CENTRO UNIVERSITÁRIO FEI

GABRIEL GONÇALVES PASQUARELLI

GABRIEL VIEIRA LIMA

KAREN NATALLY MORAES

VITOR AUGUSTO LEMES PINHEIRO DOS SANTOS

CRIAÇÃO E A AVALIAÇÃO DE UM SISTEMA DE PERCEPÇÃO MULTIMODAL PARA VEÍCULOS AUTÔNOMOS

São Bernardo do Campo

2023

RESUMO

O presente Trabalho de Conclusão de Curso aborda o desenvolvimento e avaliação de um sistema de percepção multimodal para veículos autônomos. A introdução destaca a importância crescente dos veículos autônomos e a relevância de sistemas de percepção avançados, enfrentando desafios como variações climáticas e condições de tráfego complexas. O estudo se concentra em aprimorar a precisão e robustez desses sistemas para garantir segurança e eficiência.

Na revisão da literatura, são explorados avanços recentes na área, incluindo o uso de Redes Neurais Convolucionais (CNNs) e técnicas de *deep learning*, essenciais para o reconhecimento de padrões em imagens e percepção ambiental. Estudos destacados, como os de YOLO e Viola-Jones, demonstram a eficácia de modelos baseados em aprendizado profundo em comparação com métodos tradicionais, fornecendo *insights* valiosos para o desenvolvimento do sistema proposto.

A metodologia empregada foca na segmentação semântica fracamente supervisionada, utilizando dados multimodais provenientes de sensores como câmeras e LiDAR. O estudo analisa um conjunto de dados fornecido pela Audi, contendo informações detalhadas de sensores e imagens, para treinar e avaliar o sistema de percepção.

A proposta experimental se baseia no uso da linguagem Python 3 e de bibliotecas como NumPy e Pandas para processamento e análise de dados. A abordagem inclui mapeamento e análise exploratória do conjunto de dados, visando identificar padrões e otimizar a percepção dos veículos autônomos em diferentes condições e ambientes.

Este trabalho, portanto, oferece uma contribuição significativa para o campo dos veículos autônomos, apresentando soluções inovadoras para desafios de percepção ambiental e destacando a importância de sistemas de percepção multimodal avançados para a segurança e eficiência dos veículos autônomos.

Palavras-chave: Veículos Autônomos; Sistema de Percepção Multimodal; *Deep Learning*; Segmentação Semântica Fracamente Supervisionada; Dados Multimodais; Python 3;

ABSTRACT

The present Final Paper focuses on the development and evaluation of a multimodal perception system for autonomous vehicles. The introduction emphasizes the growing importance of autonomous vehicles and the relevance of advanced perception systems, facing challenges such as climatic variations and complex traffic conditions. The study concentrates on enhancing the accuracy and robustness of these systems to ensure safety and efficiency.

In the literature review, recent advances in the field are explored, including the use of Convolutional Neural Networks (CNNs) and deep learning techniques, essential for pattern recognition in images and environmental perception. Highlighted studies, such as those of YOLO and Viola-Jones, demonstrate the effectiveness of deep learning-based models compared to traditional methods, providing valuable insights for the development of the proposed system.

The employed methodology focuses on weakly supervised semantic segmentation, using multimodal data from sensors such as cameras and LiDAR. The study analyzes a dataset provided by Audi, containing detailed sensor and image information, to train and evaluate the perception system.

The experimental proposal is based on the use of the Python 3 language and libraries like NumPy and Pandas for data processing and analysis. The approach includes mapping and exploratory analysis of the dataset, aiming to identify patterns and optimize the perception of autonomous vehicles in different conditions and environments.

Thus, this work offers a significant contribution to the field of autonomous vehicles, presenting innovative solutions to environmental perception challenges and highlighting the importance of advanced multimodal perception systems for the safety and efficiency of autonomous vehicles.

Keywords: Autonomous Vehicles; Multimodal Perception System; Deep Learning; Weakly Supervised Semantic Segmentation; Multimodal Data; Python 3;

LISTA DE ILUSTRAÇÕES

Figura 1 – Níveis de Automação Veicular 9
Figura 2 – Mensurando distância usando sensores LiDAR 11
Figura 3 – Segmentação semântica fracamente supervisionada 13
Figura 4 – Rede Neural Convolucional 14

Figura 5 – Sensores e câmeras do conjunto de dados 21
Figura 6 – Etapas empregadas para execução do trabalho proposto 22

LISTA DE TABELAS

Tabela 1 – Níveis de Automação Veicular 10

Tabela 2 – Arquivo de configuração: Vehicle Configuration 24
Tabela 3 – Arquivo de configuração: Lidar Configuration 24
Tabela 4 – Arquivo de configuração: Camera Configuration 25
Tabela 5 – Arquivo de configuração: Image Frame Information 25
Tabela 6 – Arquivo de configuração: Lidar Frame Information 26

SUMÁRIO

1	**INTRODUÇÃO**	7
1.1	OBJETIVO	8
2	**CONCEITOS FUNDAMENTAIS**	9
2.1	AUTONOMIA VEICULAR	9
2.2	SISTEMA DE PERCEPÇÃO MULTIMODAL PARA VEÍCULOS AUTÔNOMOS	10
2.3	SEGMENTAÇÃO SEMÂNTICA E APRENDIZADO FRACO	12
2.4	FUSÃO DE DADOS MULTIMODAIS	13
2.5	REDES NEURAIS CONVOLUCIONAIS	14
3	**TRABALHOS RELACIONADOS**	16
3.1	SEGMENTAÇÃO SEMÂNTICA COM BASE EM APRENDIZADO PROFUNDO	16
3.2	SEGMENTAÇÃO SEMÂNTICA FRACAMENTE SUPERVISIONADA	17
3.3	OUTRAS ABORDAGENS	18
3.4	CONSIDERAÇÕES FINAIS DO CAPÍTULO	19
4	**METODOLOGIA**	20
5	**PROPOSTA EXPERIMENTAL**	24
6	**CONCLUSÃO**	27
	REFERÊNCIAS	28

1 INTRODUÇÃO

A busca por soluções de mobilidade mais eficientes e seguras está impulsionando a pesquisa e o desenvolvimento dos veículos autônomos (VA), desencadeando uma revolução tecnológica com a promessa de aprimorar a mobilidade urbana, reduzir acidentes de trânsito e otimizar o uso de recursos (JEVAMIKYOUS; KASHEF, 2022). No âmbito da Ciência da Computação, a criação e avaliação de sistemas de percepção multimodal para VA emergem como um campo de estudo de extrema relevância e complexidade.

