

DESENVOLVIMENTO DE PERCEPÇÃO PARA VEÍCULOS AUTÔNOMOS

José Gabriel de Jesus Flores¹ (IC), Giovani Bernardes Vitor (PQ)¹¹Universidade Federal de Itajubá – Campus Itabira**Palavras-chave:** Detecção de pessoas, Integração de sistemas, Robótica aplicada, Visão computacional.**Introdução**

A tecnologia de veículos autônomos tem avançado significativamente nos últimos anos, com aplicações que vão do transporte individual a soluções logísticas em larga escala. Esses veículos se deslocam sem a intervenção de um condutor, utilizando sensores, câmeras e algoritmos de inteligência artificial para perceber o ambiente e tomar decisões em tempo real. O objetivo dessa tecnologia é aumentar o conforto e a segurança no trânsito, reduzindo acidentes causados por falha humana.

Uma das áreas de pesquisa mais importantes no desenvolvimento de veículos autônomos é a percepção do ambiente, que permite ao veículo detectar e reagir a elementos como outros veículos, ciclistas, sinais de trânsito e pedestres. A detecção de pessoas em tempo real é crucial para evitar colisões e garantir uma navegação eficiente em ambientes urbanos, onde há uma alta presença de pedestres e obstáculos inesperados.

Neste contexto, o projeto de pesquisa intitulado "Desenvolvimento de Percepção para Veículos Autônomos" tem como principal objetivo a criação de um sistema de detecção de pedestres. A proposta deste trabalho é integrar o modelo de detecção YOLOv10 (You Only Look Once), uma rede neural utilizada para tarefas de detecção em tempo real, com o framework ROS2 (Robot Operating System), que é um middleware utilizado para comunicação entre os diversos componentes do veículo.

A justificativa deste projeto reside na necessidade de aprimorar a detecção de pedestres em veículos autônomos, especialmente em ambientes urbanos, onde a interação com pessoas é constante e imprevisível. A abordagem busca analisar os resultados da detecção e avaliar sua viabilidade para futuros testes em veículos autônomos.

O método adotado para alcançar os objetivos do projeto foi estruturado em quatro etapas principais. A primeira etapa envolveu a aquisição de dados, com a construção de um dataset de imagens de pessoas, que foi rotulado por meio da plataforma Roboflow. Em seguida, a segunda etapa focou no treinamento do modelo YOLOv10, utilizando técnicas de aprendizado profundo para ajustar os parâmetros da rede neural, visando

aumentar a precisão da detecção. A terceira etapa consistiu na integração do modelo de detecção com o framework ROS2, permitindo a comunicação entre os componentes do sistema autônomo. Por fim, a quarta etapa abrangeu a validação do sistema, com testes realizados em diferentes cenários utilizando gravações em vídeo, permitindo avaliar a eficácia da solução em variados contextos de detecção de pessoas.

Metodologia**Desenvolvimento do Modelo de Detecção de Pessoas com YOLOv10**

O processo de desenvolvimento do modelo de detecção de pessoas utilizando o YOLOv10 foi dividido em quatro etapas principais, conforme ilustrado na Figura 1. As etapas são: aquisição de imagens, construção do dataset, treinamento do modelo e análise.

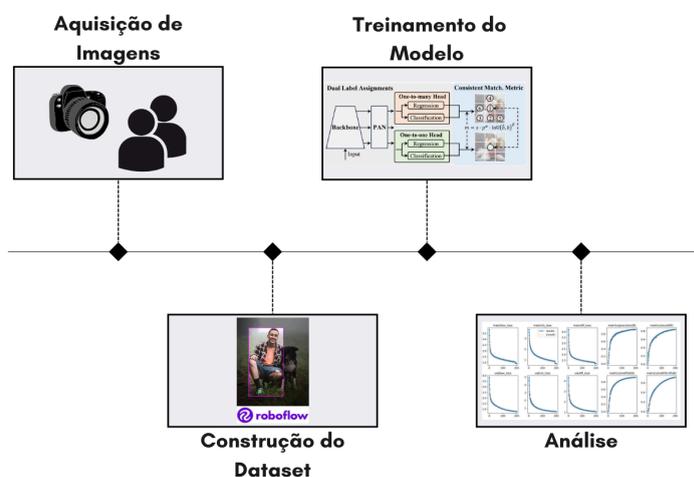


Figura 1 – Fluxograma do desenvolvimento do modelo de detecção de pessoas. Autor (2024).

