

Implementação e avaliação do YOLO11 para detecção de objetos em tempo real em ambientes dinâmicos

Vitor Amadeu Souza¹; 0009-00-02-1857-6799

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

Resumo: A detecção de objetos em tempo real tem se tornado uma área fundamental da visão computacional, com aplicações que abrangem desde sistemas de segurança até automação industrial. Este trabalho apresenta uma implementação e análise da versão mais recente do algoritmo You Only Look Once (YOLO11) para detecção de objetos em tempo real através de captura de vídeo. O estudo foi conduzido utilizando uma implementação baseada na biblioteca Ultralytics, com processamento de imagens através do OpenCV. Os experimentos foram realizados com diferentes tipos de objetos, incluindo dispositivos eletrônicos (telefones celulares e controles remotos) e recipientes (garrafas), demonstrando a versatilidade e precisão do modelo. Os resultados obtidos mostram taxas de confiança variando entre 0.67 e 0.97, indicando alta precisão na detecção de objetos cotidianos. A análise comparativa com versões anteriores do YOLO revela melhorias significativas em termos de velocidade de processamento e precisão de detecção, confirmando o avanço tecnológico representado pelo YOLO11. O sistema desenvolvido demonstrou capacidade de processamento em tempo real, mantendo estabilidade mesmo em condições de captura diversas. As contribuições deste trabalho incluem a validação do desempenho do YOLO11 em cenários práticos e a demonstração de sua aplicabilidade em sistemas de detecção automatizada.

Palavras-chave: Detecção de objetos. YOLO11. Visão computacional. Tempo real. Aprendizado profundo.

INTRODUÇÃO

A detecção de objetos representa um dos campos mais dinâmicos e desafiadores da visão computacional contemporânea, constituindo a base para inúmeras aplicações tecnológicas que impactam diretamente a sociedade moderna. Desde os primórdios da área, quando técnicas rudimentares baseadas em características manuais eram predominantes, até os avanços revolucionários proporcionados pelo aprendizado profundo, a evolução desta disciplina tem sido marcada por inovações que continuamente expandem as fronteiras do possível em termos de reconhecimento automático de padrões visuais (Zhao *et al.*, 2019).

O paradigma tradicional de detecção de objetos, baseado em abordagens de duas etapas, enfrentava limitações significativas em termos de velocidade de processamento, tornando-se inadequado para aplicações que exigem resposta em tempo real. Métodos como o R-CNN (Region-based Convolutional Neural Networks) e suas variantes, apesar de demonstrarem precisão considerável, apresentavam latências incompatíveis com as demandas de sistemas automatizados modernos (Girshick *et al.*, 2014). Esta limitação motivou o desenvolvimento de arquiteturas inovadoras que pudessem conciliar precisão e velocidade, culminando no surgimento da família de algoritmos YOLO (You Only Look Once), que revolucionou o campo ao introduzir uma abordagem de detecção em uma única etapa.

A arquitetura YOLO, proposta inicialmente por Redmon *et al.* (2016), representou um marco paradigmático ao reformular o problema de detecção de objetos como uma questão de regressão espacialmente separada, permitindo que a localização e classificação de objetos fossem realizadas simultaneamente em uma única passagem pela rede neural. Esta abordagem não apenas acelerou drasticamente o processo de detecção, mas também simplificou significativamente a arquitetura do sistema, eliminando a necessidade de múltiplas etapas de processamento características dos métodos anteriores. A evolução subsequente da família YOLO tem sido caracterizada por melhorias consistentes tanto em termos de precisão quanto de eficiência computacional, com cada nova versão incorporando avanços teóricos e práticos que ampliam suas capacidades e aplicabilidade.

A motivação para este estudo surge da necessidade de fornecer uma avaliação sistemática do YOLO11, contribuindo para o corpo de conhecimento científico na área de visão computacional e oferecendo insights para pesquisadores e profissionais que consideram a implementação deste algoritmo em suas aplicações específicas. A metodologia adotada neste trabalho enfatiza a reprodutibilidade, utilizando equipamentos e configurações acessíveis que refletem cenários típicos de implementação, garantindo que os resultados obtidos sejam representativos de aplicações do mundo real.

MÉTODOS

A implementação do sistema de detecção foi realizada utilizando a linguagem de programação Python 3.12, escolhida por sua ampla adoção na comunidade científica e pela disponibilidade de bibliotecas especializadas em processamento de imagens e aprendizado de máquina. A biblioteca Ultralytics foi selecionada como framework principal para a implementação do YOLO11. Para o processamento de imagens e operações de visão computacional, foi utilizada a biblioteca OpenCV na versão 4.5.0, reconhecida como padrão da indústria para este tipo de aplicação (Bradski, 2000).

