

De Olho na Segurança: Como o Safe-Drive Detecta Distrações e Comportamentos Inseguros de Motoristas

Ângela Brígida Albarello¹, Edna Dias Canedo¹, Vinícius P. Gonçalves¹,
Fábio L. L. de Mendonça¹, Francisco Airton Silva², Rodolfo I. Meneguette³,
Geraldo P. Rocha Filho⁴

¹Universidade de Brasília – UnB

²Universidade Federal do Piauí – UFPI

³Universidade de São Paulo – USP

⁴Universidade Estadual do Sudoeste Bahia – UESB

angela.albarello@aluno.unb.br, {ednacanedo, vpgvinicius}@unb.br,

fabio.mendonca@redes.unb.br, faps@ufpi.edu.br, meneguette@icmc.usp.br

geraldo.rocha@uesb.edu.br

Abstract. *Urban traffic safety is one of the main challenges faced by smart cities, especially given the increasing number of accidents caused by driver distractions. Identifying and mitigating distracted and unsafe driving behaviors in real time remains an open problem due to scene complexity and the limitations of existing solutions. To address this scenario, this work proposes Safe-Drive, a hybrid computer vision solution based on two YOLO convolutional architectures: one dedicated to detecting distracted behaviors and another focused on segmenting seat belt usage. When compared to other approaches in the literature, Safe-Drive achieved a high accuracy rate and a 52% reduction in inference time in the worst-case scenario. These results highlight Safe-Drive as an effective and scalable solution for real-time detection of distracted and unsafe driver behaviors.*

Resumo. *A segurança no trânsito urbano é um dos principais desafios enfrentados por cidades inteligentes, especialmente diante do crescente número de acidentes causados por distrações ao volante. Identificar e mitigar comportamentos distraídos e inseguros de motoristas em tempo real ainda representa um problema em aberto, devido à complexidade das cenas e às limitações das soluções existentes. Para enfrentar esse cenário, este trabalho propõe o Safe-Drive, uma solução híbrida de visão computacional baseada em duas arquiteturas convolucionais do YOLO: uma dedicada à detecção de comportamentos distraídos e outra voltada à segmentação do uso do cinto de segurança. Quando comparado com outros trabalhos da literatura, o Safe-Drive mostrou alta taxa de acertos com uma redução no tempo de inferência de 52% no pior caso. Esses resultados destacam o Safe-Drive como uma solução eficaz e escalável na detecção de comportamentos distraídos e inseguros de motorista em tempo real.*

1. Introdução

O trânsito urbano permanece um dos maiores desafios para cidades inteligentes, com implicações significativas para a segurança pública e os custos econômicos [Rocha Filho et al. 2020, Ribeiro Jr et al. 2023]. Estima-se que comportamentos distraídos estejam entre as principais causas de acidentes, especialmente com o aumento do uso de dispositivos móveis ao volante [Bouhsissin et al. 2023, Qu et al. 2024, Kumar and Raja 2025]. De acordo com a Organização Mundial da Saúde (OMS), motoristas que utilizam o celular enquanto dirigem têm aproximadamente quatro vezes mais chances de se envolver em um acidente [World Health Organization 2022]. Mesmo os dispositivos com viva-voz não oferecem segurança significativa, e o ato de enviar mensagens de texto aumenta consideravelmente o risco de colisões.

O avanço das arquiteturas baseadas em visão computacional, como as redes neurais convolucionais (CNNs), representa uma alternativa promissora para o desenvolvimento de soluções inteligentes voltadas ao monitoramento de comportamentos de motoristas. Dentre esses modelos, destaca-se o *You Only Look Once* (YOLO) [Redmon et al. 2016, Ferrante et al. 2024] cujo objetivo é realizar a detecção de objetos em tempo real com precisão e eficiência computacional. O YOLO foi desenvolvido para identificar múltiplos objetos em uma única passada pela imagem, tornando-o adequado para aplicações que exigem respostas rápidas. Entretanto, no contexto da detecção de distrações ao volante, um desafio reside na sua dependência de caixas delimitadoras do YOLO, as quais podem não ser precisas para identificar elementos visuais mais sutis, por exemplo, o uso adequado do cinto de segurança. Esse desafio é explorado nesta pesquisa por meio da integração de mecanismos de segmentação semântica, com o intuito de aprimorar a capacidade de análise comportamental do modelo.

Diversos trabalhos têm sido propostos para lidar com o problema da detecção de comportamentos distraídos ao volante. A maioria desses trabalhos [Zhao et al. 2011, Abouelnaga et al. 2017, Du et al. 2023, Debsi et al. 2024, Abbass and Ban 2024, Shirole et al. 2025] foca na classificação de distrações específicas, utilizando modelos baseados em algoritmos tradicionais ou arquiteturas convolucionais. Embora tais trabalhos tragam avanços, eles geralmente tratam o problema de maneira isolada, desconsiderando elementos complementares à segurança, como o uso adequado do cinto de segurança. Estudos mais recentes [Bouhsissin et al. 2023, Shen et al. 2024] exploram modelos de detecção mais robustos, como o *Single Shot multibox Detector* (SSD) [Liu et al. 2016] e o *Faster Region-based Convolutional Neural Network* (Faster R-CNN) [Ren et al. 2015, Kshatri and Rathore 2025], priorizando a acurácia da detecção de posturas de risco, mas sem incorporar mecanismos de segmentação ou análise combinada de múltiplos comportamentos. Ainda, trabalhos como [Elshamy et al. 2024, Al-Mahbashi et al. 2025] destacam a eficácia do YOLO para tarefas de detecção de distrações, mas não abordam sua aplicação integrada com segmentação semântica para análise do uso do cinto de segurança, característica presente nesta pesquisa.

