

**Redes Neurais Convolucionais para detecção de
impressões digitais em prontuários civis**

Rafaela Nunes Bonacorso

Monografia - MBA em Inteligência Artificial e Big Data

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Rafaela Nunes Bonacorso

Redes Neurais Convolucionais para detecção de impressões digitais em prontuários civis

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Prof. Dr. Ricardo Cerri

Versão original

São Carlos
2023

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi, ICMC/USP, com os dados fornecidos pelo(a) autor(a)

| | |
|-------|---|
| S856m | Bonacorso, Rafaela Nunes Redes Neurais Convolucionais para detecção de impressões digitais em prontuários civis / Rafaela Nunes Bonacorso ; orientador Ricardo Cerri. – São Carlos, 2023. 42 p. : il. (algumas color.) ; 30 cm. Monografia (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2023. 1. Aprendizado de Máquina. 2. YOLO. 3. Impressões Digitais. 4. Detecção de Objetos. 5. Processamento de Imagem. I. Cerri, Ricardo, orient. II. Redes Neurais Convolucionais para detecção de impressões digitais em prontuários civis. |
|-------|---|

Rafaela Nunes Bonacorso

Convolutional Neural Networks for fingerprint detection on civil records

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Prof. Dr. Ricardo Cerri

Original version

**São Carlos
2023**

RESUMO

Bonacorso, R. N. **Redes Neurais Convolucionais para detecção de impressões digitais em prontuários civis.** 2023. 42p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

A impressão digital é um tipo de característica biométrica popular devido a sua singularidade, aceitabilidade e baixo custo. É comumente utilizada em aplicações comerciais e oficiais, podendo ser coletada por meio de leitores biométricos ou, na falta destes, em papel e tinta. No último caso, o prontuário deve ser digitalizado e suas informações extraídas para a inserção do registro nos sistemas automatizados de identificação biométrica. Devido à média diária de prontuários civis poder variar de aproximadamente 1.000 a 5.000 unidades a depender do estado, a extração manual das 10 impressões digitais, fotografia e informações descritivas por prontuário não é viável. Este processo é lento, não eficaz e depende de um grande número de pessoas para ser realizado. Desta forma, a inserção pode ser realizada de forma massiva e automatizada. Este trabalho tem como objetivo a detecção de impressões digitais de acordo com seus respectivos dedos em imagens de prontuários civis digitalizados. Assim, foram utilizadas técnicas de processamento de imagem para a detecção de objetos em áreas específicas dos prontuários, obtendo um conjunto de imagens com anotações de coordenadas para cada dedo. Em seguida, este conjunto foi utilizado para o treinamento, teste e validação do modelo YOLOv8x com 1.483 imagens. Os resultados obtidos foram precisão de 0,995, mAP50 de 0,993 e mAP50-95 de 0,929, demonstrando a eficácia do modelo.

Palavras-chave: Aprendizado de Máquina. YOLO. Impressões Digitais. Detecção de Objetos. Processamento de Imagem.

ABSTRACT

Bonacorso, R. N. **Convolutional Neural Networks for fingerprint detection on civil records**. 2023. 42p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

Fingerprint is a popular type of biometric characteristic due to its uniqueness, acceptability and low cost. It is commonly used in commercial and official applications and can be collected using biometric scanners for liveness detection or on paper with ink. In the latter case, the document must be scanned and its information extracted in order to be inserted into automated biometric identification systems. Since the daily average of civil records can vary from approximately 1,000 to 5,000, units depending on the state, manual extraction of the 10 fingerprints, photographs and descriptive information per record is not viable. This process is slow, not effective and depends on a large number of people to be carried out. Thus, the insertion can be accomplished in a massive and automated way. This work aims to automate the detection of fingerprints in images of scanned civil records and classify them into their respective fingers. To achieve this, image processing techniques were used to detect objects in specific areas of the records in order to create a dataset with coordinate annotations. Then, this dataset was used to train, test and validate the YOLOv8 model with 1,483 images. The results obtained were accuracy of 0.995, mAP50 of 0.993 and mAP50-95 of 0.929, which demonstrates the effectiveness of the model.

Keywords: Machine Learning. YOLO. Fingerprints. Object Detection. Image Processing.

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1 – Exemplos de biometrias | 21 |
| Figura 2 – Redução de ruídos com filtro Gaussiano | 23 |
| Figura 3 – Detecção de bordas Canny | 23 |
| Figura 4 – Operações morfológicas | 24 |
| Figura 5 – Comparação das últimas versões do YOLO (JOCHER; CHAURASIA; QIU, 2023) | 27 |
| Figura 6 – Exemplo fictício de prontuário civil utilizado. | 29 |
| Figura 7 – Prontuário civil fictício com as caixas detectadas do processamento da imagem | 31 |
| Figura 8 – Prontuário civil fictício com as detecções do YOLOv8x treinado | 34 |
| Figura 9 – Matriz de confusão | 35 |
| Figura 10 – Curva precisão x confiança | 35 |
| Figura 11 – Curva revocação x confiança | 36 |
| Figura 12 – Curva F1 x confiança | 36 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 – Informações extraídas do processamento da imagem | 32 |
| Tabela 2 – Classe dos dedos | 33 |
| Tabela 3 – Métricas do modelo YOLOv8x treinado | 37 |

SUMÁRIO

| | | |
|----------|---------------------------------------|-----------|
| 1 | INTRODUÇÃO | 17 |
| 1.1 | Contextualização, Motivação e Lacunas | 17 |
| 1.2 | Objetivos | 18 |
| 1.3 | Contribuições | 18 |
| 1.4 | Organização do texto | 19 |
| 2 | FUNDAMENTAÇÃO TEÓRICA | 21 |
| 2.1 | Biometrias | 21 |
| 2.2 | Processamento de Imagens | 22 |
| 2.3 | Modelo de Detecção de Objetos - YOLO | 24 |
| 2.4 | YOLOv8 | 26 |
| 2.5 | Trabalhos Relacionados | 27 |
| 3 | AVALIAÇÃO EXPERIMENTAL | 29 |
| 3.1 | Conjuntos de Dados | 29 |
| 3.2 | Configuração Experimental | 30 |
| 3.3 | Resultados e Discussões | 32 |
| 4 | CONCLUSÕES | 39 |
| | Referências | 41 |

1 INTRODUÇÃO

1.1 Contextualização, Motivação e Lacunas

Biometria é uma forma de identificação pessoal utilizando características biológicas ou comportamentais. Existem diversos tipos de amostras biométricas como padrão de íris, DNA, voz, face e impressões digitais. A impressão digital é uma das características biométricas dominantes que continua se espalhando devido a sua singularidade, aceitabilidade e baixo custo. Atualmente essas características têm sido populares na identificação de pessoas em várias aplicações comerciais e autenticações em equipamentos eletrônicos (JAIN; BOLLE; PANKANTI, 1999).

Entretanto, sua importância vai além dessas aplicações. As impressões digitais podem identificar cadáveres desconhecidos, vítimas de desastres, autoria de crimes a partir de latentes encontradas em locais de crime e autoria de fraudes dos mais diversos tipos, além de ajudar a encontrar pessoas desaparecidas. Sistemas automatizados de identificação biométrica são comuns pelos órgãos de segurança pública em vários países, podendo conter ou não a biometria facial além das impressões digitais. Neste sistema, as impressões digitais do registro que está entrando são comparadas com todas as imagens armazenadas no banco de dados. Para isso, é necessário que os órgãos de segurança pública possuam um banco de dados robusto e povoado com o maior número de registros possível.

A inserção massiva de registros nos bancos de dados de sistemas automatizados de identificação biométrica passa por algumas etapas. Antigamente, e até atualmente em alguns estados, a coleta de impressões digitais era realizada em papel e tinta. O prontuário era arquivado e hoje é digitalizado. A extração das informações dos prontuários pode ser realizada de forma manual, entretanto, considerando o alto volume de prontuários, não é o método mais eficiente. Desta forma, modelos de detecção de objetos podem ser utilizados para a detecção das impressões digitais e métodos de reconhecimento óptico de caracteres (OCR) podem ser utilizados para a extração dos dados descritivos. Esses dados são, então, compilados em arquivos nist ou json para sua inserção no sistema.

Existem vários modelos de detecção de objetos em imagens e todos possuem vantagens e desvantagens. Os modelos YOLO são conhecidos por seu rápido desempenho, podendo detectar objetos em tempo real, e são considerados uma das escolhas mais comuns na produção devido a sua arquitetura simples, baixa complexidade e fácil implementação (DIWAN; ANIRUDH; TEMBHURNE, 2023). Comparações iniciais da versão mais recente (YOLOv8) demonstram sua superioridade como o novo estado da arte do YOLO, garantindo maior precisão em relação a seus antecessores (HUSSAIN, 2023). O YOLOv8 possui 5 modelos de detecção pré-treinados, que variam em precisão e velocidade. O modelo

YOLOv8x é o mais complexo, com um tempo maior para detecção e uma precisão maior, comparado aos outros da mesma linha (JOCHER *et al.*, 2023).

