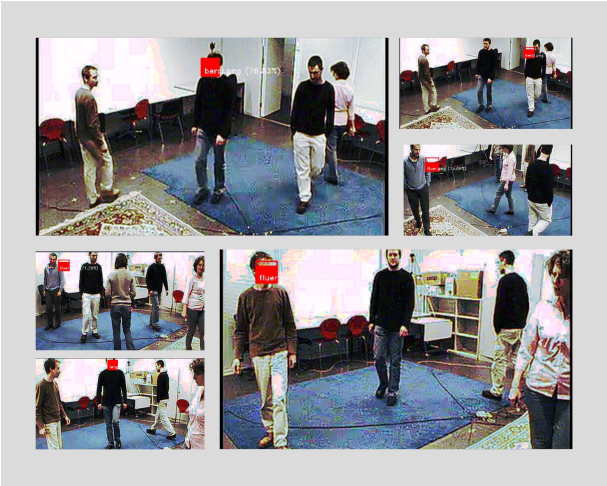
Figura 12: Dataset de Rostos

Vídeos de baixa qualidade

Os resultados deste teste foram obtidos a partir de várias configurações usadas no *script* de reconhecimento facial Python sobre uma compilação de vídeos de baixa qualidade. Os vídeos consistem em seis indivíduos que caminham em frente a uma câmara fixa e as experiências foram realizadas com diferentes intervalos de fotogramas, fatores de aumento de escala e um limiar de confiança fixo. Abaixo

serão apresentados os resultados obtidos durante estes testes em tabelas, sendo igualmente tecidos alguns comentários acerca das mesmas, bem como os metadados que comprovam de certa forma a 'baixa qualidade' em termos da resolução da imagem. Todos os quatro vídeos tem resolução de 360x288 com 0.104 megapíxeis (Tabela 9). Alguns fotogramas dos vídeos são mostrados na Figura 13, sendo observável a reduzida qualidade dos mesmos. As tabelas 10, 11 e 12 apresentam os resultados para o vídeo facial #1, variando-se os parâmetros de dimensão FX e FY.

Figura 13: FacialRecognition SetV1



Vídeos	Resolução	Megapixels	Fonte
Facial #1	360x288	0.104	Lab, 2006a
Facial #2	360x288	0.104	Lab, 2006b
Facial #3	360x288	0.104	Lab, 2006c
Facial #4	360x288	0.104	Lab, 2006d

Tabela 9: Metadados Vídeos (Rostos #1)

Nome Dataset	Facial #1 - FX 1 FY 1				
Duração(s)	472				
FPS	25 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	11660	3932	2360	1684	1308
Confiablidade (%)	70	70	70	70	70
Deteção Positiva (Correta)	23	6	6	6	3
Deteção Positiva (Incorreta)	26	10	4	4	2
TP%	46,94	37,50	60	60	60
FP%	53,06	62,5	40	40	40
Precisão	0,469	0,375	0,60	0,60	0,60

Tabela 10: Resultados Reconhecimento Facial (FX 1 & FY 1)

Nome Dataset	Facial #1 - FX 1.5 FY 1.5				
Duração(s)	472				
FPS	25 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	11660	3932	2360	1684	1308
Confiablidade (%)	70	70	70	70	70
Deteção Positiva (Correta)	76	24	14	16	3
Deteção Positiva (Incorreta)	74	28	15	11	10
TP%	50,66	46,15	48,27	59,25	23,07
FP%	49,33	53,84	51,72	40,74	76,92
Precisão	0,5066	0,4615	0,4827	0,5925	0,2307

Tabela 11: Resultados Reconhecimento Facial (FX 1.5 & FY 1.5)

Nome Dataset	Facial #1 - FX 2 FY 2				
Duração(s)	472				
FPS	25 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	11660	3932	2360	1684	1308
Confiablidade (%)	70	70	70	70	70
Deteção Positiva (Correta)	86	25	22	17	5
Deteção Positiva (Incorreta)	82	28	15	10	10
TP%	51,19	47,16	59,45	62,96	33,33
FP%	48,80	52,83	40,54	37,03	66,66
Precisão	0,5119	0,4716	0,5945	0,6296	0,3333

Tabela 12: Resultados Reconhecimento Facial (FX 2 & FY 2)

1. **Principais observações e perceções**:

   a) **Intervalo de fotogramas e desempenho de deteção**:

i. **Intervalos de fotogramas curtos (1 a 3)**: Como esperado, os intervalos mais curtos resultaram em deteções mais frequentes, porém é necessário reter que devido ao intervalo de fotogramas ser menor, o tempo de execução é exponencialmente maior.

   - **Intervalo de fotogramas 1, escala 1x1**: Obteve um grande número de resultados positivos, mas a custo de um aumento de falsos positivos. Nomeadamente, "subject8.png" foi detetado várias vezes como um falso positivo, mesmo com pontuações de confiança elevadas (acima de 80%). (consultar tabela 10).
   - **Intervalo de fotogramas 3, escala 1,5x1,5**: Embora a taxa de TP tenha melhorado, os falsos positivos continuaram a ser um problema significativo, indicando que o modelo se debate com identificações incorretas apesar do aumento da resolução (consultar tabela 11).

ii. **Intervalos de fotogramas moderados (5 a 7)**: Este intervalo ofereceu um equilíbrio entre a identificação de fotogramas suficientes e a redução de falsos positivos:

   - **Intervalo de fotogramas 5, escala 1,5x1,5**: Apresentou um bom desempenho na deteção de positivos verdadeiros, como "subject3.png" e "subject5.png", minimizando os falsos positivos em comparação com intervalos de fotogramas inferiores. (consultar tabela 11)
   - **Intervalo de fotogramas 7, escala 2x2**: Surgiu como uma das configurações ideais, com um forte desempenho de TP (por exemplo, deteções de "subject5.png" com níveis de confiança acima de 80%) e uma redução de falsos positivos. (consultar tabela 12)

iii. **Intervalos de fotogramas longos (9)**: À medida que os intervalos de fotogramas aumentavam, a frequência de deteção diminuía:

   - **Intervalo de fotogramas 9, escala 2x2**: Apesar de terem sido analisados menos fotogramas, foram ainda registadas descobertas significativas de TP, particularmente em "subject4.png" e "subject2.png". No entanto, ainda se registou um aparecimento consistente de falsos positivos de elevada confiança, como "subject8.png" e "subject1.png". (consultar tabela 12)

b) **Aumento de escala e resolução de imagem**:

   i. **Sem aumento de escala (1x1)**: Existiram dificuldades em identificar corretamente os rostos devido à baixa resolução do vídeo. Falsos positivos, como "subject8.png", apareceram consistentemente em diferentes intervalos de fotogramas. (consultar tabela 10)

   ii. **Aumento de escala moderado (1,5x1,5)**: Proporcionou uma melhoria notável na precisão da deteção. Por exemplo, "subject3.png" e "subject5.png" foram corretamente identificados com níveis de confiança superiores a 75%. No entanto, isto também introduziu novos falsos positivos, indicando que a resolução por si só não pode resolver o problema das deteções incorretas. (consultar tabela 11)

   iii. **Aumento de escala elevado (2x2)**: Esta configuração levou a uma maior clareza nas deteções, com resultados TP significativos (por exemplo, "subject5.png" com 81,83% de confiança). No entanto, o sistema ainda mostrou uma tendência para identificar erradamente certos rostos com elevada confiança, sugerindo que o aumento excessivo da escala pode introduzir artefactos ou ruído que o modelo interpreta erradamente. (consultar tabela 12)

c) **Análise de falsos positivos**:

   i. Em todas as configurações, o modelo identificou consistentemente "subject8.png", "subject11.png.png" e "subject1.png" como falsos positivos, muitas vezes com pontuações de confiança elevadas (mais de 80%). Isto sugere que estes rostos podem partilhar caraterísticas faciais ou padrões semelhantes aos rostos identificados no vídeo, levando a identificações erradas consistentes.

   ii. Curiosamente, a confiança nos falsos positivos manteve-se elevada mesmo quando os intervalos de fotogramas foram aumentados, indicando que poderá existir potencial problema com a sensibilidade do modelo e a calibração do limiar.

d) **Análise de verdadeiros positivos**

   i. Os rostos principais como "subject3.png", "subject5.png" e "subject2.png" foram corretamente identificadas em várias configurações, com níveis de confiança que variam normalmente entre 70% e 90%.

   ii. A melhor configuração para resultados consistentes de TP foi observada com o intervalo de fotogramas 7 e um *upscale* de 2x2, onde o

modelo alcançou um equilíbrio entre a deteção de rostos verdadeiros e a limitação de falsos positivos.

2. **Resumo das observações**:

   a) **Configuração otimizada para precisão**: O intervalo de fotogramas 7 com *upscale* 2x2 forneceu consistentemente os melhores resultados, equilibrando as taxas de deteção de verdadeiros positivos e minimizando os falsos positivos. Esta configuração teve um bom desempenho em vários vídeos, mesmo com baixa resolução.

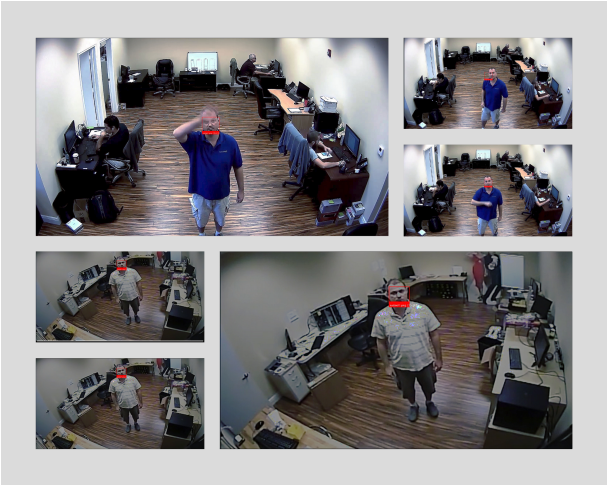
   b) **Compensação entre TP% e FP%**: Os testes destacaram um compromisso consistente entre o aumento da precisão de TP e a gestão de ocorrências de FP. O aumento de escala melhorou a deteção de TP, mas também trouxe mais falsos positivos.

   c) **Sensibilidade do modelo e falsos positivos**: A deteção persistente de falsos positivos (especialmente "subject8.png") indica que a sensibilidade do modelo pode precisar de ser recalibrada.

Vídeos de alta qualidade

Os resultados deste teste foram obtidos a partir de várias configurações utilizadas no *script* de reconhecimento facial Python sobre uma compilação de vídeos de alta qualidade. Os vídeos consistem num indivíduo que caminha em frente a uma câmara fixa e as experiências foram realizadas com diferentes intervalos de fotogramas, fatores de aumento de escala e um limiar de confiança fixo. Abaixo serão apresentados os resultados obtidos durante estes testes em tabelas e tecidos alguns comentários acerca das mesmas, bem como os metadados que comprovam de certa forma a "elevada qualidade"em termos da resolução da imagem. Alguns fotogramas dos vídeos ditos de alta qualidade empregues encontram-se na Figura 14. Os vídeos têm uma resolução de 1920 x 1080, o que corresponde ao formato dito de full HD, apresentando cada fotograma 2.1 megapíxeis, conforme indicando na Tabela 13. As Tabelas 14, 15 e 16 apresentam os resultados para o vídeo Facial #2, com fatores de escala 1.0, 1.5 e 2.0.

Figura 14: FacialRecognition SetV2



Vídeos	Resolução	Megapixels	Fonte
Facial #1	1920x1080	2.1	YouTube, 2022a
Facial #2	1920x1080	2.1	YouTube, 2022b

Tabela 13: Metadados Vídeos (Rostos #2)

Nome Dataset	Facial #2 - FX 1 FY 1				
Duração(s)	104				
FPS	30 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	3120	1040	624	445	346
Confiablidade (%)	70	70	70	70	70
Deteção Positiva (Correta)	195	64	39	26	21
Deteção Positiva (Incorreta)	33	12	8	5	4
TP%	85,52	84,21	82,98	83,87	84,00
FP%	14,47	15,78	17,02	16,12	16,00
Precisão	0,8552	0,8421	0,8298	0,8387	0,84

Tabela 14: Resultados Reconhecimento Facial V2 (FX 1 & FY 1)

Nome Dataset	Facial #2 - FX 1.5 FY 1.5				
Duração(s)	104				
FPS	30 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	3120	1040	624	445	346
Confiablidade (%)	70	70	70	70	70
Deteção Positiva (Correta)	189	65	38	27	22
Deteção Positiva (Incorreta)	21	10	10	1	4
TP%	90	86,66	79,17	96,43	84,62
FP%	10	13,33	20,83	3,57	15,38
Precisão	0,90	0,8666	0,7917	0,9643	0,8462

Tabela 15: Resultados Reconhecimento Facial V2 (FX 1.5 & FY 1.5)

Nome Dataset	Facial #2 - FX 2 FY 2				
Duração(s)	104				
FPS	30 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	3120	1040	624	445	346
Confiablidade (%)	70	70	70	70	70
Deteção Positiva (Correta)	190	63	49	26	23
Deteção Positiva (Incorreta)	133	45	24	8	7
TP%	58,82	58,33	67,12	76,47	76,66
FP%	41,17	41,66	32,87	23,52	23,33
Precisão	0,5882	0,5833	0,6712	0,7647	0,7666

Tabela 16: Resultados Reconhecimento Facial V2 (FX 2 & FY 2)

1. **Principais observações e perceções**:

   a) **Intervalo de fotogramas e desempenho de deteção**:

      i. **Intervalos de fotogramas curtos (1 a 3)**: Como esperado, os intervalos mais curtos resultaram em deteções mais frequentes. Obviamente, o tempo de execução está diretamente ligado ao número de fotogramas a analisar, sendo substancialmente maior quando se procede à análise de todos os fotogramas.

         - **Intervalo de fotogramas 1, *upscale* 1x1**: Conseguiu um elevado número de resultados positivos, mas à custa de um aumento de falsos positivos. Foram observados níveis de confiança elevados, muitas vezes superiores a 90%, indicando a forte convicção do modelo nestas deteções. (consultar tabela 14)

- **Intervalo de fotogramas 3, escala superior 1,5x1,5**: Melhorou a taxa de TP, mas os falsos positivos continuaram a ser um problema, especialmente quando estavam presentes no vídeo rostos de aspeto semelhante aos encontrados no dataset criado. (consultar tabela 15)

ii. **Intervalos de fotogramas moderados (5 a 7)**: Este intervalo ofereceu um equilíbrio entre a captura de fotogramas suficientes e a redução de falsos positivos:

- **Intervalo de fotogramas 5, escala superior 1,5x1,5**: Apresentou um bom desempenho na deteção de verdadeiros positivos, ao mesmo tempo que reduziu os falsos positivos em comparação com intervalos de fotogramas inferiores. (consultar tabela 15)
- **Intervalo de fotogramas 7, *upscale* 2x2**: Surgiu como uma das configurações otimizadas, com um forte desempenho de TP e uma redução de falsos positivos. (consultar tabela 16)

iii. **Intervalos de fotogramas**: À medida que os intervalos entre fotogramas aumentavam, a frequência de deteção diminuía:

- **Intervalo de fotogramas 9, *upscale* 2x2**: Embora tenham sido analisados menos fotogramas, foram ainda registados resultados significativos de TP. No entanto, registou-se um aparecimento consistente de falsos positivos com elevado nível de confiança, indicando a necessidade de uma recalibração do modelo. (consultar tabela 16)

b) **Aumento de escala e resolução de imagem**:

i. **Sem aumento de escala (1x1)**: O script teve um bom desempenho dado o vídeo de origem de alta qualidade, embora continuassem a existir falsos positivos, tais como identificações incorretas com níveis de confiança elevados.

ii. **Aumento de escala moderado (1,5x1,5)**: Proporcionou uma melhoria notável na precisão da deteção. Os verdadeiros positivos foram identificados com elevada confiança, excedendo frequentemente 95%. No entanto, também introduziu novos falsos positivos, o que sugere que o aumento de escala pode, por vezes, exacerbar as identificações incorretas.

iii. **Aumento de escala elevado (2x2)**: Esta configuração conduziu a uma maior clareza nas deteções, com resultados TP significativos. No entanto, o sistema ainda mostrou uma tendência para identificar erradamente certos rostos com elevada confiança, indicando que o aumento excessivo da escala pode introduzir ruído que o modelo interpreta erradamente.

c) **Análise de falsos positivos**:

i. Em todas as configurações, o modelo identificou erroneamente certos rostos (por exemplo, "subject5.png" e "subject10.png") como falsos positivos, muitas vezes com pontuações de confiança elevadas. Isto sugere que estes rostos partilham caraterísticas faciais ou padrões semelhantes com os rostos no vídeo, levando a identificações erradas consistentes.

ii. A confiança do script em falsos positivos permaneceu alta, mesmo quando os intervalos de fotogramas foram aumentados, indicando um problema potencial com a sensibilidade do modelo e a sua calibração.

d) **Análise de Verdadeiros Positivos**:

i. Os rostos principais, como "subject1.png", foram identificadas corretamente em várias configurações, com níveis de confiança que variam normalmente entre 80% e 99%.

ii. A melhor configuração para resultados TP consistentes foi observada com o intervalo de fotogramas 7 e um *upscale* de 2x2, onde o modelo alcançou um equilíbrio entre a deteção de rostos verdadeiros e a limitação de falsos positivos.

