

COMBATENDO A DISTRAÇÃO AO VOLANTE: UMA ANÁLISE DE PRECISÃO E LATÊNCIA NA COMPARAÇÃO DE MODELOS PARA DETECÇÃO EM TEMPO REAL¹

Yago Oliveira Silva², Geraldo P. Rocha Filho³

RESUMO

A desatenção do condutor permanece como uma das principais causas de acidentes fatais no trânsito. Com o avanço das tecnologias de assistência ao motorista, de visão computacional e aprendizado de máquina, assim como do número exponencial de veículos em circulação, tornam-se cruciais os investimentos em pesquisas que envolvam soluções tecnológicas para reduzir os riscos no trânsito e promover um ambiente viário mais seguro. Este trabalho compara três modelos de detecção de objetos — YOLOv8 (You Only Live Once), Faster R-CNN (Faster Region-based Convolutional Neural Network) e SSD (Single Shot Detection) — na tarefa de detectar distrações visuais de condutores. Para isso, foi utilizado um subconjunto manualmente anotado do StateFarm Distracted Driver, com 2853 imagens distribuídas em sete categorias e divididas em 70% treino, 20% validação e 10% teste. Os modelos foram treinados no ambiente Google Colab com GPU T4 e o desempenho dos algoritmos foi aferido por meio das métricas de precisão e tempo de inferência. O estudo apresenta uma avaliação qualitativa dos três detectores no cenário de distração, destacando fatores de implementação que impactam no desempenho de cada um dos modelos.

PALAVRAS-CHAVE: Distração ao volante, Faster R- CNN, Segurança viária, SSD, Visão computacional, YOLOv8,

DISTRACTION AT THE DRIVER'S DRIVE: AN ANALYSIS OF PRECISION AND LATENCY IN THE COMPARISON OF MODELS FOR REAL-TIME DETECTION

ABSTRACT

Driver inattention remains one of the leading causes of fatal traffic accidents. With the advancement of driver assistance technologies, computer vision, and machine learning, as well as the exponential increase in the number of vehicles on the road, investment in research involving technological solutions to reduce traffic risks and promote a safer road environment is crucial. This work compares three object detection models—YOLOv8 (You Only Live Once), Faster R-CNN (Faster Region-based Convolutional Neural Network), and SSD (Single Shot Detection)—in the task of detecting driver visual distractions. To this end, we used a manually annotated subset of the StateFarm Distracted Driver dataset, with 2,853 images distributed across seven categories and divided into 70% training, 20% validation, and 10% testing. The models were trained in the Google Colab environment with a T4 GPU, and the algorithms' performance was measured using accuracy and inference time metrics. The study presents a qualitative

¹UESB - Universidade Estadual do Sudoeste da Bahia.

² Pesquisador de Iniciação Científica, Colegiado de Ciência da Computação, Universidade Estadual do Sudoeste da Bahia (UESB), Vitória da Conquista - BA, Brasil. 202110839@uesb.edu.br.

³ Professor Orientador, Colegiado de Ciência da Computação, Universidade Estadual do Sudoeste da Bahia (UESB), Vitória da Conquista - BA, Brasil. geraldo.rocha@uesb.edu.br.

evaluation of the three detectors in the distraction scenario, highlighting implementation factors that impact the performance of each of the models.

KEYWORDS: Computer vision, Distracted driving, Faster R- CNN, Road safety, SSD, YOLOv8

INTRODUÇÃO

Os Acidentes de Transporte Terrestre (ATT) representam uma crise global de saúde pública, com cerca de 1,35 milhões de óbitos anuais, sendo a principal causa de morte entre jovens [Andrade and Antunes 2020]. A distração ao volante, potencializada pelo uso de dispositivos móveis, é um fator responsável por 59% dos acidentes no Brasil [Observatório Nacional de Segurança Viária 2016]. As tecnologias como a frenagem autônoma [Hyundai Motor Brasil 2025] mitigam as consequências, mas não abordam a causa raiz: o comportamento do motorista.