Em geral, os modelos são treinados com objetos mais comuns, como carros, pessoas, placas, etc., e devem ser treinados para a correta detecção de outros objetos, como impressões digitais. Para o treinamento dos modelos de detecção de objetos utilizando aprendizado de máquina, são necessários um conjunto de imagens com as respectivas anotações dos objetos a serem identificados. Essas anotações podem ser feitas de forma manual, com auxílio de algum software de imagem para a verificação das coordenadas dos objetos, ou automatizada por meio de técnicas de processamento de imagens. Ressalta-se que o processamento de imagem para detecção de objetos possui eficácia limitada, uma vez que busca quaisquer objetos na imagem, sem obter classificação adequada para cada objeto encontrado.

Diversos estudos sobre impressões digitais são realizados constantemente, como, por exemplo, descoberta de novos reveladores (BARROS *et al.*, 2023); determinação de características a partir da densidade de linhas (CARVALHO *et al.*, 2022); detecção, classificação e contagem de minúcias utilizando aprendizado de máquina (GAO *et al.*, 2023); e melhoria de modelos de comparação de impressões digitais baseados em aprendizagem profunda (SHAFAGHI *et al.*, 2023). Entretanto, são escassos os trabalhos referentes a detecção de impressões digitais em imagens.

1.2 Objetivos

O objetivo do trabalho é treinar o modelo YOLOv8x para a detecção de impressões digitais em imagens, com a finalidade de ser utilizado para a extração das informações biométricas em prontuários civis.

Os objetivos específicos são: criar conjuntos de treinamento, teste e validação por meio de técnicas de processamento de imagem para a detecção de objetos em áreas específicas; treinar modelo YOLOv8x com o conjunto criado; testar e analisar o modelo treinado; e utilizar o modelo treinado na prática.

1.3 Contribuições

As principais contribuições deste trabalho são o desenvolvimento de um método de processamento de imagem para a detecção de objetos em áreas específicas; e o treinamento de um modelo YOLOv8x para a extração de impressões digitais em prontuários civis com a finalidade de serem inseridas em sistemas automatizados de identificação biométrica dos órgãos de segurança pública.

1.4 Organização do texto

No capítulo 2 é feita uma revisão bibliográfica, apresentando os tipos de biometrias, etapas de processamento de imagens, um breve histórico do modelo de detecção de objetos YOLO e as modificações feitas em cada versão e, por fim, os trabalhos que relacionam o uso de aprendizado de máquina com impressões digitais.

Na sequência, o capítulo 3 descreve a metodologia utilizada para a criação do conjunto de treinamento, teste e validação e o treinamento do modelo YOLOv8x, e a discussão dos resultados obtidos.

O capítulo 4 finaliza o trabalho, discutindo o uso do modelo treinado na prática e apontando as melhorias que podem ser realizadas no futuro.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Biometrias

De acordo com a Organização Internacional de Normalização (ISO), o termo biometria refere-se ao reconhecimento automatizado de indivíduos baseado em suas características biológicas e comportamentais (Biometrics, 2002). Existem vários tipos de amostras biométricas como a face, impressão digital, DNA, íris, entre outras, que podem ser mensuradas por características físicas, como, por exemplo, o formato dos elementos da face, as linhas das impressões digitais, a sequência genética e a textura da íris. A Figura 1 mostra exemplos mais comuns de biometrias.

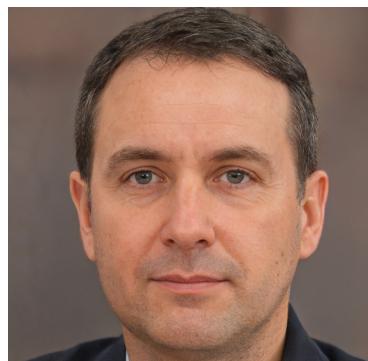
A Organização Internacional de Polícia Criminal (Interpol) classifica as biometrias como primárias e secundárias. Os métodos primários de identificação humana apresentam elementos singulares e individualizadores. São eles a papiloscopia, odontologia legal e a análise genética. Os métodos secundários podem ser utilizados como fonte investigativa, entretanto, não são decisivos e únicos.

A identificação humana por meio de impressões digitais é popular, rápida, de coleta não invasiva e precisa. Os desenhos consistem em um padrão de textura regular composta de cumes (cristas) e vales (sulcos) encontrados nas superfícies internas das mãos e pés. Esses desenhos apresentam características únicas e individuais, são perenes, imutáveis e de fácil classificação em padrões. Desta forma, a impressão digital é uma das biometrias mais utilizadas para a distinção de indivíduos em todo o mundo, seja para fins de identificação civil como para fins de investigação criminal (CARVALHO; ALMEIDA, 2020).

Existem diversas áreas de estudos sobre as impressões digitais, como, por exemplo, descoberta de novos reveladores (BARROS *et al.*, 2023); determinação de características a partir da densidade de linhas (CARVALHO *et al.*, 2022); detecção, classificação e contagem



(a) Impressão digital



(b) Face



(c) Iris

Figura 1 – Exemplos de biometrias

de minúcias utilizando aprendizado de máquina (GAO *et al.*, 2023); e melhoria de modelos de comparação de impressões digitais baseados em aprendizagem profunda (SHAFAGHI *et al.*, 2023). Entretanto, são escassos os trabalhos referentes a detecção de impressões digitais em imagens.

2.2 Processamento de Imagens

A análise de imagens, com intuito de extrair informações significativas de seu conteúdo, pode ser bem mais complexa que apenas alterar ou realçar a aparência de uma imagem. Para isso, são necessárias técnicas de processamento de imagens. As imagens ilustradas nessa seção foram retiradas da documentação da biblioteca OpenCV (DOXYGEN, 2023).

A conversão de imagens para escala de cinza é uma técnica ordinária e amplamente utilizada, que simplifica o processamento de imagens, reduz o tamanho dos dados e pode aprimorar recursos específicos da imagem, tornando-a uma ferramenta valiosa. Imagens em escalas de cinza são mais simples de processar computacionalmente, pois contêm apenas valores de intensidade, enquanto as imagens coloridas têm três canais de cores (normalmente vermelho, verde e azul) (GONZALEZ; WOODS, 2002). Devido à quantidade de canais de cores, as imagens coloridas possuem mais informações e maior dimensionalidade. Reduzir uma imagem para tons de cinza simplifica os dados, ao mesmo tempo que retém informações visuais importantes. Além disso, a conversão da imagem para escala de cinza pode aprimorar recursos ou padrões específicos, como bordas, textura ou contraste, o que simplifica o processo de detecção de objetos e segmentação de imagens (SONKA; HLAVAC; BOYLE, 2014).

A redução de ruídos em imagens é necessária para melhorar a qualidade dos dados, aumentar a confiabilidade da análise e sua interpretação e facilitar a tomada de decisões. A técnica permite extrair informações significativas de dados e sinais, ao mesmo tempo que minimiza o impacto de distorções indesejadas. O desfoque gaussiano (Gaussian Blur), ilustrado na Figura 2, é um método utilizado no pré-processamento de imagens, que reduz efetivamente o ruído de alta frequência enquanto preserva a estrutura subjacente e os detalhes da imagem. É um filtro passa-baixa, que enfatiza transições mais suaves nos valores dos pixels através da convolução da imagem com um kernel gaussiano (GONZALEZ; WOODS, 2002). Ao contrário de outras técnicas, não favorece o desfoque em uma direção específica, preservando detalhes importantes da imagem, o que resulta em uma imagem visualmente mais agradável e mais suave (SONKA; HLAVAC; BOYLE, 2014).

O algoritmo de detecção de bordas "Canny", ilustrado na Figura 3, é conhecido por sua versatilidade e capacidade de identificar com precisão as bordas das imagens, minimizando detecções falsas. Canny utiliza uma combinação de informações de gradiente e supressão não máxima para localizar bordas verdadeiras (CANNY, 1986). Além disso, o



Figura 2 – Redução de ruídos com filtro Gaussiano

algoritmo incorpora suavização gaussiana para reduzir o ruído na imagem, o que ajuda a melhorar a qualidade dos resultados de detecção de bordas (GONZALEZ; WOODS, 2002).