É válido salientar, entretanto, que apesar dos avanços no problema da detecção de comportamentos distraídos ao volante, a integração de múltiplas formas de análise visual, tais como a classificação de distrações e a verificação do uso do cinto de segurança, ainda impõe desafios, os quais esta pesquisa busca explorar. Destacam-se, entre esses desafios, dois principais: (i) como garantir precisão na detecção de diferentes comportamentos do

condutor; e (ii) como assegurar a eficiência do processamento em tempo real, mesmo em dispositivos com recursos computacionais escassos, sem comprometer a capacidade de resposta do sistema.

Este trabalho apresenta o Safe-Drive, uma solução voltada à detecção de comportamentos distraídos e inseguros de motoristas em tempo real. O Safe-Drive foi modelado com base em duas arquiteturas do YOLO: o YOLO-Detect, responsável pela classificação de posturas de distração ao volante; e o YOLO-Seg, responsável pela segmentação semântica para identificar o uso do cinto de segurança. Essa solução híbrida visa superar as limitações das abordagens anteriores que se baseiam exclusivamente em caixas delimitadoras. Para validar o Safe-Drive, foram conduzidos experimentos utilizando o *dataset State Farm Distracted Driver Detection* [StateFarm 2016], contemplando diferentes aspectos do problema, como a precisão por classe e o tempo médio de inferência. Os resultados obtidos demonstram que o Safe-Drive alcançou níveis satisfatórios de acurácia, com tempos de inferência reduzidos mesmo em ambientes com restrições computacionais, destacando-se como uma solução eficiente e escalável para aplicações embarcadas em veículos inteligentes.

O restante deste artigo está organizado da seguinte maneira. Na Seção 2, são apresentados os principais trabalhos relacionados à detecção de comportamentos inseguros e distraídos. A Seção 3 apresenta como o Safe-Drive foi modelado, enquanto a Seção 4 apresenta os resultados obtidos dos experimentos realizados. Por fim, a Seção 5 apresenta as considerações finais e os trabalhos futuros.

2. Trabalhos Relacionados

A detecção de comportamentos distraídos no trânsito tem se consolidado como uma área de pesquisa essencial, especialmente diante do impacto desses comportamentos na segurança urbana. Esta seção apresenta a evolução das principais abordagens propostas na literatura, destacando os avanços obtidos, os desafios enfrentados e as lacunas existentes. Em [Zhao et al. 2011], é proposta uma abordagem baseada em aprendizado de máquina para a classificação automática de comportamentos de motoristas, com foco na identificação de diferentes posturas ao volante. Para tanto, foi utilizado o *Support Vector Machine* (SVM). Os resultados foram comparados com quatro classificadores tradicionais: perceptron linear, *k-nearest neighbor* (k-NN), *Multi-Layer Perceptron* (MLP) e o classificador de Parzen. Apesar dos avanços, o estudo apresenta limitações como o uso de apenas quatro categorias de comportamento e baixa robustez em condições dinâmicas e em tempo real.

Já em [Sajid et al. 2021], é proposto um framework de aprendizado profundo para detecção de comportamentos distraídos de motoristas, com o objetivo de aumentar a segurança no trânsito por meio da identificação automatizada de posturas inadequadas ao volante. O framework é baseado no uso das redes *EfficientDet*, com *backbone EfficientNet*, treinadas por meio de aprendizado por transferência. Com foco na detecção em tempo real, o trabalho de [Fresta et al. 2025] propõe um sistema voltado à identificação de distração cognitiva. Para isso, foi conduzida uma campanha de coleta de dados com 42 participantes em um simulador de direção, enquanto os sujeitos realizavam uma tarefa cognitiva composta por vinte perguntas. Ambos os estudos apresentam limitações, como a ausência de análise multimodal, a exclusividade do uso de imagens estáticas e a restrição

das classes comportamentais a categorias visuais específicas, o que pode comprometer a generalização dos modelos para contextos mais dinâmicos e variados de direção.

Os autores no trabalho [Du et al. 2023] propõem uma arquitetura baseada no modelo YOLO para detecção em tempo real de comportamentos de direção distraída, com foco na redução do custo computacional e na ampliação da aplicabilidade em sistemas embarcados. A proposta se baseia na integração de estratégias de *lightweight design* ao YOLOv8, inspiradas na *GhostConv*, que tem se destacado por seu desempenho eficiente em computação de borda. Para isso, o trabalho introduz o módulo GhostC2f, que aplica transformações lineares para gerar mapas de características adicionais sem aumentar a carga computacional. No entanto, o estudo apresenta limitações. Há forte dependência de um conjunto específico de dados, o que pode limitar a generalização do modelo para outros contextos geográficos ou condições adversas não contempladas no treinamento. Além disso, o foco exclusivo na classificação de distrações visuais desconsidera aspectos importantes da segurança veicular, como o uso do cinto de segurança. Diferente do Safe-Drive, que combina detecção e segmentação para avaliar múltiplos fatores de risco, a abordagem de [Du et al. 2023] não utiliza segmentação semântica, o que pode afetar a acurácia em situações com sutilezas visuais.

O trabalho de [Abbass and Ban 2024] propõe uma arquitetura leve para a detecção de distração de motoristas, com foco em aplicações embarcadas em diferentes níveis de veículos autônomos. Para alcançar este objetivo, a arquitetura foi baseada no uso da MobileNet como extratora de características, seguida por uma camada de *pooling global*, múltiplas camadas densas e camadas de *dropout* para reduzir *overfitting*. Como limitação, há dificuldades específicas na classificação correta entre algumas classes visuais semelhantes. Além disso, o modelo ainda não foi avaliado em cenários reais com sensores integrados em tempo de execução contínua.

Com foco em um conjunto de dados multimodal, o trabalho de [Wang et al. 2024] apresentou o MultiFuser, um modelo baseado em transformadores para integrar dados RGB, profundidade e infravermelho. O objetivo é aprimorar a identificação precisa de comportamentos no interior de veículos e, consequentemente, reforçar a segurança no trânsito e a interação entre condutor e sistema veicular. A abordagem é baseada em um transformador de fusão multimodal que realiza a modelagem das relações espaço-temporais e a integração adaptativa de diferentes modalidades sensoriais. No entanto, uma limitação do estudo é a dependência de sensores especializados, como câmeras de profundidade, que ainda não são amplamente integrados em veículos comerciais. Além disso, o estudo não discute o desempenho do modelo em condições adversas como ruídos sensoriais, falhas de sincronização entre modalidades ou ambientes externos mais dinâmicos.