Os sistemas de visão computacional são promissores para detectar essas distrações em tempo real, mas enfrentam desafios de precisão e eficiência computacional. Estudos anteriores relacionados à área concentram-se em cenários limitados, como o uso de celular [Abouelnaga et al. 2017], ou carecem de avaliações comparativas robustas que considerem métricas essenciais como o tempo de inferência, de Du et al. (2023).

Este trabalho realiza uma análise comparativa de três modelos de detecção de objetos — YOLOv8, Faster R-CNN e SSD — aplicados à identificação de sete classes distintas de distrações. Utilizando o conjunto de dados *StateFarm Distracted Driver*, os modelos foram avaliados não apenas pela precisão, mas também por seu tempo de inferência. O intuito é determinar o modelo mais viável para sistemas de monitoramento em tempo real e contribuir para avanços na segurança viária.

MATERIAIS E MÉTODOS

O objetivo desta pesquisa é realizar uma análise comparativa de desempenho no reconhecimento de distrações de motoristas, considerando três modelos de detecção de objetos: YOLOv8, SSD e Faster R-CNN. Essa avaliação foi baseada nas métricas de precisão e de tempo de inferência, para determinar o modelo com o melhor desempenho geral.

Nesse estudo foram utilizadas 2.853 imagens do dataset State Farm, disponível no Kaggle⁴. As imagens foram anotadas manualmente utilizando a ferramenta

⁴ Disponível em: www.kaggle.com. Acesso em: 05 out. 2024.

Labellmg⁵, para o formato do YOLO, e posteriormente processadas no Roboflow⁶ para padronização e conversão aos demais modelos. As anotações consistiram na criação de caixas delimitadoras que contornam as áreas em que os comportamentos distrativos ocorrem. O dataset foi disponibilizado com 10 categorias de comportamentos distrativos, que foram consolidadas em 7 classes nesta pesquisa. Essas classes de comportamentos distintos podem ser visualizadas na Tabela I, com suas respectivas instâncias. O conjunto final foi segmentado em proporções de 70% para treino (1.997 imagens), 20% para validação (571 imagens) e 10% para teste (285 imagens).

Classe	Comportamento do Condutor	Instâncias
C0	Direção normal	329
C1	Usando o celular	956
C2	Operando o rádio	262
C3	Bebendo	309
C4	Alcançando objetos no banco traseiro	122
C5	Cuidando do cabelo ou maquiagem	479
C6	Conversando com o Passageiro	396

Tabela I: Comportamentos de condutores e suas instâncias.

Fonte: Do autor.

A pesquisa compreendeu cinco fases: (i) coleta e anotação, (ii) divisão de dados, (iii) treinamento, (iv) previsão, e (v) avaliação. No treinamento, os hiperparâmetros de cada modelo foram ajustados para maximizar a precisão. Os algoritmos iteraram sobre lotes de imagens, usando as anotações para calcular a função de perda, realizar retropropagação e atualizar os pesos da rede. A inferência no conjunto de teste mediu a capacidade de generalização dos modelos.

O Faster R-CNN foi implementado em PyTorch⁷, adaptando sua camada preditora. O SSD foi construído com a API de Detecção de Objetos do TensorFlow⁸. O YOLOv8 usou a biblioteca Ultralytics⁹, partindo do modelo yolov8n.pt. Os procedimentos ocorreram no Google Colab com aceleração por GPU T4.

⁵ Disponível em: <https://github.com/HumanSignal/labellmg>. Acesso em: 05 out. 2024.

⁶ Disponível em: <https://roboflow.com>. Acesso em: 13 out. 2024.

⁷ Disponível em <https://pytorch.org>. Acesso em: 15 dez. 2024.

⁸ Disponível em <https://www.tensorflow.org>. Acesso em: 15 dez. 2024.

⁹ Disponível em <https://www.ultralytics.com>. Acesso em: 15 dez. 2024.