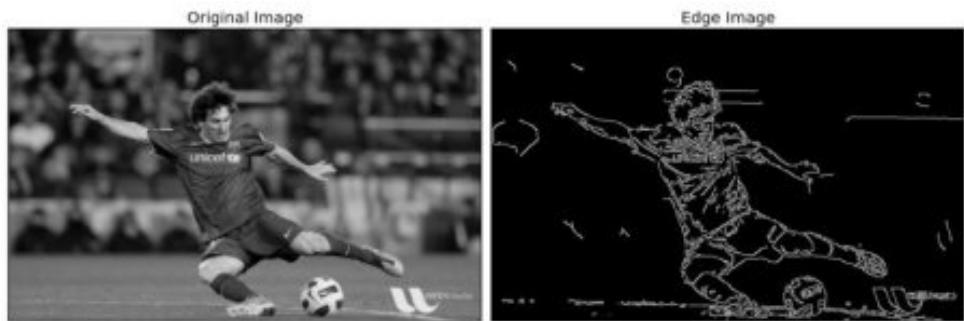


Figura 3 – Detecção de bordas Canny

Os pixels ao longo da borda de uma região binária podem ser identificados usando operações morfológicas simples. Operações morfológicas, ilustradas na Figura 4, são um conjunto de técnicas de processamento de imagens utilizadas para análise de formas, extração de características e aprimoramento de imagens. Essas operações modificam a forma e a estrutura dos objetos em uma imagem, por meio de elementos estruturantes (kernels). As operações morfológicas fundamentais são dilatação e erosão, que podem ser combinadas para criar operações mais complexas, como abertura, fechamento e gradiente morfológico. A dilatação expande ou engrossa as regiões brancas em uma imagem binária. A erosão, por sua vez, reduz ou afina essas regiões. A operação de abertura combina erosão seguida de dilatação e é utilizada para remover pequenos detalhes e ruídos, preservando as estruturas essenciais da imagem. Em contrapartida, o fechamento combina dilatação seguida de erosão e serve para fechar pequenas lacunas nas regiões brancas e conectar estruturas próximas (SOILLE *et al.*, 1999). Por fim, a operação de gradiente morfológico calcula a diferença entre a dilatação e a erosão de uma imagem, destacando as bordas e os limites dos objetos.

Uma vez encontradas as regiões em uma imagem binária, o próximo passo geralmente é encontrar os contornos das regiões. A detecção de contorno permite a extração de informações significativas das imagens, possibilitando análises adicionais e tomadas de

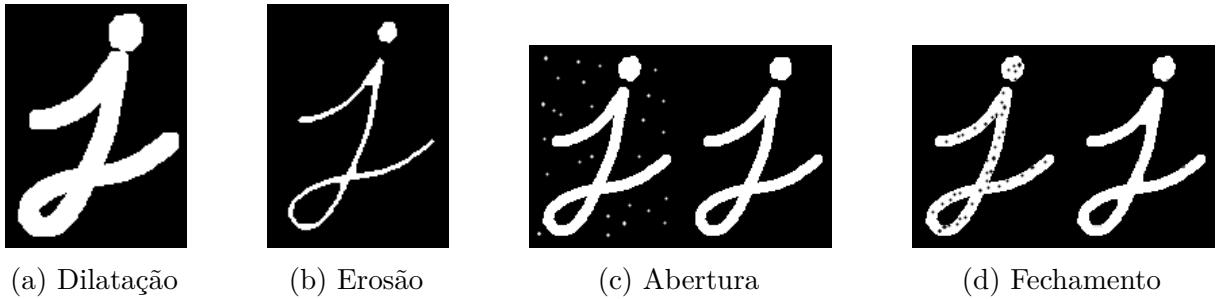


Figura 4 – Operações morfológicas

decisão com base nos objetos e formas detectados. Desta forma, delimitam-se as áreas nas quais foram encontradas as regiões de interesse.

A combinação das etapas descritas pode ser utilizada para a extração de características desejadas em uma imagem. Essas características auxiliam nas análises de imagens e na aplicação de outros métodos. No presente trabalho, o processamento de imagem foi utilizado para localização de impressões digitais em locais específicos das imagens, com a finalidade de criar as anotações para o treinamento do modelo de detecção de objetos YOLO.

2.3 Modelo de Detecção de Objetos - YOLO

You Only Look Once (YOLO) é um algoritmo comumente utilizado para detecção de objetos, seja em imagens ou vídeos. Desde 2015, quando a primeira versão do YOLO foi lançada, o algoritmo rapidamente ganhou popularidade por sua alta velocidade e precisão. Em consequência, vários pesquisadores estudaram mudanças e melhorias, criando novas versões. Atualmente a versão mais atualizada é YOLOv8 (JOCHER *et al.*, 2023).

A arquitetura inicial do YOLO foi inspirada na GoogleNet e composta por 24 camadas convolucionais, seguidas de duas camadas completamente conectadas. A capacidade de realizar a tarefa de detecção com apenas uma única passagem da imagem completa, tornou-o muito mais rápido quando comparado aos detectores de objetos existentes, permitindo desempenho em tempo real. Assim, YOLO é utilizado para a predição simultânea de caixas delimitadoras dos objetos e suas probabilidades de classe. Entretanto, a primeira versão do YOLO apresentava limitações, como o posicionamento impreciso, incapacidade de prever objetos muito próximos e dificuldade para prever objetos com proporções não vistas nos dados de treinamento (TERVEN; CORDOVA-ESPARZA, 2023).

A segunda versão do YOLO (YOLOv2) incluiu várias melhorias, podendo detectar mais de 9.000 categorias. Essa versão aumentou desempenho da rede para entradas de resolução mais alta, utilizou a normalização em lote em todas as camadas convolucionais, caixas de âncora para caixas limitadoras e clusters de dimensão, e removeu as camadas densas, utilizando uma arquitetura totalmente convolucional, entre outras melhorias.

YOLOv2 utilizou um novo modelo de classificação (Darknet-19), que utiliza 19 camadas convolucionais e 5 camadas de MaxPooling, reduzindo a quantidade de operações a serem realizadas e aumentando a velocidade do processamento (JOCHER *et al.*, 2023; TERVEN; CORDOVA-ESPARZA, 2023).

A terceira versão do YOLO (YOLOv3) incluiu mudanças significativas, como o uso de recursos multiescala para detecção de objetos e o ajuste da estrutura básica da rede, mantendo o desempenho em tempo real. A rede de extração de atributos do YOLOv3 utilizou o modelo residual e o algoritmo continha 53 camadas convolucionais (Darknet-53). Todas as camadas de maxpooling foram substituídas por convolucionais e conexões residuais foram adicionadas (TERVEN; CORDOVA-ESPARZA, 2023).

YOLOv4 teve uma mudança significativa, com mais foco na comparação de dados. Dentre suas principais contribuições pode-se destacar a introdução de Mosaic data augmentation, um modelo de detecção mais eficiente e sem âncora, que permite treinamento rápido e com precisão e a influência dos métodos bag-of-freebies e bag-of-specials durante o treinamento, aumentando sua eficácia (TERVEN; CORDOVA-ESPARZA, 2023; JOCHER *et al.*, 2023).

YOLOv5 é uma das versões com mais flexibilidade de uso, o desempenho do modelo foi aprimorado com a adição de alguns recursos como otimização de hiperparâmetros, rastreamento integrado de experimentos e exportação automática. A estrutura CSPDarknet53 modificada facilitou o treinamento do conjunto de dados, tornando mais fácil colocá-lo em produção do que a estrutura usada no YOLOv4. Ademais, uma ferramenta de pré-treinamento (AutoAnchor) foi implementada, a qual verifica e ajusta as caixas de âncora para o conjunto de dados e configurações de treinamento. YOLOv5 é dividida em cinco versões para atender aplicações específicas e requisitos de hardware, são elas: YOLOv5n (nano), YOLOv5s (pequeno), YOLOv5m (médio), YOLOv5l (grande) e YOLOv5x (extra grande). Entretanto, sua detecção de objetos pequenos não é tão eficaz quanto a de objetos maiores (JOCHER *et al.*, 2023; TERVEN; CORDOVA-ESPARZA, 2023).

As versões 6 (YOLOv6) e 7 (YOLOv7) foram lançadas quase concomitantes. A arquitetura do YOLOv6 foi modificada e baseada em RepVGG, com alto paralelismo. Para obter um detector mais rápido, foi utilizado um esquema de quantização para detecção usando RepOptimizer e destilação por canal. Esta versão foi utilizada em vários robôs de entregas e dividida em 8 modelos (YOLOv6-N ao YOLOv6-L6). O YOLOv7 propôs algumas mudanças na arquitetura e uma série de bag-of-freebies, aumentando a precisão sem afetar a velocidade de inferência, apenas o tempo de treinamento, e adicionou outras tarefas como estimativa de pose, dentre outras mudanças (TERVEN; CORDOVA-ESPARZA, 2023).

Finalmente, YOLOv8 é um modelo de última geração, introduzindo novos recursos e melhorias para aprimorar o desempenho, a flexibilidade e a eficiência. O modelo pode

ser utilizado em uma gama completa de tarefas de inteligência artificial, como detecção, segmentação, estimativa de pose, rastreamento e classificação. A próxima seção explica com mais detalhes o modelo YOLOv8.

2.4 YOLOv8

YOLOv8 foi lançado com 6 versões: YOLOv8n (nano), YOLOv8s (pequeno), YOLOv8m (médio), YOLOv8l (grande), YOLOv8x (extra grande) e YOLOv8-Seg (segmentação semântica) (JOCHER *et al.*, 2023).