2. **Resumo das observações**:

a) **Configuração otimizada para a precisão**:O intervalo de fotogramas 7 com *upscale* 2x2 forneceu consistentemente os melhores resultados, equilibrando as taxas de deteção de verdadeiros positivos e minimizando os falsos positivos. Esta configuração teve um bom desempenho em vários vídeos.

b) **Compensação entre TP% e FP%**: O teste destacou um compromisso consistente entre o aumento da precisão de TP e a gestão de ocorrências de FP. O aumento de escala melhorou a deteção de TP, mas também trouxe mais falsos positivos.

c) **Sensibilidade do modelo e falsos positivos**: A deteção persistente de falsos positivos, especialmente certos rostos mal identificados, indica que a sensibilidade do modelo pode precisar de ser recalibrada.

Tempos de Execução

Importa ter em conta que o aumento do intervalo entre fotogramas leva a diminuição da quantidade de fotogramas analisados. Tal significa que a carga computacional irá ser reduzida e o processamento será mais rápido. Por outro lado, caso exista um intervalo de fotogramas reduzidos ou um aumento de escala aplicado aos fotogramas irá haver um acréscimo do tempo de processamento. Estas afirmações podem ser confirmadas nas Tabelas 17 e 18 onde são detalhados os tempos de execução de todos os testes efetuados.

Dataset	Facial #1														
Escala	1x1					1.5x1.5					2x2				
Intervalo de Fotogramas	1	3	5	7	9	1	3	5	7	9	1	3	5	7	9
Tempo de Execução (s)	834	278	167	140	100	1 860	622	377	315	240	2 801	890	532	450	316
Fotogramas Analisados	11660	3932	2360	1684	1308	11660	3932	2360	1684	1308	11660	3932	2360	1684	1308

Tabela 17: Tempos de execução (Reconhecimento Facial #1)

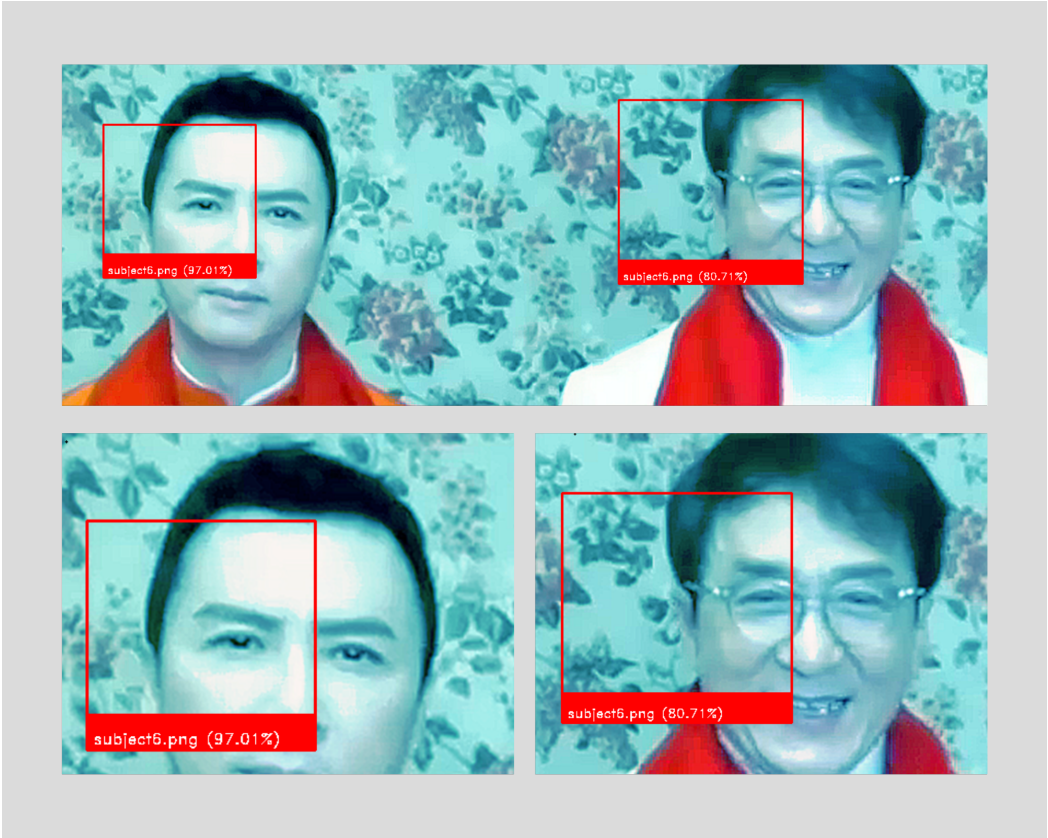
Dataset	Facial #2														
Escala	1x1					1.5x1.5					2x2				
Intervalo de Fotogramas	1	3	5	7	9	1	3	5	7	9	1	3	5	7	9
Tempo de Execução (s)	1441	437	281	200	165	3213	974	626	446	367	4553	1380	887	632	521
Fotogramas Analisados	3120	1040	624	445	346	3120	1040	624	445	346	3120	1040	624	445	346

Tabela 18: Tempos de execução (Reconhecimento Facial #2)

Conclusões

Em suma, o modelo demonstrou resultados bastante positivos e foi capaz de superar alguns dos desafios que poderiam tornar o reconhecimento facial mais difícil. Assim, o modelo conseguiu identificar corretamente rostos em condições de baixa qualidade e a partir de diversos ângulos. Contudo, importa notar que o modelo de reconhecimento facial apresenta sérias limitações em etnias asiáticas, ocorrendo num número significativo de falsos positivos, alguns deles com elevado grau de confiança. A Figura 15 apresenta dois rostos asiáticos em que o modelo identificou, erradamente, como sendo o indivíduo "subject6.png". Note-se que a menor precisão da dlib em faces não-caucasianas é uma situação conhecida davisking, 2024.

Figura 15: Exemplo Etnia



6.4 TESTES DE RECONHECIMENTO DE OBJETOS

Os testes para avaliação de desempenho da deteção de objetos referentes às quatro classes suportadas incidiram na precisão e no tempo de execução dos algoritmos. Em particular, os tempos de execução dos algoritmos são de grande relevância em casos que envolvam vídeos de longa duração, pois pretende-se com a ferramenta de deteção assente em visão computacional, auxiliar o analista forense, por forma a que numa primeira triagem obtenha uma filtragem com a indicação das secções do vídeo que contêm as classes de objetos pretendidas (e.g. , facas, espingardas, etc.). Com essa informação, o analista forense pode focar os seus esforços nessas secções de interesse identificadas pelos algoritmos de IA.

Esta secção apresenta os resultados referentes à deteção de quatro classes de objetos matrículas de veículos, pistola, espingarda e faca em vídeos de testes. Assim, apresentam-se, sob forma de tabelas, os resultados detalhados para cada classe de

objeto/vídeo. Por fim, é feita uma análise geral dos resultados, destacando-se as principais observações.

Para cada classe de objeto serão apresentados e explicados os dados contidos numa tabela para cada vídeo que demonstra dados como o grau de precisão em que os objetos foram identificados, bem como a quantidade fotogramas que existiam. Por último e em forma de conclusão serão tecidos alguns comentários a termos de observações anotadas durante o processo.

6.4.1 *Matrículas de veículos*

Os resultados dos testes servirão para avaliar o desempenho do sistema de reconhecimento relativo à classe matrícula em dois vídeos distintos, em termos de deteções de verdadeiros positivos (TP) e falsos positivos (FP). A Tabela 19 apresenta os metadata dos vídeos empregues nos testes. A Figura 16 ilustra a forma como a aplicação identifica matrículas nos fotogramas.

Datasets	Resolução	Megapixels	Fonte
License #1	3840x2160	8.3	Shandilya, 2021
License #2	1280x720	0.922	Sveyek, 2021

Tabela 19: Metadados Videos (Classe matrícula)

Figura 16: Classe Matrícula Exemplos

Os resultados referentes à classe matrículas de veículos encontram-se nas tabelas 20 e 21.

1. **Principais perceções**:

a) **Vídeo 1**:

Nome Dataset	license #1				
Duração(s)	31				
FPS	30 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	930	310	186	132	103
Confiablidade (%)	75	75	75	75	75
Deteção Positiva (Correta)	574	191	114	83	64
Deteção Positiva (Incorreta)	6	2	1	1	1
TP%	98,97	98,96	99,13	98,81	98,46
FP%	1,03	1,03	0,86	1,19	1,15
Precisão	0,9897	0,9896	0,9913	0,9881	0,9846

Tabela 20: Classe Matrícula (Video 1 Resultados)

i. À medida que o intervalo de fotogramas aumenta, a TP% diminui ligeiramente e a FP% aumenta ligeiramente. Isto é esperado num ambiente mais complexo, onde saltar demasiados fotogramas pode reduzir a precisão e aumentar a possibilidade de falsos positivos.

ii. Apesar destas alterações, o modelo continua a ter um desempenho razoável, com a TP% a manter-se acima dos 98%, mesmo no maior intervalo de fotogramas, e a FP% a aumentar apenas cerca de meio ponto percentual.

b) **Vídeo 2**:

Nome Dataset	license #2				
Duração(s)	60				
FPS	60 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	3600	1200	720	514	400
Confiablidade (%)	75	75	75	75	75
Deteção Positiva (Correta)	79	28	17	15	10
Deteção Positiva (Incorreta)	0	0	0	0	0
TP%	100	100	100	100	100
FP%	0	0	0	0	0
Precisão	1	1	1	1	1

Tabela 21: Classe Matrícula (Vídeo 2 Resultados)

i. O modelo de deteção tem um desempenho perfeito, mantendo 100% de TP% e 0% de FP%, independentemente do intervalo de fotogramas. Isto sugere que, em cenários menos complexos neste vídeo apenas

existem objetos da classe matrícula, o aumento do intervalo de fotogramas não prejudica o desempenho do modelo.

2. **Observações finais**: Em relação a classe das matrículas, o modelo demonstrou uma elevada precisão. Tal se percebe dado as características únicas das matrículas de veículos, que têm sido adaptadas para que possam ser facilmente detetadas por visão computacional. Acresce-se que as matrículas apresentam propriedades substancialmente distintas das outras classes de objetos consideradas neste trabalho, o que ajuda a compreender que não ocorreram falsos positivos de outras classes de objetos. Porém, em situações de grande distâncias e ângulos mais desfavoráveis, algumas das matrículas não eram detetadas.

6.4.2 *Pistola*

Os resultados dos testes servirão para avaliar o desempenho do sistema de reconhecimento relativo à classe pistola em dois vídeos distintos, em termos de deteções de verdadeiros positivos (TP) e falsos positivos (FP). A Tabela 22 apresenta a resolução e o número de megapíxeis dos dois vídeos empregues nos testes para a deteção de objetos da classe "pistola". Por sua vez, a Figura 17 apresenta alguns o resultado do processamento de alguns dos fotogramas, sendo visíveis as *bounding boxes* e o respetivo valor de confiança.

Datasets	Resolução	Megapixels	Fonte
Pistol #1	1280x720	0.92	YouTube, 2022f
Pistol #2	1280x720	0.92	YouTube, 2022e

Tabela 22: Metadados Videos (Classe Pistola)

Figura 17: Classe Pistola Exemplos

Os resultados referentes à classe pistola encontram-se nas tabelas 23 e 24.

1. **Principais perceções**:

   a) **Vídeo 1 (Assalto à loja de conveniência)**:

Nome Dataset	pistol #1				
Duração(s)	22				
FPS	29.97 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	659	219	131	94	73
Confiablidade (%)	75	75	75	75	75
Deteção Positiva (Correta)	17	4	4	3	0
Deteção Positiva (Incorreta)	4	2	1	0	1
TP%	80,95	66,67	80	100	0
FP%	19,04	33,33	20	0	100
Precisão	0,8095	0,6667	0,80	1	0

Tabela 23: Classe Pistola (Vídeo 1 Resultados)

   i. À medida que o intervalo entre fotogramas analisados aumenta, a TP% diminui significativamente, começando em 80,95% no intervalo 1 e caindo para 0% no intervalo 9. Este declínio acentuado na precisão da deteção mostra que o modelo tem dificuldade em detetar pistolas quando são analisados menos fotogramas neste ambiente mais complexo.

   ii. A FP% aumenta moderadamente, começando em 19,05% e atingindo um pico de 33,33% no intervalo 5, mas não há falsos positivos

nos intervalos 7 e 9, provavelmente devido ao facto de terem sido detetados menos objetos em geral.

iii. Apesar da diminuição da precisão da deteção, o número total de deteções começa relativamente baixo, com 17 objetos no intervalo 1, e desce para 1 objeto no intervalo 9.

b) **Vídeo 2 (Assalto a restaurante)**:

Nome Dataset	pistol #2				
Duração(s)	61				
FPS	30 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	1830	610	366	261	203
Confiablidade (%)	75	75	75	75	75
Deteção Positiva (Correta)	42	13	7	6	2
Deteção Positiva (Incorreta)	19	8	4	3	3
TP%	68,85	61,90	63,63	66,67	40
FP%	31,14	38,09	36,36	33,33	60
Precisão	0,6885	0,6190	0,6363	0,6667	0,40

Tabela 24: Classe Pistola (Vídeo 2 Resultados)

i. A precisão da deteção mantém-se razoável, com uma TP% a começar em 68,85% para o intervalo 1 e a diminuir para 40% no intervalo 9. Isto mostra que o aumento do intervalo de fotogramas tem impacto na deteção de pistolas, mas não tão severamente como no Vídeo 1.

ii. A FP% aumenta acentuadamente, de 31,15% no intervalo 1 para 60% no intervalo 9. Isto sugere que o modelo produz mais deteções incorretas à medida que o intervalo entre fotogramas aumenta neste ambiente complexo.

iii. O número total de deteções é maior neste vídeo, começando com 42 deteções no intervalo 1 e caindo para 3 deteções no intervalo 9.

2. **Observações finais**: A classe pistola demonstra uma precisão e identificação satisfatória. Os objetos desta classe são identificados em diversos ângulos onde poderia ser difícil identificar estes. Contudo, existem fotogramas onde são claramente identificáveis pistolas por um observador humano, sem que o modelo seja capaz de identificar o objeto.

6.4.3 *Espingarda*

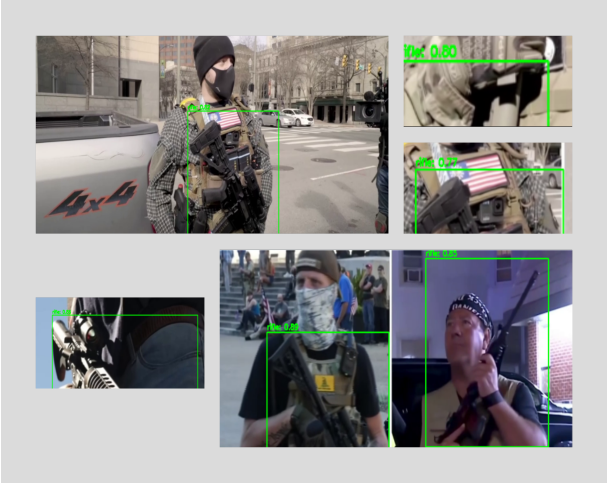
Os resultados dos testes servirão para avaliar o desempenho do sistema de reconhecimento relativo à classe espingarda em dois vídeos distintos, em termos de deteções de verdadeiros positivos (TP) e falsos positivos (FP). A Tabela 25 apresenta a metadata dos dois vídeos empregues nos testes à classe "espingarda".

