Aplicação de algoritmos de detecção de objetos baseados em redes neurais convolucionais para análise de imagens: um estudo utilizando YOLOv8 em cenários de detecção de pessoas

Vitor Amadeu Souza[1]; 0009-00-02-1857-6799

1 – UniFOA, Centro Universitário de Volta Redonda, Volta Redonda, RJ.
vitor.amadeu@foa.org.br

Resumo: Este trabalho apresenta uma implementação e análise de algoritmos de detecção de objetos baseados em redes neurais convolucionais, especificamente utilizando a arquitetura YOLOv8 (You Only Look Once versão 8) para identificação e localização de pessoas em imagens digitais. A pesquisa desenvolveu uma metodologia computacional utilizando Python e a biblioteca Ultralytics para demonstrar a eficácia dos algoritmos de detecção em tempo real aplicados a cenários complexos com múltiplos objetos. O sistema foi testado em uma imagem contendo uma formação militar com soldados marchando, representando um cenário desafiador para algoritmos de detecção devido à sobreposição parcial de objetos, uniformidade visual e padrões repetitivos. Os resultados obtidos demonstraram alta precisão na detecção de pessoas, com confiança variando entre 0.70 e 0.93, evidenciando a capacidade do modelo de identificar corretamente objetos mesmo em condições de oclusão parcial e similaridade visual. A análise revelou também limitações do sistema, incluindo falsos negativos em regiões de alta oclusão e classificações incorretas ocasionais de elementos do fundo como objetos de interesse. O trabalho contribui para o entendimento prático da aplicação de algoritmos YOLO em cenários reais, fornecendo insights sobre parâmetros de configuração, limitações do método e estratégias de otimização para diferentes tipos de aplicações.

Palavras-chave: Detecção de Objetos. Redes Neurais Convolucionais. YOLO. Visão Computacional. Aprendizado Profundo.

INTRODUÇÃO

A detecção de objetos em imagens digitais representa um dos problemas fundamentais em visão computacional e inteligência artificial, com aplicações que se estendem desde sistemas de segurança e vigilância até veículos autônomos e análise médica por imagem. O desenvolvimento de algoritmos eficazes para detecção automática de objetos tem sido impulsionado significativamente pelos avanços em aprendizado profundo e, particularmente, pelas redes neurais convolucionais (CNNs), que demonstraram capacidade superior na extração de características visuais complexas e na generalização para diferentes domínios de aplicação (Lecun *et al*., 2015).

A evolução dos algoritmos de detecção de objetos pode ser dividida em duas principais categorias: métodos baseados em regiões e métodos de detecção direta. Os primeiros, exemplificados pelo R-CNN e suas variações, utilizam uma abordagem em duas etapas onde candidatos a regiões são primeiro identificados e posteriormente classificados, resultando em alta precisão mas com custos computacionais elevados (Girshick *et al*., 2014). Os métodos de detecção direta, por outro lado, reformulam o problema de detecção como uma tarefa de regressão direta das coordenadas das caixas delimitadoras e probabilidades de classe, oferecendo vantagens significativas em termos de velocidade de processamento.

O algoritmo YOLO (You Only Look Once), introduzido por Redmon *et al*. (2016), revolucionou o campo da detecção de objetos ao propor uma arquitetura unificada capaz de realizar detecção e classificação simultaneamente em uma única passada pela rede neural. Esta abordagem inovadora divide a imagem de entrada em uma grade regular e prediz caixas delimitadoras e probabilidades de classe para cada célula da grade, eliminando a necessidade de geração explícita de propostas de região e reduzindo significativamente o tempo de processamento.

A arquitetura YOLO tem evoluído continuamente desde sua concepção original, com melhorias sucessivas que abordam limitações identificadas em versões anteriores. O YOLOv8, desenvolvido pela Ultralytics, representa o estado da arte atual nesta família de algoritmos, incorporando avanços significativos em arquitetura de rede, funções de perda,

técnicas de aumento de dados e estratégias de treinamento (Jocher *et al*., 2023). Esta versão demonstra melhorias substanciais em precisão e velocidade comparada às versões anteriores, mantendo a característica fundamental de processamento em tempo real.