A oitava versão da série de modelos YOLO possui arquitetura semelhante a do YOLOv5 com algumas modificações no módulo C2f (gargalo parcial de estágio cruzado com duas convoluções), que combina recursos de alto nível com informações contextuais para melhorar a precisão da detecção.

YOLOv8 utiliza em sua arquitetura um backbone CSPDarknet53 modificado. O módulo C2f substitui a camada CSPLayer do YOLOv5. Cada convolução passa por normalização em lote. A camada de saída possui uma função sigmóide como função de ativação para representar a probabilidade de a caixa delimitadora conter um objeto e uma função softmax para as probabilidades das classes.

Além disso, YOLOv8 é um detector sem âncora. Isso significa que não depende de caixas de âncora predefinidas, mas prevê diretamente as caixas delimitadoras e as probabilidades de classe para cada objeto na imagem de entrada. Esta abordagem reduz a complexidade do modelo e permite mais flexibilidade na detecção de objetos de vários tamanhos e proporções. Assim, essa arquitetura permite que cada segmento se concentre em sua tarefa, melhorando a precisão geral do modelo (TERVEN; CORDOVA-ESPARZA, 2023).

Ademais, YOLOv8 utiliza as funções de perda CIoU (interseção sobre união completa) e DFL (perda focal distributiva) para caixa delimitadora e entropia cruzada binária para classificação, melhorando o desempenho da detecção de objetos, principalmente de objetos menores. Em comparação com o modelo YOLOv5, YOLOv8 alcançou uma precisão média maior para o conjunto de dados MS COCO test-dev 2017 com uma velocidade de 280 FPS em um NVIDIA A100 e TensorRT. A Figura 5 ilustra a comparação entre as últimas versões do YOLOv8, na qual é possível verificar maior precisão do modelo YOLOv8 em relação a suas versões anteriores. Por se tratar de um modelo com maior precisão e facilidade de uso, o presente trabalho optou por utilizar o YOLOv8x (JOCHER *et al.*, 2023).

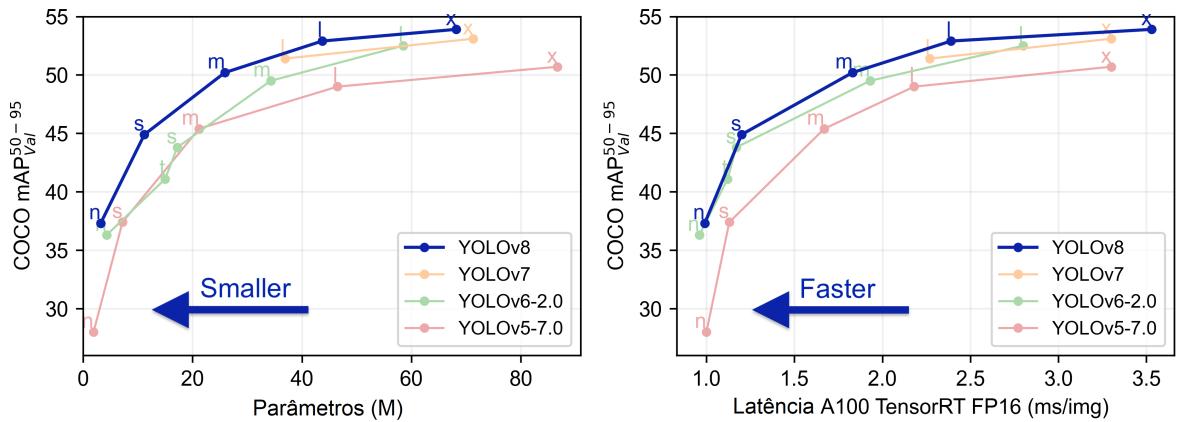


Figura 5 – Comparação das últimas versões do YOLO (JOCHER; CHAURASIA; QIU, 2023)

2.5 Trabalhos Relacionados

Com o desenvolvimento da inteligência artificial, técnicas de aprendizado de máquina tem sido gradualmente aplicadas em diversas áreas. No âmbito das biometrias, o aprendizado profundo com a construção de modelos para a classificação, detecção de minúcias e comparação de impressões digitais, além de detecção e reconhecimento facial vem sendo utilizado em estudos e aplicações diárias.

Anusha e Kumar (ANUSHA; KUMAR, 2023) desenvolveram um sistema automatizado de identificação de latentes por meio de redes neurais convolucionais (CNN) de um algoritmo de aprendizado de máquina profundo. Neste trabalho foram aplicadas técnicas de processamento de imagem para extração de minúcias, as quais foram utilizadas para o treinamento do modelo. Jian et al. (JIAN; ZHOU; LIU, 2021) utilizaram algoritmo genético para otimizar a estrutura de uma rede convolucional densamente conectada (DenseNet) para detecção de vivacidade de impressões digitais.

Por serem modelos conhecidos e de fácil compreensão e aplicação, diversos estudos optam pela utilização de modelos YOLO. Trusiak e Saeed (TRUSIAK; SAEED, 2023) treinaram o modelo YOLOv5 para identificar e localizar minúcias em imagens de impressões digitais e avaliaram sua agilidade entre os métodos de detecção de minúcias, no qual aferiram que o YOLO foi superior nos testes realizados. De forma semelhante, Gao et al. (GAO *et al.*, 2023) desenvolveram um algoritmo de detecção automatizado com base no modelo YOLOv5s para a classificação de minúcias em impressões digitais. Os autores identificaram 6 tipos de minúcias e as quantificaram para realizar sua análise estatística. Rim et al. (RIM; KIM; HONG, 2021) apresentaram um método de classificação para a identificação detalhada de informações de impressões digitais, com o objetivo de distinguir informações como classificação da mão esquerda e direita, dos poros do suor e das cicatrizes, utilizando os modelos Alexnet, VGG-16, Yolo-v2 e Resnet-50. Entretanto, não foi possível

encontrar trabalhos de detecção de impressões digitais em imagens utilizando aprendizado de máquina.

3 AVALIAÇÃO EXPERIMENTAL

3.1 Conjuntos de Dados

As imagens utilizadas foram prontuários civis digitalizados. Por questões legais, nenhuma imagem apresentada nesse trabalho contém dados reais e as impressões digitais foram extraídas aleatoriamente de imagens da Internet. A Figura 6 representa o modelo de prontuário utilizado.

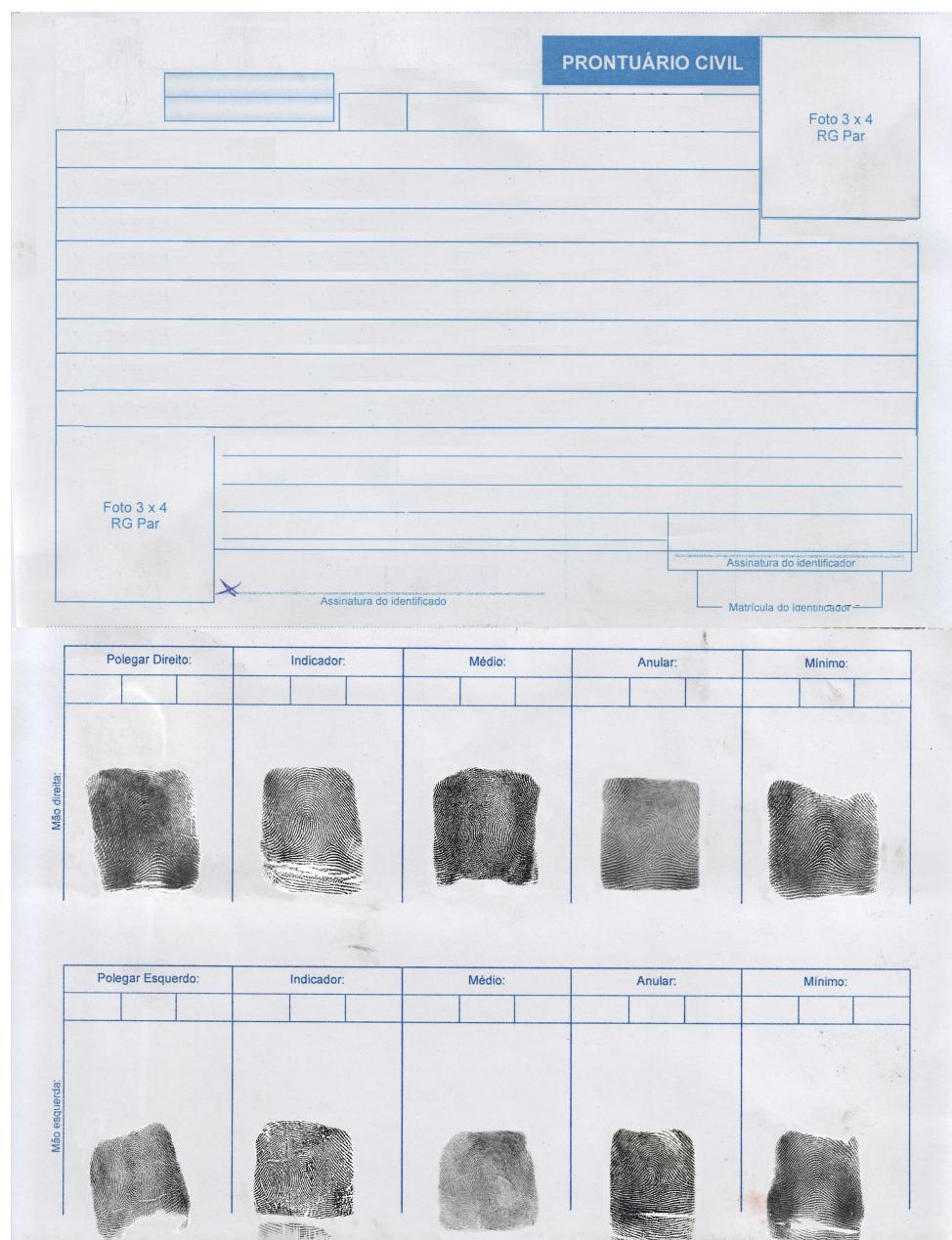


Figura 6 – Exemplo fictício de prontuário civil utilizado.