Datasets	Resolução	Megapixels	Fonte
Rifle #1	1280x720	0.92	YouTube, 2022g
Rifle #2	1280x720	0.92	YouTube, 2022h

Tabela 25: Metadados Videos (Classe Espingarda)

A Figura 18 apresenta alguns dos fotogramas nos quais foram identificadas objetos da classe espingarda, anotados através dos retângulos de cor verde e o respetivo índice de confiança.

Figura 18: Classe Espingarda Exemplos



1. **Principais perceções**:

    a) **Vídeo 1 (Demonstração de armas)**:

Nome Dataset	rifle #1				
Duração(s)	16				
FPS	30 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	480	160	96	69	53
Confiablidade (%)	75	75	75	75	75
Deteção Positiva (Correta)	15	5	3	1	0
Deteção Positiva (Incorreta)	0	0	0	0	0
TP%	100	100	100	100	0
FP%	0	0	0	0	0
Precisão	1	1	1	1	0

Tabela 26: Classe Espingarda (Vídeo 1 Resultados)

i. A TP% mantém-se a 100% em todos os intervalos de fotogramas, mostrando que o modelo tem um desempenho excecional neste vídeo, independentemente da frequência com que os fotogramas são analisados.

ii. A FP% mantém-se em 0% em todos os intervalos de fotogramas, o que significa que não existem falsos positivos.

iii. Apesar da precisão perfeita, o número total de deteções diminui à medida que o intervalo de fotogramas aumenta, com 15 deteções no intervalo 1 e apenas 1 deteção no intervalo 9. Essa queda é esperada à medida que menos fotogramas são analisados.

b) **Vídeo 2 (Protesto Americano)**:

Nome Dataset	rifle #2				
Duração(s)	40				
FPS	30 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	1200	400	240	171	133
Confiablidade (%)	75	75	75	75	75
Deteção Positiva (Correta)	733	250	136	100	83
Deteção Positiva (Incorreta)	5	3	2	0	0
TP%	99,18	98,81	98,55	100	100
FP%	0,67	1,18	1,45	0	0
Precisão	1	1	1	1	0

Tabela 27: Classe Espingarda (Vídeo 2 Resultados)

i. A TP% é consistentemente alta, começando em 99,19% no intervalo 1 e caindo ligeiramente para 98,55% no intervalo 5, antes de diminuir mais notavelmente para 83% no intervalo 9. Isto indica que a precisão da deteção permanece forte, mas diminui à medida que o intervalo de fotogramas aumenta.

ii. A FP% começa baixa, com 0,68% para o intervalo 1 e aumenta ligeiramente para 1,45% no intervalo 5, antes de voltar a cair para 0% no intervalo 9. Isto sugere que são feitas menos deteções incorretas em intervalos maiores, possivelmente porque são detetados menos objetos em geral.

iii. O número total de deteções diminui significativamente, começando com 733 deteções no intervalo 1 e caindo para 83 deteções no intervalo 9, o que reflete um menor número de fotogramas a serem analisados e menos oportunidades para detetar espingardas.

2. **Observações finais**: A classe espingarda é a classe com a qual a aplicação CyberSync apresenta melhor desempenho. Contudo, e ao observar os resultados é possível identificar que está poderia similarmente as outras possuir um melhor desempenho em termos da quantidade de identificação de objetos.

6.4.4 *Faca*

Os resultados dos testes servirão para avaliar o desempenho do sistema de reconhecimento relativo à classe faca em dois vídeos distintos, em termos de deteções de

verdadeiros positivos (TP) e falsos positivos (FP). A Tabela 28 lista a resolução e os megapíxeis dos vídeos empregues para os testes da classe faca. A Figura 19 apresenta alguns das deteções de objetos da classe faca efetuadas pela aplicação CyberSync.

Datasets	Resolução	Megapixels	Fonte
Knife #1	1280x720	0.92	YouTube, 2022c
Knife #2	1280x720	0.92	YouTube, 2022d

Tabela 28: Metadados Vídeos (Classe Faca)

Figura 19: Classe Faca Exemplos

1. **Principais perceções**:

   a) **Vídeo 1 (Vídeo de defesa pessoal)**:

Nome Dataset	knife #1				
Duração(s)	14				
FPS	30 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	420	140	84	60	47
Confiablidade (%)	75	75	75	75	75
Deteção Positiva (Correta)	71	21	13	9	5
Deteção Positiva (Incorreta)	3	0	1	0	0
TP%	95,94	100	92,85	100	100
FP%	4,05	0	7,14	0	0
Precisão	0,9594	1	0,9285	0	0

Tabela 29: Classe Faca (Video 1 Resultados)

i. À medida que o intervalo de fotogramas aumenta, a TP% mantém-se consistentemente elevada. Começa em 95,95% no intervalo 1, mantém-se em 100% nos intervalos 3 e 7 e desce ligeiramente para 92,86% no intervalo 5. Isto mostra que a precisão da deteção permanece forte em geral, mesmo em intervalos de fotogramas mais elevados.

ii. A FP% é baixa, começando em 4,05% no intervalo 1 e subindo para 7,14% no intervalo 5, mas cai para 0% nos intervalos 3, 7 e 9. Isto indica que o modelo gera mais falsos positivos quando são analisados mais fotogramas.

iii. O número total de deteções diminui à medida que o intervalo de quadros aumenta, de 71 deteções no intervalo 1 para 5 deteções no intervalo 9, demonstrando a diminuição esperada de objetos detetados quando menos quadros são analisados.

b) **Vídeo 2 (Vídeo de defesa pessoal)**:

Nome Dataset	knife #2				
Duração(s)	41				
FPS	30 fps				
Intervalo de Análise	1	3	5	7	9
Fotogramas Analisadas	1230	410	246	176	137
Confiablidade (%)	75	75	75	75	75
Deteção Positiva (Correta)	218	70	38	36	25
Deteção Positiva (Incorreta)	2	1	2	0	0
TP%	99,09	98,59	95	100	100
FP%	0,90	1,40	5	0	0
Precisão	0,9909	0,9859	0,95	1	1

Tabela 30: Classe Faca (Vídeo 2 Resultados)

i. A TP% é alta, começando em 99,09% no intervalo 1 e caindo ligeiramente para 95% no intervalo 5. No entanto, a precisão da deteção mantém-se em 100% nos intervalos 7 e 9, indicando um desempenho estável com intervalos de fotogramas maiores.

ii. A FP% mantém-se baixa, começando em 0,91% no intervalo 1, aumentando ligeiramente para 5% no intervalo 5, mas regressa a 0% nos intervalos 7 e 9.

iii. O número total de deteções diminui significativamente à medida que o intervalo de fotogramas aumenta, de 218 deteções no intervalo 1

para 25 deteções no intervalo 9. Isto segue o padrão esperado de menos deteções à medida que menos quadros são analisados.

2. **Observações finais**: A classe faca demonstra uma precisão e identificação bastante positiva, tendo em conta que existe alguma proximidade entre a classe faca e a classe pistola. Os objetos desta classe são identificados em diversos ângulos onde poderia ser difícil identificar estes, mas tal e qual como a classe anterior carece de um acréscimo na deteção das ocorrências, sendo que existem por vezes fotogramas onde é claramente identificável um objeto desta classe a olho humano, porém o modelo não é capaz de identificar este.

6.4.5 *Tempos de execução*

Os tempos de execução referentes aos testes de deteção das quatro classes de objetos encontram-se documentados nas Tabelas 31 e 32. Novamente se denota que um maior intervalo entre fotogramas analisados leva a um menor tempo de execução em virtude da diminuição da carga computacional.

Dataset	Matricula de Veiculo					Pistola				
Intervalo de Fotogramas	1	3	5	7	9	1	3	5	7	9
Tempo de Execução (s)	1101	443	300	260	225	551	204	135	106	90
Fotogramas Analisados	4530	1510	906	646	503	2489	829	497	355	276

Tabela 31: Tempos de execução (Reconhecimento Objetos #1)

Dataset	Espingarda					Faca				
Intervalo de Fotogramas	1	3	5	7	9	1	3	5	7	9
Tempo de Execução (s)	359	119	72	51	41	349	122	75	52	41
Fotogramas Analisados	1680	560	336	240	186	1650	550	330	236	184

Tabela 32: Tempos de execução (Reconhecimento Objetos #2)

6.5 DISCUSSÃO

Os resultados obtidos indicam que a abordagem de integração de tecnologias de IA para a análise de vídeos forenses é promissora. A capacidade de detetar e reconhecer elementos-chave pertencentes as classes destacadas acima nos resultados em vídeos poderá fornecer um forte apoio aos investigadores, ajudando assim a resolver os

casos mais rapidamente. No entanto, é importante compreender que através dos resultados apresentados é também possível chegar a duas conclusões chave:

1. **Reconhecimento Facial**: Apesar dos resultados do modelo de reconhecimento facial serem satisfatórios e demonstrarem que o modelo é capaz de enfrentar desafios como a qualidade de vídeo, é também necessário repensar no modelo utilizado visto que existem modelos mais atuais que podem eventualmente obter uma taxa de precisão superior ao modelo utilizado (i.e *deepface* 2024). Acresce-se que o modelo é sensível à etnia dos indivíduos que se pretende identificar.
2. **Reconhecimento de objetos**: Os resultados dos testes referentes ao reconhecimento de objetos demonstram bons níveis de precisão, observando-se baixas taxas de falsos positivos. Porém, e de forma a obter uma maior taxa de deteção poderá ser necessário proceder à criação manual de um novo *dataset* bem como ao seu treino, de forma a que o modelo consiga detetar com maior frequência objetos que são visíveis nos fotogramas e conhecidos pelo modelo numa maior variedade de ângulos.

6.6 DESAFIOS ENCONTRADOS

Seguidamente, descrevem-se os principais desafios encontrados.

1. **Formação de um *dataset***: A criação de um conjunto de dados para proceder ao treino de modelo do YOLOV5 de forma a que este demonstrasse um grande grau de fiabilidade elevado apresentou um desafio, sendo que se chegou a conclusão que seria necessário recorrer à criação manual de um *dataset* completo e específico para o propósito. Dentro deste tópico foram identificados os seguintes desafios a ter em conta:
    - **Anotação Precisa e Consistente**: Um dos principais desafios foi a anotação manual do elevado conjunto de dados. Este processo exigia muitos recursos e todas as anotações tinham de ser muito precisas e consistentes. O modelo poderia correr o risco de perder o seu desempenho se não fosse corretamente anotado, sendo necessário dedicar tempo e esforço suficientes para fazer a anotação da forma mais meticulosa possível. No entanto, foi por vezes trabalhoso, por vezes propenso a erros e, consequentemente necessário uma revisão constante para eliminar as incoerências.

- **Diversidade e Amplitude dos Dados**: Outra preocupação constante foi o facto do *dataset* não ser muito diversificado, criando assim o risco de um modelo que funcionaria muito bem em situações particulares, mas que teria dificuldade em generalizar para novos contextos.

2. **Treino da IA** : O treino da IA foi tambem ele desafiador, visto que os *datasets* criados a partir da junção de *datasets* de classes singulares careciam de imagens que cobrissem todos os ângulos dos objetos a detetar, fazendo com que a identificação de algumas classes ficasse aquém do expectável não em termos de precisão, mas sim em termos da quantidade de identificações. Outro ponto deste desafio foi também o hardware disponível para o treino que por vezes se demonstrou insuficiente para correr modelos de treino de grandes dimensões.

3. **Qualidade de Vídeo**: Por último, outro desafio foi a qualidade dos vídeos, que afetou diretamente a eficácia dos modelos de IA treinados. Notou-se que ao fazer os testes com vídeos que possuíssem alguns dos problemas enumerados abaixo seria árduo obter resultados com uma alta precisão:

   - **Baixa qualidade e ou definição de vídeo**: Todos os dados são recolhidos através de imagens de CCTV, clips de vídeo de telemóveis, etc., como elementos de prova para análise forense de vídeo. Todas estas fontes de vídeo não cumprem necessariamente as normas de qualidade exigidas para uma análise de vídeo eficiente. A partir de filmagens de baixa qualidade, os pequenos pormenores não podem ser identificados, mesmo com o redimensionamento do vídeo ou da imagem. O potencial de melhoramento dos conteúdos de baixa resolução é limitado, o que conduz a problemas de análise forense de vídeo. Javed et al., 2021

   - **Compressão**: Os ficheiros de vídeo são normalmente comprimidos para utilizar o espaço de forma eficiente, mas isso pode levar à perda de detalhes como a qualidade do vídeo. Estes pormenores não recuperados dificultam a realização de uma análise eficiente por parte dos investigadores. Javed et al., 2021

   - **Luminosidade**: Se o vídeo tiver imagens de baixa resolução, o brilho pode induzir em erro a análise. Por exemplo, se houver duas filmagens de CCTV do mesmo cenário, o vídeo resultante pode estar sobre ou subexposto. Para o corrigir, o investigador terá de ajustar manualmente o brilho de cada vídeo para obter mais detalhes. Eventualmente, o processo de ajuste do brilho poderá ser automatizado. Javed et al., 2021

4. **Identificação de vídeos "oficiais"**: Outra limitação foi a impossibilidade de obter conjuntos de dados oficiais de vídeos relativos à análise forense digital para as classes de objetos que iam ser identificadas. Isto implicou que fossem usados vídeos obtidos em outras fontes como o youtube que mantivessem alguma relação com o que poderia vir a ser identificado num contexto real

6.6.1 *Possíveis Melhorias*

Tendo em conta os desafios encontrados durante o desenvolvimento e implementação da aplicação, apresentam-se algumas sugestões de melhoria para versões futuras.

1. **Otimização de Hardware**: A atualização para hardware mais avançado, poderá melhorar significativamente a eficiência relativa ao processamento da aplicação. A otimização do código no que diz respeito ao paralelismo e à utilização de recursos computacionais pode reduzir ainda mais o tempo de análise, dando assim a oportunidade de aplicar a análise de vídeo em tempo real ou quase real, aumentando assim consideravelmente a aplicabilidade prática nas investigações forenses.
2. **Melhorias na Interface**: A usabilidade global da aplicação depende da interface Web. Um maior desenvolvimento da interface, melhorando o fluxo de trabalho, acrescentando visualizações mais intuitivas e tornando os guias e instruções mais claros, tornará a experiência do utilizador ainda mais eficiente. Estas melhorias na interface não só tornarão a aplicação muito mais fácil de utilizar pelo investigador forense, como também poderão reduzir o tempo necessário para a análise, proporcionando uma exploração mais rápida e eficaz dos dados.
3. **Melhoria dos Datasets**: A eficácia da aplicação reside na criação de conjuntos de dados que sejam simultaneamente completos e representativos.Esta construção baseia-se na criação de conjuntos de dados que contenham variáveis tão distintas quanto possível em todos os parâmetros, como tipos de crimes, ambientes e etnias, para treinar modelos de IA inclusivo para aplicabilidade em diversas situações. Um conjuntos de dados de elevada qualidade com rotulagem exata evitarão enviesamentos e proporcionarão melhores intervalos de confiança nos resultados.
4. **Treino Adicional dos Modelos**: O treino adicional dos modelos de IA é, por consequente, uma das principais oportunidades de melhoria. É expectável que,

com dados de treino mais extensos, a precisão e a capacidade de generalização do modelo possam aumentar significativamente. Isto permitirá, por sua vez, que a aplicação produza resultados mais fiáveis em casos forenses.

5. **Pré-Processamento de vídeo**: A resolução dos desafios identificados acima referentes a qualidade de vídeo é algo a tomar em consideração no desenvolvimento desta metodologia, pois pode contribuir para que esta aplicação alcance maiores taxas de deteção e precisão.

6. **Condições Noturnas**: Será interessante também conseguir que o modelo consiga identificar objetos em vídeos onde as condições sejam de baixa luminosidade, evitando que o modelo não fique condicionado a deteções apenas com condições de luminosidade favoráveis.

7. **Deteção de Matrículas**: Uma caraterística que poderia ser melhorada é a deteção de matrículas. Atualmente, o CyberSync não efetua Reconhecimento Ótico de Caracteres (OCR) (reconhecimento ótico de caracteres) para ler matrículas. No entanto, é um bom ponto a ter em conta para futuras melhorias, uma vez que o OCR permitirá à ferramenta detetar automaticamente o texto das matrículas nos vídeos em análise, aumentando a sua capacidade de investigação.