Embora os trabalhos analisados apresentem contribuições relevantes para a detecção de comportamentos distraídos ao volante, muitos ainda se limitam à análise isolada de ações visuais, dependem de sensores especializados ou enfrentam dificuldades de generalização para contextos reais e dinâmicos. Nesse cenário, o Safe-Drive se diferencia por integrar, de forma eficiente, a detecção de distrações com a segmentação semântica do uso do cinto de segurança, promovendo uma análise mais completa e precisa do comportamento do condutor. Essa combinação não apenas amplia a cobertura dos fatores de risco monitorados, como também assegura desempenho compatível com aplicações embarcadas e em tempo real, posicionando o Safe-Drive como uma alternativa promissora

frente às limitações observadas nas abordagens existentes.

3. Safe-Drive

Esta seção apresenta o Safe-Drive, uma solução para tratar do problema de detecção de comportamento de motoristas em tempo real, com foco na identificação de distrações ao volante e no uso adequado do cinto de segurança. Para tanto, foi modelado um mecanismo de detecção de comportamento com base em duas arquiteturas convolucionais do YOLO. O objetivo é aumentar a precisão na identificação de comportamentos de risco, ao mesmo tempo em que se assegura baixa latência na resposta do sistema, viabilizando alertas imediatos e ações preventivas que contribuem para uma direção mais segura e eficiente.

A Figura 1 apresenta uma visão geral do funcionamento do Safe-Drive, destacando desde a captura das imagens pela câmera até a detecção do comportamento do motorista em níveis de risco. Assume-se que cada veículo esteja equipado com uma câmera interna, responsável por registrar imagens do condutor em tempo real. Por meio das imagens capturadas por uma câmera instalada no interior do veículo, o Safe-Drive processa e detecta os comportamentos do motorista com base em duas arquiteturas convolucionais do YOLO. Para tanto, o Safe-Drive integra o YOLO-Detect e YOLO-Seg em sua arquitetura. O YOLO-Detect realiza a detecção de caixas delimitadoras e a classificação associada às categorias de distração do motorista de forma a otimizar o tempo de processamento. Já o YOLO-Seg é responsável pela segmentação, focando na identificação da presença do cinto de segurança através da análise de regiões específicas da imagem. Essa combinação permite ao Safe-Drive abordar diferentes aspectos do comportamento do motorista de forma eficiente em tempo real, detectando comportamentos como uso de celular, operação do rádio ou condução sem cinto de segurança, classificados em três escalas de riscos: Seguro, Distraído e Inseguro. As categorias utilizadas são apresentadas na Tabela 1, que relaciona cada classe identificada à respectiva escala de risco.

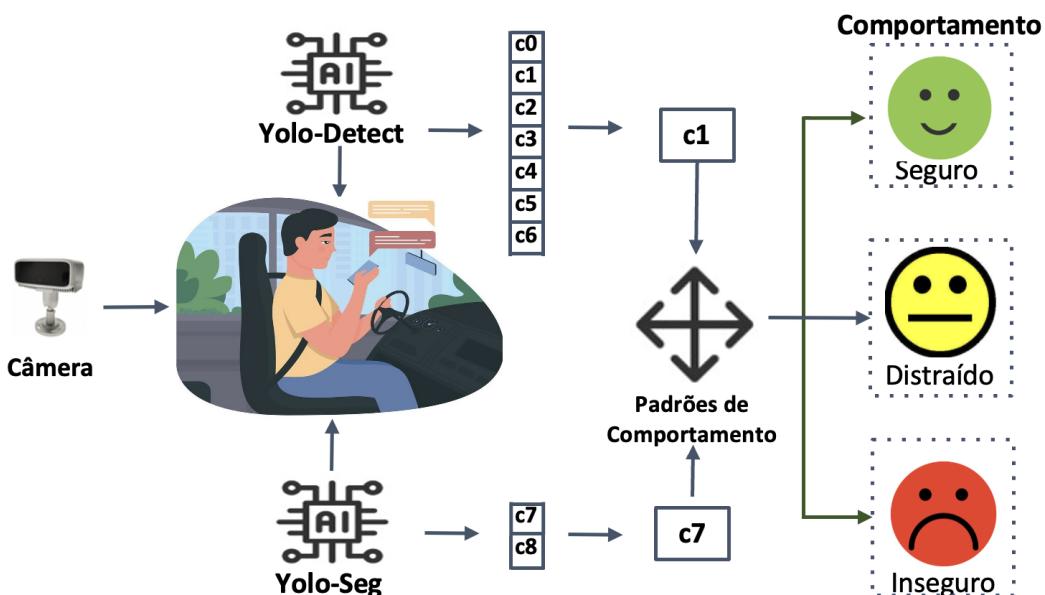


Figura 1. Cenário de funcionamento do sistema Safe-Drive.

Tabela 1. Classificação do comportamento do motorista e escala de risco do comportamento

Classe	Descrição do Comportamento	Escalas de Risco
c0	Condução adequada - sem distração (mãos ao volante e olhar frontal)	Seguro
c1	Uso do celular	Inseguro
c2	Interação com rádio	Distraído
c3	Consumo de bebida enquanto dirige	Inseguro
c4	Movimentos exagerados para alcançar objetos	Inseguro
c5	Cuidado com aparência enquanto dirige	Distraído
c6	Conversa que reduz a atenção ao trânsito	Distraído
c7	Uso do cinto de segurança	Seguro
c8	Ausência de cinto de segurança	Inseguro

3.1. Formulação da Solução

Para a construção do mecanismo de avaliação de risco do sistema Safe-Drive, é necessário formalizar a forma como os comportamentos do condutor e o uso do cinto de segurança são interpretados a partir das imagens capturadas no interior do veículo. Essa interpretação combina duas fontes de informação: (i) a classificação do comportamento do motorista, realizada pelo modelo YOLO-Detect, e (ii) a verificação do uso do cinto de segurança, realizada pelo modelo YOLO-Seg.