3.2 Configuração Experimental

Uma amostra de aproximadamente 3.000 imagens de prontuários civis foi disponibilizada para estudo e todo trabalho foi programado em Python. Em um primeiro script, foi realizado um processamento de imagens para detecção de quaisquer objetos nos prontuários, aplicando filtros e condições para selecionar apenas as caixas desejadas. Utilizando-se da biblioteca OpenCV (cv2), as imagens foram convertidas em escala de cinza, para aprimorar recursos ou padrões específicos, como bordas, textura ou contraste, simplificando os dados e o processo de detecção de objetos. Em seguida o filtro gaussiano foi aplicado para a redução de ruídos, permitindo extrair informações significativas de dados e sinais, ao mesmo tempo que minimizando o impacto de distorções indesejadas. A seguir, Canny Edge Detection foi aplicado para identificar com precisão as bordas das imagens, minimizando detecções falsas. Com o uso da função bitwisenot, os pixels foram invertidos na imagem, para adequar a imagem à próxima etapa que foi a utilização de uma função de gradiente morfológico, combinando dilatação e erosão da imagem, a fim de destacar as bordas e os limites dos objetos. Por fim, uma função localização de contornos foi utilizada para detectar os objetos na imagem e criar as caixas nos objetos encontrados. Nesta função, foi aplicado um filtro para remover todas caixas que não atingiram um tamanho mínimo, isto é, removendo pontos e letras que foram identificados como objetos. O processo foi executado em uma imagem de cada vez e a sequência foi definida empiricamente a partir de tentativa e erro. Para a visualização da imagem processada com todas as caixas identificadas, foram desenhados os retângulos nos objetos identificados nas imagens, como mostra a Figura 7.

Após realizado o processamento da imagem, foram estipuladas condições de localização e áreas de acordo com a localização de cada dedo na imagem e um tamanho médio de impressões digitais, para a identificação das caixas encontradas. As caixas já identificadas foram salvas em um banco de dados previamente preparado para auxiliar no processo. Por mais que houvesse outras formas de armazenar as informações, o banco de dados MySQL foi escolhido pela facilidade de uso através do Python. A tabela 'bboxes' continha 3 colunas (numero do rg, dedo, caixa), exemplificada na Tabela 1. O número do registro geral foi utilizado como identificador, a variável dedo refere-se ao dedo coletado na região de interesse e a variável caixa configura as coordenadas da caixa delimitadora (esquerda, topo, direita, base).

Em um segundo script, foram utilizados filtros nas informações salvas no banco de dados, como: apenas uma caixa referente a cada dedo e dez caixas por imagem. As imagens que não seguiram esse padrão foram descartadas. Esse método de detecção de objetos a partir do processamento de imagem provou-se suficientemente eficiente, pois mesmo com a grande quantidade caixas faltantes, como o exemplo da Figura 7, ou mais de uma caixa para o mesmo dedo, foi possível reservar 1.483 imagens para o treinamento

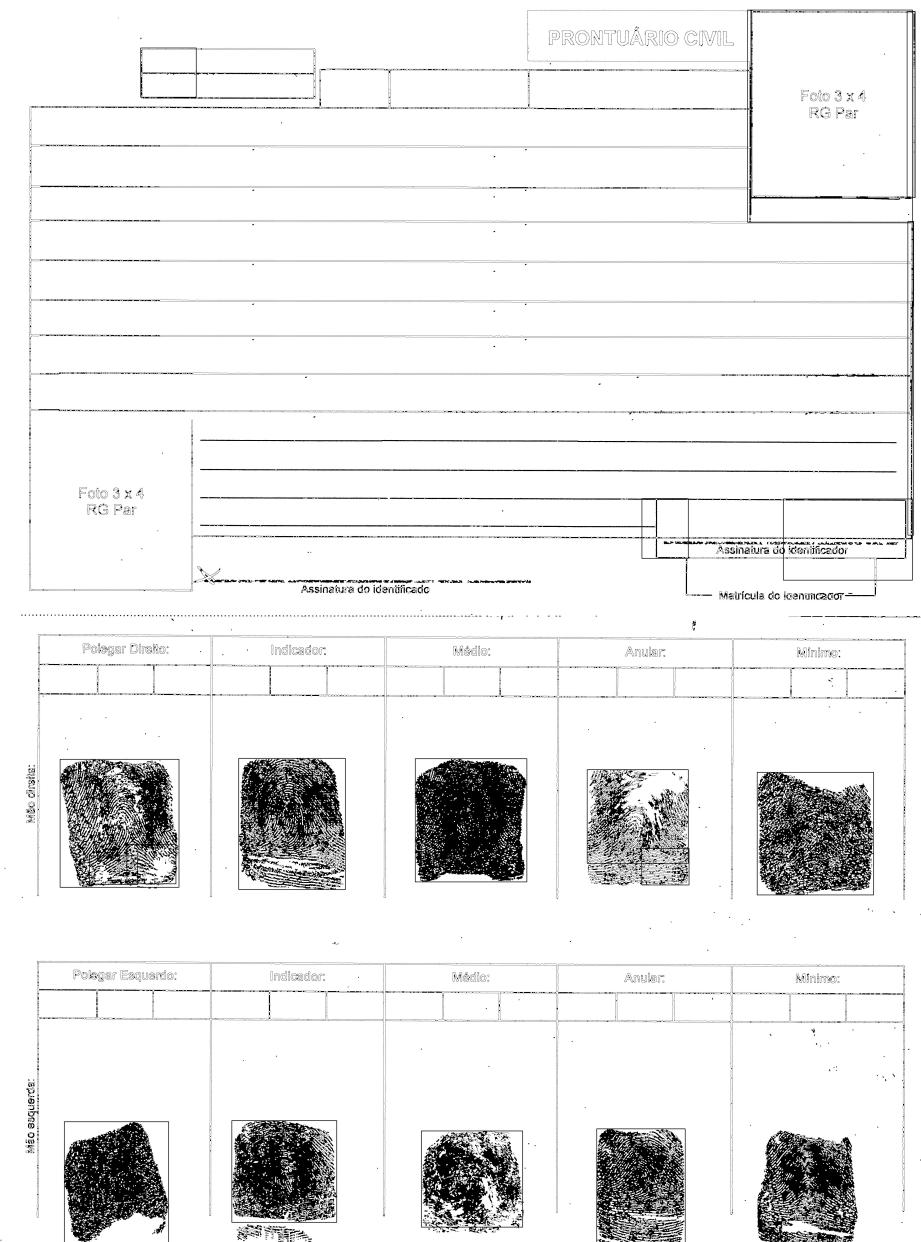


Figura 7 – Prontuário civil fictício com as caixas detectadas do processamento da imagem

do modelo YOLOv8x.

