

## **Implementação de Faster R-CNN para detecção automática de veículos em ambientes de tráfego: uma abordagem baseada em deep learning**

Vitor Amadeu Souza<sup>1</sup>; 0009-0002-1857-6799

1 – UniFOA, Centro Universitário de Volta Redonda, Volta Redonda, RJ.  
[vitor.amadeu@foa.org.br](mailto:vitor.amadeu@foa.org.br)

**Resumo:** A detecção automática de objetos em imagens tem se tornado uma área de crescente interesse na comunidade científica, especialmente com o avanço das técnicas de deep learning. Este estudo apresenta a implementação do algoritmo Faster R-CNN (Region-based Convolutional Neural Network) para detecção de veículos em ambientes de tráfego urbano. O trabalho utiliza o modelo pré-treinado FasterRCNN ResNet50 FPN disponível na biblioteca PyTorch, aplicado sobre imagens contendo veículos em movimento. A metodologia empregada consistiu na implementação de um pipeline completo de detecção, incluindo carregamento de imagens via URL, pré-processamento, inferência do modelo e visualização dos resultados com bounding boxes e scores de confiança. Os resultados demonstraram precisão na identificação de veículos, com o modelo conseguindo detectar corretamente um veículo esportivo vermelho com score de confiança de 0.97, evidenciando a eficácia da arquitetura Faster R-CNN para aplicações em sistemas de transporte inteligentes e monitoramento de tráfego urbano.

**Palavras-chave:** Detecção de objetos. Faster R-CNN. Deep learning. Visão computacional. Veículos autônomos.

## INTRODUÇÃO

A detecção automática de objetos em imagens representa uma das aplicações mais promissoras da visão computacional contemporânea, com particular relevância para sistemas de transporte inteligentes e veículos autônomos. Desde os trabalhos pioneiros de Viola e Jones (2001) sobre detecção de faces usando classificadores em cascata, até as modernas arquiteturas baseadas em redes neurais convolucionais, a evolução dos algoritmos de detecção tem sido notável. A introdução das Convolutional Neural Networks (CNNs) por LeCun *et al.* (1998) revolucionou o campo da visão computacional, estabelecendo as bases para os avanços subsequentes em tarefas de classificação e detecção de objetos.

O desenvolvimento de arquiteturas específicas para detecção de objetos ganhou impulso com a criação do R-CNN (Region-based Convolutional Neural Network) por Girshick *et al.* (2014), que introduziu uma abordagem de duas etapas combinando propostas de regiões com classificação por CNNs. Esta metodologia foi posteriormente aprimorada com o Fast R-CNN (Girshick, 2015), que otimizou o processo de treinamento e inferência através de um esquema de pooling de características compartilhadas. A evolução natural dessa linha de pesquisa culminou no Faster R-CNN, proposto por Ren *et al.* (2015), que integrou a geração de propostas de regiões diretamente na arquitetura da rede neural, eliminando a dependência de algoritmos externos.

A arquitetura Faster R-CNN representa um marco na detecção de objetos por sua capacidade de operar em tempo real mantendo alta precisão. O modelo utiliza uma Region Proposal Network (RPN) para gerar propostas de objetos candidatos, seguida por uma rede de classificação e refinamento de bounding boxes. Esta abordagem de dois estágios tem demonstrado performance superior em diversos benchmarks, incluindo o PASCAL VOC (Everingham *et al.*, 2010) e o Microsoft COCO (Lin *et al.*, 2014), tornando-se uma referência para aplicações em sistemas de visão computacional.

No contexto específico da detecção de veículos, a aplicação de técnicas de deep learning tem mostrado resultados promissores para sistemas de monitoramento de tráfego e veículos autônomos. Trabalhos como os de Redmon *et al.* (2016) com YOLO e Liu *et al.* (2016) com

SSD demonstraram a viabilidade de sistemas de detecção em tempo real para aplicações veiculares.