As imagens utilizadas no treinamento foram extraídas do Person Detection Dataset, criado por Titulacin (2023) e disponibilizado na plataforma Roboflow. Esse conjunto de dados contém uma variedade de imagens de indivíduos em diferentes cenários, possibilitando ao modelo aprender a identificar pessoas em múltiplos ambientes. A maioria das imagens foi obtida de fontes públicas, como Unsplash e Google, e previamente

rotuladas com caixas delimitadoras.

A construção do dataset foi dividido em três subconjuntos: 4.407 imagens para treinamento, 1.071 para validação e 5 imagens para teste. Essa divisão é para uma avaliação do desempenho do modelo em dados inéditos, prevenindo problemas como o overfitting. Utilizou-se a plataforma Roboflow para realizar o pré-processamento das imagens, aplicando técnicas como rotação, filtros gaussianos e mosaico, o que aumentou a diversidade do conjunto de dados.

O modelo foi treinado com a arquitetura YOLOv10s, otimizada para dispositivos com recursos limitados. As configurações de treinamento estão na Tabela 1.

Hyperparameters	Values
Number of Epochs	200
Batch Size	16
Image Size	640
Initial Learning Rate	0.01
Data Augmentation	Random
Deep Learning Framework	PyTorch

Tabela 1 – Configurações de treinamento do modelo YOLOv10s. Autor(2024)

Para acompanhar o desempenho do modelo, foram utilizados métodos como a matriz de confusão e as curvas de perda. As principais métricas avaliadas foram a precisão e o recall, com o objetivo de garantir a detecção correta de pedestres e minimizar erros de classificação e localização. Essas análises foram essenciais para verificar a capacidade do modelo de generalizar em novos dados, assegurando sua eficácia em diferentes cenários.

Integração do Modelo YOLOv10 com o ROS2

A integração do modelo de detecção de pessoas YOLOv10 com o ROS2 seguiu uma série de etapas, como demonstrado na Figura 2. O fluxograma descreve o processo completo, desde a captura dos dados de vídeo até a exibição e armazenamento dos resultados, permitindo que o sistema realize a detecção de pessoas em tempo real e exiba os resultados.

O processo de captura de vídeo foi realizado utilizando uma câmera conectada ao sistema ou a partir de vídeos previamente gravados. Os frames de vídeo foram capturados por um nó Publisher, cuja função é publicar os frames em um tópico específico do ROS2. Esse nó estabelece a comunicação com os outros componentes

do sistema, garantindo que os dados sejam enviados em tempo real para o tópico configurado.

Após a publicação dos frames no tópico, eles são processados por um nó Subscriber, que utiliza o modelo YOLOv10 para realizar a inferência. Durante esse processo, o sistema identifica as pessoas presentes nas imagens e os marca com caixas delimitadoras diretamente sobre os frames processados. Após a detecção, os resultados são exibidos em tempo real em uma interface gráfica e simultaneamente gravados em arquivos de vídeo, possibilitando uma análise posterior.