Para as imagens reservadas, as coordenadas de cada caixa foram convertidas para o padrão de anotações do modelo YOLO (classe, ponto central da coordenada x, ponto central da coordenada y, largura da caixa, altura da caixa), com todos valores normalizados entre 0 e 1. Para cada caixa, foram encontrados os pontos centrais das coordenadas x e y das caixas, e divididas pela largura e altura da imagem respectivamente. De forma semelhante, a altura e largura de cada caixa foram divididas pela altura e largura da imagem respectivamente. A amostra foi dividida em 1.131 imagens para treino, 144 imagens para validação e 208 imagens para teste com suas respectivas anotações no padrão YOLO. As classes foram divididas de acordo com cada dedo, seguindo a Tabela 2.

| Número do RG | Dedo | Caixa (esquerda, topo, direita, base) |
|--------------|------|---------------------------------------|
| ####### | f1 | (397, 4036, 1018, 4704) |
| ####### | f3 | (2252, 4033, 2835, 4676) |
| ####### | f5 | (4038, 4104, 4646, 4744) |
| ####### | f2 | (1330, 4031, 1887, 4716) |
| ####### | f7 | (1294, 5922, 1842, 6453) |
| ####### | f6 | (419, 5928, 964, 6568) |
| ####### | f9 | (3200, 5965, 3662, 6559) |
| ####### | f8 | (2284, 5976, 2813, 6478) |
| ####### | f4 | (3150, 4090, 3678, 4579) |

Tabela 1 – Informações extraídas do processamento da imagem

A quantidade de épocas determina o número de vezes que o conjunto de treino passa pelo modelo. A escolha do número de épocas depende de vários fatores e pode ser determinada experimentalmente, verificando os resultados obtidos pela validação do modelo. Os desenvolvedores dos modelos YOLO recomendam começar um treinamento utilizando 300 épocas e verificar se houve overfitting ou não, para ajustar a quantidade de épocas ideal para cada caso (JOCHER *et al.*, 2023). Quanto mais épocas forem utilizadas e quanto maior o conjunto de treinamento, mais tempo o modelo demora para finalizar o treinamento, ou seja, é necessário disponibilidade de tempo e uma máquina com memória suficiente que suporte o treinamento. Desta forma, neste trabalho, os hiperparâmetros utilizados foram 50 épocas de treinamento e 8 minibatches. A partir da validação do modelo, foi possível verificar que a quantidade de épocas foi suficiente para este caso.

3.3 Resultados e Discussões

A Figura 8 apresenta a imagem do prontuário fictício com as detecções realizadas pelo modelo YOLO. Pode-se verificar que todas as classes foram bem definidas e, de forma geral, com um bom nível de confiança. Uma das soluções para aumentar o nível de confiança, é aumentar a base de treinamento, pois quanto mais imagens são fornecidas para o treinamento do modelo, melhor será sua precisão.

Várias métricas e visualizações são usadas para avaliar o desempenho do modelo. A matriz de confusão serve para avaliar o desempenho de um algoritmo de classificação, mostrando o número de resultados verdadeiros positivos (objetos detectados corretamente), verdadeiros negativos (fundo ignorado corretamente), falsos positivos (objetos detectados incorretamente) e falsos negativos (objetos perdidos). A Figura 9 representa a matriz de confusão gerada pela validação do modelo utilizando 144 imagens, onde aproximadamente

| Classe | Nome | Dedo |
|--------|------|--------------------|
| 0 | f1 | Polegar Direito |
| 1 | f2 | Indicador Direito |
| 2 | f3 | Médio Direito |
| 3 | f4 | Anular Direito |
| 4 | f5 | Mínimo Direito |
| 5 | f6 | Polegar Esquerdo |
| 6 | f7 | Indicador Esquerdo |
| 7 | f8 | Médio Esquerdo |
| 8 | f9 | Anular Esquerdo |
| 9 | f10 | Mínimo Esquerdo |

Tabela 2 – Classe dos dedos

98% das amostras foram corretamente detectadas e classificadas.

A precisão é a razão entre o número de verdadeiros positivos e o número total de previsões positivas feitas pelo modelo, ou seja, a quantidade de vezes que o modelo acerta em relação ao total de vezes que tenta acertar. A curva precisão x confiança, representada na Figura 10, ilustra como a precisão muda à medida que o limite de confiança varia, quanto mais próximo de 1 for a confiança, maior será a precisão.

De forma semelhante, a revocação é a razão entre o número de verdadeiros positivos e o número total de casos positivos, ou seja, a quantidade de vezes que o modelo acerta em relação ao total de vezes que deveria ter acertado. A curva revocação x confiança, representada na Figura 11 mostra o comportamento da revocação com a variação do limiar de confiança. É possível visualizar que a partir de confiança igual a 0,6, o valor da revocação começa a decair, e após aproximadamente confiança igual a 0,8, há uma queda significativa da revocação.

O modelo segue o que normalmente ocorre: a precisão aumenta e a revocação diminui com o aumento da confiança. Entender os gráficos das Figuras 10 e 11 ajuda na seleção de um limite de confiança apropriado para controlar a compensação entre precisão e revocação e auxilia a compreensão de quão bem o modelo captura todos os objetos na imagem em diferentes níveis de confiança. Ambas as curvas combinadas podem ser utilizadas para a escolha do limite de confiança. Para alta precisão, pode-se aumentar o limite de confiança, ou, para maior revocação, pode-se escolher um limite de confiança mais baixo para capturar mais instâncias positivas ao custo de aceitar mais falsos positivos. A pontuação F1 é a média harmônica entre a precisão e a revocação, proporcionando um equilíbrio entre essas duas métricas. A curva F1 x confiança, representada na Figura 12,

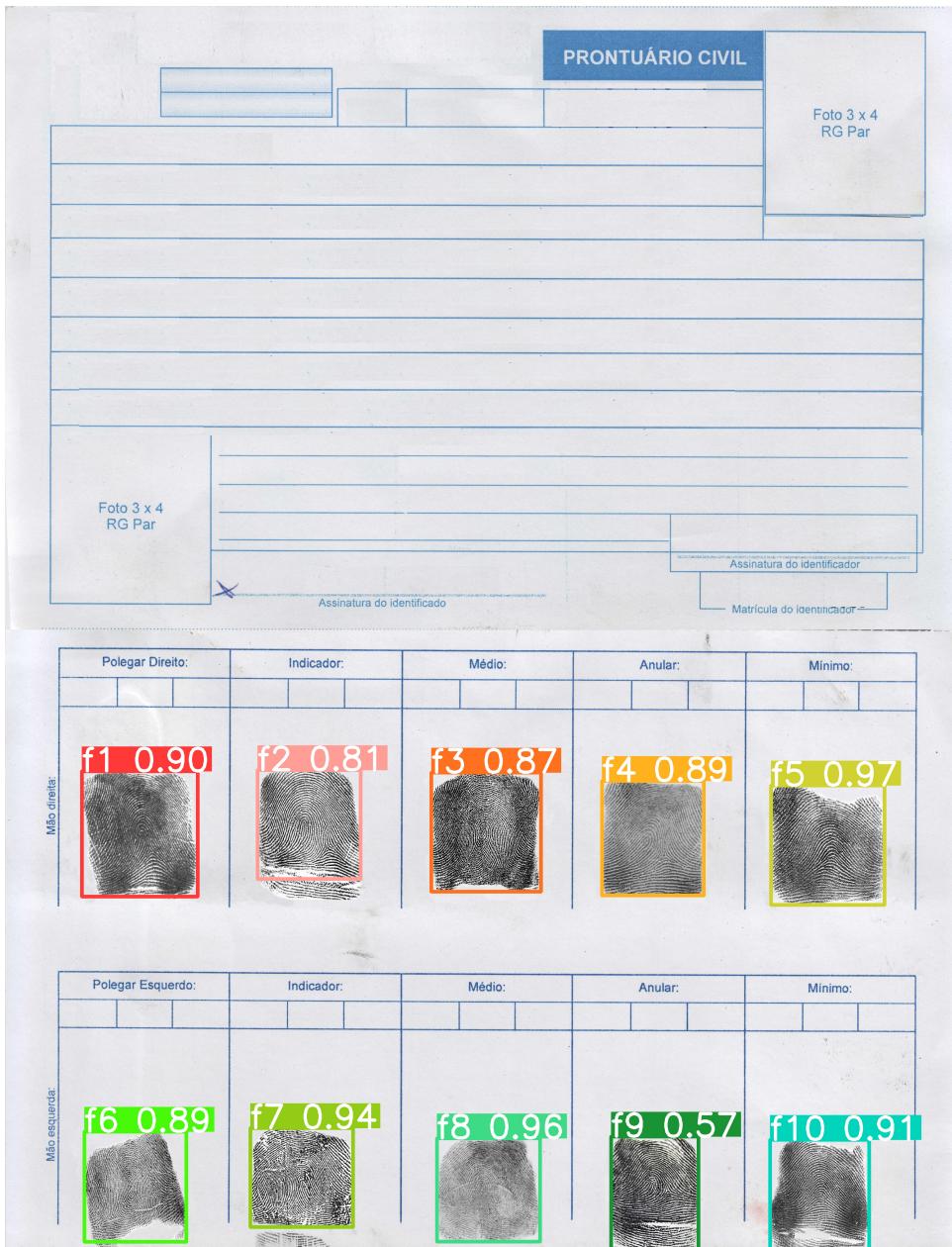


Figura 8 – Prontuário civil fictício com as detecções do YOLOv8x treinado

ilustra como a métrica varia em diferentes níveis de confiança. É possível perceber que esta curva combina as curvas representadas nas figuras 10 e 11, auxiliando na escolha do limite de confiança e fornecendo um modelo melhor.