8. **Vídeos Rotulados**: Seria interessante obter um conjunto de vídeos devidamente rotulados para a avaliação dos resultados do modelo. Idealmente, esses vídeos devem conter informações precisas sobre os fotogramas em que os objetos de interesse aparecem, facilitando a validação e a melhoria contínua da aplicação.

9. **Inclusão da Métrica de *Recall***: A inclusão de métricas mais abrangentes, como o *recall* (ou sensibilidade), nos resultados será essencial para entender melhor o desempenho da aplicação. O *recall* é particularmente importante para avaliar a capacidade do modelo em identificar corretamente todas as ocorrências dos objetos de interesse, evitando falsos negativos.

Estas melhorias poderão auxiliar na criação de uma ferramenta mais eficaz e acessível para a análise de vídeos forenses, de forma a contribuir significativa para a resolução de casos que exigissem um maior grau de complexidade.

7

CONCLUSÕES

Os resultados obtidos a partir da elaboração desta dissertação servem como ponto de partida para a criação de uma aplicação que pode auxiliar de forma positiva e eficiente os investigadores forenses na área da análise forense digital de vídeos. Os resultados obtidos através da aplicação desenvolvida como caso de estudo, permitem ao investigador traçar um plano para a investigação forense de um vídeo, visto que estes dados possibilitam a elaboração de um pré-relatório que demonstra os intervalos de tempo onde se localizam os rostos/objetos que são necessários detetar num espaço de tempo inferior ao de uma investigação sem recurso a Inteligência Artificial (IA).

Através dos desafios mencionados ao longo do processo de desenvolvimento da prova de conceito e a partir dos resultados demonstrados foram detetadas algumas das falhas que a metodologia executada poderia demonstrar caso fosse aplicada a um caso real. Porem, é necessário ter em conta que o desenvolvimento de uma aplicação real com metodologia semelhante requererá uma maior ponderação na abordagem, de forma a especificar concretamente o seu caso de uso final em termos das classes a detetar.

Tendo em conta as observações indicadas acima e olhando também para os resultados positivos obtidos ao longo da implementação da metodologia desenvolvida, é possível concluir que a existência de métodos que recorressem ao uso da inteligência artificial de forma ao auxiliar no processo da análise forense digital poderiam trazer consigo melhorias, quer no tempo despendido para a execução dos métodos tradicionais, quer também pela precisão que um modelo de IA poderá apresentar de forma a detetar eventos ou objetos que possam ser cruciais para o desfecho de uma análise forense de vídeo. Porem é de notar que por se tratar de uma área que poderá ser crítica para a resolução de processos delicados, tais como crimes, será necessária a intervenção humano para confirmar os resultados obtidos.

A prova de conceito desenvolvida apresenta de forma simplificada o potencial e as melhorias que um método destes poderá trazer em relação a avanços na área da automatização do processo da análise tendo em mente a diminuição do esforço

manual e aumento da eficiência da análise, através das seguintes funcionalidades principais:

- **Extração de informação** Antes de realizar uma identificação manual exaustiva, o investigador pode obter uma visão geral dos padrões de ocorrência dos eventos de interesse, como a presença de objetos ou rostos. A aplicação facilita a pré-visualização desses eventos, dando ao investigador uma orientação clara sobre o que pode esperar encontrar ao longo da investigação. Esta visão global inicial permite uma preparação mais direcionada e estratégica, tendo em conta um processo de análise mais estruturado e eficaz.
- **Criação de gráficos** Através dos gráficos criados pela aplicação, é possível identificar com precisão os intervalos de tempo do vídeo em análise durante os quais é reconhecido um rosto ou objetos. Esses gráficos fornecem ao investigador um relatório visual e detalhado, permitindo que este concentre a sua atenção nos momentos de maior incidência. Tal permite um mais rápido processamento de vídeos.
- **Possível elaboração de um pré-relatório**: Com os dados obtidos a partir da análise prévia e automatizada, torna-se possível de certa forma a elaboração de um pré-relatório antes do início de uma análise forense detalhada de um determinado vídeo. Este pré-relatório fornece uma visão inicial dos eventos e objetos detetados, reduzindo o tempo total da análise ao permitir que o investigador inicie a sua investigação já com uma base de dados relevante. O pré-processamento de informações como rostos ou objetos presentes no vídeo agiliza o processo e pode guiar o investigador para as áreas de maior interesse, aumentando a eficácia da análise forense.

Para trabalho futuro e de forma a que a metodologia apresentada seja embutida numa aplicação pronta para ser usada em contexto real será necessário rever e repensar os desafios apresentados, bem como ter em conta principalmente as melhorias anteriormente discutidas que poderão contribuir de forma significativa para simplificar a visualização e uma maior precisão na obtenção dos resultados.

Em conclusão, a inteligência artificial demonstra um grande potencial para ser aplicada na área da análise forense digital sendo que poderão ser criadas melhorias através desta tecnologia que influenciaram de forma positiva esta área de um ponto de vista geral. Apesar deste grande potencial esta também por sua vez traz grandes desafios nomeadamente relacionados com problemas éticos e carece sempre de uma monitorização humana de forma a que sejam identificados os possíveis erros que possam ocorrer.

BIBLIOGRAFIA

(Fev. de 2011). URL: https://rishikeshpansare.wordpress.com/wp-content/uploads/2016/02/digital-evidence-and-computer-crime-third-edition.pdf.

About (nov. de 2020). URL: https://opencv.org/about.

ActiveAI Security Platform | Darktrace (jul. de 2024). URL: https://darktrace.com/platform.

Al, Shaikha et al. (mai. de 2023). *Maximizing the Potential of Artificial Intelligence in Digital Forensics Investigations*. DOI: 10.13140/RG.2.2.32764.72320.

Al Amin, Md. e Bikash Kumar Paul (2024). «A Smart Surveillance System to Detect Modern Gun Using YOLOv5 Algorithm: A Deep Learning Approach». Em: *Proceedings of International Joint Conference on Advances in Computational Intelligence*. Ed. por Mohammad Shorif Uddin e Jagdish Chand Bansal. Singapore: Springer Nature Singapore, pp. 235–244. ISBN: 978-981-97-0180-3.

API Cloud Natural Language | Cloud Natural Language API | Google Cloud (abr. de 2021). URL: https://cloud.google.com/natural-language/docs/reference/rest?hl=pt-br.

Apruzzese, Giovanni et al. (mar. de 2023). «The Role of Machine Learning in Cybersecurity». Em: *Digital Threats* 4.1. DOI: 10.1145/3545574. URL: https://doi.org/10.1145/3545574.

Autopsy (jul. de 2024). [Online; accessed 16. Sep. 2024]. URL: https://www.autopsy.com.

Bangar, Siddhesh (jun. de 2022). «AlexNet Architecture Explained - Siddhesh Bangar - Medium». Em: *Medium*. ISSN: 6240-5285. URL: https://medium.com/@siddheshb008/alexnet-architecture-explained-b6240c528bd5.

Bhagyalakshmi., P., P. Indhumathi. e Lakshmi. R. | Dr. Bhavadharini (abr. de 2019). *Real Time Video Surveillance for Automated Weapon Detection*. DOI: 10.31142/ijtsrd22791.

Big Multimedia Data and Applications (set. de 2024). URL: https://www.frontiersin.org/research-topics/49519/big-multimedia-data-and-applications.

Bishop, Christopher (jan. de 2006). *Pattern Recognition and Machine Learning*. Vol. 16, pp. 140–155. DOI: 10.1117/1.2819119.

Boyko, Nataliya, Oleg Basystiuk e Nataliya Shakhovska (2018). «Performance Evaluation and Comparison of Software for Face Recognition, Based on Dlib and Opencv Library». Em: *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, pp. 478–482. DOI: 10.1109/DSMP.2018.8478556.

Carrier (jun. de 2005). *Books*. URL: https://github.com/Urinx/Books/blob/master/Forensic/File%20System%20Forensic%20Analysis.pdf.

Cath, Corinne (nov. de 2018). «Governing artificial intelligence: ethical, legal and technical opportunities and challenges». Em: *Philos. Trans. Royal Soc. A* 376.2133. ISSN: 1471-2962. DOI: 10.1098/rsta.2018.0080.

Davenport, Thomas e Ravi Kalakota (jun. de 2019). «The potential for artificial intelligence in healthcare». Em: *Future Healthcare Journal* 6.2, p. 94. DOI: 10.7861/futurehosp.6-2-94.

davisking (set. de 2024). *dlib*. [Online; accessed 24. Sep. 2024]. URL: https://github.com/davisking/dlib/issues/1407.

deepface (set. de 2024). URL: https://github.com/serengil/deepface.

Digital Forensics Used to Help Law Enforcement, Employers Defend Against Cybercrime (jun. de 2024). URL: https://news.erau.edu/headlines/digital-forensics-used-to-help-law-enforcement-employers-defend-against-cybercrime.

Dul, Camilla (2022). «Facial Recognition Technology vs Privacy: The Case of Clearview AI». Em: *QMLJ*, p. 1.

EDI Weekly: Engineered Design Insider (set. de 2024). URL: https://www.ediweekly.com/the-three-different-types-of-artificial-intelligence-ani-agi-and-asi.

Express - Node.js web application framework (jul. de 2024). URL: https://expressjs.com.

Face Recognition in Python - Javatpoint (jul. de 2024). URL: https://www.javatpoint.com/face-recognition-in-python.

FTK Imager (jun. de 2024). URL: https://www.exterro.com/digital-forensics-software/ftk-imager.

Garfinkel, Simson L. (ago. de 2010). «Digital forensics research: The next 10 years». Em: *Digital Invest.* 7, S64–S73. ISSN: 1742-2876. DOI: 10.1016/j.diin.2010.05.009.

Goodfellow, Ian, Yoshua Bengio e Aaron Courville (2016). *Deep Learning*. http://www.deeplearningbook.org. MIT Press.

Graves, Alex, Abdel-rahman Mohamed e Geoffrey Hinton (2013). «Speech recognition with deep recurrent neural networks». Em: *2013 IEEE International*

Conference on Acoustics, Speech and Signal Processing, pp. 6645–6649. DOI: 10.1109/ICASSP.2013.6638947.

Hastie, T., R. Tibshirani e J.H. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer. ISBN: 9780387848846. URL: https://books.google.pt/books?id=eBSgoAEACAAJ.

Hill, Kashmir (nov. de 2021). «The Secretive Company That Might End Privacy as We Know It». Em: *New York Times*. ISSN: 0362-4331. URL: https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html.

James, Joshua I e Pavel Gladyshev (mar. de 2013). «Challenges with Automation in Digital Forensic Investigations». Em.

Jarrett, Aaron e Kim-Kwang Raymond Choo (nov. de 2021). «The impact of automation and artificial intelligence on digital forensics». Em: *WIREs Forensic Sci.* 3.6, e1418. ISSN: 2573-9468. DOI: 10.1002/wfs2.1418.

Javed, Abdul Rehman et al. (nov. de 2021). «A comprehensive survey on digital video forensics: Taxonomy, challenges, and future directions». Em: *Eng. Appl. Artif. Intell.* 106, p. 104456. ISSN: 0952-1976. DOI: 10.1016/j.engappai.2021.104456.

Jiang, Peiyuan et al. (2022). «A Review of Yolo Algorithm Developments». Em: *Procedia Computer Science* 199. The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020/2021): Developing Global Digital Economy after COVID-19, pp. 1066–1073. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2022.01.135. URL: https://www.sciencedirect.com/science/article/pii/S1877050922001363.

Joshua Saxe, Konstantin Berlin (jan. de 2023). «Deep Neural Network Based Malware Detection Using Two Dimensional Binary Program Features». Em: *arXiv*. URL: https://arxiv.org/abs/1508.03096.

Khan, Maliha et al. (2019). *Face Detection and Recognition Using OpenCV*. DOI: 10.1109/ICCCIS48478.2019.8974493.

Lab, CV (2006a). *Facial Recognition Video 1*. EPFL CV Lab. URL: https://documents.epfl.ch/groups/c/cv/cvlab-pom-video1/www/6p-c0.avi.

— (2006b). *Facial Recognition Video 2*. EPFL CV Lab. URL: https://documents.epfl.ch/groups/c/cv/cvlab-pom-video1/www/6p-c1.avi.

— (2006c). *Facial Recognition Video 3*. EPFL CV Lab. URL: https://documents.epfl.ch/groups/c/cv/cvlab-pom-video1/www/6p-c2.avi.

— (2006d). *Facial Recognition Video 4*. EPFL CV Lab. URL: https://documents.epfl.ch/groups/c/cv/cvlab-pom-video1/www/6p-c3.avi.

Lillis, David et al. (2016). «Current Challenges and Future Research Areas for Digital Forensic Investigation». Em: *CoRR* abs/1604.03850. arXiv: 1604.03850. URL: http://arxiv.org/abs/1604.03850.

Medium (fev. de 2022). «An Introduction to Convolutional Neural Network (CNN)». Em: *Medium*. URL: https://medium.com/sfu-cspmp/an-introduction-to-convolutional-neural-network-cnn-207cdb53db97.

— (ago. de 2023). «Recurrent Neural Network (RNN) Architecture Explained». Em: *Medium*. URL: https://medium.com/@poudelsushmita878/recurrent-neural-network-rnn-architecture-explained-1d69560541ef.

Menghani, Gaurav (mar. de 2023). «Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better». Em: *ACM Comput. Surv.* 55.12. ISSN: 0360-0300. DOI: 10.1145/3578938. URL: https://doi.org/10.1145/3578938.

Mitchell, Tom M e Tom M Mitchell (1997). *Machine learning*. Vol. 1. 9. McGraw-hill New York.

Mobile data traffic forecast (set. de 2024). URL: https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/mobile-traffic-forecast.

Nguyen, Thanh Thi et al. (2019). «Deep Learning for Deepfakes Creation and Detection». Em: *CoRR* abs/1909.11573. arXiv: 1909.11573. URL: http://arxiv.org/abs/1909.11573.

Nuix Neo: uma plataforma para todos os seus desafios de dados complexos | Nuix (jul. de 2024). URL: https://www.nuix.com/nuix-neo.

Oriwoh, Edewede et al. (out. de 2013). «Internet of Things Forensics: Challenges and Approaches». Em: DOI: 10.4108/icst.collaboratecom.2013.254159.

Oxygen Forensics (jun. de 2024). «How Artificial Intelligence empowers digital forensics». Em: *Oxygen Forensics*. URL: https://oxygenforensics.com/en/resources/digital-investigations-with-ai.

PhotoREC (mai. de 2023). «Digital Picture and File Recovery». Em: *CGSecurity*. URL: https://www.cgsecurity.org/wiki/PhotoRec.

Rajkomar, Alvin, Jeffrey Dean e Isaac Kohane (2019). «Machine Learning in Medicine». Em: *The New England Journal of Medicine* 380.14, pp. 1347–1358. ISSN: 1533-4406. DOI: 10.1056/NEJMra1814259. URL: https://www.nejm.org/doi/full/10.1056/NEJMra1814259.

Rajpurkar, Pranav (2017). «CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning». Em: *CoRR* abs/1711.05225. arXiv: 1711.05225. URL: http://arxiv.org/abs/1711.05225.

Recent Developments in Forensic Trace Evidence Analysis (jun. de 2024). URL: https://nij.ojp.gov/library/publications/toward-locards-exchange-principle-recent-developments-forensic-trace-evidence.

Redmon, Joseph et al. (2016). «You Only Look Once: Unified, Real-Time Object Detection». Em: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788. DOI: 10.1109/CVPR.2016.91.

Russel e Norvig (mar. de 1996). «Artificial intelligence—a modern approach by Stuart Russell and Peter Norvig, Prentice Hall. Series in Artificial Intelligence, Englewood Cliffs, NJ.» Em: *Knowledge Engineering Review* 11.1, pp. 78–79. ISSN: 1469-8005. DOI: 10.1017/S0269888900007724.

Santos, Tomás Moreira (set. de 2024). *FEUP - Real-Time Weapon Detection in Surveillance Footage*. URL: https://sigarra.up.pt/feup/en/pub_geral.pub_view?pi_pub_base_id=616376.