A escolha do tema para esse Trabalho de Conclusão de Curso (TCC) está fundamentada na compreensão das limitações presentes nos sistemas de VA em operação na atualidade. A percepção do ambiente continua a representar um desafio significativo para a ampla adoção desses veículos. O sistema de percepção, como componente central da autonomia veicular, desempenha papel essencial na interpretação e resposta a informações sensoriais provenientes de diversas fontes, tais como câmeras, sensores LIDAR e RADAR (VARGAS et al., 2021). Dessa forma, a busca por soluções que aprimorem a precisão, a robustez e a capacidade de percepção desses sistemas torna-se crucial para garantir a segurança e o êxito dos veículos autônomos.

No cenário atual, os VA enfrentam desafios significativos relacionados à percepção do ambiente. A complexidade das variações climáticas, condições de tráfego, presença de objetos imprevisíveis e necessidade de tomar decisões em tempo real representam obstáculos substanciais. Entretanto, é importante destacar que já foram alcançados avanços notáveis na área de percepção multimodal (JEVAMIKYOUS; KASHEF, 2022). O desenvolvimento de Redes Neurais Convolucionais (CNNs) de alto desempenho, por exemplo, se configura como uma das principais prioridades, viabilizando a construção de sistemas de percepção mais confiáveis para veículos autônomos. Além disso, melhorias em sensores de alta resolução e o uso de técnicas de fusão de dados complementam o progresso significativo nesse campo.

Portanto, o propósito deste trabalho é desenvolver e avaliar um sistema de percepção multimodal. Essa abordagem é fundamental para aprimorar a segurança, eficiência e capacidade operacional de VA em diferentes condições e ambientes, implementaremos a técnica de segmentação semântica fracamente supervisionada e integrando dados provenientes dos sensores previamente mencionados.

1.1 OBJETIVO

O objetivo desse trabalho é analisar o conjunto de dados do Audi A2D2 (GEYER et al., 2020) e desenvolver um sistema de percepção multimodal para VA, aplicando a técnica de *deep learning*. Esse é um subcampo da aprendizagem de máquina que se concentra no treinamento de redes neurais artificiais com múltiplas camadas para aprender e extrair representações hierárquicas de dados, com foco específico em segmentação semântica fracamente supervisionada (MO et al., 2022).

O objetivo principal é abordar desafios específicos relacionados à percepção ambiental, interpretação e resposta, visando aprimorar a robustez. Nesse sentido, integrando soluções existentes, como segmentação semântica com supervisão fraca, adaptação de domínio, redes neurais convolucionais (CNN), fusão de dados, processamento em tempo real e consideração do contexto (MO et al., 2022). Essas soluções visam enfrentar diversas situações, como condições meteorológicas adversas, cenários urbanos complexos e interações com outros veículos, proporcionando uma percepção eficaz em ambientes desafiadores.

Para alcançar esses resultados, serão realizadas as seguintes etapas:

a) Analisar e compreender as limitações dos sistemas de veículos autônomos, especialmente em relação à percepção do ambiente.
b) Desenvolver algoritmos com segmentação semântica fracamente supervisionada através do deep learning.
c) Executar testes para avaliar a eficácia do sistema em diferentes cenários e condições.
d) Realizar uma análise detalhada dos resultados obtidos durante os testes.
e) Comparar os resultados com padrões de desempenho estabelecidos.
f) Fazer ajustes no processo de desenvolvimento conforme necessário para melhorar continuamente a qualidade e eficácia do sistema de percepção multimodal.

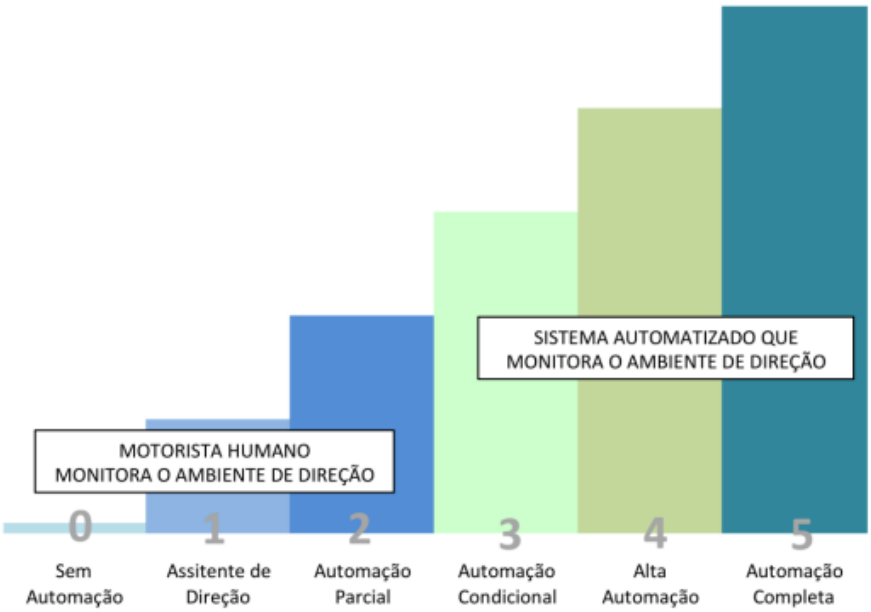
2 CONCEITOS FUNDAMENTAIS

Esta seção explora os componentes essenciais relacionados à autonomia veicular, juntamente com os conceitos associados à percepção multimodal desta tecnologia e ao processamento de máquina para alcançar tal finalidade.

2.1 AUTONOMIA VEICULAR

A tecnologia que a indústria automobilística mais discute atualmente é a relacionada aos Veículos Autônomos (HÖRL et al., 2016). Com importante potencial para mudar drasticamente não só o ambiente de trânsito, mas também a forma como vemos a mobilidade, projetamos cidades, comportamento humano e, consequentemente, trabalhamos e vivemos em uma sociedade cada vez mais conectada com enormes implicações econômicas. A fim de padronizar as discussões sobre a tecnologia, a SAE (Society of Automotive Engenineers – Sociedade de Engenheiros Automobilísticos) definiu 6 níveis (de 0 a 5) de automação em VAs, conforme mostrado na Figura 1.

Figura 1 – Níveis de Automação Veicular



Fonte: NHTSA 2018

A Tabela 1 oferece uma visão detalhada da automação veicular em cada nível, conforme definido pela Arquitetura de Sistema Eletrônico (SEA).