Mean Average Precision at IoU 0,5 (mAP50) e Mean Average Precision at IoU 0,5 to 0,95 (mAP50-95) são métricas de avaliação de modelos treinados para tarefas como classificação de imagens e localização de objetos dentro de uma imagem. Essas métricas são utilizadas na comparação de diferentes modelos de detecção de objetos e são importantes para avaliar o desempenho do modelo em localizar e classificar com precisão objetos dentro de uma imagem. A precisão média avalia o desempenho do modelo através da área sob a curva precisão x revocação. Em uma tarefa de detecção multiclasse, mAP

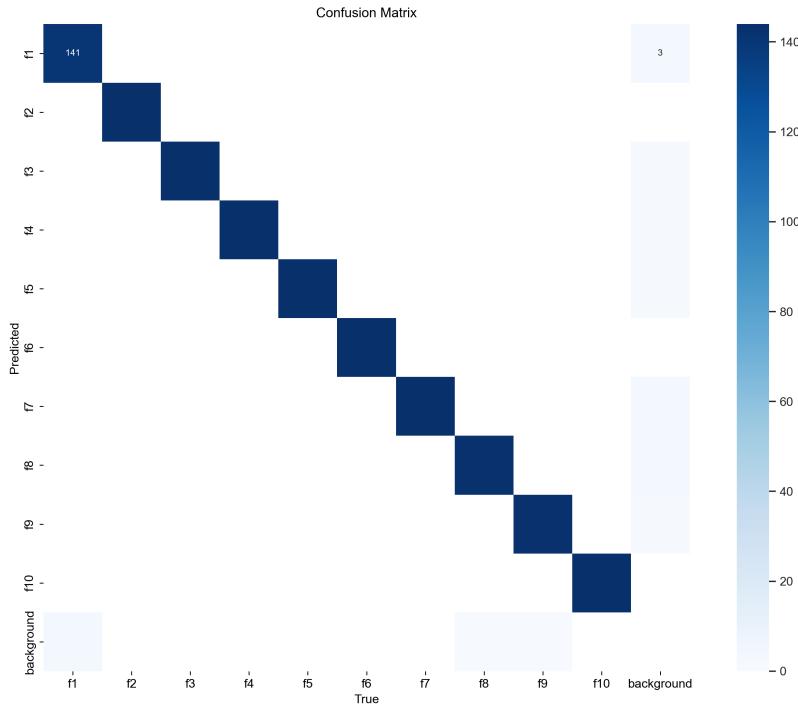


Figura 9 – Matriz de confusão

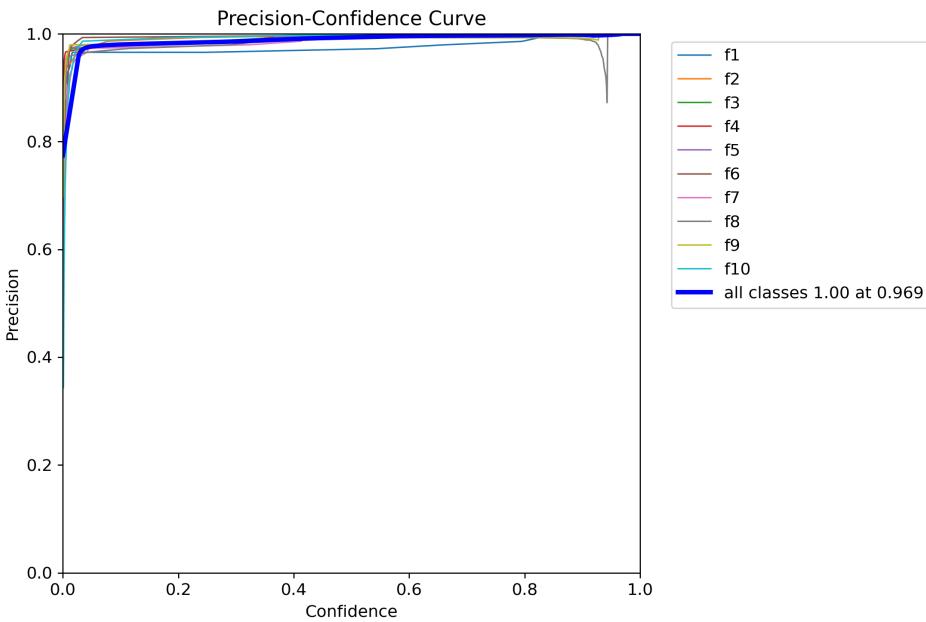


Figura 10 – Curva precisão x confiança

calcula a média dos valores de precisão média calculados para cada classe, o que pode ser visualizado na Figura 12. Intersection over Union (IoU) mede a sobreposição entre as caixas delimitadoras preditas e as caixas delimitadoras verdadeiras. Um IoU de 0,5 significa que a caixa delimitadora predita deve ter pelo menos 50% de sobreposição com a caixa delimitadora verdadeira para ser considerada uma detecção correta. No caso da mAP50-95, são calculados vários limites de IoU variando de 0,5 a 0,95 em pequenos incrementos. A

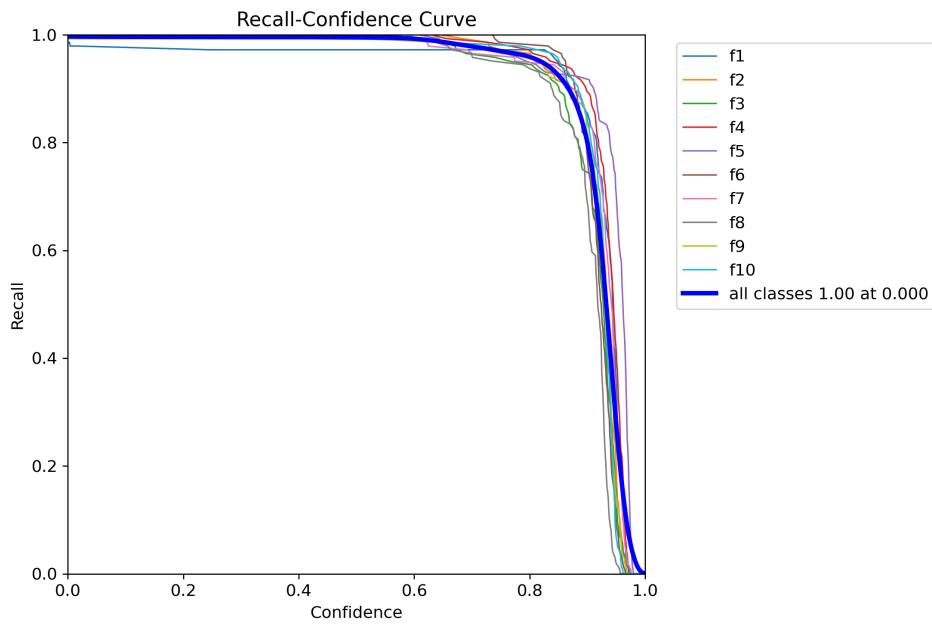


Figura 11 – Curva revocação x confiança

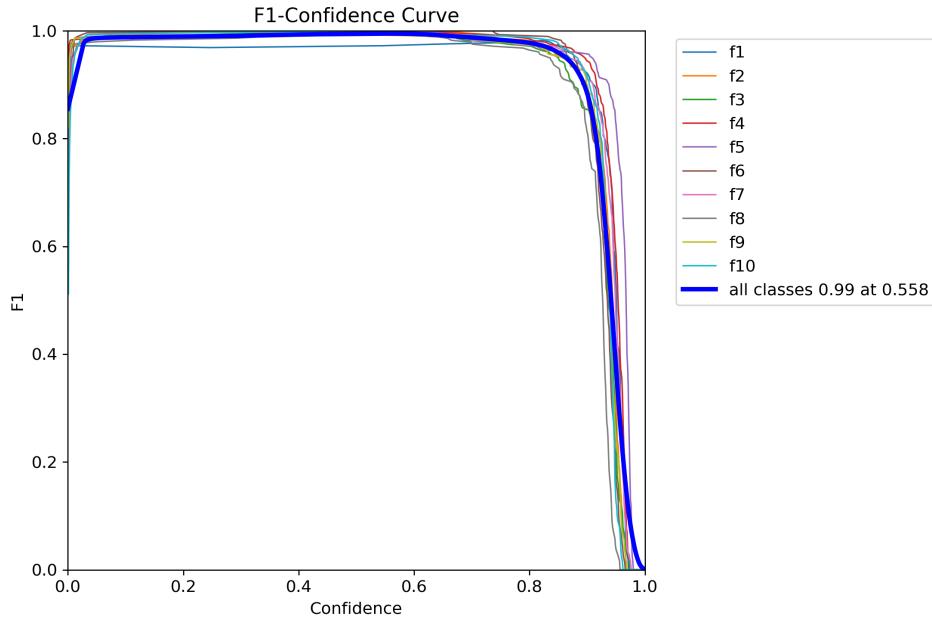


Figura 12 – Curva F1 x confiança

Tabela 3 mostra os valores de precisão, revocação, mAP50 e mAP50-95 para o modelo YOLOv8x treinado.