Shandilya, Nimish (2021). *Car Number Plate Video Dataset*. Kaggle. URL: https://www.kaggle.com/datasets/nimishshandilya/car-number-plate-video.

Spencer, Fm (dez. de 2018). «Digital Forensics with Artificial Intelligence Internet of Things». Em: DOI: 10.13140/RG.2.2.36612.17280.

Sutskever, Ilya, Oriol Vinyals e Quoc V. Le (2014). *Sequence to Sequence Learning with Neural Networks*. arXiv: 1409.3215 [cs.CL]. URL: https://arxiv.org/abs/1409.3215.

Sutton, Richard S (2018). «Reinforcement learning: An introduction». Em: *A Bradford Book*.

Sveyek (2021). *Video ANPR*. GitHub Repository. URL: https://github.com/sveyek/Video-ANPR.

Taigman (2014). «DeepFace: Closing the Gap to Human-Level Performance in Face Verification». Em: *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708. DOI: 10.1109/CVPR.2014.220.

Talaei Khoei, Tala, Hadjar Ould Slimane e Naima Kaabouch (nov. de 2023). «Deep learning: systematic review, models, challenges, and research directions». Em: *Neural Computing and Applications* 35.31, pp. 23103–23124. ISSN: 1433-3058. DOI: 10.1007/s00521-023-08957-4. URL: https://doi.org/10.1007/s00521-023-08957-4.

Tang, Duyu e Ting Liu (jan. de 2015). «Document Modeling with Gated Recurrent Neural Network for Sentiment Classification». Em: pp. 1422–1432. DOI: 10.18653/v1/D15-1167.

Turing, A. M. (out. de 1950). «I.—COMPUTING MACHINERY AND INTELLIGENCE». Em: *Mind* LIX.236, pp. 433–460. ISSN: 0026-4423. DOI: 10.1093/mind/LIX.236.433. eprint: https://academic.oup.com/mind/article-

pdf/LIX/236/433/30123314/lix-236-433.pdf. URL: https://doi.org/10.1093/mind/LIX.236.433.

Ultralytics (jul. de 2024a). *Casa*. URL: https://docs.ultralytics.com/pt#yolo-licenses-how-is-ultralytics-yolo-licensed.

— (set. de 2024b). *YOLO Métricas de desempenho*. URL: https://docs.ultralytics.com/pt/guides/yolo-performance-metrics/#class-wise-metrics.

Upadhyay, Saurabh e Sanjay Kumar Singh (nov. de 2020). «Video Authentication-An Overview». Em: *International Journal of Computer Science & Engineering Survey* 2.4. ISSN: 0976-3252. DOI: 10.5121/ijcses.2011.2406.

VeraCrypt (jun. de 2024). URL: https://www.veracrypt.fr/code/VeraCrypt.

Volatility (jun. de 2024). *volatility3*. URL: https://github.com/volatilityfoundation/volatility3.

Vue.js (jul. de 2024). URL: https://vuejs.org/guide/introduction.html.

Wallach, Izhar, Michael Dzamba e Abraham Heifets (2015). «AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery». Em: *CoRR* abs/1510.02855. arXiv: 1510.02855. URL: http://arxiv.org/abs/1510.02855.

Warsi, Arif et al. (jan. de 2020). *Automatic Handgun and Knife Detection Algorithms: A Review*. DOI: 10.1109/IMCOM48794.2020.9001725.

What are Convolutional Neural Networks? | IBM (set. de 2024). URL: https://www.ibm.com/topics/convolutional-neural-networks.

What Is a Recurrent Neural Network (RNN)? | IBM (set. de 2024). URL: https://www.ibm.com/topics/recurrent-neural-networks.

What is Flask Python - Python Tutorial (jul. de 2024). URL: https://pythonbasics.org/what-is-flask-python.

Wireshark (jun. de 2024). URL: https://www.wireshark.org.

XRY Mobile Data Forensic Phone Extraction Recovery (mar. de 2024). URL: https://www.msab.com/product/xry-extract.

YouTube (2022a). *Facial Recognition Video 1*. YouTube. URL: https://www.youtube.com/watch?v=M4NecQmbVuk.

— (2022b). *Facial Recognition Video 2*. YouTube. URL: https://www.youtube.com/watch?v=9wxEmqyVlB8.

— (2022c). *Knife Video 1*. YouTube. URL: https://www.youtube.com/watch?v=wb2oh2skrHA&t.

— (2022d). *Knife Video 2*. YouTube. URL: https://www.youtube.com/watch?v=DYA8fkZAOFs&t.

— (2022e). *Pistol Video 1*. YouTube. URL: https://www.youtube.com/watch?v=2xdtZ-ZksDc.

— (2022f). *Pistol Video 2*. YouTube. URL: https://www.youtube.com/watch?v=LWnx3u8gDs0.

— (2022g). *Rifle Video 1*. YouTube. URL: https://www.youtube.com/watch?v=bqoJyCWNDYo&t.

— (2022h). *Rifle Video 2*. YouTube. URL: https://www.youtube.com/watch?v=Zczmz4850wM&t.

APÊNDICES

A

APÊNDICE A

A.1 ARQUITETURA

Este capítulo apresenta o protótipo desenvolvido em detalhe, integrando tecnologias de IA para análise de dados de vídeo e dando um indício do que pode ser feito para ajudar os investigadores forenses. A solução é implementada utilizando uma interface Web construída sobre a estrutura Vue.js, que interage com dois scripts Python responsáveis por alojar modelos de IA YOLOv5 (para deteção de objetos) e FaceRecognition (para reconhecimento de indivíduos).

A.1.1 *YOLOV5*

Script de python permite aos utilizadores enviar vídeos para deteção de objetos. Ele processa os vídeos usando o YOLOv5, filtra as caixas delimitadoras de saída dos resultados do modelo e, em seguida, padroniza e retorna os dados no formato JSON.

A.1.1.1 *Dependências*

1. **Flask and Flask-CORS**: Usado para criar a aplicação web e tratar a partilha de recursos entre origens (CORS).
2. **Torch**: Usado para carregar o modelo do YOLOV5
3. **CV2 OpenCV**: Usado para processamento de vídeo
4. **JSON & OS**: Usado para tratamento de ficheiros e operações JSON

A.1.1.2 *Configuração da aplicação*

1. **Inicialização da aplicação Flask**:

```
app = Flask(\_\_name\_\_)
```

2. **Ativação do CORS**:

```
CORS(app)
```

A.1.1.3 *Endpoints*

1. **/process_video [POST]**
   - Extrai dados JSON do pedido, incluindo:
     - **videoPath**: Diretório de vídeos a processar.
     - **investigationName**: Nome para o arquivo JSON de saída.
     - **investigationOutput**: Diretório para o arquivo de saída.
     - **confidentiality**: Limite de confiança para a deteção de objetos.
     - **selected_classes**: Classes a serem incluídas na deteção.
     - **frames**: Intervalo de fotogramas para processamento
   - Chama a função '*main*' para processar vídeos.
2. **/available_classes [GET]**
   - Carrega o modelo YOLOv5 e recupera os nomes das classes.
   - Devolve os nomes das classes como uma resposta JSON, para que possam ser selecionados na aplicação Web

A.1.1.4 *Funções auxiliares*

1. **load_yolov5_model()**:
   - Carrega o modelo YOLOv5 a partir de pesos especificados.
2. **detect_objects(model, frame)**:
   - Efetua a deteção de objetos num determinado fotograma de vídeo utilizando o modelo YOLOv5.
   - Devolve os resultados da deteção.
3. **process_video(video_path, yolo_model, confidence_threshold, selected_classes, frame_interval=3)**:
   - Processa um ficheiro de vídeo:

a) Abre o vídeo e lê as *frames*.

b) Realiza a deteção de objetos em intervalos de *frames* especificados.

c) Filtra as deteções com base no limite de confiança e nas classes selecionadas.

d) Desenha caixas delimitadoras e rótulos nas *frames*.

e) Armazena os resultados da deteção através da data e hora.

4. **main(folder_path, json_file_name, output_directory, confidence_threshold, selected_classes, frame_interval=3)**
   - Carrega o modelo YOLOv5.
   - Processa todos os vídeos MP4 na pasta especificada.
   - Coleciona e guarda os resultados da deteção num ficheiro JSON.

A.1.2 *Face Recognition*

Em resumo, este script cria um serviço web para reconhecimento facial por vídeo. Ele recebe vídeos como entrada, processa-os usando a biblioteca face_recognition, aprimora as *frames*, deteta e identifica faces e salva a saída em formato JSON.

A.1.2.1 *Dependências*

1. **face_recognition**: deteção e reconhecimento de faces
2. **os, sys, cv2, numpy, math, time, json e datetime**: acesso a funções utilitárias.
3. **Flask e Flask-CORS**: para criar a aplicação web e lidar com o CORS (Cross-Origin Resource Sharing).

A.1.2.2 *Configuração da aplicação*

1. **Inicialização da aplicação Flask**:

```
app = Flask(\_\_name\_\_)
```

2. **Ativação do CORS**:

```
CORS(app)
```

A.1.2.3 *Endpoints*

1. **/start-face-recognition [POST]**:
    - Extrai dados JSON do pedido, incluindo:
        - **videoPath**: Diretório de vídeos a processar.
        - **investigationName**: Nome para o arquivo JSON de saída.
        - **investigationOutput**: Diretório para o arquivo de saída.
        - **confidentiality**: Limite de confiança para a deteção de objetos.
        - **frames**: Intervalo de fotogramas para processamento
        - **upscale**: Escala dos frames a processar
    - Inicializa um objeto FaceRecognition e processa cada vídeo na pasta especificada.
    - Evoca a função save_summary para guardar o resumo dos vídeos processados.

A.1.2.4 *Funções auxiliares*

1. **face_confidence(face_distance, face_match_threshold=0.6)**
    - Calcula e devolve uma pontuação de confiança para o reconhecimento facial com base na distância do rosto.
2. **enhance_frame(frame)**
    - Melhora a *frame* de vídeo redimensionando, aplicando equalização de histograma, nitidez e redução de ruído.

A.1.2.5 *Classe FaceRecognition*

Esta classe trata da funcionalidade principal de reconhecimento facial.

1. **Atributos**:
    - **Parâmetros de inicialização**: investigation_id, save_directory, selected_frames, selected_confidentiality, selected_upscale

- **Outros parâmetros**: face_locations, face_encodings, face_names, known_face_encodings, known_face_names, process_current_frame, directory_name, last_recognized, processed_videos

2. **Métodos**:

   a) ***init()***:

      - Inicializa a classe com os parâmetros fornecidos.
      - Cria um diretório para guardar os resultados e codifica a face conhecida

   b) ***create_directory()***:

      - Cria um diretório para guardar imagens processadas e devolve o seu caminho.

   c) ***encode_faces()***:

      - Codifica faces de imagens armazenadas no diretório 'faces' e armazena-as em known_face_encodings e known_face_names.

   d) ***run_recognition(video_path)***:

      - Processa um ficheiro de vídeo para reconhecimento facial.
      - Melhora os *frames*, deteta rostos, associa-os a rostos conhecidos e guarda os fotogramas com rostos reconhecidos.

   e) ***save_summary()***:

      - Salva um resumo dos vídeos processados num arquivo JSON.

A.1.3 *Node.JS*

Este script Node.js cria um servidor que pode listar arquivos, servir imagens, vídeos e arquivos JSON e interagir com o VLC media player. O servidor também é configurado com CORS para permitir solicitações de uma origem especificada e inclui algumas verificações de segurança para garantir que apenas os tipos de arquivo permitidos sejam servidos.

A.1.3.1 *Dependências*

1. **express**: Criar servidor web

2. **cors**: Para lidar com a partilha de recursos entre origens (CORS)
3. **fs**: Para operações do sistema de ficheiros.
4. **path**: Para lidar com caminhos de ficheiros e diretórios.
5. **child_process**: Para execução de comandos shell.

A.1.3.2 *Configuração da aplicação*

1. **Inicializar aplicação express**:

```
const app = express()
```

2. **Definir a porta de serviço**:

```
const port = 3000;
```

3. **Ativação de *parsing* JSON e o CORS**:

```
app.use(express.json());
app.use(cors({ origin: 'http://localhost:5173' }));
```

A.1.3.3 *Endpoints*

1. **/start-face-recognition [POST]**:
   - ***/files [GET]***
     - Finalidade: Listar os ficheiros no caminho especificado.
     - Parâmetro de consulta: *path*.
     - Resposta: Objeto JSON que contem o nome respetivos ficheiros.
   - ***/get-json-files [GET]***
     - Finalidade: Listar os ficheiros JSON no caminho especificado.
     - Parâmetro de consulta: *path*.
     - Resposta: Objeto JSON com nomes dos ficheiros JSON presentes.
   - ***/read-json [GET]***
     - Finalidade: Ler e devolver o conteúdo de um ficheiro JSON especificado.

  - Parâmetro de consulta: *path*.
  - Resposta: Conteúdo JSON presente no ficheiro.

- ***/image [GET]***
  - Finalidade: Fazer o *localhosting* de uma imagem para que esta possa ser acessível pela aplicação web.
  - Parâmetro de consulta: *path*.
  - Resposta: Envia o ficheiro de imagem se este existir e for de um tipo permitido.

- ***/video [GET]***
  - Finalidade: Fazer o *localhosting* de um video para que este possa ser acessível pela aplicação web.
  - Parâmetro de consulta: *path*.
  - Resposta: Envia o ficheiro de vídeo se este existir e for de um tipo permitido.

- ***/json [GET]***
  - Finalidade: Fazer o *localhosting* de um ficheiro JSON para que este possa ser acessível pela aplicação web.
  - Parâmetro de consulta: *path*.
  - Resposta: Envia o ficheiro JSON se existir e for de um tipo permitido.

- ***/open-vlc [GET]***
  - Finalidade: Abre um ficheiro de vídeo através do VLC em um timestamp específico.
  - Parâmetro de consulta: *timestamp*, *filePath*.
  - Resposta: Envia uma mensagem indicando se o VLC foi aberto com sucesso.

- ***/open-vlc2 [POST]***
  - Finalidade: Abre um ficheiro de vídeo no VLC com legendas geradas a partir dos *timestamps* fornecidos.
  - Parâmetro de consulta: *timestamps*, *filePath*.
  - Resposta: Envia uma mensagem indicando se o VLC foi aberto com sucesso.

A.1.3.4 *Funções auxiliares*

1. **readDirectories(dirPath)**
    - Lê diretorias e devolve uma lista de ficheiros para cada diretoria, filtrando opcionalmente pela extensão do ficheiro.
2. **allowedImageFile(filePath)**
    - Verifica se o ficheiro é um tipo de imagem permitido (.png, .jpg, .jpeg).
3. **allowedVideoFile(filePath)**
    - Verifica se o ficheiro é de um tipo de vídeo permitido (.mp4, .webm, .ogg).
4. **allowedJsonFile(filePath)**
    - Verifica se o ficheiro é um ficheiro JSON.
5. **generateSRTContent(timestamps)**
    - Gera legendas SRT a partir de um array de timestamps.
6. **formatTime(time)**
    - Formata o tempo para o arquivo SRT.
7. **padZero(number)**
    - Preenche números de um dígito com um zero à esquerda.

B

APÊNDICE B

B.1 CRIAÇÃO DE *DATASETS*

A criação de *datasets* para proceder ao treino de modelos de IA capazes de analisar vídeos forenses demonstrou-se uma tarefa importante e difícil. De seguida serão demonstrados os três métodos utilizados ao longo de todo o processo de desenvolvimento da aplicação, sendo que mesmo utilizando o último método que irá ser identificado seria necessário recorrer ao melhoramento do *dataset* como descrito na secção dos desafios encontrados ao longo do desenvolvimento.

1. **Experimentos com *datasets* Pré-Criados**: Inicialmente, o método utilizado foi recorrer ao uso de *datasets* já existentes disponíveis nomeadamente em sites como o roboflow e também o kaggle. No entanto, esses *datasets* mostraram-se inadequados para as necessidades específicas deste projeto. As principais limitações encontradas neste método foram as seguintes:

   - **Pouca Diversidade de Classes**: Os *dataset*s disponíveis geralmente incluíam um número limitado de classes, o que restringia a capacidade do modelo de IA de reconhecer uma ampla gama de objetos ou indivíduos relevantes em contextos forenses.
   - **Falta de Inclusividade**: Os *dataset*s não eram inclusivos o suficiente para cobrir os casos específicos que precisavam ser abordados, o que levou a uma baixa precisão e generalização do modelo em cenários reais.