Define-se que o conjunto de comportamentos detectados é formado por nove classes, denotadas de c_0 a c_8 , conforme apresentado na Tabela 1. Para fins de avaliação de risco, essas classes são organizadas em três subconjuntos principais. O primeiro conjunto, D_{seguro} , é composto apenas pela classe c_0 , que representa a condução adequada, sem sinais de distração. O segundo conjunto, $D_{distraido}$, agrupa os comportamentos que desviam parcialmente a atenção do condutor e é composto pelas classes c_2 , c_5 e c_6 . Já o conjunto $D_{inseguro}$ representa comportamentos de alto risco à segurança e inclui as classes c_1 , c_3 e c_4 . Adicionalmente, o uso do cinto de segurança é avaliado a partir das classes c_7 , que representa a presença do cinto, e c_8 , que representa sua ausência.

Com base nessas definições, o mecanismo de avaliação de risco pode ser modelado como uma função $f(C, S)$ que recebe como entrada a classe de comportamento C e a classe de segurança S (relativa ao cinto de segurança), retornando como saída um dos níveis de risco:

$$f(C, S) = \begin{cases} \text{Seguro,} & \text{se } C \in D_{seguro} \wedge S = c_7 \\ \text{Distraído,} & \text{se } C \in D_{distraido} \wedge S = c_7 \\ \text{Inseguro,} & \text{se } C \in D_{inseguro} \vee S = c_8 \\ \text{Indefinido,} & \text{caso contrário} \end{cases} \quad (1)$$

Essa função reflete diretamente a lógica heurística implementada no Safe-Drive, permitindo avaliar o nível de risco associado à condição do condutor.

3.2. Mecanismo de Detecção e Avaliação de Risco do Safe-Drive

O objetivo do mecanismo de detecção do sistema Safe-Drive é identificar comportamentos de risco do motorista em tempo real, por meio da análise de imagens capturadas no

interior do veículo. Para isso, são utilizados dois modelos distintos baseados na arquitetura convolucional YOLO: o YOLO-Detect, responsável pela detecção e classificação de comportamentos distraídos, e o YOLO-Seg, destinado à segmentação semântica com foco na verificação do uso do cinto de segurança. O funcionamento do mecanismo de detecção do Safe-Drive é apresentado no Algoritmo 1.

Algorithm 1 Mecanismo de Detecção e Avaliação de Risco do Safe-Drive

Entrada: Imagem I capturada do interior do veículo

Saída: Nível de risco: Seguro, Distraído, Inseguro ou Indefinido

```

 $D_{\text{seguro}} \leftarrow \{c_0\}$ 
 $D_{\text{distraído}} \leftarrow \{c_2, c_5, c_6\}$ 
 $D_{\text{inseguro}} \leftarrow \{c_1, c_3, c_4\}$ 
 $I_{\text{proc}} \leftarrow \text{PREPROCESSARIMAGEM}(I)$ 
 $feat_{\text{det}} \leftarrow \text{BACKBONEDETECT}(I_{\text{proc}})$ 
 $neck_{\text{det}} \leftarrow \text{NECKDETECT}(feat_{\text{det}})$ 
 $out_{\text{det}} \leftarrow \text{HEADDETECT}(neck_{\text{det}})$ 
 $feat_{\text{seg}} \leftarrow \text{BACKBONESEG}(I_{\text{proc}})$ 
 $neck_{\text{seg}} \leftarrow \text{NECKSEG}(feat_{\text{seg}})$ 
 $out_{\text{seg}} \leftarrow \text{HEADSEG}(neck_{\text{seg}})$ 
 $C \leftarrow \text{None}$ 
 $S \leftarrow \text{None}$ 
if CONFIANÇA( $out_{\text{det}}$ ) > 0.25 then
     $C \leftarrow \text{CLASSIFICARCOMPORTAMENTO}(out_{\text{det}})$ 
end if
if CONFIANÇA( $out_{\text{seg}}$ ) > 0.20 then
     $S \leftarrow \text{CLASSIFICARCINTO}(out_{\text{seg}})$ 
    if  $C = \text{None}$  then
         $C \leftarrow S$ 
    end if
end if
end if
 $risco \leftarrow \begin{cases} \text{Seguro,} & \text{se } C \in D_{\text{seguro}} \text{ e } S = c_7 \\ \text{Distraído,} & \text{se } C \in D_{\text{distraído}} \text{ e } S = c_7 \\ \text{Inseguro,} & \text{se } C \in D_{\text{inseguro}} \text{ ou } S = c_8 \\ \text{Indefinido,} & \text{caso contrário} \end{cases}$ 
return  $risco$ 

```

O Algoritmo 1 é iniciado a partir da chegada de uma imagem capturada no interior do veículo. Essa imagem é inicialmente submetida a um processo de pré-processamento e, em seguida, analisada simultaneamente por dois modelos especializados: o YOLO-Detect e o YOLO-Seg, ambos estruturados em três estágios. O estágio de *Backbone* realiza a extração inicial de características visuais relevantes, como contornos, formas e texturas. Em seguida, o estágio de *Neck* promove a fusão multiescala dessas características, otimizando a representação semântica da imagem. Por fim, o estágio de *Head* executa a detecção final, associando as regiões identificadas às suas respectivas classes. No YOLO-Detect, essa estrutura é empregada para classificar o comportamento do motorista em uma

das categorias c_0 a c_6 , desde que a confiança da predição ultrapasse o limiar de 0,25. Paralelamente, o YOLO-Seg utiliza a mesma estrutura para realizar a segmentação semântica, determinando a presença ou ausência do cinto de segurança, rotulando a imagem como c_7 ou c_8 com base em uma confiança mínima de 0,20. É válido salientar que os valores 0,25 e 0,20 foram selecionados com base em estudos empíricos. Os limiares de 0,25 e 0,20 foram definidos com base em estudos empíricos, de forma a equilibrar sensibilidade e precisão nos módulos de detecção e segmentação. Caso o comportamento do condutor não seja identificado com precisão, a informação da segmentação torna-se o principal critério de análise.