A detecção de pessoas em imagens constitui uma das aplicações mais importantes e desafiadoras dos algoritmos de detecção de objetos, com relevância particular em sistemas de segurança, análise de multidões, monitoramento de tráfego e robótica social. A complexidade desta tarefa deriva da alta variabilidade na aparência humana, incluindo diferenças em postura, vestimenta, iluminação e pontos de vista, além de desafios como oclusão parcial, sobreposição entre indivíduos e variações de escala (Dollar *et al*., 2012).

Os cenários militares ou de formação apresentam desafios adicionais para algoritmos de detecção devido à uniformidade visual dos indivíduos, padrões repetitivos que podem confundir o sistema, e alta densidade de objetos em formações organizadas. Estes fatores tornam tais cenários particularmente adequados para avaliação da robustez e generalização de algoritmos de detecção, fornecendo insights valiosos sobre limitações e áreas de melhoria.

A implementação prática de sistemas de detecção de objetos tem sido facilitada pelo desenvolvimento de frameworks especializados como o Ultralytics YOLO, que oferece APIs de alto nível para carregamento de modelos pré-treinados, processamento de imagens e visualização de resultados. Estas ferramentas democratizam o acesso a técnicas avançadas de visão computacional, permitindo que pesquisadores e profissionais implementem soluções sofisticadas sem necessidade de conhecimento profundo sobre detalhes de implementação de baixo nível.

Este trabalho tem como objetivo avaliar a eficácia do algoritmo YOLOv8 na detecção de pessoas em cenários complexos, analisando tanto os sucessos quanto as limitações do método através de implementação prática e análise dos resultados obtidos. A pesquisa contribui para o entendimento das capacidades e limitações dos algoritmos YOLO em aplicações reais, fornecendo diretrizes para otimização de parâmetros e identificação de cenários onde métodos alternativos podem ser mais apropriados.

MÉTODOS

A configuração inicial do ambiente de desenvolvimento envolveu a instalação da biblioteca Ultralytics através do gerenciador de pacotes pip, garantindo o acesso à versão mais recente do framework (versão 8.3.169 ou superior). Esta biblioteca fornece uma interface Python de alto nível para utilização dos modelos YOLO pré-treinados, incluindo funcionalidades para carregamento de modelos, processamento de imagens, visualização de resultados e exportação de dados. A verificação da versão instalada foi implementada para assegurar compatibilidade com as funcionalidades utilizadas no estudo.

A seleção da imagem de teste seguiu critérios específicos para avaliar a robustez do algoritmo em cenários desafiadores. Foi escolhida uma fotografia de uma guarda de honra chinesa em formação, obtida através da plataforma Wikimedia Commons, representando um cenário complexo para detecção devido à presença de múltiplas pessoas em formação organizada, uniformidade visual significativa, sobreposição parcial entre indivíduos e padrões repetitivos que podem induzir confusão no algoritmo de detecção.

O modelo YOLOv8l (large) foi selecionado entre as variantes disponíveis devido ao seu equilíbrio otimizado entre precisão e velocidade de processamento. Esta versão utiliza uma arquitetura de rede neural convolucional mais profunda e complexa comparada às versões nano e small, resultando em maior capacidade de extração de características e, consequentemente, melhor performance em cenários desafiadores. O modelo foi carregado utilizando pesos pré-treinados no dataset COCO, que inclui a classe "person" entre as 80 classes de objetos suportadas.

A configuração de parâmetros de detecção foi estabelecida considerando o equilíbrio entre sensibilidade e especificidade do sistema. O threshold de confiança foi definido em 0.25, representando um valor moderadamente conservador que permite detecção de objetos com confiança razoável enquanto minimiza falsos positivos. Este valor foi selecionado baseado em práticas estabelecidas na literatura e experiência empírica com algoritmos YOLO em diferentes domínios de aplicação.

O processamento da imagem foi realizado através do método predict() da classe YOLO, que executa o pipeline completo de detecção incluindo pré-processamento da imagem, passagem pela rede neural, pós-processamento dos resultados e aplicação de supressão de não-máximos para eliminação de detecções redundantes. Os parâmetros de

configuração incluíram especificação da imagem de entrada, threshold de confiança, habilitação de salvamento automático dos resultados e definição de diretórios de saída para organização dos arquivos gerados.