Conforme aferido na Tabela 3, o menor valor de mAP é equivalente a 0,9 para a classe f1, o que reflete em uma ótima precisão do modelo treinado. Valores de mAP acima de 0,8 são considerados bons e geralmente suficientes para várias aplicações do mundo real, valores de mAP acima de 0,9, como o que foi obtido pelo modelo, são considerados excelentes e normalmente exigidos para tarefas de alta precisão, como imagens médicas ou direção autônoma, onde a precisão é crítica. Desta forma, combinando todos os gráficos

| Classe | Precisão | Revocação | mAP50 | mAP50-95 |
|--------|----------|-----------|-------|----------|
| all | 0.995 | 0.995 | 0.993 | 0.929 |
| f1 | 0.973 | 0.971 | 0.988 | 0.9 |
| f2 | 0.998 | 1 | 0.995 | 0.933 |
| f3 | 0.999 | 1 | 0.995 | 0.925 |
| f4 | 0.998 | 1 | 0.995 | 0.937 |
| f5 | 0.998 | 1 | 0.995 | 0.951 |
| f6 | 0.998 | 1 | 0.995 | 0.933 |
| f7 | 0.997 | 0.993 | 0.995 | 0.925 |
| f8 | 0.993 | 0.993 | 0.988 | 0.926 |
| f9 | 0.993 | 0.991 | 0.992 | 0.93 |
| f10 | 0.999 | 1 | 0.995 | 0.921 |

Tabela 3 – Métricas do modelo YOLOv8x treinado

e valores obtidos pelo modelo treinado, é possível afirmar que o modelo possui um alto nível de desempenho para a tarefa de detecção de impressões digitais em prontuários e a classificação dos dedos de acordo com a posição que se encontra na imagem.

4 CONCLUSÕES

As impressões digitais são o tipo mais comum de amostra biométrica e servem para identificar pessoas, seja em cadáveres desconhecidos e vítimas de desastres, na autoria de crimes e fraudes ou em autenticações em equipamentos eletrônicos. Atualmente diversos países possuem sistemas automatizados de identificação biométrica, onde as impressões digitais de uma pessoa são conectadas a seus dados descritivos. Esses sistemas são de grande importância para encontrar autoria em crimes e identificar fraudes, como por exemplo fraudes de benefícios. Entretanto, nem todos os estados possuem sistemas de coleta de impressões digitais in vivo, ou seja, por meio de um leitor biométrico, sendo a coleta ainda feita em papel e tinta. Desta forma, o prontuário é digitalizado para ter seus dados extraídos e ser posteriormente inseridos no banco de dados. O presente trabalho teve como objetivo principal a detecção de impressões digitais em imagens de prontuários civis para possibilitar a inclusão massiva nos bancos de dados de sistemas automatizados de identificação biométrica dos órgãos de segurança pública.

Foram tratadas aproximadamente 3.000 imagens utilizando técnicas de processamento de imagens, a fim de identificar objetos em áreas específicas. As etapas utilizadas no processamento das imagens e sua ordem foram cautelosamente planejadas e otimizadas para atingir melhores resultados. O método de processamento de imagens descrito neste trabalho pode ser utilizado para localizar quaisquer objetos em imagens com a finalidade de obter conjuntos de treinamento para modelos de detecção de objetos mais complexos. Um total de 1.483 imagens foram reservadas, com suas respectivas anotações, para o treinamento, teste e validação do modelo YOLOv8x para detecção das impressões digitais e sua classificação por dedo. Após o treinamento, o modelo foi testado e validado, atingindo uma precisão de 0,995 e um mAP50-95 de 0.929, o que pode ser considerado um ótimo resultado.

Apesar de os dados mostrarem um bom desempenho do modelo, foram observadas algumas limitações. Como a base de treinamento foi de apenas um tipo de prontuário, o modelo teve dificuldade em classificar corretamente as impressões digitais em outros tipos. Além disso, em alguns casos de pessoas com dedos faltantes houve uma confusão entre as classes. Para solucionar esses limites do modelo, sugere-se um novo treinamento, com uma base de dados maior, com diversos tipos de prontuários e com dedos faltantes. Ademais, para aumentar a confiança das detecções, sugere-se um novo treinamento com uma amostra de treinamento superior a 5.000 imagens. O processamento de imagem utilizado nesse trabalho pode servir de apoio para a criação de outras anotações para treinamento de um modelo. Independentemente das limitações observadas e considerando que, em geral, cada estado apresenta apenas um ou dois tipos de prontuários civis, o modelo treinado está

apto para a detecção das impressões digitais de acordo com cada dedo (classes) para uma futura inserção em massa em sistemas automatizados de identificação biométrica.

REFERÊNCIAS

- ANUSHA, K.; KUMAR, P. S. Fingerprint image enhancement for crime detection using deep learning. *In: SPRINGER. Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 2.* [S.l.: s.n.], 2023. p. 257–268.
- BARROS, R. M. *et al.* Surface-assisted laser desorption/ionization mass spectrometry analysis of latent fingermarks using greenly synthesized silver nanoparticles. **Surfaces**, MDPI, v. 6, n. 4, p. 341–350, 2023.
- Biometrics. 2002. ISO/IEC JTC 1/SC 37:2002. International Organization for Standardization and International Electrotechnical Commission.
- CANNY, J. A computational approach to edge detection. **IEEE Transactions on pattern analysis and machine intelligence**, Ieee, n. 6, p. 679–698, 1986.
- CARVALHO, D. da S. *et al.* Outcome of sex determination from ulnar and radial ridge densities of brazilians' fingerprints: applying an existing method to a new population. **Science & Justice**, Elsevier, v. 62, n. 2, p. 181–192, 2022.
- CARVALHO, L. A. M. N. de; ALMEIDA, S. dos S. Impacto dos laudos de perícia papiloscópica em locais de crime nas decisões judiciais. **Research, Society and Development**, v. 9, n. 7, p. e110973658–e110973658, 2020.
- DIWAN, T.; ANIRUDH, G.; TEMBHURNE, J. V. Object detection using yolo: Challenges, architectural successors, datasets and applications. **multimedia Tools and Applications**, Springer, v. 82, n. 6, p. 9243–9275, 2023.
- DOXYGEN. **OpenCV modules**. 2023. <<https://docs.opencv.org/3.4/>>. Accessed: 2023-10-26.
- GAO, M. *et al.* Statistics of fingerprint minutiae frequency and distribution based on automatic minutiae detection method. **Forensic Science International**, Elsevier, v. 344, p. 111572, 2023.
- GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 2nd ed. ed. [S.l.: s.n.]: Prentice Hall, 2002. ISBN 9780201180756,0201180758,0130946508,9780130946508.
- HUSSAIN, M. Yolo-v1 to yolo-v8, the rise of yolo and its complementary nature toward digital manufacturing and industrial defect detection. **Machines**, MDPI, v. 11, n. 7, p. 677, 2023.
- JAIN, A.; BOLLE, R.; PANKANTI, S. **Biometrics: personal identification in networked society**. [S.l.: s.n.]: Springer Science & Business Media, 1999. v. 479.
- JIAN, W.; ZHOU, Y.; LIU, H. Densely connected convolutional network optimized by genetic algorithm for fingerprint liveness detection. **IEEE Access**, v. 9, p. 2229–2243, 2021.
- JOCHER, G.; CHAURASIA, A.; QIU, J. **Ultralytics YOLOv8**. 2023. Available at: <<https://github.com/ultralytics/ultralytics>>.

JOCHER glenn *et al.* **Ultralytics: YOLO: A Brief History.** 2023. <<https://docs.ultralytics.com/>>. Accessed: 2023-10-20.

RIM, B.; KIM, J.; HONG, M. Fingerprint classification using deep learning approach. **Multimedia Tools and Applications**, Springer, v. 80, p. 35809–35825, 2021.

SHAFAGHI, H. *et al.* A fast and light fingerprint-matching model based on deep learning approaches. **Journal of Signal Processing Systems**, Springer, v. 95, n. 4, p. 551–558, 2023.

SOILLE, P. *et al.* **Morphological image analysis: principles and applications.** [S.l.: s.n.]: Springer, 1999. v. 2.

SONKA, M.; HLAVAC, V.; BOYLE, R. **Image Processing, Analysis, and Machine Vision.** 4. ed. [S.l.: s.n.]: Cengage, 2014. ISBN 1133593607,9781133593607.

TERVEN, J.; CORDOVA-ESPARZA, D. A comprehensive review of yolo: From yolov1 and beyond. **arXiv preprint arXiv:2304.00501**, 2023.

TRUSIAK, K.; SAEED, K. Finger minutiae extraction based on the use of yolo-v5. *In: SPRINGER. International Conference on Computer Information Systems and Industrial Management.* [S.l.: s.n.], 2023. p. 36–48.