Na etapa final, os resultados produzidos pelos modelos YOLO-Detect e YOLO-Seg são combinados por uma heurística integrada (Equação 1), que avalia conjuntamente o comportamento do condutor e o uso do cinto de segurança. Com base nessa análise, o Safe-Drive classifica o motorista em três categorias de risco:

- Seguro: motorista com uso adequado do cinto de segurança (c_7) e sem indícios de distração (c_0);
- Distraído: motorista utilizando o cinto (c_7), mas executando ações que desviam parcialmente sua atenção, como interagir com o rádio, conversar ou cuidar da aparência (c_2, c_5, c_6);
- Inseguro: motorista sem o cinto de segurança (c_8) ou envolvido em comportamentos de alto risco, como uso de celular, ingestão de bebidas ou movimentos bruscos para alcançar objetos (c_1, c_3, c_4).

4. Avaliação de Desempenho

Nesta seção, é avaliado o desempenho do Safe-Drive na identificação de comportamentos de distração do motorista. Para fins comparativos, o Safe-Drive foi confrontado com dois algoritmos utilizados na literatura: Faster R-CNN [Kshatri and Rathore 2025] e SSD [Liu et al. 2016]. A seguir, são apresentados o cenário modelado, as métricas utilizadas, os parâmetros adotados e os resultados obtidos.

4.1. Configurações dos Experimentos

Para avaliar o desempenho do Safe-Drive, foi utilizado o conjunto de dados *State Farm Distracted Driver Detection Dataset* [StateFarm 2016], que contém mais de 22 mil imagens organizadas em 10 categorias de distração de motoristas. Neste trabalho, optou-se por utilizar sete classes, com o objetivo de consolidar categorias semelhantes. Por exemplo, as classes que distinguem o uso do celular na mão direita ou esquerda foram unificadas na categoria “dirigindo usando celular”. Adicionalmente, um subconjunto do conjunto de dados foi anotado manualmente para identificação do uso do cinto de segurança. Essa amostragem consistiu em 200 imagens, sendo 100 com cinto e 100 sem cinto, anotadas com a ferramenta LabelImg¹. Para viabilizar a detecção baseada em segmentação semântica, um segundo conjunto de 109 imagens foi anotado manualmente com máscaras que representassem a presença do cinto de segurança, utilizando a ferramenta Roboflow².

Com o objetivo de aumentar a diversidade dos dados e melhorar o treinamento, foram aplicadas técnicas de *data augmentation*, incluindo inversão horizontal, rotações de

¹<https://pypi.org/project/labelImg/>

²<https://roboflow.com/>

90 graus e operações de ampliação. Todas as imagens foram redimensionadas para 640×640 pixels. O treinamento foi realizado em uma GPU Tesla T4 com 15 GB de memória dedicada, utilizando as seguintes configurações: 200 épocas, tamanho de lote (*batch size*) de 16 e taxa de aprendizado inicial de 0,01. A técnica de *hold-out* foi empregada para a divisão dos dados, sendo 70% para treinamento, 20% para validação e 10% para teste. O objetivo dos experimentos foi validar a capacidade do Safe-Drive em detectar comportamentos distraídos e o uso (ou não) do cinto de segurança em tempo real, avaliando tanto a acurácia do modelo quanto sua viabilidade computacional para aplicação embarcada. A Tabela 2 apresenta o conjunto de parâmetros utilizados para realizar os experimentos com o Safe-Drive.

Para avaliar o desempenho do Safe-Drive, foram utilizadas as seguintes métricas:

- Acurácia: Esta métrica avalia a proporção de previsões corretas em relação ao total de previsões realizadas.
- Precisão: Esta métrica mede a proporção de previsões positivas corretas entre todas as previsões positivas feitas pelo modelo.
- Recall: Esta métrica indica a proporção de comportamentos de risco corretamente identificados pelo modelo entre todos os casos reais dessa classe.
- mAP (mean Average Precision): Esta métrica combina precisão e recall ao longo de diferentes limiares de confiança. Isto é, avalia o desempenho considerando uma sobreposição mínima entre previsão e anotação real.
- Tempo médio de inferência: Esta métrica representa o tempo médio necessário para o modelo processar uma imagem e produzir uma previsão.

Tabela 2. Parâmetros utilizados nos experimentos com o Safe-Drive

Parâmetro	Valor padrão
Conjunto de dados	State Farm Distracted Driver Detection
Número de classes utilizadas	7 distrações + 2 para o uso do cinto
Técnicas de <i>data augmentation</i>	Inversão, rotação de 90°, ampliação
Tamanho das imagens	640×640 pixels
Divisão dos dados	70% treino, 20% validação, 10% teste
Épocas de treinamento	200
Tamanho do <i>batch</i>	16
Taxa de aprendizado	0,01
Ambiente de execução	Google Colab Pro
Hardware utilizado	GPU Tesla T4 com 15 GB de memória