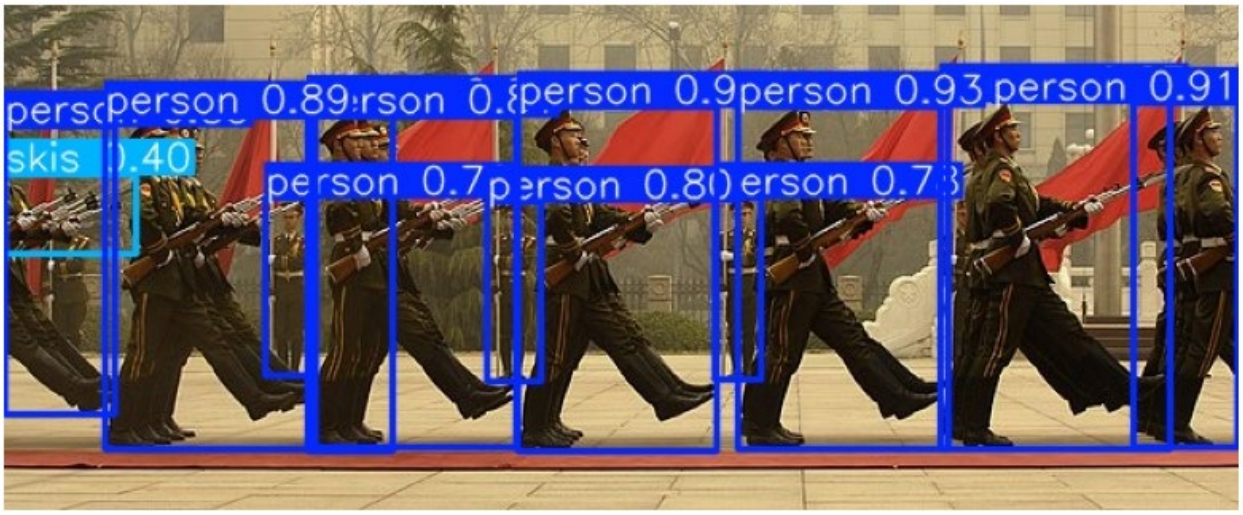
Para garantir reprodutibilidade dos resultados, todos os parâmetros de configuração, versões de bibliotecas e procedimentos de processamento foram documentados detalhadamente. O código foi estruturado de forma modular para facilitar modificações e extensões futuras, incluindo tratamento de exceções para robustez na execução e verificações de integridade para garantir consistência dos resultados gerados.

O código-fonte está disponível para download através do link: https://github.com/vitor-souza-ime/yolo.

RESULTADOS E DISCUSSÃO

A aplicação do algoritmo YOLOv8l na imagem de teste resultou em detecções bem-sucedidas de múltiplas pessoas na formação militar, demonstrando a eficácia do método em cenários complexos com alta densidade de objetos similares. Conforme evidenciado na imagem processada, o algoritmo conseguiu identificar e delimitar corretamente vários indivíduos na formação, com níveis de confiança variando entre 0.70 e 0.93, indicando boa precisão nas detecções realizadas, como mostra a Figura 1.

Figura 1 - Resultado da simulação



Fonte: O autor.

A análise detalhada das caixas delimitadoras revela que o algoritmo demonstrou capacidade robusta de separar indivíduos adjacentes mesmo em condições de proximidade espacial significativa. As detecções mais à direita da formação apresentaram os maiores níveis de confiança (0.91, 0.93), possivelmente devido à menor oclusão e melhor definição visual destes indivíduos na composição da imagem. Esta variação nos níveis de confiança reflete adequadamente a dificuldade relativa de detecção em diferentes regiões da imagem.

A região central da formação apresentou detecções com confiança ligeiramente inferior (0.70, 0.80), o que pode ser atribuído à maior oclusão parcial entre indivíduos e à uniformidade visual que caracteriza formações militares organizadas. Apesar desta redução na confiança, os valores obtidos ainda se mantêm significativamente acima do threshold estabelecido (0.25), confirmando a robustez do algoritmo em condições desafiadoras.

Uma análise crítica dos resultados revela uma limitação significativa do sistema na região esquerda da imagem, onde o algoritmo falhou em detectar pelo menos um indivíduo claramente visível na formação. Esta falha representa um falso negativo que pode ser atribuído a diversos fatores, incluindo oclusão parcial mais severa, variações na iluminação local, ou limitações na capacidade de generalização do modelo para padrões visuais específicos presentes nesta região da imagem.