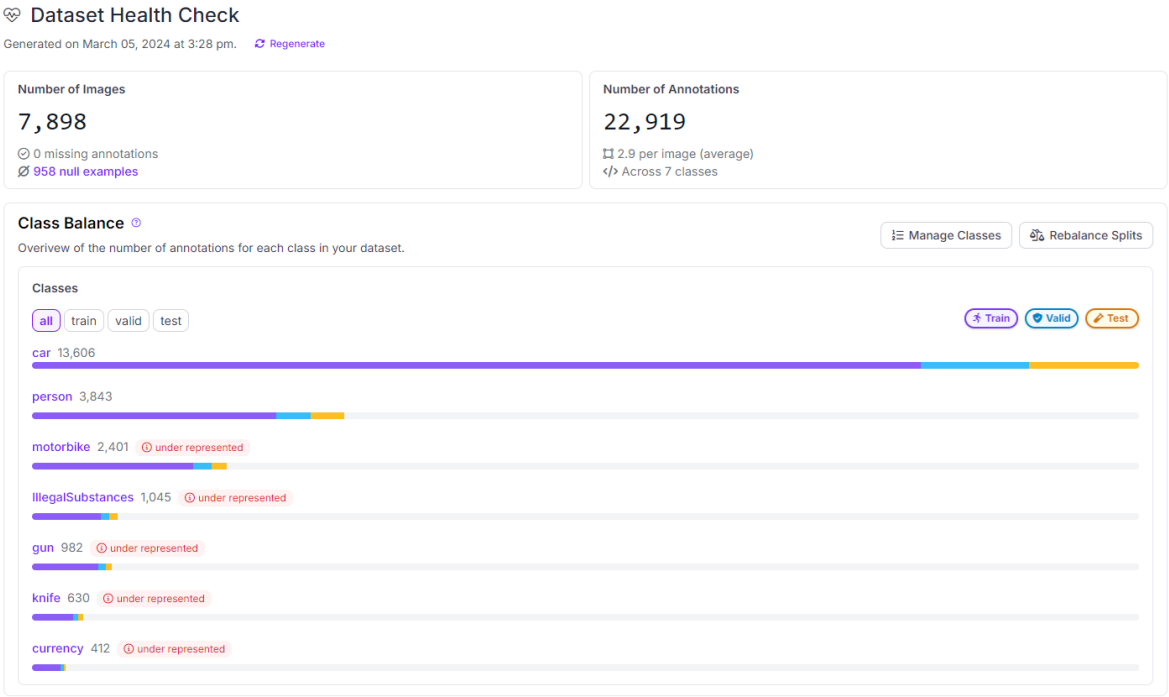
   Devido às limitações apresentadas, foi então necessário explorar outras abordagens para criar um *dataset* mais adequado.

2. **Criação de um *dataset* Manual com o Roboflow**: Como alternativa aos *dataset*s pré-criados, decidi proceder à construção manual de um novo *dataset* utilizando a plataforma Roboflow algo que se mostrou ser um processo bastante moroso e inadequado também para o objetivo final do trabalho a desenvolver. Esse processo envolveu:

- **Marcação Manual de Classes**: Cada imagem do *dataset* teve que ser anotada manualmente, identificando e rotulando os objetos de interesse. Esse processo foi extremamente demorado, pois exigia uma atenção meticulosa para garantir que todas as classes fossem corretamente identificadas.
- **Limitações de Escala**: Devido à natureza morosa do processo, tornou-se inviável criar um *dataset* suficientemente grande para treinar um modelo de IA altamente robusto. O tempo necessário para marcar manualmente um grande número de imagens tornaria o processo impraticável para a criação de um *dataset* em grande escala.

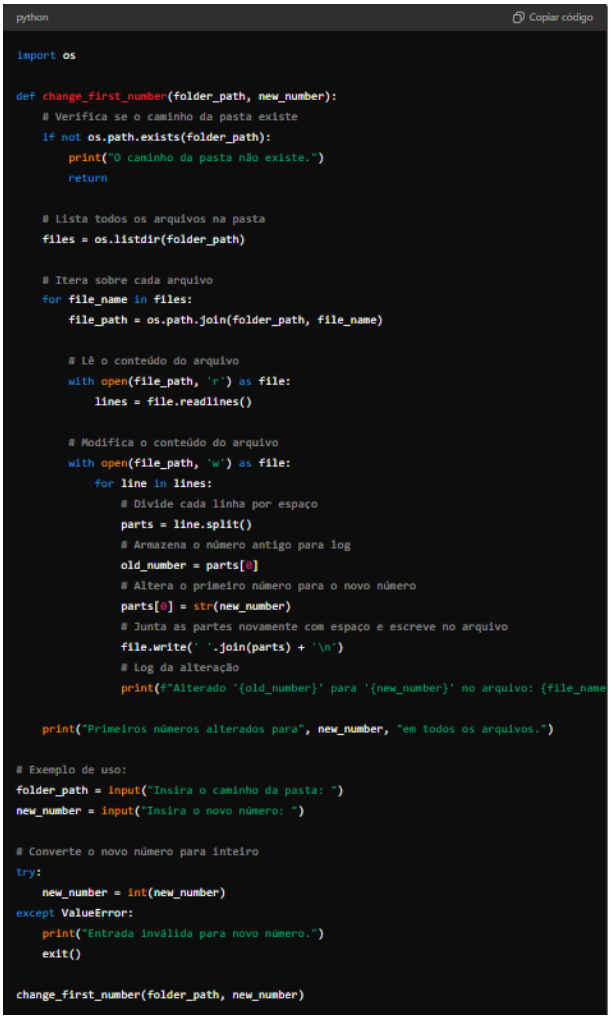
Apesar das dificuldades encontradas, esta abordagem permitiu a criação de um *dataset* personalizado que atendesse às necessidades específicas do projeto, ainda que numa escala limitada.

Figura 20: Criação *dataset* via Roboflow



3. **Uso de Script Python para Unificação de *datasets***: Para superar as limitações mencionadas anteriormente, foi desenvolvido um script Python, descrito abaixo, que permitiu a conversão e unificação de vários *datasets* num único *dataset* coeso. Esse script foi utilizado para alterar a numeração das classes em arquivos de anotação, facilitando a integração de diferentes *datasets* que utilizavam esquemas de rotulagem diferentes.

Figura 21: Python Script *datasets*



```python
import os

def change_first_number(folder_path, new_number):
    # Verifica se o caminho da pasta existe
    if not os.path.exists(folder_path):
        print("O caminho da pasta não existe.")
        return

    # Lista todos os arquivos na pasta
    files = os.listdir(folder_path)

    # Itera sobre cada arquivo
    for file_name in files:
        file_path = os.path.join(folder_path, file_name)

        # Lê o conteúdo do arquivo
        with open(file_path, 'r') as file:
            lines = file.readlines()

        # Modifica o conteúdo do arquivo
        with open(file_path, 'w') as file:
            for line in lines:
                # Divide cada linha por espaço
                parts = line.split()
                # Armazena o número antigo para log
                old_number = parts[0]
                # Altera o primeiro número para o novo número
                parts[0] = str(new_number)
                # Junta as partes novamente com espaço e escreve no arquivo
                file.write(' '.join(parts) + '\n')
                # Log da alteração
                print(f"Alterado '{old_number}' para '{new_number}' no arquivo: {file_name

    print("Primeiros números alterados para", new_number, "em todos os arquivos.")

# Exemplo de uso:
folder_path = input("Insira o caminho da pasta: ")
new_number = input("Insira o novo número: ")

# Converte o novo número para inteiro
try:
    new_number = int(new_number)
except ValueError:
    print("Entrada inválida para novo número.")
    exit()

change_first_number(folder_path, new_number)
```

Esse script foi fundamental para facilitar a integração de vários *datasets*, permitindo a criação de um *dataset* mais diversificado e adequado para o treino dos modelos de IA. A capacidade de modificar automaticamente os rótulos das classes permitiu a unificação de diferentes fontes de dados, superando, assim, algumas das limitações iniciais enfrentadas com os *datasets* pré-criados.

Em suma, esta secção resume as estratégias e ferramentas utilizadas para abordar os desafios relacionados ao treino de IA, destacando as soluções implementadas para otimizar o processo e melhorar os resultados obtidos.

B.2 ANALISE RESULTADOS TREINO YOLOV5

Nesta secção, serão delineadas as principais visualizações e métricas utilizadas para avaliar a eficácia dos modelos de deteção de objetos. O exame enfatiza a compreensão da precisão geral e do desempenho específico da classe, a fim de discernir os pontos fortes e identificar as áreas que precisam de ser melhoradas.

Irão ser investigadas para os modelos treinados as seguintes áreas principais:

Matriz de confusão: A matriz de confusão fornece uma visão detalhada dos resultados, mostrando as contagens de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos para cada classe. Ultralytics, 2024b

Curva F1: Esta curva representa a pontuação F1 em vários limiares. A interpretação desta curva pode oferecer informações sobre o equilíbrio do modelo entre falsos positivos e falsos negativos em diferentes limiares. Ultralytics, 2024b

Resultados de treino/validação: Os resultados de formação e validação fornecem informações sobre o processo de aprendizagem do modelo ao longo do tempo. O acompanhamento das perdas (como a perda de caixa, a perda de classificação) durante a formação permite detetar sinais de sob reajuste ou subajuste. Se for observada estabilidade na convergência destas métricas, o modelo generalizou-se bem. A instabilidade é uma indicação de que devem ser efetuados alguns refinamentos adicionais nas estratégias de formação. A análise detalhada oferece um esquema claro para compreender o desempenho dos modelos e orienta o processo de melhoramento da precisão da deteção para todas as classes. Cada gráfico e medida estatística é útil para estimar a sustentabilidade do modelo, a fiabilidade e as potenciais áreas de melhoria. Ultralytics, 2024b

1. **Modelo 1**
   Classes: Carro, Motociclo, Numerário, Estupefacientes, Arma, Pessoas
   a) **Matriz de confusão**:
      i. **Desempenho**: O modelo é testado num conjunto de dados com várias classes. Embora tenha um bom desempenho para muitas categorias, a matriz de confusão mostra uma diagonal clara, que é positiva.
      ii. **Observações**: Existe uma confusão considerável presente na classe "numerário"sendo que, por exemplo, a distinção entre frente e verso

é difícil para algumas imagens. O modelo também tem dificuldade em distinguir entre as categorias de "estupefacientes"e "arma".

b) **Curva F1**:

i. **Limiar**: A pontuação F1 é de 0,87 com um limiar de confiança de 0,689, o que é um valor relativamente elevado.

ii. **Perceção**: A curva F1 mostra uma variabilidade significativa entre as classes, indicando que algumas categorias têm um desempenho muito melhor do que outras.

c) **Resultados da formação/validação**:

i. **Convergência**: As curvas de perda indicam uma convergência mais lenta devido à complexidade do conjunto de dados, mas acabam por estabilizar. A perda de validação é ligeiramente superior à perda de treino, algo que sugere algum sob reajuste.

ii. **Precisão e Recuperação**: Ambas as métricas mostram melhorias ao longo do tempo, com a precisão a estabilizar-se ligeiramente acima da recuperação.

iii. **Pontuações mAP**: O mAP@0.5 é de aproximadamente 0,83, enquanto o mAP@0.5:0.95 é de cerca de 0,8, o que reflete um desempenho razoável nas previsões.

O modelo 1 destaca os desafios da adaptação a um conjunto de dados heterogéneo e desequilibrado. Embora o modelo funcione bem com certas classes, como “carro”, o seu desempenho é muito fraco em diferenciações mais precisas para classes como, "numerário", "estupefacientes"e "arma".

2. **Modelo 2**
Classes: Estupefacientes, Carro, Numerário, Arma, Faca, Moto, Pessoa

a) **Matriz de confusão**:

i. **Desempenho**: A matriz de confusão revela um bom desempenho na maioria das classes, com uma precisão particularmente elevada para “carro” (97%) e “numerário” (98%).

ii. **Observações**: O modelo funciona razoavelmente bem com algumas classes, como “arma” (80%) e “mota” (95%). No entanto, existe confusão entre categorias semelhantes, como “faca” (35% erradamente classificada como “arma”) e a classe “pessoa”, que tem 22%

erradamente classificada. O modelo também tem dificuldades com "estupefacientes", que apresenta 18% de erros de classificação.

b) **Curva F1**:

i. **Limiar**:A pontuação F1 é de 0,84 com um limiar de confiança de 0,462.

ii. **Perceção**: A curva F1 varia consoante a classe, com "numerário" e "carro" a apresentarem as pontuações F1 mais estáveis e mais elevadas. Em contrapartida, classes como "faca" e "estupefacientes" têm pontuações F1 mais baixas, refletindo a dificuldade em distinguir estas classes.

c) **Resultados da formação/validação**:

i. **Convergência**: As curvas de perda indicam uma convergência estável, com um mínimo de sob reajuste. A perda de validação mantém-se perto à perda de treino, sugerindo que o modelo generaliza bem.

ii. **Precisão e Recuperação**: Ambas as métricas estabilizam acima de 0,8, com a precisão ligeiramente superior à recuperação.

iii. **Pontuações mAP**: O mAP@0.5 é de 0,71, enquanto o mAP@0.5:0.95 é de aproximadamente 0,7, o que reflete um bom desempenho nas previsões.

O modelo 2 tem um desempenho muito mais equilibrado numa vasta gama de classes, com um bom tratamento das classes, como "arma", "faca" e "pessoa". No entanto, este modelo ainda sofre na diferenciação entre categorias relacionadas (por exemplo, "faca" versus "arma") devido ao desequilíbrio das classes.

3. **Modelo 3**
Classes: Pessoa, Veículo, Motociclo, Faca, Pistola, Espingarda, Documentos, Numerário

a) **Matriz de Confusão**:

i. **Desempenho**: A matriz de confusão indica um desempenho variado entre as classes. As classes "documentos" e "numerário" apresentam uma precisão elevada (87% e 99%, respetivamente), enquanto as classes "faca" e "pistola" apresentam maior confusão.

ii. **Observações**: A classe "motociclo" apresenta erros de classificação significativos e a classe "faca" é erradamente classificada como "pistola" em 20% das vezes.

b) **Curva F1**:

i. **Limiar**: A pontuação F1 é de 0,82 com um limiar de confiança de 0,336.

ii. **Perceção**: As curvas F1 mostram uma distribuição significativa do desempenho entre as classes, com as classes "numerário" e "documentos" a obterem as pontuações mais elevadas. Por outro lado, a classe "motociclo" tem o desempenho F1 mais fraco, o que indica que é difícil distinguir esta classe das outras.

c) **Resultados da formação/validação**:

i. **Convergência**: As curvas de perda indicam convergência, embora a perda de validação mostre algum ruído e potencial sob reajuste, particularmente na perda de objetos.

ii. **Precisão e Recuperação**: A precisão estabiliza em torno de 0,75, enquanto a recuperação fica um pouco atrás, principalmente no início do treino.

iii. **Pontuações mAP**: O mAP@0.5 é de 0,52 e o mAP@0.5:0.95 é de 0,5, refletindo um desempenho moderado com uma queda notável na capacidade de deteção de precisão.

O modelo 3 demonstra um bom desempenho em determinadas classes, como "numerário" e "espingarda", mas tem dificuldades significativas noutras, como "motociclo" e "faca". O modelo é afetado pelo desequilíbrio de classes e pela confusão entre classes visualmente semelhantes (por exemplo, "faca" e "pistola").