A configuração experimental foi projetada para simular condições típicas de operação em ambientes internos, utilizando iluminação artificial padrão de escritório e uma câmera web convencional com resolução de 640x480 pixels, representando um cenário comum de implementação em aplicações comerciais e acadêmicas. Esta escolha visa demonstrar a aplicabilidade do YOLO11 em contextos onde recursos computacionais e equipamentos de captura sofisticados podem não estar disponíveis, ampliando o espectro de aplicações potenciais do sistema desenvolvido.

O modelo YOLO11 utilizado nos experimentos corresponde à versão nano (yolo11n.pt), caracterizada por sua otimização para aplicações que priorizam velocidade de processamento sobre precisão máxima. Esta escolha metodológica justifica-se pela necessidade de demonstrar que mesmo versões computacionalmente menos intensivas do algoritmo podem produzir resultados satisfatórios para uma ampla gama de aplicações. A arquitetura nano do YOLO11 apresenta aproximadamente 3,2 milhões de parâmetros,

significativamente menos que as versões medium e large, mas mantendo capacidades robustas de detecção conforme demonstrado em estudos preliminares (Jegham, 2024).

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

RESULTADOS E DISCUSSÃO

Os experimentos conduzidos com o sistema de detecção baseado em YOLO11 produziram resultados que demonstram a eficácia e robustez do algoritmo em condições práticas de operação. A análise quantitativa dos dados coletados revela padrões consistentes de desempenho que confirmam as expectativas teóricas sobre as capacidades desta nova versão do algoritmo, ao mesmo tempo em que fornecem insights valiosos sobre suas características operacionais específicas.

A detecção de telefones alcançou nível de confiança de 0.86, valor que indica clareza na identificação e classificação do objeto. Este resultado é significativo considerando a complexidade visual dos dispositivos móveis modernos, que frequentemente apresentam superfícies reflexivas, geometrias curvas e detalhes visuais que podem desafiar algoritmos de detecção menos sofisticados. A capacidade do YOLO11 de manter estabilidade na detecção destes objetos reflete os avanços arquiteturais incorporados nesta versão, particularmente no que se refere ao processamento de características visuais complexas e à robustez contra variações de iluminação e reflexão (Khanam; Hussain, 2024).

A análise da detecção de controle remoto produziu o resultado com scores de confiança alcançando 0.97, representando o mais alto nível observado em todos os experimentos conduzidos. Este desempenho pode ser atribuído às características visuais distintivas dos controles remotos, que tipicamente apresentam geometrias regulares, padrões de botões bem definidos e contrastes cromáticos que facilitam sua identificação por algoritmos de visão computacional. A consistência destes resultados ao longo de múltiplas capturas sugere que o YOLO11 desenvolveu representações internas robustas para esta categoria de objetos, demonstrando a eficácia do processo de treinamento utilizado no desenvolvimento do modelo.

O desempenho na detecção de garrafa apresentou score de confiança de 0.67, que embora ainda dentro de faixas aceitáveis para aplicações práticas, foi inferior aos obtidos com os outros objetos eletrônicos testados. A natureza cilíndrica e superfícies curvas, apresenta desafios únicos para algoritmos de detecção, particularmente em relação à definição precisa das bordas do objeto e à distinção entre o recipiente e seu conteúdo. A Figura 1 apresenta os objetos detectados assim como os níveis de confiança.

Figura 1 - Objetos detectados e níveis de confiança



Fonte: O autor.

A análise qualitativa das detecções produzidas pelo sistema revela aspectos adicionais sobre o comportamento do algoritmo que complementam os dados quantitativos apresentados. As caixas delimitadoras (bounding boxes) geradas pelo YOLO11 demonstraram precisão na demarcação dos contornos dos objetos detectados. Esta precisão na localização espacial é vital para aplicações que requerem não apenas identificação dos objetos, mas também informações precisas sobre sua posição e dimensões.

A eficiência computacional observada representa outro aspecto significativo dos resultados obtidos. O consumo de recursos do sistema, monitorado durante a execução dos experimentos, manteve-se em níveis compatíveis com hardware convencional, incluindo processadores de especificações modestas e ausência de unidades de processamento gráfico dedicadas. Esta eficiência torna o YOLO11 uma opção viável para implementações em dispositivos com limitações computacionais, ampliando significativamente seu potencial de adoção em aplicações embarcadas e sistemas com restrições de energia.