4.2. Impactos dos Resultados Obtidos

A Figura 2 apresenta as métricas Precisão, Recall e mAP durante o treinamento das duas arquiteturas que compõem o Safe-Drive: o YOLO-Detect e o YOLO-Seg. Observa-se que o YOLO-Detect (Figuras 2(a), 2(b), e 2(c)) responsável pela classificação dos comportamentos distraídos, atingiu precisão superior a 0.85 (Figura 2(a)) e recall estabilizado em torno de 0.75 (Figura 2(b)). A métrica de mAP também demonstrou resultado adequado, com valores próximos de 0.80 (Figura 2(c)), indicando que o modelo foi capaz de generalizar sob critérios mais rigorosos de sobreposição. Isso significa que o Safe-Drive

não apenas reconhece corretamente os comportamentos, mas também consegue localizar com precisão as regiões específicas na imagem onde esses comportamentos ocorrem. Isso é um fator essencial para diferenciar ações visuais semelhantes, como manusear o celular ou interagir com o rádio. Por sua vez, o YOLO-Seg (Figuras 2(d), 2(e), e 2(f)), responsável pela segmentação do cinto de segurança, também apresentou desempenho consistente. A precisão se manteve acima de 0.85 ao longo das épocas (Figura 2(d)), com recall próximo de 0.75 (Figura 2(e)). As métricas de mAP demonstraram estabilidade, com aproximadamente 0.78 em mAP. Esses resultados reforçam a eficácia da abordagem integrada do Safe-Drive, demonstrando que ambos os modelos foram capazes de aprender representações visuais para seus respectivos objetivos, mantendo desempenho estável.

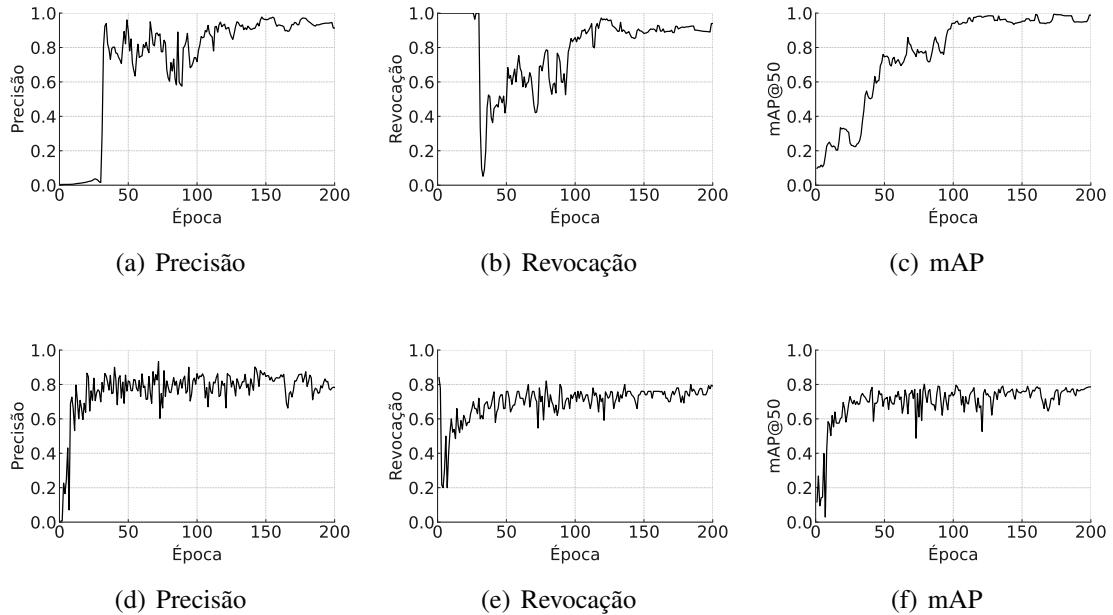


Figura 2. Desempenho das métricas Precisão, Revocação, mAP para as arquiteturas YOLO-Detect e YOLO-Seg.

A Tabela 3 apresenta os resultados de desempenho do Safe-Drive discriminados por classe, destacando a efetividade do modelo na identificação de diferentes comportamentos ao volante. O módulo YOLO-Detect, responsável pela classificação das distrações visuais (classes c_0 a c_6), demonstrou desempenho consistente, com valores de precisão e recall superiores a 0,96 em todas as classes. O mAP médio de 0,96 reforça a capacidade do modelo em não apenas identificar corretamente os comportamentos, mas também localizar com precisão as regiões da imagem associadas a esses comportamentos. Esse resultado é particularmente relevante em aplicações de segurança veicular, em que a correta delimitação espacial do comportamento é essencial para evitar falsos positivos. Por exemplo, diferenciar o uso de celular da simples movimentação das mãos próximas ao volante. Já o YOLO-Seg voltado à segmentação semântica do uso do cinto de segurança (classe c_7), apresentou desempenho considerado satisfatório, com precisão de 0,884, recall de 0,759 e mAP de 0,799. Embora esses valores sejam ligeiramente inferiores aos observados no YOLO-Detect, eles refletem a maior complexidade inerente à tarefa de segmentação, que exige o reconhecimento de padrões visuais sutis, como a presença parcial do cinto, variações de ângulo da câmera, roupas sobrepostas e condições

de iluminação adversas. Ainda assim, ao combinar as duas arquiteturas, o Safe-Drive apresentou uma média geral de precisão de 0,973, recall de 0,965 e mAP de 0,969, o que ratifica sua efetividade e confiabilidade para uso em tempo real.

Tabela 3. Desempenho do Safe-Drive por classe

Classe	Precisão	Recall	mAP
c0	0.942	1.000	0.990
c1	1.000	0.997	0.995
c2	0.996	0.982	0.989
c3	0.995	1.000	0.995
c4	0.998	1.000	0.995
c5	0.995	1.000	0.995
c6	0.978	0.982	0.994
c7	0.884	0.759	0.799
Média Geral	0.973	0.965	0.969

A Figura 3 apresenta a comparação do Safe-Drive, comparando-o com outros dois modelos da literatura o Faster R-CNN e o SSD. O Safe-Drive apresenta o menor tempo médio de inferência (Figura 3(a)), com 0,0346 segundos por imagem, sendo aproximadamente 52% mais rápido que o SSD (0,0724 s) e cerca de 62% mais rápido que o Faster R-CNN (0,0911 s). Esse desempenho superior em tempo de resposta pode ser atribuído à arquitetura eficiente do Safe-Drive, que realiza a detecção de objetos em uma única etapa (*one-stage detector*), diferentemente do Faster R-CNN, que adota uma abordagem em duas etapas (*two-stage detector*), o que naturalmente introduz maior latência. Além disso, o Safe-Drive foi projetado com otimizações que reduzem a complexidade computacional sem comprometer a precisão, como a especialização dos modelos YOLO-Detect e YOLO-Seg para tarefas específicas, simplificando cada rede. Essas características tornam o Safe-Drive eficiente para uso em dispositivos embarcados, destacando sua aplicabilidade na identificação de comportamentos de risco em tempo real no contexto veicular.