Tabela 1 – Níveis de Automação Veicular

Nível	Descrição
0	A maioria dos veículos hoje nas ruas são controlados manualmente. Os motoristas possuem total controle.
1	Esse é o nível mais baixo de automação. Alguns controles individuais são automatizados, tais como estabilidade eletrônica e "Cruise Control"adaptativo, no qual o veículo mantém distância segura do veículo da frente.
2	Pelo menos dois controles podem ser automatizados em conjunto, aceleração e direção. Nesse nível, a velocidade é controlada com sensores que identificam o veículo da frente e as faixas das pistas são monitoradas para manter o veículo na pista, liberando o motorista de colocar a mão na direção.
3	O motorista concede controle completo do carro em certas circunstâncias. Mas, o computador avisa quando o motorista precisa voltar ao comando em tempo hábil.
4	O veículo faz as funções críticas de segurança durante toda a viagem, sem a expectativa de que o motorista precise controlar o veículo em algum momento. Como o veículo nessa fase dirige o automóvel em todos os momentos, até o estacionamento, este poderá rodar sem pessoas.
5	Veículo totalmente autônomo sem a necessidade do pedal de aceleração e volante. Possuem capacidade de ir a qualquer lugar e fazer qualquer movimento que um motorista experiente faz.

Fonte: NHTSA 2018

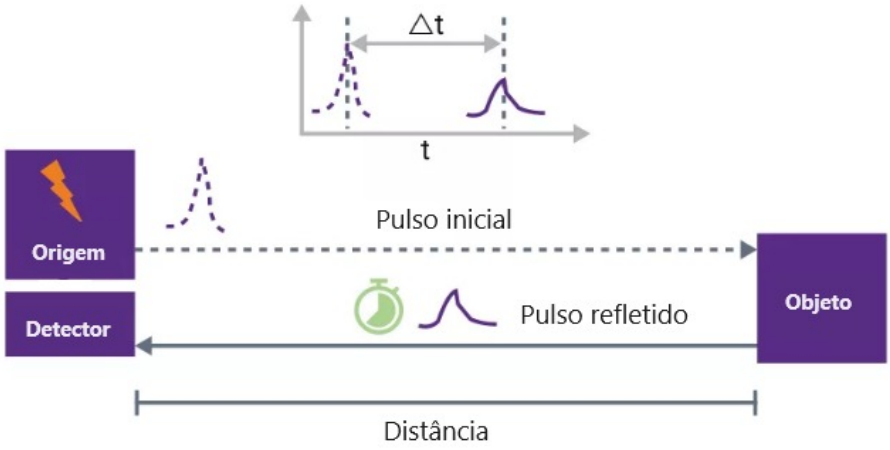
2.2 SISTEMA DE PERCEPÇÃO MULTIMODAL PARA VEÍCULOS AUTÔNOMOS

Um sistema de percepção destinado a veículos autônomos constitui um conjunto de tecnologias e sensores que desempenham um papel crucial no funcionamento seguro e eficaz desses veículos, dispensando a necessidade de intervenção humana. Dentro da composição primordial desse sistema, destacam-se as câmeras, encarregadas de capturar imagens do entorno. Essas imagens, por sua vez, são processadas por meio de técnicas de visão computacional, permitindo a identificação de objetos, sinais de trânsito, faixas de rodagem, pedestres e outros veículos (MO et al., 2022).

Para além das câmeras, o LiDAR (Light Detection and Ranging) assume um papel central. Por meio da utilização de feixes de laser, o LiDAR mensura a distância entre o veículo e os elementos circundantes, gerando um mapa tridimensional minuciosamente detalhado do

ambiente, como mostrado na Figura 2. Esta tecnologia revela-se especialmente eficaz em condições de iluminação reduzida ou em situações adversas, como chuva, neve ou neblina (MO et al., 2022).

Figura 2 – Mensurando distância usando sensores LiDAR



Fonte: NHTSA 2018

Os sensores de radar e ultrassônicos também compõem uma parte crucial do sistema de percepção. Estes emitem ondas eletromagnéticas e sonoras, respectivamente, para detectar a presença de objetos e determinar a velocidade e direção desses elementos em movimento. As informações obtidas são então utilizadas para calcular a proximidade dos objetos em relação ao veículo, sendo particularmente úteis em condições climáticas desafiadoras e a baixas velocidades (CUI; GE, 2003).

A incorporação de um sistema de GPS (Global Positioning System) desempenha um papel fundamental ao fornecer informações precisas sobre a localização do veículo em relação às coordenadas geográficas (CUI; GE, 2003). Isso se revela essencial para a navegação autônoma e para a determinação da rota mais eficiente.

Todas essas informações são processadas por uma unidade de controle central, que integra os dados provenientes dos diferentes sensores em um modelo coerente do ambiente ao redor do veículo. Esta representação virtual é vital para a tomada de decisões em tempo real. Além disso, algoritmos de fusão de dados são empregados para combinar e integrar as informações provenientes dos diversos sensores, criando uma visão unificada e abrangente do ambiente (MO et al., 2022).

2.3 SEGMENTAÇÃO SEMÂNTICA E APRENDIZADO FRACO

A segmentação semântica, também conhecida como segmentação de instâncias ou rotulagem a nível de pixel, constitui uma subárea essencial do aprendizado profundo (deeplearning). Seu propósito fundamental é atribuir um rótulo a cada pixel de uma imagem, visando identificar distintos objetos na cena e delinear suas fronteiras. Esta tarefa de visão computacional desempenha um papel crucial na interpretação de informações sensoriais provenientes de diversas fontes, como imagens, dados do sensor LiDAR e vídeos. O resultado dessa análise proporciona uma identificação precisa de objetos e sua localização (MO et al., 2022).

No capítulo 3 de (MO et al., 2022), discute-se o paradigma de aprendizado de máquina. Nesse paradigma, modelos treinados com supervisão limitada ou rótulos de qualidade inferior possibilitam uma abordagem de aprendizado mais flexível, com rótulos atribuídos a nível de imagem e região. Essa abordagem é particularmente útil em situações em que obter rótulos detalhados para cada pixel de uma imagem é desafiador, o que é comum em conjuntos de dados complexos.

A aplicação do aprendizado fraco na segmentação semântica auxilia na mitigação da complexidade e dos custos associados à obtenção de rótulos precisos em conjuntos de dados extensos. Isso permite que os modelos se adaptem de maneira mais flexível e realista a diferentes cenários. Essa abordagem é especialmente valiosa em aplicações práticas, onde, em vez de fornecer rótulos precisos para cada pixel em um conjunto de dados, é possível fornecer rótulos aproximados ou menos detalhados. Por exemplo, ao invés de rotular cada pixel de um veículo, é possível rotular toda a região correspondente ao veículo como uma única classe semântica, conforme discutido no trabalho de (MO et al., 2022) no contexto do estado-da-arte em tecnologias de segmentação semântica, no Capítulo 3, '*Weakly-supervised semantic segmentation*'.