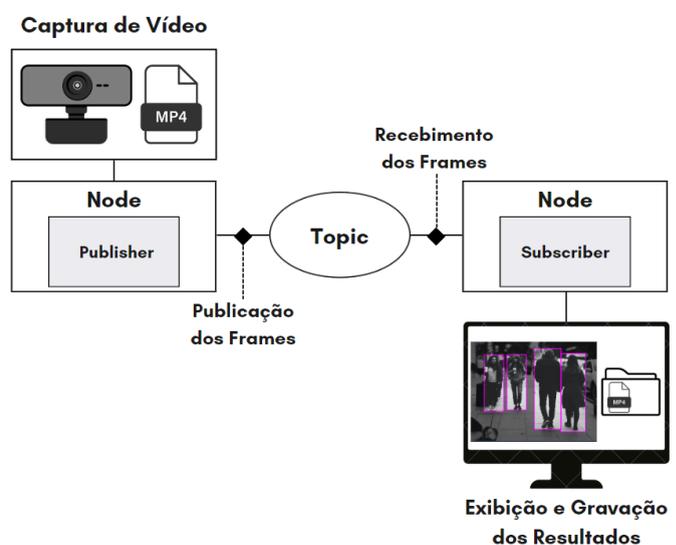


Figura 2 – Fluxograma da integração do YOLOv10 com o ROS2. Autor (2024).

Resultados e discussão

Nesta seção, são apresentados os resultados obtidos durante o desenvolvimento do projeto, com foco nas fases principais: treinamento do modelo de detecção de pessoas, a integração desse modelo com o ROS2, e a análise dos resultados inesperados. O projeto foi desenvolvido utilizando a seguinte configuração de hardware e software mostrado na Tabela 2.

Componente Avaliado	Especificação Utilizada
Sistema operacional	Ubuntu 22.04 LTS
GPU	RTX 3070 Ti Laptop
CPU	AMD Ryzen 6800H
Versão YOLO	Yolov10s
Framework ROS2	Humble

Tabela 2 – Configurações de hardware e software utilizadas no projeto. Autor(2024)

Resultados do Treinamento do Modelo de Detecção de Pessoas com YOLOv10

Durante o treinamento do modelo YOLOv10, foram obtidos resultados que mostram seu desempenho em diferentes métricas. A Figura 3 apresenta a matriz de confusão normalizada, que avalia a capacidade do modelo em classificar corretamente as pessoas em comparação com o fundo. A matriz demonstra uma precisão de 74% para a classe "Pessoa", indicando que, na maioria dos casos, as detecções foram corretas. No entanto, o modelo apresentou uma taxa de 26% de falsos negativos, falhando em identificar algumas pessoas. Embora o desempenho geral tenha sido satisfatório, ainda há margem para melhorias.

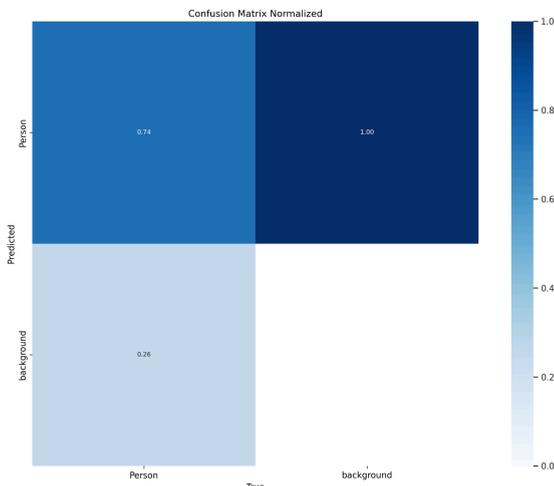


Figura 3 – Matriz de confusão normalizada para a detecção de pessoas pelo YOLOv10. Autor (2024).

A Figura 4 mostra a evolução das curvas de perda durante o treinamento e validação. As curvas de "Bounding Box Loss" e "Classification Loss" indicam uma redução constante dos erros ao longo das 200 épocas, demonstrando um bom ajuste do modelo aos dados. A precisão e o recall também melhoraram, atingindo 74% e 70%, respectivamente. Isso indica que o YOLOv10 atingiu um bom equilíbrio entre a minimização de falsos positivos e a maximização da detecção de pedestres. As métricas mAP50 e mAP50-95, com valores de 0.75 e 0.50, mostram que o modelo tem uma capacidade razoável de detecção, embora ainda haja margem para otimizações, especialmente em cenários mais desafiadores.