As implicações práticas destes resultados estendem-se além das métricas técnicas, influenciando diretamente as possibilidades de aplicação do sistema em contextos reais. A combinação de alta precisão, processamento em tempo real e eficiência computacional posiciona o YOLO11 como uma solução adequada para uma ampla gama de aplicações, desde sistemas de segurança residencial até aplicações industriais de controle de qualidade. A versatilidade demonstrada na detecção de objetos com características visuais distintas sugere potencial para customização e adaptação a domínios específicos através de treinamento adicional com datasets especializados.

CONCLUSÕES

A análise quantitativa dos dados coletados revela que o YOLO11 mantém consistentemente altos níveis de confiança na detecção de objetos diversos, com scores variando entre 0.67 e 0.97 dependendo das características específicas dos objetos analisados. Esta variação, longe de representar uma limitação do sistema, reflete a sensibilidade apropriada do algoritmo às diferentes complexidades visuais apresentadas pelos objetos testados, indicando um comportamento algorítmico sofisticado e bem calibrado. A capacidade demonstrada de alcançar scores de confiança superiores a 0.85 para a maioria dos objetos testados posiciona o YOLO11 como uma solução confiável para aplicações críticas onde a precisão na detecção constitui um requisito fundamental.

O desempenho temporal observado durante os experimentos confirma a adequação do YOLO11 para aplicações que demandam processamento em tempo real, uma característica fundamental para sistemas automatizados modernos. A manutenção de taxas de processamento superiores a 20 quadros por segundo, mesmo utilizando hardware de especificações modestas e a versão nano do algoritmo, demonstra otimizações arquiteturais significativas que tornam esta tecnologia acessível para implementações com restrições computacionais. Esta eficiência computacional representa um avanço importante na democratização de tecnologias avançadas de visão computacional, permitindo sua adoção em contextos anteriormente limitados por restrições de recursos.

A versatilidade demonstrada na detecção de objetos com características visuais distintas sublinha o potencial de generalização do algoritmo para domínios específicos. A capacidade

de detectar adequadamente tanto objetos com superfícies reflexivas e geometrias complexas, quanto recipientes com características visuais desafiadoras, sugere que o YOLO11 desenvolveu representações internas suficientemente flexíveis para adaptar-se a uma ampla variedade de contextos aplicativos. Esta versatilidade é fundamental para a viabilidade econômica de implementações, reduzindo a necessidade de desenvolvimentos customizados específicos para cada aplicação.

As implicações dos resultados obtidos estendem-se além das métricas técnicas imediatas, influenciando as perspectivas de desenvolvimento futuro na área de visão computacional. A demonstração de que algoritmos de detecção de objetos podem alcançar simultaneamente alta precisão e eficiência computacional abre possibilidades para aplicações anteriormente consideradas impraticáveis devido a limitações tecnológicas. Sistemas de monitoramento distribuído, interfaces homem-máquina baseadas em gestos, automação residencial inteligente e aplicações de realidade aumentada são exemplos de domínios que podem beneficiar-se diretamente dos avanços representados pelo YOLO11.

A contribuição científica deste estudo reside não apenas na validação do desempenho do YOLO11, mas também na demonstração de metodologias apropriadas para avaliação sistemática de algoritmos de detecção de objetos em condições reais. Os protocolos experimentais desenvolvidos e as métricas de avaliação empregadas podem servir como referência para estudos futuros na área, contribuindo para a padronização de procedimentos de avaliação e facilitando comparações entre diferentes algoritmos e implementações.

REFERÊNCIAS

BRADSKI, G. The OpenCV Library. Dr. Dobb's Journal of Software Tools, v. 25, n. 11, p. 120-126, 2000. Disponível em: <https://opencv.org>. Acesso em: 15 ago. 2024.

KHANAM, Rahima; HUSSAIN, Muhammad. Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725, 2024.

GIRSHICK, R.; DONAHUE, J.; DARRELL, T.; MALIK, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, p. 580-587, 2014. DOI: 10.1109/CVPR.2014.81. Acesso em: 18 jul. 2024.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and

Pattern Recognition, p. 779-788, 2016. DOI: 10.1109/CVPR.2016.91. Acesso em: 12 jun. 2024.

ULTRALYTICS. YOLO11 Documentation and Performance Benchmarks. Ultralytics Technical Report, 2024. Disponível em: <https://docs.ultralytics.com/models/yolo11/>. Acesso em: 30 ago. 2024.

JEGHAM, Nidhal et al. Yolo evolution: A comprehensive benchmark and architectural review of yolov12, yolo11, and their previous versions. arXiv preprint arXiv:2411.00201, 2024.

ZHAO, Z. Q.; ZHENG, P.; XU, S. T.; WU, X. Object Detection with Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems, v. 30, n. 11, p. 3212-3232, 2019. DOI: 10.1109/TNNLS.2018.2876865. Acesso em: 08 jul. 2024.