A abordagem da segmentação pode ser realizada de diversas maneiras, incluindo métodos por pontos, por rabiscos e por bounding boxes. Na segmentação por pontos, cada pixel da imagem é individualmente rotulado, proporcionando uma precisão extremamente alta na identificação de classes semânticas. Por outro lado, a segmentação por rabiscos envolve a marcação manual de áreas na imagem para indicar diferentes regiões ou objetos. Esses rabiscos servem como pistas para o algoritmo de segmentação, permitindo a identificação e rotulagem automática das áreas correspondentes. Já na segmentação por bounding boxes, retângulos delimitadores são desenhados ao redor dos objetos na imagem. Embora essa técnica seja menos precisa em termos

de contorno, é computacionalmente eficiente. Na Figura 3, pode-se observar um exemplo de como cada abordagem é aplicada em uma imagem real.

Figura 3 – Segmentação semântica fracamente supervisionada

Fonte: MO et al. 2022

2.4 FUSÃO DE DADOS MULTIMODAIS

A fusão de dados multimodais refere-se à prática de combinar informações provenientes de diferentes tipos de sensores, proporcionando uma melhoria significativa no desempenho e na robustez dos algoritmos de segmentação semântica. Em cenários onde as condições de iluminação se mostram desafiadoras, como em ambientes escuros ou adversos devido às condições climáticas, a confiança nos dados obtidos de uma única modalidade sensorial, como uma câmera, pode ser insuficiente para embasar com precisão as decisões de um VA (LAHAT; ADALI; JUTTEN, 2015).

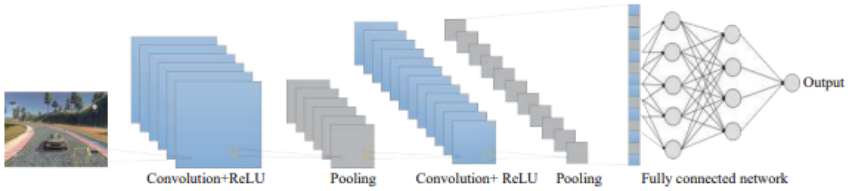
Estudos indicam que imagens capturadas por câmeras termográficas demonstram uma notável resiliência em condições climáticas adversas, como ressaltado no artigo "Review the state-of-the-art technologies of semantic segmentation based on deep learning", desenvolvido pela Tongji University em 2018. Ao integrar informações provenientes de câmeras termográficas e de imagem padrão (RGB) em uma única estrutura de segmentação, os resultados obtidos são substancialmente aprimorados. Câmeras RGB (Red-Green-Blue) têm a capacidade de fornecer informações sólidas em longas distâncias, principalmente em ambientes com alta luminosidade. Por outro lado, sensores LiDAR utilizam pulsos de luz laser para determinar a distância dos objetos, tornando-os menos suscetíveis a variações nas condições de luminosidade ou adversidades climáticas. Ao combinar as informações provenientes desses dois tipos de sensores, obtemos um conjunto de dados que oferece uma visão mais rica e precisa do ambiente.

A fusão de dados multimodais, portanto, representa uma estratégia fundamental para aprimorar a compreensão e interpretação do ambiente em aplicações como a segmentação semântica, onde a integração de diferentes fontes de informação se mostra crucial para a tomada de decisões precisas e confiáveis.

2.5 REDES NEURAIS CONVOLUCIONAIS

Redes neurais convolucionais são um tipo de arquitetura de rede neural profunda que se mostrou especialmente eficaz em tarefas de visão computacional, como reconhecimento de padrões em imagens. As CNNs (Convolutional Neural Networks), são projetadas para processar dados que têm uma estrutura de grade, como imagens. Elas consistem em camadas de convolução que aplicam filtros para detectar características locais, como bordas e texturas, em regiões específicas da entrada. (KEBRIA et al., 2020). Na Figura 4, está esquematizada uma típica rede neural convolucional.

Figura 4 – Rede Neural Convolucional



Fonte: KEBRIA et al. 2020

No estudo sobre aprendizado por imitação profunda para veículos autônomos, Kebria, Khosravi, Salaken e Nahavandi (2020) exploram as camadas de convolução em sua arquitetura de rede neural. Essas camadas, fundamentais em redes neurais convolucionais (CNNs), operam através da aplicação de filtros que visam detectar características locais, tais como bordas e texturas, em regiões específicas da entrada (KEBRIA et al., 2020). Esse processo permite que a rede identifique padrões visuais cruciais para tarefas como reconhecimento de objetos em imagens.

Em seguida, as camadas de *pooling* entram em ação. Como discutido por Kebria, essas camadas são essenciais na redução da dimensionalidade do dado processado, mantendo ao mesmo tempo as características mais significativas da imagem. Ao agrupar informações em regiões específicas e preservar os aspectos mais proeminentes, as camadas de *pooling* contribuem para a eficiência computacional da rede e evitam o excesso de processamento de detalhes menos relevantes (KEBRIA et al., 2020).

A aplicação dessas camadas em dados multimodais é particularmente intrigante, pois as CNNs podem processar distintos tipos de entrada, como imagens e dados textuais, de forma integrada (KEBRIA et al., 2020)). Em cenários nos quais as informações provêm de diversas

fontes sensoriais ou modalidades, como nos veículos autônomos, isso proporciona uma abordagem unificada para a interpretação e extração de padrões em dados heterogêneos.

3 TRABALHOS RELACIONADOS

Para avaliar o progresso na literatura em relação ao nosso objetivo definido, a pesquisa de artigos desempenha um papel central na construção da base de conhecimento e na coleta de informações relevantes. Com o intuito de garantir uma abordagem abrangente a partir de fontes confiáveis, implementamos uma estratégia de busca bem planejada. Essa estratégia englobou a utilização de palavras-chave específicas, a saber: "Autonomous vehicles", "Environmental perception", "RADAR", "LIDAR", "Camera", "Deep learning", "CNN"e "A2D2".

Os critérios de inclusão foram considerados relevantes os artigos publicados nos últimos dois anos, e requisitaram que os estivessem disponíveis em inglês ou português e relacionados às palavras-chave mencionadas. Os artigos que não satisfaziam esses critérios foram descartados.