4. **Modelo 4**
Classes: Matrícula, Faca, Pistola, Espingarda, Estupefacientes, Numerário

a) **Matriz de Confusão**:

i. **Desempenho**: A matriz de confusão indica um bom desempenho para as classes "matricula" (98%) e "numerário" (100%). No entanto, o modelo mostra uma confusão significativa na distinção entre as classes "pistola" e "espingarda", em que 23% das instâncias de

"pistola" são erradamente classificadas como "espingarda" e 21% das instâncias de "espingarda" são erradamente classificadas como "pistola".

ii. **Observações**: O modelo tem um melhor desempenho na deteção de "matricula" e "numerário", enquanto tem dificuldades em diferenciar "pistola" e "espingarda".

b) **Curva F1**:

i. **Limiar**: A pontuação F1 ideal é 0,91 com um limite de confiança de 0,237.

ii. **Perceção**: As curvas F1 mostram que classes como "matricula" e "numerário" atingem pontuações elevadas numa vasta gama de limiares de confiança. Em contraste, "pistola" e "espingarda" mostram mais variabilidade, com as pontuações F1 a cair rapidamente à medida que a confiança aumenta, refletindo a confusão destacada na matriz de confusão.

c) **Resultados de treino/validação**:

i. **Convergência**: As curvas de perda demonstram uma convergência suave, com as perdas de treino e validação a diminuírem de forma constante. A perda de objetos apresenta algum ruído, mas, em geral, o modelo apresenta uma boa convergência sem sinais de sob reajuste significativo.

ii. **Precisão e Recuperação**: A precisão estabiliza em torno de 0,9, enquanto a recuperação fica um pouco atrás. O modelo parece favorecer ligeiramente a precisão, levando a melhores resultados para classes com alta representação (por exemplo, "matricula").

iii. **Pontuações mAP**: O mAP@0.5 é de 0,71 e o mAP@0.5:0.95 é de 0,69, refletindo um desempenho sólido com uma pequena queda na deteção.

O modelo 4 apresenta um excelente desempenho na deteção das classes "matricula" e "numerário", mas mostra alguma confusão entre "pistola" e "espingarda". O modelo é robusto na maioria das classes, mas poderia melhorar caso a confusão em classes de aparência semelhante fosse resolvida.

5. **Modelo 5**

   Classes: Matrícula, Faca, Pistola, Espingarda, Estupefacientes, Numerário

   a) **Matriz de Confusão**:

      i. **Desempenho**: A matriz de confusão mostra um desempenho variado entre as classes. As classes "matricula" e "estupefacientes" apresentam uma precisão muito elevada (98% e 100%, respetivamente).

      ii. **Observações**: A classe "pistola" é particularmente problemática, com uma sobreposição significativa de previsões para "espingarda". O modelo também tem dificuldade em distinguir entre "faca" e "pistola", o que leva a erros de classificação.

   b) **Curva F1**:

      i. **Limiar**: A pontuação F1 ideal é 0,92 com um limite de confiança de 0,250.

      ii. **Perceção**: As curvas F1 demonstram um forte desempenho para as classes "matricula" e "estupefacientes". No entanto, as classes "pistola" e "espingarda" apresentam um desempenho F1 mais fraco, o que indica desafios na distinção dessas classes das outras, particularmente em níveis de confiança mais altos.

   c) **Resultados da formação/validação**:

      i. **Convergência**: As curvas de perda mostram uma convergência consistente, embora a perda de validação ainda apresente algum ruído.

      ii. **Precisão e Recuperação**: A precisão estabiliza em torno de 0,90, com a recuperação ligeiramente atrás, especialmente nas primeiras fases de treino.

      iii. **Pontuações mAP**: O mAP@0.5 é de 0,71 e o mAP@0.5:0.95 é de 0,65, o que reflete um bom desempenho geral, mas indica desafios na deteção de determinadas classes.

   O modelo 5 consegue um bom desempenho em muitas áreas, apresentando ainda espaço para aperfeiçoamento, particularmente no tratamento de desequilíbrios de classe e na redução de erros de classificação em categorias estreitamente relacionadas.

6. **Modelo 6**

Classes: Matrícula, Faca, Pistola, Espingarda, Estupefacientes, Numerário

a) **Matriz de Confusão**:

i. **Desempenho**: A matriz de confusão indica um bom desempenho global, particularmente para as classes "matrícula" (98%) e "estupefacientes" (100%). As classes "faca" e "pistola" revelam alguma confusão, com erros de classificação moderados entre elas e as outras classes.

ii. **Observações**: Existe uma confusão notável entre as classes "pistola" e "espingarda", com 22% das instâncias de "pistola" a serem incorretamente previstas como "espingarda".

b) **Curva F1**:

i. **Limiar**: A pontuação F1 ótima é de 0,92 com um limiar de confiança de 0,274.

ii. **Perceção**: As curvas F1 revelam um desempenho consistente na maioria das classes, com "estupefacientes" e "numerário" a obterem pontuações quase perfeitas. No entanto, as classes "pistola" e "espingarda" apresentam maior variação nas suas pontuações F1, indicando dificuldade em distingui-las com precisão em determinados níveis de confiança.

c) **Resultados do treino/validação:**

i. **Convergência**: As curvas de perda indicam uma convergência suave, com as perdas de validação a estabilizarem à medida que a formação progride. O modelo mostra sinais mínimos de sob reajuste, e as perdas de treino e validação alinham-se estreitamente, indicando um modelo regularizado.

ii. **Precisão e Recuperação**: Tanto a precisão como a recuperação atingem cerca de 90%, com um desempenho estável ao longo do processo de formação. A ligeira diferença entre a precisão e a recuperação sugere que, embora o modelo dê prioridade às deteções exatas, ainda há espaço para melhorias na redução das deteções falhadas.

iii. **Pontuações mAP**: O mAP@0.5 é 0,70, enquanto o mAP@0.5:0.95 permanece em 0,67, refletindo um desempenho decente numa gama de limiares de IoU.

O modelo 6 tem uma eficácia elevada, uma vez que apresenta bons resultados na identificação da classe "matrículas" e "estupefacientes", embora tenha dificuldades na diferenciação das classes "pistola" e "espingarda", que são semelhantes em termos de aparência. Apresenta uma consistência muito boa durante o processo de treino, com um sob reajuste limitado e uma convergência acentuada.

7. **Modelo 7**

Classes: Matrícula, faca, pistola, espingarda, estupefacientes, numerário, telemóvel, cartão de crédito

a) **Matriz de Confusão**:

i. **Desempenho**: A matriz de confusão revela um forte desempenho na identificação das classes "matrícula" (98%) e "numerário" (100%).

ii. **Observações**: A classe "pistola" tem um problema significativo de classificação incorreta, bem como também a classe "espingarda" que é incorretamente classificada em 22% das vezes. Existe também uma confusão notória entre as classes "faca" e "arma de fogo".

b) **Curva F1**:

i. **Limiar**: A pontuação F1 ótima é de 0,92 com um limiar de confiança de 0,338.

ii. **Perceção**: As curvas F1 mostram que as classes "numerário" e "estupefacientes" mantêm pontuações consistentemente elevadas, enquanto "pistola" e "espingarda" apresentam quedas significativas no desempenho F1, indicando desafios nestas classes.

c) **Resultados do treinamento/validação**:

i. **Convergência**: As curvas de perda demonstram convergência, embora haja algum ruído na perda de objetos durante a validação, o que pode estar a levar a uma classificação incorreta de classes mais pequenas ou menos frequentes.

ii. **Precisão e Recuperação**: A precisão estabiliza em torno de 0,90, com a recuperação mostrando tendências semelhantes, mas com um

pouco mais de variabilidade. A diferença entre estas métricas sugere que, ocasionalmente, falham deteções em determinadas classes.

iii. **Pontuações mAP**: O mAP@0.5 é 0,73 e o mAP@0.5:0.95 é 0,71, refletindo um desempenho consistente em todo o conjunto de dados, mas ainda com espaço para melhorias, especialmente em classes como “pistola” e “espingarda”.

O modelo 7 demonstra um bom desempenho geral, particularmente em classes como “matrícula”, “numerário” e “estupefacientes”. No entanto, tem dificuldades na deteção de “arma de fogo” e “espingarda”, provavelmente devido à confusão e desequilíbrios entre classes.

8. **Modelo 8**

Classes: Faca, pistola, espingarda, matricula

a) **Matriz de Confusão**:

i. **Desempenho**: A matriz de confusão indica que o modelo tem um bom desempenho em algumas classes. Especificamente, a classe "matricula"apresenta uma precisão muito alta (98%), seguida por "espingarda"com 93% e "pistola"com 89%.

ii. **Observações**: O modelo enfrenta dificuldades com a classe "faca", que apresenta 36% de erros.

b) **Curva F1**:

i. **Limiar**: A pontuação F1 geral é de 0,88, com um limiar de confiança de 0,236.

ii. **Perceção**: As curvas F1 variam consideravelmente entre as classes. "matricula", "espingarda"e "pistola"mostram as curvas F1 mais estáveis e altas, enquanto "faca"apresenta uma curva F1 mais baixa, refletindo a dificuldade do modelo em distinguir corretamente essa classe.

c) **Resultados do treinamento/validação**:

i. **Convergência**: As curvas de perda sugerem uma convergência estável, com mínima diferença entre as perdas de treino e validação.

ii. **Precisão e Recuperação**: Ambas as métricas estabilizam-se acima de 0,8, com a precisão ligeiramente superior à recuperação.

iii. **Pontuações mAP**: O mAP@0.5 é de 0,71, enquanto o mAP@0.5:0.95 é de aproximadamente 0,7, algo que sugere um desempenho aceitável

e consistente em prever corretamente as classes, embora apresente alguma dificuldade com classes menos representadas ou mais ambíguas.

O Modelo 8 tem um desempenho razoavelmente bom, ilustrando um equilíbrio adequado entre as classes "matricula", "espingarda" e "pistola". No entanto, continua a enfrentar desafios significativos na distinção entre classes semanticamente semelhantes ou visualmente semelhantes, que é o caso da classe "faca".

A evolução do modelo da v1 para a v8 sublinha o aumento gradual da complexidade tanto dos conjuntos de dados como dos próprios modelos. De um conjunto de dados geral, capaz de diferenciar as classes o mais amplamente possível, o processo passou a consistir em ultrapassar alguns desafios para obter um conjunto de dados equilibrado e reduzir o número de classes devido a restrições de tempo, na ótica de que isso contribuísse para o desenvolvimento de um modelo mais fiável. À medida que os conjuntos de categorias cresciam de categorias simples como "carros" para categorias muito mais pormenorizadas, como "numerário", "armas" e "estupefacientes", também cresciam as dificuldades das tarefas para as quais os modelos estavam a ser testados. Esta dificuldade acrescida deu origem a muitas classificações erradas entre categorias visualmente semelhantes sobretudo a partir da versão 3, em que "pistola" e "espingarda" eram normalmente mal identificadas. O principal problema foi o desequilíbrio de classes sendo que categorias importantes eram bastante dominantes em comparação com algumas sub-representas, apontando assim para a necessidade de escolhas seletivas nas classes para que se obtivesse uma melhoria da representação para um melhor desempenho global do modelo.

APÊNDICE C

C.1 RESULTADOS DOS TESTES

C.1.1 *Reconhecimento de Faces*

Figura 22: Resultados FaceRecognition V1

Facial Recon #1

Frames	FY	FX	Conf.
1	1	1	70

Detections		TP%	FP%
Positive	False		
23	26	46,9388	53,0612

Frames	FY	FX	Conf.
3	1	1	70

Detections		TP%	FP%
Positive	False		
6	10	37,5	62,5

Frames	FY	FX	Conf.
5	1	1	70

Detections		TP%	FP%
Positive	False		
6	4	60	40

Frames	FY	FX	Conf.
7	1	1	70

Detections		TP%	FP%
Positive	False		
6	4	60	40

Frames	FY	FX	Conf.
9	1	1	70

Detections		TP%	FP%
Positive	False		
3	2	60	40

Video Summary: 6 persons walking in a room with a fixed camera

Frames	FY	FX	Conf.
1	1.5	1.5	70

Detections		TP%	FP%
Positive	False		
76	74	50,6667	49,3333

Frames	FY	FX	Conf.
3	1.5	1.5	70

Detections		TP%	FP%
Positive	False		
24	28	46,1538	53,8462

Frames	FY	FX	Conf.
5	1.5	1.5	70

Detections		TP%	FP%
Positive	False		
14	15	48,2759	51,7241

Frames	FY	FX	Conf.
7	1.5	1.5	70

Detections		TP%	FP%
Positive	False		
16	11	59,2593	40,7407

Frames	FY	FX	Conf.
9	1.5	1.5	70

Detections		TP%	FP%
Positive	False		
3	10	23,0769	76,9231

Frames	FY	FX	Conf.
1	2	2	70

Detections		TP%	FP%
Positive	False		
86	82	51,1905	48,8095

Frames	FY	FX	Conf.
3	2	2	70

Detections		TP%	FP%
Positive	False		
25	28	47,1698	52,8302

Frames	FY	FX	Conf.
5	2	2	70

Detections		TP%	FP%
Positive	False		
22	15	59,4595	40,5405

Frames	FY	FX	Conf.
7	2	2	70

Detections		TP%	FP%
Positive	False		
17	10	62,963	37,037

Frames	FY	FX	Conf.
9	2	2	70

Detections		TP%	FP%
Positive	False		
5	10	33,3333	66,6667

Metadata of video's:

6p-c0	Image Size : 360x288
	Megapixels : 0.104
6p-c1	Image Size : 360x288
	Megapixels : 0.104
6p-c2	Image Size : 360x288
	Megapixels : 0.104
6p-c3	Image Size : 360x288
	Megapixels : 0.104

Figura 23: Resultados FaceRecognition V2

Test Facial Recon Part 2

Frames	FY	FX	Conf.
1	1	1	70

Detections		TP%	FP%
Positive	False		
195	33	85,52632	14,47368

Frames	FY	FX	Conf.
3	1	1	70

Detections		TP%	FP%
Positive	False		
64	12	84,21053	15,78947

Frames	FY	FX	Conf.
5	1	1	70

Detections		TP%	FP%
Positive	False		
39	8	82,97872	17,02128

Frames	FY	FX	Conf.
7	1	1	70

Detections		TP%	FP%
Positive	False		
26	5	83,87097	16,12903

Frames	FY	FX	Conf.
9	1	1	70

Detections		TP%	FP%
Positive	False		
21	4	84	16

Video Summary: person walking infront o CCTV camera

Frames	FY	FX	Conf.
1	1.5	1.5	70

Detections		TP%	FP%
Positive	False		
189	21	90	10

Frames	FY	FX	Conf.
3	1.5	1.5	70

Detections		TP%	FP%
Positive	False		
65	10	86,66667	13,33333

Frames	FY	FX	Conf.
5	1.5	1.5	70

Detections		TP%	FP%
Positive	False		
38	10	79,16667	20,83333

Frames	FY	FX	Conf.
7	1.5	1.5	70

Detections		TP%	FP%
Positive	False		
27	1	96,42857	3,571429

Frames	FY	FX	Conf.
9	1.5	1.5	70

Detections		TP%	FP%
Positive	False		
22	4	84,61538	15,38462

Frames	FY	FX	Conf.
1	2	2	70

Detections		TP%	FP%
Positive	False		
190	133	58,82353	41,17647

Frames	FY	FX	Conf.
3	2	2	70

Detections		TP%	FP%
Positive	False		
63	45	58,33333	41,66667

Frames	FY	FX	Conf.
5	2	2	70

Detections		TP%	FP%
Positive	False		
49	24	67,12329	32,87671

Frames	FY	FX	Conf.
7	2	2	70

Detections		TP%	FP%
Positive	False		
26	8	76,47059	23,52941

Frames	FY	FX	Conf.
9	2	2	70

Detections		TP%	FP%
Positive	False		
23	7	76,66667	23,33333

Metadata of video's:

sinv_1	Image Size : 1920x1080
	Megapixels : 2.1
sinv_2	Image Size : 1920x1080
	Megapixels : 2.1

C.1.2 *Reconhecimento de Objetos*

C.1.2.1 *Matricula*

Figura 24: Resultados Classe Matricula

License Plate Recognition

Video Summary: compilation of 2 videos

Metadata of video's:

licenset1	Image Size : 3840x2160
	Megapixels : 8.3
licenset2	Image Size : 1280x720
	Megapixels : 0.922