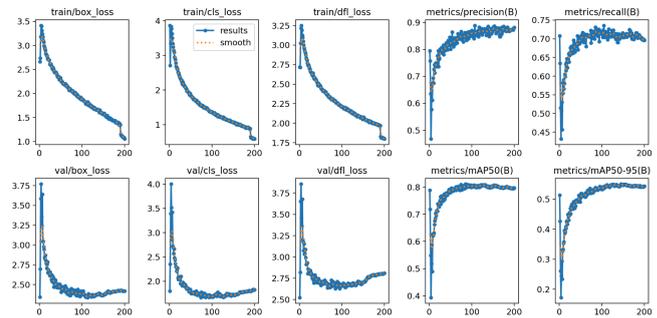


Figura 4 – Curvas de perda e métricas de avaliação durante o treinamento e validação. Autor (2024).

Em relação aos resultados visuais, a Figura 5 apresenta as detecções realizadas pelo modelo, onde as pessoas são destacadas por caixas delimitadoras com valores de confiança atribuídos a cada detecção. Esses resultados mostram que o modelo é capaz de identificar pedestres com precisão em diferentes cenários e condições

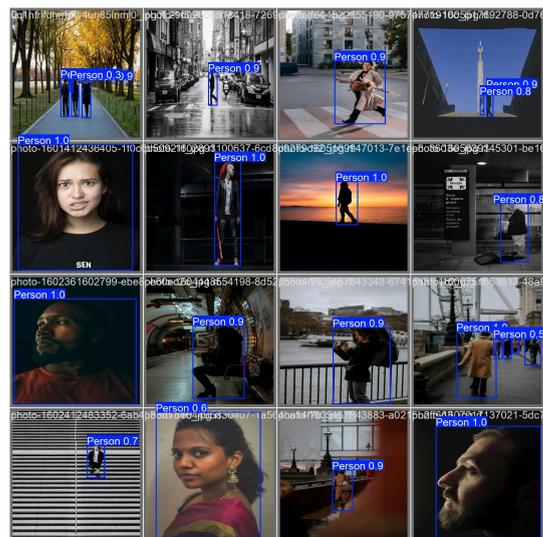


Figura 5 – Detecções de pessoas pelo YOLOv10 com caixas e valores de confiança. Autor (2024).

Resultados da Integração do Modelo YOLOv10 com o ROS2

A integração do modelo YOLOv10 com o ROS2 permitiu a detecção em tempo real de pessoas, tanto em vídeos gravados quanto em transmissões ao vivo via webcam. A Figura 6 ilustra a detecção bem-sucedida de pessoas em um vídeo gravado, com valores de confiança variando entre 0.89 e 0.92, confirmando a precisão do modelo em identificar múltiplos indivíduos na cena.

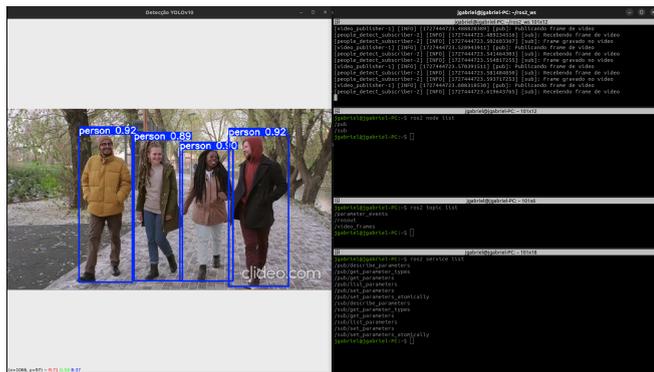


Figura 6 – Detecção de pessoas com YOLOv10 integrado ao ROS2. Autor (2024).