Em termos de precisão (Figura 3(b)), o Safe-Drive atingiu 98,7%, superando o Faster R-CNN (86,36%) e mantendo desempenho similar com o SSD (97,6%). Esse resultado pode ser atribuído à capacidade do Safe-Drive de realizar detecções mais refinadas com menor taxa de falsos positivos devido a sua arquitetura otimizada com camadas de detecção ancoradas em múltiplas escalas. Já no quesito recall (Figura 3(c)), o SSD obteve uma leve vantagem com 98,5%, seguido do Safe-Drive com 98,2%. Isso demonstra que ambos os modelos são eficazes em recuperar os comportamentos relevantes, sendo o desempenho do Safe-Drive impulsionado por sua estrutura de detecção unificada.

Quanto à acurácia (Figura 3(d)), o Faster R-CNN alcançou o melhor resultado (98,95%), seguido do Safe-Drive (98,67%). Isso ocorre pois o Faster R-CNN aproveita a abordagem em duas etapas que tende a ser mais criteriosa na seleção e refinamento das regiões de interesse. No entanto, essa vantagem vem com o custo de maior tempo de processamento e inferência como pode ser revisto na Figura 3(a). Portanto, o Safe-Drive apresenta um equilíbrio entre desempenho e eficiência computacional, consolidando-se como uma alternativa promissora para aplicações em tempo real de detecção de comportamentos de risco no trânsito.

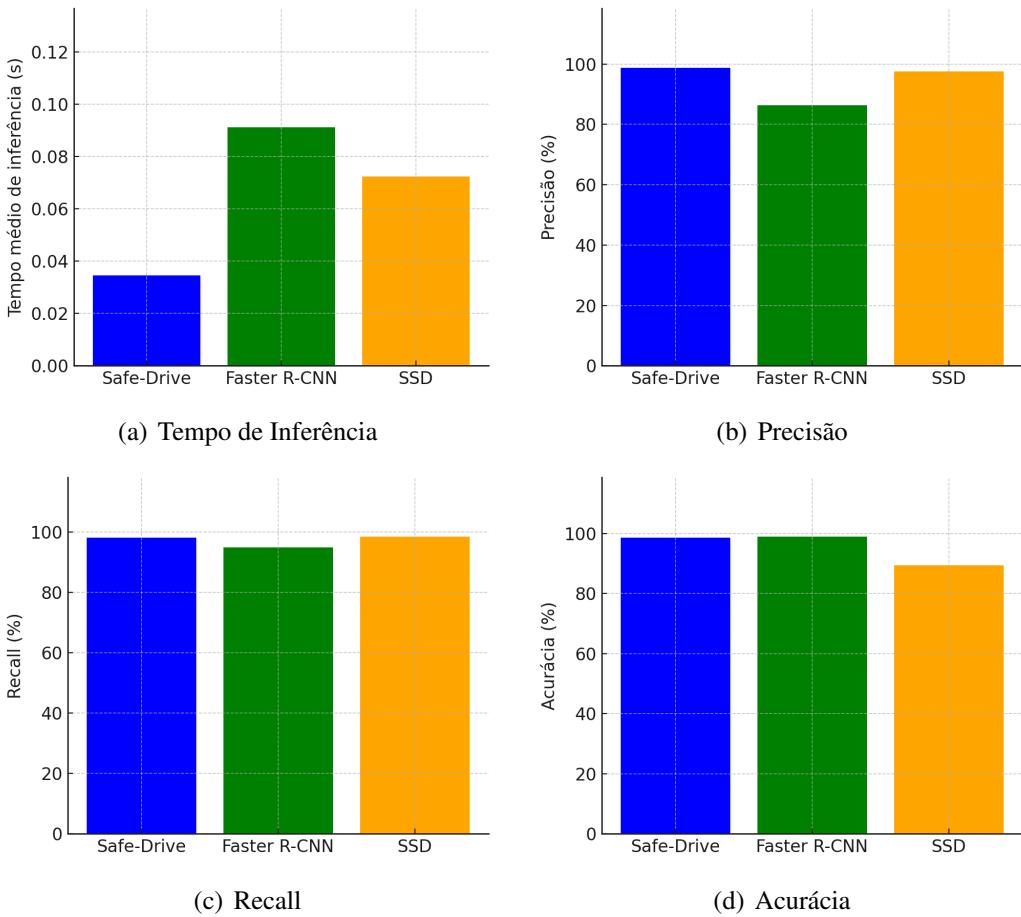


Figura 3. Comparando o Safe-Drive com o Faster R-CNN e SSD

5. Conclusão e Trabalhos Futuros

Este artigo apresentou o Safe-Drive, uma solução híbrida baseada em visão computacional para detecção de comportamentos inseguros e distraídos ao volante. O Safe-Drive integra duas arquiteturas convolucionais. O YOLO-Detect voltado à detecção de comportamentos visuais e o YOLO-Seg voltado à segmentação semântica para identificação do uso do cinto de segurança. Essa abordagem combinada permite uma análise mais abrangente do estado do condutor, indo além da simples classificação de ações e incorporando aspectos essenciais de segurança que são frequentemente negligenciados, como o uso do cinto.