Para determinar a importância dos artigos, aplicamos uma estratégia de leitura dinâmica que envolveu a avaliação do resumo, da conclusão, da introdução e da primeira frase de cada parágrafo, antes de procedermos à leitura integral do artigo.

Esse método de leitura dinâmica permitiu uma rápida avaliação dos aspectos essenciais do conteúdo de cada artigo, incluindo os principais resultados. Após a análise dos artigos, criamos uma tabela de sumarização de resultados utilizando o software Excel [1], que possibilitou a organização e a comparação das informações mais relevantes de cada artigo.

A seleção final dos artigos para a pesquisa foi conduzida após discussões e análises detalhadas entre os membros da equipe, garantindo a pertinência e a contribuição dos artigos escolhidos para o estudo em questão. Os trabalhos mais relevantes serão abordados detalhadamente a seguir.

3.1 SEGMENTAÇÃO SEMÂNTICA COM BASE EM APRENDIZADO PROFUNDO

No trabalho de (JEVAMIKYOUS; KASHEF, 2022), os métodos de destaque na literatura, como o YOLO e o algoritmo Viola-Jones, oferecem abordagens distintas para essa finalidade. O YOLO é um algoritmo de aprendizado profundo que divide a imagem em uma grade e prevê caixas delimitadoras e probabilidades de classe para cada célula, enquanto o algoritmo Viola-Jones utiliza recursos semelhantes ao HAAR e uma cascata de classificadores para a detecção de objetos. Os resultados demonstram que modelos baseados em aprendizado profundo, como YOLOv2 e YOLOv3, superam métodos tradicionais, como o Viola-Jones, em termos de

[1] https://docs.google.com/spreadsheets/d/1YhSYuSAA3oq8phWeD9ilJ707fbw_JZaj/edit?usp=sharing&ouid=103409432561298068129&rtpof=true&sd=true

precisão, recuperação e velocidade de processamento, desempenhando um papel essencial no componente de percepção de veículos autônomos, contribuindo para a segurança e eficiência na navegação.

Por outro lado, no trabalho de (MO et al., 2022), apresenta uma análise abrangente sobre as abordagens em segmentação semântica, fundamentadas em aprendizado profundo. Os autores examinam a pesquisa mais atual nas áreas de segmentação semântica fracamente supervisionada, adaptabilidade de domínio, fusão de dados multimodais e segmentação em tempo real. Além disso, o estudo avalia o desempenho em tempo real dos modelos de segmentação e destaca desafios e direções promissoras de pesquisa para otimizar a segmentação semântica por meio do aprendizado profundo.

Os dois estudos mencionados acima oferecem contribuições valiosas que estabelecem uma base sólida e abrangente para a implementação de técnicas de aprendizado profundo. Enquanto (JEVAMIKYOUS; KASHEF, 2022) destaca a eficácia de modelos como YOLO na percepção de veículos autônomos, o trabalho de (MO et al., 2022) oferece uma análise abrangente sobre as estratégias em segmentação semântica, fundamentadas em aprendizado profundo.

3.2 SEGMENTAÇÃO SEMÂNTICA FRACAMENTE SUPERVISIONADA

A segmentação semântica fracamente supervisionada visa segmentar imagens em diferentes categorias semânticas sem depender de anotações pixel a pixel, reduzindo assim os requisitos refinados e o custo associado à anotação manual. Utilizando anotações em nível de imagem, que indicam apenas a presença ou ausência de uma categoria, e anotações de caixa delimitadora para fornecer a localização aproximada de objetos, métodos como mapas de ativação de classes e mecanismos de atenção são empregados para inferir máscaras de segmentação em nível de pixel. Essa abordagem, também explorando anotações de rabisco e parciais, tem mostrado resultados promissores na redução dos requisitos de anotação, tornando a segmentação semântica mais acessível e viável em cenários desafiadores ou dispendiosos em termos de anotação pixel a pixel (MO et al., 2022). Com isso, a aplicação da técnica de segmentação semântica fracamente supervisionada assume uma importância crucial, proporcionando uma solução inovadora para os desafios enfrentados.

3.3 OUTRAS ABORDAGENS

No trabalho de (LAHAT; ADALI; JUTTEN, 2015) destaca-se a preferência por métodos baseados em dados devido à complexidade e às relações desconhecidas entre as modalidades. Neste contexto, é mencionado o uso de modelos simples, como relações lineares entre variáveis, e a incorporação de antecedentes independentes do modelo, como dispersão, não negatividade, independência estatística, classificação baixa e suavidade. Além disso, diversos métodos e estruturas são referenciados, incluindo análise fatorial simultânea baseada modelos estocásticos baseados em aprendizado de máquina, regressão, análise fatorial de grupo bayesiana, decomposições tensoriais, aprendizado de dicionário, codificação esparsa, modelos generativos probabilísticos e aprendizado múltiplo.

Por outro lado, no trabalho de (CUI; GE, 2003) aborda os desafios no posicionamento de veículos autônomos em ambientes urbanos densos, conhecidos como "urban canyon environments". Ao enfrentar as limitações do sistema de posicionamento global (GPS) devido ao bloqueio de sinais por edifícios altos, o artigo propõe uma abordagem restrita, modelando a trajetória do veículo como segmentos de linhas. Essa inovação reduz o número mínimo de satélites necessários para dois, tornando-a aplicável em ambientes urbanos. Diversas estratégias, como um método de aumento de estado para estimar simultaneamente as posições do receptor GPS e os parâmetros da linha, são desenvolvidas, demonstrando eficácia por meio de resultados de simulação na resolução de desafios em ambientes urbanos complexos.

Já no trabalho (VARGAS et al., 2021) apresenta uma abordagem abrangente sobre os fundamentos físicos, o espectro eletromagnético e o princípio de operação de diversos sensores para veículos autônomos (AVs), como RADAR, LiDAR, ultrassom, câmera e sistema global de navegação por satélite (GNSS). A pesquisa quantifica os efeitos de diferentes condições climáticas no desempenho desses sensores, destacando a vulnerabilidade dos sensores ultrassônicos em aplicações automotivas. Além disso, aborda a operação de sistemas RADAR para AVs, com ênfase em RADARs de onda contínua modulada por frequência linear (L-FMCW), explicando o funcionamento por meio de um diagrama de blocos de nível superior. A discussão inclui o uso de microantenas e arquiteturas de sistema em chip (SOC) em sistemas de RADAR AV, explorando suas aplicações em controle de cruzeiro adaptativo, sistemas de prevenção de colisões, detecção de ponto cego e assistência para mudança de faixa.