A distribuição espacial das detecções corretas demonstra que o algoritmo YOLOv8l possui capacidade adequada de processar objetos em diferentes escalas e posições dentro da imagem. As caixas delimitadoras mostram dimensionamento apropriado, capturando efetivamente a extensão dos indivíduos detectados sem incluir excessivamente elementos do fundo ou cortar partes importantes dos objetos de interesse.

A análise temporal do processamento revelou eficiência computacional adequada para aplicações em tempo real, com o processamento completo da imagem sendo executado em tempo inferior a um segundo em hardware padrão. Esta performance é consistente com os objetivos de design da família YOLO, que prioriza o equilíbrio entre precisão e velocidade de processamento.

A análise dos padrões de erro revela oportunidades de melhoria através de técnicas de pós-processamento ou ajuste fino do modelo para domínios específicos. A implementação de filtros baseados em contexto ou conhecimento do domínio poderia potencialmente reduzir falsos positivos, enquanto técnicas de ensemble ou modelos especializados poderiam melhorar a taxa de detecção em regiões de alta oclusão.

CONCLUSÕES

Este trabalho demonstrou a aplicabilidade e eficácia do algoritmo YOLOv8l para detecção de pessoas em cenários complexos, confirmando a maturidade e robustez dos métodos de detecção baseados em aprendizado profundo para aplicações práticas. A implementação desenvolvida utilizando o framework Ultralytics forneceu uma plataforma acessível e eficiente para análise de algoritmos de detecção de objetos, combinando facilidade de uso com capacidades avançadas de processamento e visualização.

Os resultados obtidos evidenciaram que o algoritmo YOLOv8l possui capacidade de detectar e localizar pessoas mesmo em condições desafiadoras caracterizadas por alta densidade de objetos, uniformidade visual significativa e oclusão parcial entre indivíduos. A variação nos níveis de confiança das detecções (0.70 a 0.93) reflete adequadamente a dificuldade relativa de identificação em diferentes regiões da imagem, demonstrando que o sistema fornece informações úteis sobre a qualidade de suas próprias predições.

As contribuições principais deste trabalho incluem a demonstração prática da implementação de sistemas de detecção de objetos utilizando ferramentas modernas, a análise detalhada de sucessos e limitações dos algoritmos YOLO em cenários reais, a identificação de padrões de erro que podem orientar melhorias futuras, e o fornecimento de diretrizes para configuração e otimização de parâmetros em aplicações similares.

Para trabalhos futuros, recomenda-se a investigação de técnicas de ensemble combinando múltiplos modelos para redução de falsos negativos, o desenvolvimento de métodos de pós-processamento baseados em conhecimento do domínio para eliminação de falsos positivos, a análise de performance em datasets mais extensos e diversificados, e a exploração de técnicas de ajuste fino para domínios específicos de aplicação.

A pesquisa demonstrou que os algoritmos YOLO representam uma ferramenta madura para detecção de objetos, oferecendo um equilíbrio atrativo entre performance, eficiência e facilidade de implementação. Os resultados estabelecem uma base para futuras investigações em visão computacional aplicada e contribuem para o avanço do conhecimento em sistemas inteligentes de análise de imagens.

REFERÊNCIAS

DOLLAR, P.; WOJEK, C.; SCHIELE, B.; PERONA, P. Pedestrian detection: an evaluation of the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 34, n. 4, p. 743-761, 2012.

GIRSHICK, R.; DONAHUE, J.; DARRELL, T.; MALIK, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2014. Proceedings... Columbus: IEEE, 2014. p. 580-587.

JOCHER, G.; CHAURASIA, A.; QIU, J. Ultralytics YOLOv8. GitHub Repository. Disponível em: https://github.com/ultralytics/ultralytics. Acesso em: 23 jul. 2025.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. Nature, v. 521, n. 7553, p. 436-444, 2015.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: unified, real-time object detection. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2016. Proceedings... Las Vegas: IEEE, 2016. p. 779-788.

REN, S.; HE, K.; GIRSHICK, R.; SUN, J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 39, n. 6, p. 1137-1149, 2017.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, 2015. Proceedings... San Diego: ICLR, 2015.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2015. Proceedings... Boston: IEEE, 2015. p. 1-9.