A motivação para este estudo reside na necessidade de avaliar o desempenho do Faster R-CNN em cenários reais de detecção veicular, utilizando implementações que possam ser facilmente replicadas e adaptadas para diferentes contextos de aplicação. O uso de modelos pré-treinados, conforme proposto por Deng *et al.* (2009) no ImageNet, permite aproveitar conhecimento previamente adquirido em grandes datasets, reduzindo os requisitos computacionais e de dados para treinamento especializado.

## MÉTODOS

A metodologia empregada neste estudo baseia-se na implementação do algoritmo Faster R-CNN utilizando a biblioteca PyTorch e sua interface torchvision, seguindo as práticas estabelecidas por Paszke *et al.* (2019) para desenvolvimento de aplicações de deep learning. O pipeline implementado compreende quatro etapas principais: carregamento e pré-processamento de imagens, configuração do modelo pré-treinado, execução da inferência e visualização dos resultados com anotações gráficas.

O carregamento de imagens foi implementado através de uma função dedicada que utiliza a biblioteca requests para obtenção de imagens via URL, seguida pelo processamento com a Python Imaging Library (PIL) para conversão ao formato RGB. A transformação das imagens para tensores PyTorch é realizada mediante a aplicação de `transforms.ToTensor()`, que normaliza os valores dos pixels para o intervalo  $[0, 1]$  e reorganiza as dimensões para o formato padrão CHW (Channels, Height, Width) esperado pelas redes neurais convolucionais, conforme descrito por Howard *et al.* (2017).

O modelo Faster R-CNN empregado utiliza a arquitetura ResNet-50 como backbone, combinada com Feature Pyramid Networks (FPN) para extração hierárquica de características, seguindo a abordagem proposta por Lin *et al.* (2017). A escolha da ResNet-50 como extrator de recursos baseia-se em seu equilíbrio comprovado entre precisão e eficiência computacional, conforme demonstrado por He *et al.* (2016) em seus experimentos extensivos. O modelo é inicializado com pesos pré-treinados no dataset COCO, que contém

80 classes de objetos incluindo veículos, seguindo a metodologia de transfer learning estabelecida por Yosinski *et al.* (2014).

O sistema de classes utilizado corresponde ao conjunto COCO\_INSTANCE\_CATEGORY\_NAMES, que inclui 80 categorias de objetos do mundo real, incluindo diferentes tipos de veículos como carros, motocicletas, ônibus e caminhões. Esta taxonomia foi estabelecida por Lin *et al.* (2014) como padrão para avaliação de algoritmos de detecção de objetos e é amplamente adotada na comunidade científica.

A visualização dos resultados é implementada através de uma função específica que desenha bounding boxes retangulares ao redor dos objetos detectados, acompanhados de labels textuais indicando a classe predita e o score de confiança correspondente. A implementação utiliza a biblioteca ImageDraw do PIL para renderização gráfica, aplicando um threshold configurável para filtrar detecções com baixa confiança, seguindo as práticas recomendadas por Padilla *et al.* (2020) para avaliação de sistemas de detecção de objetos.

O threshold de confiança utilizado foi definido em 0.5, permitindo a visualização de detecções com diferentes níveis de certeza do modelo. Esta escolha metodológica possibilita uma análise mais abrangente do comportamento do algoritmo, incluindo casos limítrofes que podem ser relevantes para aplicações específicas. A seleção do threshold apropriado é vital para o desempenho do sistema, conforme discutido por Davis e Goadrich (2006) em sua análise sobre curvas de precisão em problemas de classificação desbalanceada.

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

## RESULTADOS E DISCUSSÃO

A implementação do sistema de detecção Faster R-CNN demonstrou resultados satisfatórios na identificação de veículos na imagem de teste utilizada. O modelo conseguiu detectar com sucesso um veículo esportivo vermelho com score de confiança de 0.97, indicando uma predição confiável por parte da rede neural. Este resultado está em consonância com os benchmarks reportados por Ren *et al.* (2015) no trabalho original do Faster R-CNN, que

demonstrou mean Average Precision (mAP) superior a 70% no dataset PASCAL VOC 2007. A Figura 1 apresenta o resultado.