No estudo (HÖRL et al., 2016) antecipa a disponibilidade de carros totalmente autônomos na próxima década, prevendo que um número expressivo de veículos terá direção totalmente

autônoma nos próximos 50 anos. A pesquisa sobre veículos autônomos é abordada por meio de diversos métodos, como análise estatística de dados históricos, técnicas de agregação de dados de sites e mídias sociais, questionários, pesquisas e simulações. O artigo ressalta os impactos ambientais positivos dos veículos autônomos, especialmente quando utilizados de forma compartilhada, enquanto destaca a importância de narrativas realistas e dados concretos para simulações significativas. Além disso, são mencionadas brevemente as colaborações entre empresas da indústria automotiva no desenvolvimento de plataformas de mobilidade autônoma.

3.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Com base na revisão bibliográfica realizada, os estudos se concentrou na segmentação semântica e na aplicação do aprendizado profundo em sistemas de percepção voltados para veículos autônomos. Essa análise realça a significativa importância e destaque das abordagens que se baseiam no aprendizado profundo em contextos relacionados a veículos autônomos. Além disso, sublinha-se a crescente relevância de abordar desafios específicos, como a adaptação de domínio e a fusão de dados, com o propósito de aprimorar tanto a eficiência quanto a precisão na percepção do ambiente por parte desses veículos. Portanto, a aplicação de estratégias de deep learning com segmentação semântica, tais como a segmentação semântica fracamente supervisionada, a adaptação de domínio, a fusão de dados multimodais e a segmentação em tempo real, representa uma abordagem essencial para o objetivo proposto.

4 METODOLOGIA

O sistema de percepção, enquanto componente central da autonomia dos veículos, desempenha um papel importante na interpretação e reação a informações sensoriais provenientes de diversas fontes, como câmeras, sensores LiDAR, radar e outros dispositivos. Dessa forma, a busca por soluções que aprimorem a precisão, robustez e capacidade de percepção desses sistemas torna-se crucial para a segurança e o êxito dos VAs. Esta seção descreve o conjunto de passos adotados para alcançar os objetivos propostos neste trabalho, que consiste no desenvolvimento e avaliação de um sistema de percepção multimodal para veículos autônomos, visando aumentar sua segurança, eficiência e capacidade de operação em diversas condições e ambientes.

Para treinar e avaliar o desempenho do nosso sistema de percepção multimodal para veículos autônomos, optamos por utilizar uma subárea essencial de *deep learning*, a segmentação semântica fracamente supervisionada. Esta abordagem representa um avanço significativo em relação às abordagens anteriores, como *random forest* e conditional *random field* (CRF), especialmente no que diz respeito à economia de tempo e custos associados à obtenção de anotações em nível de pixel.

Os recentes avanços em sensores, como câmeras e LiDAR, impulsionaram o desenvolvimento acelerado da segmentação semântica, que baseada na fusão de dados multimodais tornou-se uma nova direção de pesquisa para utilizar dados de uma variedade de sensores, cada um com características complementares, a fim de aprimorar o desempenho da segmentação. Um sistema de sensoriamento baseado em uma única câmera pode não oferecer geometria 3D confiável e se adaptar a condições de iluminação complexas ou adversas. Por sua vez, o LiDAR pode fornecer geometria 3D de alta precisão sem depender da luz ambiente.

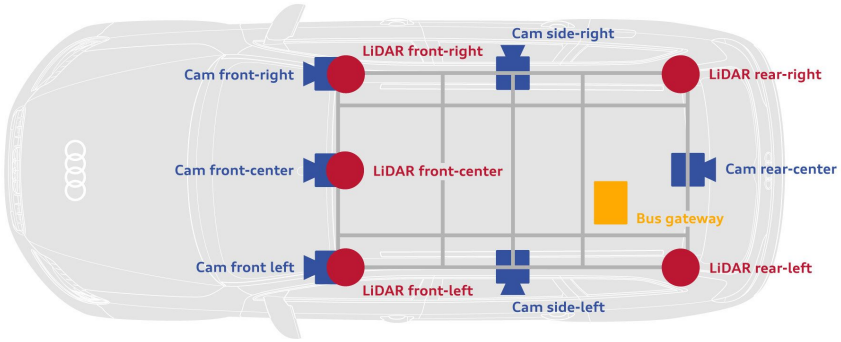
Além disso, existem diversas formas de segmentação, como por pontos, por rabisco e por *bounding boxes*, que se mostram mais simples e exigem uma menor categorização de pixels. Ao combinar essas técnicas com uma seleção apropriada de sensores, é possível obter uma segmentação semântica de alta qualidade e adaptável a diferentes ambientes.

Para o desenvolvimento dos objetivos propostos neste trabalho, inicialmente foi obtido um conjunto de dados fornecido pela empresa de automóveis Audi, composto por mais de 40.000 quadros 20 contendo imagens de segmentação semântica e rótulos de nuvem de pontos. Adicionalmente, são fornecidos dados de sensores não rotulados, totalizando aproximadamente 390.000 frames, para sequências com várias iterações, registrados em três cidades.

O conjunto de dados consiste em informações provenientes de uma variedade de sensores instalados no veículo autônomo em análise. Cada sensor, pertencente às categorias de 'lidars', 'câmeras' e 'veículo', possui um sistema de coordenadas associado denominado de 'visão' ('view'). Esse sistema é definido por uma origem, um eixo x e um eixo y, todos especificados em coordenadas cartesianas (em metros) relativos a um sistema de coordenadas externas.

Neste veículo em particular, cinco sensores LiDAR estão presentes: 'front_left', 'front_center', 'front_right', 'rear_right' e'rear_left'. Cada um destes possui uma 'view' que define sua posição e orientação no referencial do veículo. O veículo também possui seis câmeras distintas: 'front_left', 'front_center', 'front_right', 'side_right', 'rear_center' e 'side_left'. Na Figura 5, são mostrados os sensores e as câmeras do conjunto de dados fornecido pela empresa de automóveis Audi.

Figura 5 – Sensores e câmeras do conjunto de dados



Fonte: GEYER et al. 2020

Utilizaremos o dataset previamente mapeado e normalizado para o treinamento do modelo de deep learning, adotando uma abordagem que envolve a segmentação semântica fracamente supervisionada. Inicialmente, dividiremos a base em dois conjuntos distintos: o conjunto de dados de treinamento (training set) e o conjunto de testes (test set). O conjunto de treinamento será empregado para instruir o modelo a discernir padrões por meio da segmentação semântica, com o intuito de ajustar seus parâmetros e minimizar a disparidade entre a identificação de objetos e os rótulos reais presentes no conjunto.