RESULTADOS E DISCUSSÃO

Na avaliação do desempenho dos modelos YOLOv8, SSD e Faster R-CNN, foram empregadas as métricas de precisão e de tempo de inferência. Os resultados da avaliação apresentados na Tabela II revelam a existência de uma hierarquia de desempenho. O YOLOv8 destacou-se com 98,70% de precisão, indicando uma capacidade superior de classificar corretamente as ações distrativas dos motoristas. O modelo SSD seguiu com 97,60% de precisão, enquanto o Faster R-CNN obteve uma precisão de 86,36%, inferior aos demais.

A menor precisão do Faster R-CNN pode ser atribuída à sua arquitetura fundamental de detecção em duas etapas. Diferente dos modelos de estágio único (como YOLO e SSD), esse modelo propõe "regiões de interesse" e, em um segundo momento, classifica os objetos nessas regiões. Se a primeira etapa não gerar propostas de região de alta qualidade, a precisão da classificação final será comprometida.

Modelo	Precisão (%)	Tempo de Inferência (s)
Faster R-CNN	86,36	0.0911
SSD	97,60	0.0724
YOLO v8	98,70	0.0346

Tabela II: Resultados.

Fonte: Do autor.

Também é possível observar que o YOLOv8 se mostrou superior em tempo de inferência, um fator determinante para a aplicação prática em tempo real. O modelo processou cada imagem em 0,0346 segundos. O YOLO operou aproximadamente 2,1 vezes mais rápido que o SSD, que registrou 0,0724 segundos, e foi 2,6 vezes mais rápido que o Faster R-CNN e seu tempo de 0,0911 segundos. No cenário de condução real, essa diferença não é trivial; pois se traduz diretamente em um sistema de alerta ao motorista com respostas mais ágeis, em que frações de segundo são importantes para a segurança.

A discussão evidencia o YOLOv8 como a melhor solução entre os três modelos analisados. A combinação de uma boa precisão, aliada a uma latência reduzida, o consolida como o modelo mais adequado, tanto para a implementação em sistemas

embarcados, quanto para aplicações críticas de segurança viária, mostrando confiabilidade e baixo tempo de resposta.

CONCLUSÕES

Este trabalho propôs uma análise comparativa da eficiência de modelos de detecção de objetos para identificar distrações de motoristas no trânsito. Os resultados alcançados não apenas demonstram a viabilidade dos modelos de predição analisados, como também evidenciam a eficácia do YOLOv8 em lidar com múltiplas detecções em tempo real. Embora o SSD e o Faster R-CNN tenham apresentado um desempenho consistente nas métricas analisadas, o YOLOv8 se destacou como o modelo mais indicado, combinando características indispensáveis para sistemas que envolvem segurança no trânsito: alta precisão e melhor tempo de inferência.

Em trabalhos futuros, planeja-se implementar um aplicativo funcional que utilize o YOLOv8, a fim de testar o sistema em cenários reais ou simulados de trânsito, avaliando sua capacidade de identificar e diferenciar comportamentos distrativos, ainda que em condições adversas, tais como iluminação variável e ângulos de câmera distintos.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Yehya Abouelnaga, Hesham M Eraqi, and Mohamed N Moustafa. Real-time distracted driver posture classification. arXiv preprint ar-Xiv:1706.09498, 2017.

- [2] Flávia Reis de Andrade and José Leopoldo Ferreira Antunes. Falta de atenção ao conduzir veículo automotor como causa de acidentes de trânsito nas rodovias federais brasileiras. *Revista brasileira de epidemiologia*, 23:e200085, 2020.

- [3] Yingjie Du, Xiaofeng Liu, Yuwei Yi, and Kun Wei. Optimizing road safety: advancements in lightweight yolov8 models and ghostc2f design for real-time distracted driving detection. *Sensors*, 23(21):8844, 2023.

- [4] HMB Hyundai Motor Brasil. Frenagem autônoma: como funciona esse sistema de segurança, 2025.

- [5] ONSV Observatorio Nacional de Segurança Viaria. Distracao ao volante pode trazer consequencias graves, lembra observatorio, 2016. Acesso em: 27 jan. 2025.