Video 1: CCTV with cars passing by

Frames	Conf.
1	75

Detections		TP%	FP%
Positive	False		
574	6	98,96552	1,034483

Frames	Conf.
3	75

Detections		TP%	FP%
Positive	False		
191	2	98,96373	1,036269

Frames	Conf.
5	75

Detections		TP%	FP%
Positive	False		
114	1	99,13043	0,869565

Frames	Conf.
7	75

Detections		TP%	FP%
Positive	False		
83	1	98,80952	1,190476

Frames	Conf.
9	75

Detections		TP%	FP%
Positive	False		
64	1	98,46154	1,538462

Video 2: Highway camera

Frames	Conf.
1	75

Detections		TP%	FP%
Positive	False		
79	0	100	0

Frames	Conf.
3	75

Detections		TP%	FP%
Positive	False		
28	0	100	0

Frames	Conf.
5	75

Detections		TP%	FP%
Positive	False		
17	0	100	0

Frames	Conf.
7	75

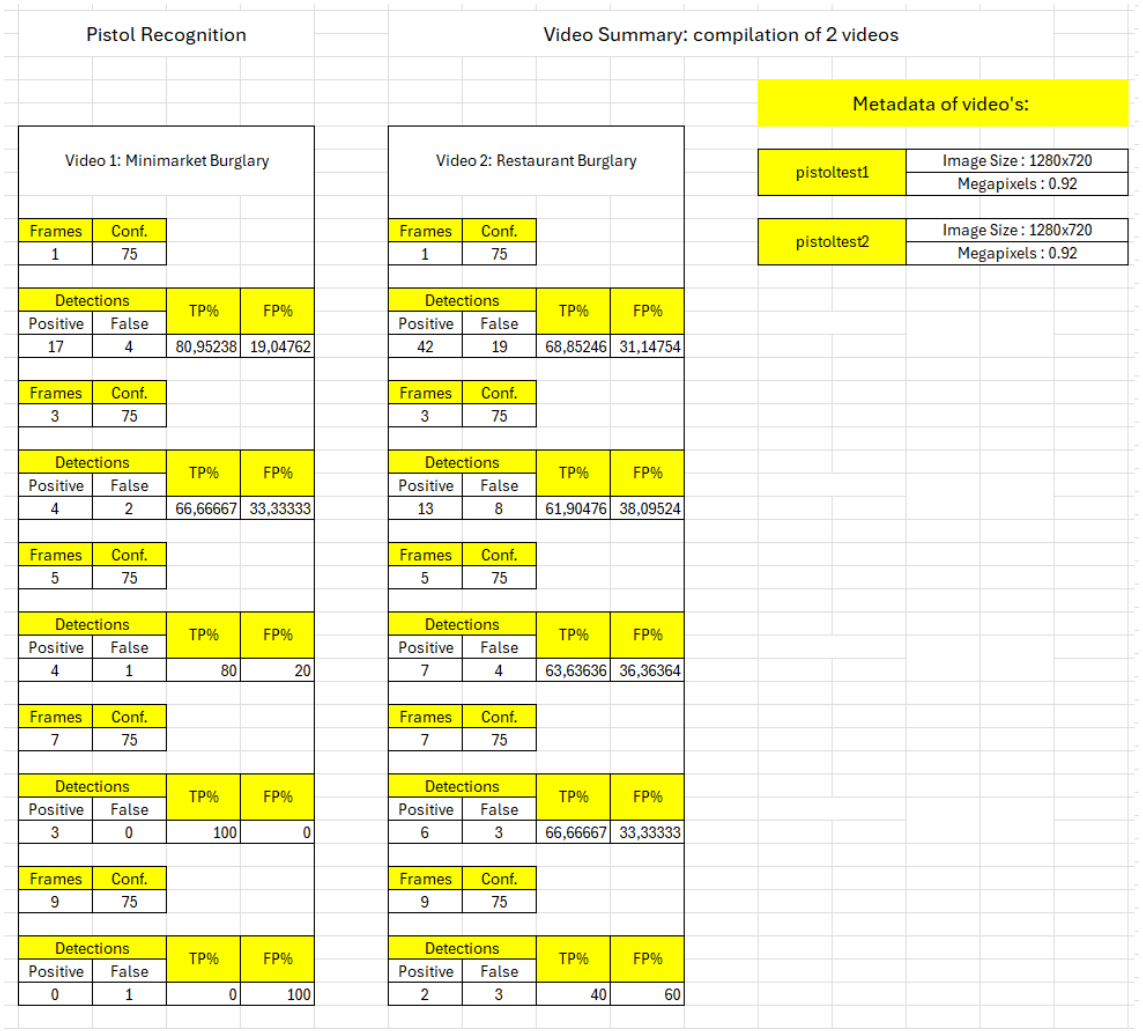
Detections		TP%	FP%
Positive	False		
15	0	100	0

Frames	Conf.
9	75

Detections		TP%	FP%
Positive	False		
10	0	100	0

C.1.2.2 *Pistola*

Figura 25: Resultados Classe Pistola



Pistol Recognition

Video 1: Minimarket Burglary

Frames	Conf.
1	75

Detections		TP%	FP%
Positive	False		
17	4	80,95238	19,04762

Frames	Conf.
3	75

Detections		TP%	FP%
Positive	False		
4	2	66,66667	33,33333

Frames	Conf.
5	75

Detections		TP%	FP%
Positive	False		
4	1	80	20

Frames	Conf.
7	75

Detections		TP%	FP%
Positive	False		
3	0	100	0

Frames	Conf.
9	75

Detections		TP%	FP%
Positive	False		
0	1	0	100

Video Summary: compilation of 2 videos

Video 2: Restaurant Burglary

Frames	Conf.
1	75

Detections		TP%	FP%
Positive	False		
42	19	68,85246	31,14754

Frames	Conf.
3	75

Detections		TP%	FP%
Positive	False		
13	8	61,90476	38,09524

Frames	Conf.
5	75

Detections		TP%	FP%
Positive	False		
7	4	63,63636	36,36364

Frames	Conf.
7	75

Detections		TP%	FP%
Positive	False		
6	3	66,66667	33,33333

Frames	Conf.
9	75

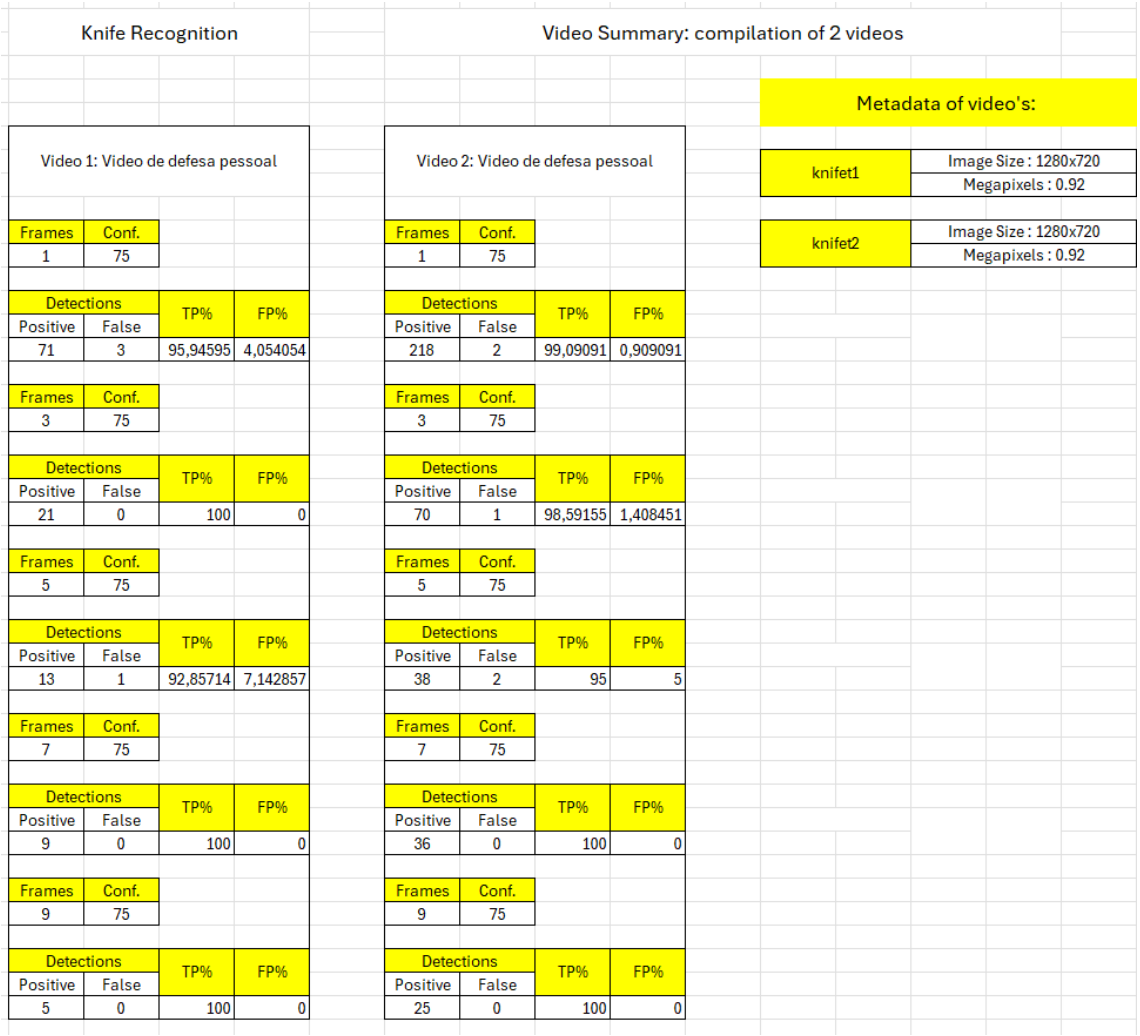
Detections		TP%	FP%
Positive	False		
2	3	40	60

Metadata of video's:

pistoltest1	Image Size : 1280x720
	Megapixels : 0.92
pistoltest2	Image Size : 1280x720
	Megapixels : 0.92

C.1.2.3 *Faca*

Figura 26: Resultados Classe Faca



Knife Recognition

Video 1: Video de defesa pessoal

Frames	Conf.
1	75

Detections		TP%	FP%
Positive	False		
71	3	95,94595	4,054054

Frames	Conf.
3	75

Detections		TP%	FP%
Positive	False		
21	0	100	0

Frames	Conf.
5	75

Detections		TP%	FP%
Positive	False		
13	1	92,85714	7,142857

Frames	Conf.
7	75

Detections		TP%	FP%
Positive	False		
9	0	100	0

Frames	Conf.
9	75

Detections		TP%	FP%
Positive	False		
5	0	100	0

Video Summary: compilation of 2 videos

Video 2: Video de defesa pessoal

Frames	Conf.
1	75

Detections		TP%	FP%
Positive	False		
218	2	99,09091	0,909091

Frames	Conf.
3	75

Detections		TP%	FP%
Positive	False		
70	1	98,59155	1,408451

Frames	Conf.
5	75

Detections		TP%	FP%
Positive	False		
38	2	95	5

Frames	Conf.
7	75

Detections		TP%	FP%
Positive	False		
36	0	100	0

Frames	Conf.
9	75

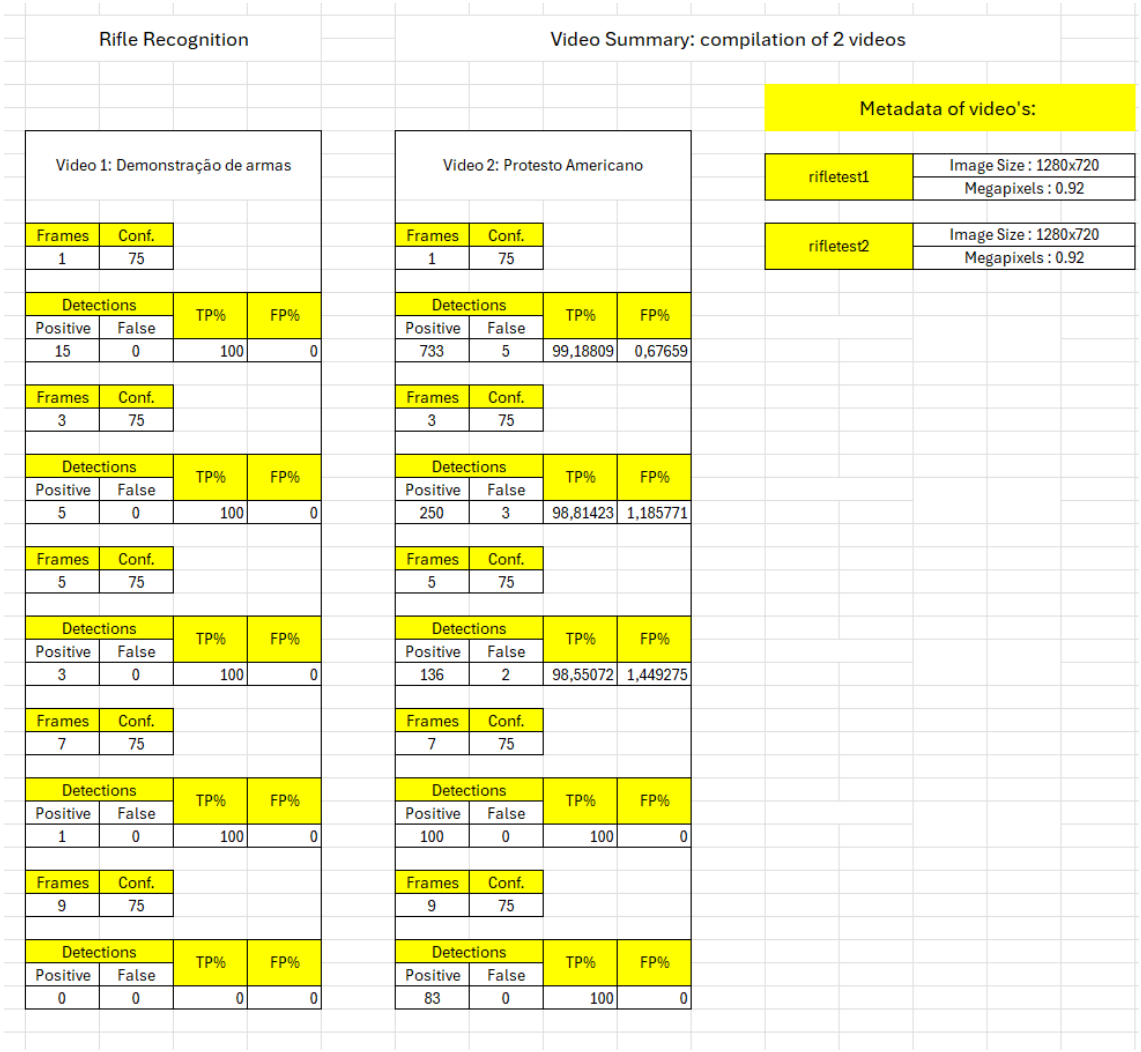
Detections		TP%	FP%
Positive	False		
25	0	100	0

Metadata of video's:

knifet1	Image Size : 1280x720
	Megapixels : 0.92
knifet2	Image Size : 1280x720
	Megapixels : 0.92

C.1.2.4 *Espingarda*

Figura 27: Resultados Classe Espingarda



Rifle Recognition

Video Summary: compilation of 2 videos

Metadata of video's:

rifletest1	Image Size : 1280x720
	Megapixels : 0.92
rifletest2	Image Size : 1280x720
	Megapixels : 0.92

Video 1: Demonstração de armas

Frames	Conf.
1	75

Detections		TP%	FP%
Positive	False		
15	0	100	0

Frames	Conf.
3	75

Detections		TP%	FP%
Positive	False		
5	0	100	0

Frames	Conf.
5	75

Detections		TP%	FP%
Positive	False		
3	0	100	0

Frames	Conf.
7	75

Detections		TP%	FP%
Positive	False		
1	0	100	0

Frames	Conf.
9	75

Detections		TP%	FP%
Positive	False		
0	0	0	0

Video 2: Protesto Americano

Frames	Conf.
1	75

Detections		TP%	FP%
Positive	False		
733	5	99,18809	0,67659

Frames	Conf.
3	75

Detections		TP%	FP%
Positive	False		
250	3	98,81423	1,185771

Frames	Conf.
5	75

Detections		TP%	FP%
Positive	False		
136	2	98,55072	1,449275

Frames	Conf.
7	75

Detections		TP%	FP%
Positive	False		
100	0	100	0

Frames	Conf.
9	75

Detections		TP%	FP%
Positive	False		
83	0	100	0

D E C L A R A Ç Ã O

Declaro, sob compromisso de honra, que o trabalho apresentado nesta dissertação, com o título *"Inteligência Artificial Aplicada à Análise Digital Forense de Vídeos"*, é original e foi realizado por Estudante Daniel António Sousinha (2222940) sob orientação de Professor Patrício Rodrigues Domingues, Professor Miguel Monteiro de Sousa Frade e Professor Miguel Cerdeira Marreiros Negrão.

Leiria, Setembro de 2024

Daniel António Sousinha

Estudante Daniel António Sousinha