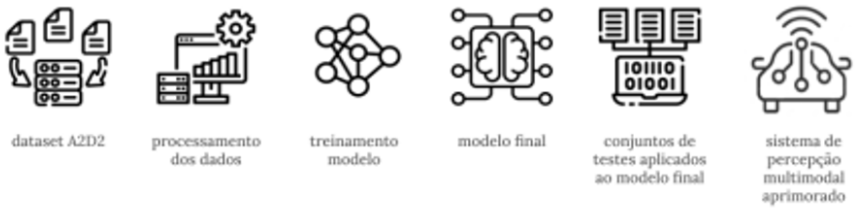
Após o período de treinamento, será fundamental avaliar o desempenho do modelo no conjunto de dados de teste. Essa etapa é crucial para examinar a capacidade do modelo em generalizar diante de um novo conjunto de dados. Além disso, nesta etapa buscaremos identificar

o conjunto ótimo de hiperparâmetros, realizando ajustes que contribuirão para o aprimoramento do desempenho do modelo no contexto da percepção multimodal de veículos autônomos.

Após a conclusão dos treinamentos e refinamentos do modelo de *deep learning* utilizando a abordagem de segmentação semântica fracamente supervisionada, estaremos prontos para empregar o modelo final para a geração de resultados. Os resultados obtidos serão submetidos a uma análise detalhada, examinando a eficácia do sistema de percepção multimodal em identificar e segmentar objetos relevantes em ambientes específicos. Esta análise permitirá avaliar não apenas a precisão do modelo na detecção de objetos, mas também sua capacidade de generalização para cenários variados e a robustez em face de condições desafiadoras.

Como destacado na Figura 6, as etapas ilustradas demonstram a progressão do treinamento do modelo até a análise dos resultados. As observações derivadas da análise dos resultados serão discutidas em relação aos objetivos estabelecidos inicialmente, destacando as contribuições específicas do trabalho para o avanço da percepção multimodal em veículos autônomos. Essa fase final do processo de pesquisa representa não apenas a validação do modelo, mas também a consolidação de conhecimento e a base para futuras melhorias e desenvolvimentos nesse campo.

Figura 6 – Etapas empregadas para execução do trabalho proposto



É importante reconhecer que esta pesquisa possui algumas limitações que podem influenciar a interpretação dos resultados. Primeiramente, a disponibilidade de dados específicos pode influenciar a abrangência da análise das limitações dos sistemas de veículos autônomos. Por exemplo, se o conjunto de dados utilizado contiver predominantemente cenários de condução urbana, as conclusões podem ser mais aplicáveis a ambientes urbanos, enquanto as situações de condução em estradas rurais ou em condições climáticas extremas podem não ser adequadamente

representadas. Além disso, a representatividade da amostra de dados pode afetar a generalização dos resultados para o conjunto mais amplo de sistemas autônomos.

A utilização do dataset fornecido pela Audi também pode implicar resultados específicos para os sistemas desenvolvidos por esta empresa, não sendo necessariamente aplicáveis a outros fabricantes de VAs. Além do desenvolvimento do sistema de percepção multimodal para veículos autônomos, é crucial abordar os desafios em segmentação semântica. A aplicação de técnicas de aprendizado profundo tem impulsionado o campo, mas questões como obtenção de conjuntos de dados mais desafiadores e o equilíbrio entre precisão e velocidade de inferência permanecem em destaque.

5 PROPOSTA EXPERIMENTAL

O avanço dos veículos autônomos exige sistemas de percepção robustos para garantir segurança e eficiência. A interpretação de dados sensoriais diversos é crucial para decisões confiáveis em variadas condições. Neste ínterim, o capítulo em questão apresenta as experimentações utilizadas para a análise de tal tema.

A linguagem escolhida para o estudo do DataSet (GEYER et al., 2020) e para todo o desenvolvimento desta pesquisa é o Python 3. Essa escolha foi motivada pela variedade de bibliotecas disponíveis, incluindo o NumPy e o Pandas. Para esta análise, essas bibliotecas foram fundamentais para abrir e manipular os arquivos de sensores LiDAR. Além disso, é importante ressaltar que o desenvolvimento ocorreu em um ambiente operacional atualizado, utilizando o MacOS 12.5.

No TCC, a realização do mapeamento do dataset foi crucial. Identificamos e catalogamos dados pertinentes, aplicando ferramentas específicas, como a biblioteca NumPy, para extrair, organizar e realizar análises exploratórias. Esse método proporcionou uma base sólida e insights valiosos para a pesquisa em questão.

Empregamos uma amostra do dataset disponibilizado no site A2D2. Os scripts utilizados nos experimentos iniciais foram compartilhados através do GitHub[1], permitindo aos leitores deste trabalho consultar e reproduzir as análises. Cada cidade fornecida possui um arquivo Configuration.json contendo informações abrangentes sobre o setup do veículo, câmeras e sensores. Na Tabela 2, 3 e 4, são apresentadas as configurações dos veículos para cada sensor

Tabela 2 – Arquivo de configuração: Vehicle Configuration

Campo	**Descrição**
origin	Cordenadas de origem do Veículo.
X-axis e Y-axis	Cordenadas da posicão atual do veículo.
ego-dimensions	Dimensões X,Y e Z do veículo utilizado (metros).

Fonte: GEYER et al. 2020

Tabela 3 – Arquivo de configuração: Lidar Configuration

Campo	**Descrição**
origin	Cordenadas do sensor em relação a origem do carro.
X-axis e Y-axis	Cordenadas do sensor em relação ao frame de referência do carro .

Fonte: GEYER et al. 2020

[1] `https://github.com/Gapasquarelli/TCC`

Tabela 4 – Arquivo de configuração: Camera Configuration

Campo	Descrição
tstamp-delay	Atraso configurado entre os frames da câmera (default 0).
origin	Cordenadas da câmera em relação a origem do carro.
X-axis e Y-axis	Cordenadas do câmera em relação ao frame de referência do carro.
CamMatrix	Matriz da câmera das imagens não distorcidas.
CamMatrixOriginal	Matriz da câmera das imagens distorcidas.
Distortion	Parametros utilizados nas imagens originais.
Resolution	Resolução da câmera.
Lens	Tipo de lente utilizado (Fisheye ou Telecam).