Figura 1 - Fazendo predição com o modelo



Fonte: AUTOMOTIVO SHOPPING

A análise qualitativa da detecção revela que o bounding box gerado pelo modelo apresenta precisão espacial adequada, englobando corretamente as dimensões do veículo. Esta precisão na localização é particularmente importante para aplicações em sistemas de monitoramento de tráfego, onde a determinação exata da posição dos veículos é essencial para análises subsequentes de fluxo e densidade veicular, conforme destacado por Guerrero-Ibáñez *et al.* (2018) em sua revisão sobre sistemas de transporte inteligentes.

O score de confiança de 0.97 obtido supera o threshold típico de 0.5 utilizado em aplicações, sugerindo que o modelo possui boa confiabilidade sobre a classificação correta do objeto detectado. Esta confiança é particularmente relevante considerando que o veículo na imagem apresenta características visuais distintas, incluindo forma aerodinâmica, cor vibrante e contexto de pista de corrida, elementos que facilitam a discriminação por parte da rede neural. Trabalhos como os de Simonyan e Zisserman (2014) demonstraram que modelos de CNN tendem a apresentar maior confiança quando os objetos alvo possuem características visuais bem definidas e contrastantes com o background.

## CONCLUSÕES

A precisão espacial do bounding box gerado pelo modelo, juntamente com a confiança da predição, indica que a abordagem é viável para aplicações que demandam localização de veículos, como sistemas de contagem automática de tráfego, análise de fluxo veicular e detecção de infrações. Esta capacidade é particularmente relevante no contexto atual de crescente demanda por sistemas de transporte inteligentes e automação veicular.

A implementação demonstrou também a importância da configuração adequada de thresholds de confiança para otimizar o equilíbrio de precisão em aplicações específicas. A flexibilidade do sistema em permitir ajustes deste parâmetro oferece adaptabilidade para diferentes cenários de uso, desde aplicações que priorizam precisão até aquelas que necessitam de maior sensibilidade na detecção.

Em perspectiva futura, a integração deste sistema com tecnologias complementares como rastreamento de objetos, análise de trajetórias e predição de movimento poderia resultar em soluções mais abrangentes para monitoramento inteligente de tráfego. A combinação com dados de múltiplas câmeras e sensores adicionais poderia ampliar as capacidades de análise e controle de sistemas de transporte urbano.

## REFERÊNCIAS

DAVIS, J.; GOADRICH, M. The relationship between Precision-Recall and ROC curves. In: PROCEEDINGS OF THE 23RD INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 2006, Pittsburgh. Proceedings... New York: ACM, 2006. p. 233-240. DOI: 10.1145/1143844.1143874. Acesso em: 13 set. 2025.

DENG, J.; DONG, W.; SOCHER, R.; LI, L. J.; LI, K.; FEI-FEI, L. ImageNet: A large-scale hierarchical image database. In: IEEE COMPUTER VISION AND PATTERN RECOGNITION, 2009, Miami. Proceedings... IEEE, 2009. p. 248-255. DOI: 10.1109/CVPR.2009.5206848. Acesso em: 13 set. 2025.

EVERINGHAM, M.; VAN GOOL, L.; WILLIAMS, C. K.; WINN, J.; ZISSERMAN, A. The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, v. 88, n. 2, p. 303-338, 2010. DOI: 10.1007/s11263-009-0275-4. Acesso em: 13 set. 2025.

GIRSHICK, R. Fast R-CNN. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2015, Santiago. Proceedings... IEEE, 2015. p. 1440-1448. DOI: 10.1109/ICCV.2015.169. Acesso em: 13 set. 2025.



GIRSHICK, R.; DONAHUE, J.; DARRELL, T.; MALIK, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE COMPUTER VISION AND PATTERN RECOGNITION, 2014, Columbus. Proceedings... IEEE, 2014. p. 580-587. DOI: 10.1109/CVPR.2014.81. Acesso em: 13 set. 2025.

GUERRERO-IBÁÑEZ, J.; ZEADALLY, S.; CONTRERAS-CASTILLO, J. Sensor technologies for intelligent transportation systems. Sensors, v. 18, n. 4, p. 1212, 2018. DOI: 10.3390/s18041212. Acesso em: 13 set. 2025.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: IEEE COMPUTER VISION AND PATTERN RECOGNITION, 2016, Las Vegas. Proceedings... IEEE, 2016. p. 770-778. DOI: 10.1109/CVPR.2016.90. Acesso em: 13 set. 2025.

HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861, 2017. Disponível em: <https://arxiv.org/abs/1704.04861>. Acesso em: 13 set. 2025.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. Proceedings of the IEEE, v. 86, n. 11, p. 2278-2324, 1998. DOI: 10.1109/5.726791. Acesso em: 13 set. 2025.

LIN, T. Y.; DOLLÁR, P.; GIRSHICK, R.; HE, K.; HARIHARAN, B.; BELONGIE, S. Feature pyramid networks for object detection. In: IEEE COMPUTER VISION AND PATTERN RECOGNITION, 2017, Honolulu. Proceedings... IEEE, 2017. p. 2117-2125. DOI: 10.1109/CVPR.2017.106. Acesso em: 13 set. 2025.

LIN, T. Y.; MAIRE, M.; BELONGIE, S.; HAYS, J.; PERONA, P.; RAMANAN, D.; DOLLÁR, P.; ZITNICK, C. L. Microsoft COCO: Common objects in context. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 2014, Zurich. Proceedings... Springer, 2014. p. 740-755. DOI: 10.1007/978-3-319-10602-1\_48. Acesso em: 13 set. 2025.

LIU, W.; ANGUELOV, D.; ERHAN, D.; SZEGEDY, C.; REED, S.; FU, C. Y.; BERG, A. C. SSD: Single shot multibox detector. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 2016, Amsterdam. Proceedings... Springer, 2016. p. 21-37. DOI: 10.1007/978-3-319-46448-0\_2. Acesso em: 13 set. 2025.

PADILLA, R.; NETTO, S. L.; DA SILVA, E. A. A survey on performance metrics for object-detection algorithms. In: INTERNATIONAL CONFERENCE ON SYSTEMS, SIGNALS AND IMAGE PROCESSING, 2020, Niterói. Proceedings... IEEE, 2020. p. 237-242. DOI: 10.1109/IWSSIP48289.2020.9145130. Acesso em: 13 set. 2025.

PASZKE, A.; GROSS, S.; MASSA, F.; LERER, A.; BRADBURY, J.; CHANAN, G.; KILLEEN, T.; LIN, Z.; GIMELSHEIN, N.; ANTIGA, L.; DESMAISON, A.; KOPF, A.; YANG, E.; DEVITO,

Z.; RAISON, M.; TEJANI, A.; CHILAMKURTHY, S.; STEINER, B.; FANG, L.; BAI, J.; CHINTALA, S. PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, v. 32, p. 8024-8035, 2019. Disponível em: <https://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>. Acesso em: 13 set. 2025.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: Unified, real-time object detection. In: *IEEE COMPUTER VISION AND PATTERN RECOGNITION*, 2016, Las Vegas. *Proceedings... IEEE*, 2016. p. 779-788. DOI: 10.1109/CVPR.2016.91. Acesso em: 13 set. 2025.

REN, S.; HE, K.; GIRSHICK, R.; SUN, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, v. 28, p. 91-99, 2015. Disponível em: <https://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf>. Acesso em: 13 set. 2025.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Disponível em: <https://arxiv.org/abs/1409.1556>. Acesso em: 13 set. 2025.

VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: *IEEE COMPUTER VISION AND PATTERN RECOGNITION*, 2001, Kauai. *Proceedings... IEEE*, 2001. p. 511-518. DOI: 10.1109/CVPR.2001.990517. Acesso em: 13 set. 2025.

YOSINSKI, J.; CLUNE, J.; BENGIO, Y.; LIPSON, H. How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, v. 27, p. 3320-3328, 2014. Disponível em: <https://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf>. Acesso em: 13 set. 2025.

AUTOMOTIVO SHOPPING. O que são carros superesportivos? Disponível em: <https://automotivoshopping.com.br/o-que-sao-carros-superesportivos/>. Acesso em: 14 ago. 2025.