Os frames processados foram gravados para posterior análise, possibilitando a identificação de possíveis melhorias, como ajustes finos nos valores de confiança e no pré-processamento dos dados. Isso facilita uma análise mais detalhada do desempenho em diferentes cenários e proporciona a base para futuras otimizações do sistema, especialmente em ambientes dinâmicos.

Embora os testes tenham sido realizados em um ambiente controlado, o sistema foi submetido a diferentes condições, como variações de iluminação e cenários com múltiplos pedestres. Em condições de baixa iluminação, por exemplo, a confiança nas detecções variou ligeiramente, mas o modelo continuou a demonstrar precisão aceitável, reforçando sua aplicabilidade em cenários urbanos. Esses resultados são promissores para futuros testes em veículos autônomos, onde a detecção em tempo real é essencial para garantir a navegação segura.

Conclusões

Este trabalho teve como principal objetivo o desenvolvimento e a integração de um sistema de detecção de pessoas utilizando o modelo YOLOv10 e o framework ROS2, com vistas à futura implementação em veículos autônomos. Durante o desenvolvimento, foram realizadas várias etapas, desde o treinamento do modelo até a sua integração com o ROS2, permitindo a detecção de pessoas em tempo real a partir de fluxos de vídeo gravados e capturados ao vivo.

Os resultados indicaram que o modelo treinado no YOLOv10 obteve uma precisão de 74% e um recall de 70%, valores que, embora satisfatórios, não são suficientes para aplicações em ambientes urbanos, onde a precisão precisa ser ainda maior para garantir a segurança dos pedestres e a confiabilidade do sistema. Um dos principais fatores que contribuíram para esses números foi a limitação de hardware utilizada durante o treinamento, que restringiu a capacidade de realizar

ajustes mais refinados no modelo.

Para trabalhos futuros, as melhorias devem se concentrar no aumento da robustez do modelo em diferentes condições ambientais, como a expansão do dataset de treinamento para abranger uma maior diversidade de cenários e condições de iluminação. Ajustes refinados nos hiperparâmetros e a implementação de técnicas avançadas de pré-processamento de imagens, como a normalização de iluminação, poderão otimizar o desempenho em cenários dinâmicos e reduzir os falsos negativos. Outro ponto importante é a necessidade de otimizar o modelo para rodar em hardware de menor capacidade, sem comprometer a precisão ou o tempo de resposta, algo crucial para a implementação em veículos reais. Além disso, o uso de hardware mais robusto, como GPUs de maior capacidade ou TPUs, permitirá que o modelo seja treinado e executado com maior eficiência e precisão, garantindo um melhor desempenho em condições mais desafiadoras.

Somente após melhorias, o sistema estará apto para testes em veículos autônomos em ambientes reais, validando sua aplicação segura e eficiente na navegação. O trabalho fornece uma base sólida para sistemas de percepção, mas reforça a necessidade de melhorias contínuas para assegurar uma detecção de pessoas precisa e confiável, um fator crítico para a segurança em operações futuras.

Agradecimentos

Gostaria de expressar minha sincera gratidão a todas as pessoas e entidades que contribuíram de forma significativa para o desenvolvimento desta pesquisa. Em especial, agradeço ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pela bolsa de estudo que possibilitou a realização deste trabalho. Agradeço também a meus orientadores e colegas, cuja orientação e apoio foram fundamentais para o sucesso desta pesquisa.

Referências

TITULACIN. **Person detection dataset**. Roboflow Universe, Roboflow, maio 2023. Disponível em: <https://universe.roboflow.com/titulacin/person-detection-9a6mk/dataset/16>. Acesso em: 02 out. 2024.

WANG, Ao; CHEN, Hui; LIU, Lihao et al. **YOLOv10: Real-Time End-to-End Object Detection**. arXiv preprint arXiv:2405.14458, 2024. Disponível em: <https://docs.ultralytics.com/models/yolov10/#exporting-yolov10>. Acesso em: 02 out. 2024.