Fonte: GEYER et al. 2020

Em cada cidade, os dados estão organizados por frame, sendo que cada captura é composta por três arquivos distintos: uma imagem em formato PNG, um arquivo JSON que referencia as configurações da imagem e dos sensores naquele instante, e um arquivo NPZ contendo as informações dos sensores LiDAR. Na Tabela 5 e 6, são apresentadas as informações dos sensores LiDAR.

Tabela 5 – Arquivo de configuração: Image Frame Information

Campo	Descrição
cam_tstamp	Timestamp do frame.
cam_name	Qual das cameras foi utilizada.
image_zoom	Zoom da imagem (default 1.0).
image_png	Nome do arquivo png referenciado.
pcld_npz	Nome do arquivo npz referenciado (arquivo com os dados dos sensores).
origin	Cordenadas da câmera em relação a origem do carro.
X-axis e Y-axis	Cordenadas da câmera em relação ao frame de referência do carro.
lidar_ids	Relação de todos os Lidar Ids.

Fonte: GEYER et al. 2020

Tabela 6 – Arquivo de configuração: Lidar Frame Information

Campo	**Descrição**
col	Coordenadas dos pontos do sensor.
row	Coordenadas dos pontos do sensor.
distance	Distância do sensor até o objeto medido.
azimuth	Ângulo do sensor até o objeto medido.
timestamp	Timestamp da aferição do sensor.
points	Pontos X, Y e Z dos pontos do sensor (col, row e distance).
depth	Profundidade do ponto de medição.
lidar_id	ID do sensor que realizou a medição.
reflectance	Mede a quantidade refletida pelo objeto que está sendo medido.

Fonte: GEYER et al. 2020

O modo estudado de treinamento do sistema, por sua vez, foi a segmentação semântica fracamente supervisionada, devido à sua eficácia em lidar com conjuntos de dados complexos e variados (neste caso, os dados dos mais diversos sensores do VA) e demonstrar ótimo resultado baseado no dataset utilizado para esta análise; isso porque utilizar outros tipos de segmentação demandam de um grande processamento computacional para uma análise pixel-a-pixel.

Sendo assim, dando ênfase ao "fracamente supervisionada", a proposta é diminuir a quantidade de dados processados tentando reconhecer os objetos através dos *bounding-boxes*, isto é, contornos delimitadores que envolvem o objeto e indicam dados como posição e tamanho aproximado. Ao incorporar esses dados de *bounding-box* no processo de treinamento, o modelo é capaz de aprender com informações mais específicas e localizadas, melhorando a precisão da segmentação semântica. Portanto, essa técnica oferece um equilíbrio entre a precisão e a eficiência, o que a torna uma opção viável para lidar com conjuntos de dados complexos e variados.

Nesta etapa, por sua vez, foram avaliadas as vantagens das *bounding-boxes* sobre outras técnicas de segmentação semântica fraca. Uma das técnicas comparadas é a *scribble*, que envolve a marcação manual aproximada dos contornos dos objetos mas, embora seja útil para tarefas mais refinadas, pode não oferecer o mesmo nível de precisão espacial que a anotação de caixa delimitadora. Da mesma forma, a técnica de ponto, que envolve marcar um único ponto dentro do objeto, pode não fornecer informações suficientes sobre a extensão do objeto em comparação com a anotação de caixa delimitadora.

A vantagem fundamental da anotação de caixa delimitadora é sua capacidade de fornecer informações espaciais detalhadas, permitindo que o modelo compreenda melhor a localização e a extensão dos objetos, o que leva a uma segmentação semântica mais precisa e confiável.

6 CONCLUSÃO

Este trabalho iniciou uma jornada intrigante no campo dos sistemas de percepção multimodal em veículos autônomos, focando em como a integração de diversos sensores pode enriquecer a percepção e a decisão desses sistemas. A pesquisa, ainda em desenvolvimento, busca explorar as sinergias entre dados visuais e de outros sensores para melhorar significativamente a segurança e eficiência dos veículos autônomos.

Até o momento, nossas investigações preliminares sugerem que a combinação dessas diferentes modalidades sensoriais pode superar as limitações individuais de cada sensor, resultando em uma compreensão mais abrangente do ambiente ao redor do veículo. Essa integração promete ser um avanço fundamental para uma navegação mais segura e inteligente.

No que diz respeito à tecnologia, começamos a avaliar o uso de algoritmos de aprendizado de máquina e redes neurais na análise dos dados multimodais. As primeiras impressões indicam que essas tecnologias têm um grande potencial para facilitar a identificação de padrões complexos e a tomada de decisões em tempo real.

Reconhecemos, contudo, que ainda há um grande caminho a percorrer. Desafios como a necessidade de grandes volumes de dados e a dependência de hardware de alto desempenho já foram identificados, e pretendemos abordá-los à medida que o projeto avança. Estas questões, juntamente com as descobertas que virão, irão moldar as etapas futuras da nossa pesquisa.

Concluindo, embora este estudo ainda esteja em andamento, os *insights* iniciais já apontam para a importância crítica dos sistemas de percepção multimodal nos veículos autônomos. Estamos entusiasmados com as possibilidades que este trabalho trará, não apenas para o campo técnico, mas também para o avanço da mobilidade urbana inteligente e segura.

REFERÊNCIAS

CUI, Youjing; GE, Shuzhi Sam. Autonomous vehicle positioning with GPS in urban canyon environments. In. Disponível em: https://ieeexplore.ieee.org/document/1177161.

GEYER, Jakob et al. A2D2: Audi Autonomous Driving Dataset. In. Disponível em: https://www.a2d2.audi.

HÖRL et al. Recent perspectives on the impact of autonomous vehicles. In. Disponível em: https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/121359/ab1216.pdf.

JEVAMIKYOUS, Hrag-Harout; KASHEF, Rasha. Autonomous Vehicles Perception (AVP) Using Deep Learning: Modeling, Assessment and Challenges. In. Disponível em: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9684905.

KEBRIA, Parham M. et al. Deep Imitation Learning for Autonomous Vehicles Based on Convolutional Neural Networks. In. Disponível em: https://www.ieee-jas.net/article/doi/10.1109/JAS.2019.1911825.

LAHAT, Dana; ADALI, Tulay; JUTTEN, Christian. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. In. Disponível em: https://ieeexplore.ieee.org/document/7214350.

MO, Yujian et al. Review the state-of-the-art technologies of semantic segmentation based on deep learning. In. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0925231222000054.

NHTSA. In. Disponível em: https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812506.

VARGAS, Jorge et al. An Overview of Autonomous Vehicles Sensors and Their Vulnerability to Weather Conditions. In. Disponível em: https://www.mdpi.com/1424-8220/21/16/53975.