Os resultados obtidos demonstram que o Safe-Drive alcançou desempenho superior quando comparado com outros algoritmos da literatura, como Faster R-CNN e SSD. O Safe-Drive obteve uma precisão média de 98,7%, recall de 98,2% e acurácia de 98,67%, com tempo médio de inferência de apenas 0,0346 segundos por imagem. Esse tempo representa uma redução de aproximadamente 62% em relação ao Faster R-CNN e 52% em relação ao SSD, evidenciando sua superioridade em cenários que exigem respostas em tempo real. Com base nesses resultados, o Safe-Drive se mostra uma solução promissora para identificar distrações de motoristas em tempo real, oferecendo precisão, velocidade e escalabilidade.

Como trabalhos futuros, pretende-se ampliar o conjunto de dados utilizado na segmentação do cinto de segurança, especialmente para cenários com iluminação desfavorável, diferentes ângulos de câmera e oclusões parciais. Planeja-se adaptar o Safe-Drive para realizar a detecção simultânea de múltiplos ocupantes no veículo. Isso inclui a identificação de comportamentos de risco tanto do condutor quanto de passageiros. Além disso, a integração de dados oriundos de sensores embarcados, tais como velocímetro, acelerômetro ou informações do painel do veículo, poderá possibilitar uma fusão multimodal que enriqueça a análise contextual e reduza falsos positivos.

Referências

- Abbass, M. A. B. and Ban, Y. (2024). Mobilenet-based architecture for distracted human driver detection of autonomous cars. *Electronics*, 13(2):365.
- Abouelnaga, Y., Eraqi, H. M., and Moustafa, M. N. (2017). Real-time distracted driver posture classification. *arXiv preprint arXiv:1706.09498*.
- Al-Mahbashi, M., Li, G., Peng, Y., Al-Soswa, M., and Debsi, A. (2025). Real-time distracted driving detection based on gm-yolov8 on embedded systems. *Journal of Transportation Engineering, Part A: Systems*, 151(3):04024126.
- Bouhsissin, S., Sael, N., and Benabbou, F. (2023). Driver behavior classification: a systematic literature review. *IEEE Access*, 11:14128–14153.
- Debsi, A., Ling, G., Al-Mahbashi, M., Al-Soswa, M., and Abdullah, A. (2024). Driver distraction and fatigue detection in images using me-yolov8 algorithm. *IET Intelligent Transport Systems*, 18(10):1910–1930.
- Du, Y., Liu, X., Yi, Y., and Wei, K. (2023). Optimizing road safety: advancements in lightweight yolov8 models and ghostc2f design for real-time distracted driving detection. *Sensors*, 23(21):8844.
- Elshamy, M. R., Emara, H. M., Shoaib, M. R., and Badawy, A.-H. A. (2024). P-yolov8: Efficient and accurate real-time detection of distracted driving. *arXiv preprint arXiv:2410.15602*.
- Ferrante, G. S., Vasconcelos Nakamura, L. H., Sampaio, S., Filho, G. P. R., and Meneguette, R. I. (2024). Evaluating yolo architectures for detecting road killed endangered brazilian animals. *Scientific reports*, 14(1):1353.
- Fresta, M., Bellotti, F., Bochenko, I., Lazzaroni, L., Merlhiot, G., Tango, F., and Berta, R. (2025). Deep learning-based real-time driver cognitive distraction detection. *IEEE Access*.
- Kshatri, S. S. and Rathore, Y. K. (2025). Real-time driver distraction detection using fast r-cnn algorithm.
- Kumar, K. S. and Raja, S. P. (2025). A comprehensive review of computer vision technology in driver monitoring system. *Data Science & Exploration in Artificial Intelligence*, pages 519–525.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I* 14, pages 21–37. Springer.

- Qu, F., Dang, N., Furht, B., and Nojoumian, M. (2024). Comprehensive study of driver behavior monitoring systems using computer vision and machine learning techniques. *Journal of Big Data*, 11(1):32.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.
- Ribeiro Jr, A., da Costa, J. B., Rocha Filho, G. P., Villas, L. A., Guidoni, D. L., Sampaio, S., and Meneguette, R. I. (2023). Harmonic: Shapley values in market games for resource allocation in vehicular clouds. *Ad Hoc Networks*, 149:103224.
- Rocha Filho, G. P., Meneguette, R. I., Neto, J. R. T., Valejo, A., Weigang, L., Ueyama, J., Pessin, G., and Villas, L. A. (2020). Enhancing intelligence in traffic management systems to aid in vehicle traffic congestion problems in smart cities. *Ad Hoc Networks*, 107:102265.
- Sajid, F., Javed, A. R., Basharat, A., Kryvinska, N., Afzal, A., and Rizwan, M. (2021). An efficient deep learning framework for distracted driver detection. *IEEE Access*, 9:169270–169280.
- Shen, Q., Zhang, L., Zhang, Y., Li, Y., Liu, S., and Xu, Y. (2024). Distracted driving behavior detection algorithm based on lightweight stardl-yolo. *Electronics*, 13(16):3216.
- Shirole, V., Shahade, A. K., and Deshmukh, P. V. (2025). A comprehensive review on data-driven driver behaviour scoring in vehicles: technologies, challenges and future directions. *Discover Artificial Intelligence*, 5(1):26.
- StateFarm (2016). State farm distracted driver detection. <https://www.kaggle.com/c/state-farm-distracted-driver-detection>. Acessado em 28 de outubro de 2024.
- Wang, R., Wang, W., Gao, J., Lin, D., Yap, K.-H., and Li, B. (2024). Multifuser: Multimodal fusion transformer for enhanced driver action recognition. *arXiv preprint arXiv:2408.01766*.
- World Health Organization (2022). Distracted driving. <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>. Acesso em: 22 mar. 2025.
- Zhao, C., Zhang, B., Lian, J., He, J., Lin, T., and Zhang, X. (2011). Classification of driving postures by support vector machines. In *2011 sixth international conference on image and graphics*, pages 926–930. IEEE.