

AIRCloud: Um Dataset Segmentado de Nuvens de Pontos 3D no Brasil

**Lucas B. Santos¹, Beatriz Pinheiro¹, Pedro Martins¹, Victor Matteus¹
Matheus Leonel¹, Iwens G. Sene¹, Lucas Araújo¹**

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
74.690-900 – Goiânia – GO – Brasil

{lucasbraga7171, biapllopes, omartins.pedro}@gmail.com,
{vitor.matt2003, matheus.leonel161}@gmail.com,
{iwens, apereiral}@ufg.br

Abstract. *Semantic segmentation of 3D point clouds is pivotal to advancing autonomous vehicles because it provides detailed information about the surrounding environment. Nevertheless, major challenges stem from the overall scarcity of annotated datasets for low-resolution LiDAR sensors and the complete absence of collections acquired specifically within Brazil. This work introduces AIRCloud, a segmented 3D point-cloud dataset captured nationally with a 16-beam LiDAR sensor. For validation, we employ the Range-Image U-Net (RIU-Net) architecture previously trained on SemanticKITTI. Multiple pre- and post-processing techniques were assessed to ease the limitations imposed by the sensor’s low resolution. Results, expressed as mean Intersection over Union (mIoU), show that targeted strategies can boost RIU-Net performance—for example, raising mIoU from 40.9% (baseline) to 44.6% with nearest neighbors interpolation. These findings underscore the potential of lower-cost sensors in Brazilian contexts, broadening the outlook for autonomous-driving research in Brazil.*

Resumo. A segmentação semântica de nuvens de pontos 3D é fundamental para o avanço dos veículos autônomos, pois fornece informações detalhadas do ambiente. Contudo, existem desafios significativos relacionados tanto à escassez geral de datasets anotados para sensores LiDAR de baixa resolução quanto à ausência de conjuntos de dados coletados especificamente no território brasileiro. Este trabalho propõe o AIRCloud, um dataset segmentado de nuvens de pontos 3D coletadas com um sensor LiDAR de 16 linhas em território nacional. Para validação, empregamos a arquitetura Range-Image U-Net (RIU-Net), previamente treinada no SemanticKITTI. Diferentes técnicas de pré-processamento e pós-processamento foram avaliadas para mitigar as limitações da baixa resolução do sensor. Os resultados, medidos pela mean Intersection over Union (mIoU), mostram que estratégias específicas podem melhorar o desempenho da RIU-Net — por exemplo, elevação da mIoU de 40,9% (baseline) para até 44,6% com interpolação “nearest neighbors”. Esses achados evidenciam o potencial de sensores de menor custo em contextos nacionais, ampliando as perspectivas para pesquisas em direção autônoma no Brasil.

1. Introdução

A busca por veículos autônomos tem se intensificado nas últimas décadas, impulsionada pelo desejo humano por eficiência, segurança e conveniência no transporte. Apesar dos avanços, a popularização dessa tecnologia ainda enfrenta desafios técnicos e econômicos significativos, especialmente relacionados à percepção, mapeamento, localização e controle do veículo [Chen et al. 2021].

Dentre esses desafios, a percepção desempenha um papel fundamental na interpretação detalhada dos dados sensoriais para identificação e classificação de objetos. Embora existam abordagens tradicionais baseadas em heurísticas, soluções que utilizam aprendizado de máquina têm demonstrado desempenho superior [Pendleton et al. 2017]. Para obter resultados eficazes, esses métodos exigem grandes conjuntos de dados sensoriais diversificados, provenientes de câmeras, radares e especialmente sensores LiDAR, que são robustos a variações ambientais [Elhousni and Huang 2020].

Apesar da importância dos sensores LiDAR para veículos autônomos, a literatura evidencia uma notável escassez de datasets especificamente voltados para sensores de baixa resolução, como o LiDAR de 16 linhas, bem como uma baixa quantidade de conjuntos de dados coletados em países emergentes, como o Brasil. Essa lacuna torna-se especialmente relevante diante das particularidades brasileiras relacionadas à infraestrutura viária, sinalização e qualidade das estradas, características que frequentemente prejudicam o desempenho de modelos treinados exclusivamente com datasets internacionais [Shinzato et al. 2016]. Além disso, restrições econômicas tornam proibitiva, em muitos casos, a aquisição de sensores LiDAR de maior resolução (64 linhas), dificultando a condução de pesquisas mais amplas nesses contextos.

Neste contexto, este trabalho propõe a construção do dataset AIRCloud, utilizando um sensor LiDAR Velodyne Puck de 16 linhas, com o objetivo de preencher essas lacunas e oferecer uma alternativa viável e econômica para estudos em direção autônoma. Busca-se avaliar tanto os desafios técnicos envolvidos na criação desse dataset, quanto o desempenho de sensores de baixa resolução comparados aos sensores mais caros e precisos utilizados em datasets internacionais, como o SemanticKITTI [Behley et al. 2019].

Assim, as principais contribuições deste estudo são: a criação de um dataset segmentado nacional para sensores LIDAR de 16 linhas; a avaliação das limitações técnicas e estratégias para melhorar o desempenho do sensor Velodyne Puck VLP-16; e a análise comparativa da performance do dataset desenvolvido frente a cenários internacionais.

2. Trabalhos Correlatos

O desenvolvimento de novos *datasets* tem desempenhado um papel crucial no avanço da área de direção autônoma, possibilitando a aplicação de modelos de inteligência artificial baseados em dados. Nos últimos anos, diversos conjuntos de dados foram criados, cada um com características e aplicações específicas, que são fundamentais para a validação e aprimoramento de algoritmos de aprendizado de máquina.

O *SemanticKITTI* [Behley et al. 2019] é um dos principais datasets voltados para a segmentação semântica ponto a ponto com dados LiDAR no contexto de direção autônoma. Derivado do *KITTI Vision Odometry Benchmark* [Geiger et al. 2012], ele disponibiliza mais de 43 mil scans anotados, abrangendo 28 classes e um campo de visão

completo de 360 graus. Por ser um dos pioneiros na segmentação semântica detalhada de nuvens de pontos, o *SemanticKITTI* estimulou o surgimento de conjuntos de dados semelhantes, como o *SemanticPOSS* [Pan et al. 2020]. Este dataset conta com 2.988 scans e 14 classes, destacando-se pela ênfase em objetos dinâmicos, especialmente pedestres e veículos em movimento, complementando limitações do *SemanticKITTI* e contribuindo significativamente para avanços em cenários urbanos mais complexos e desafiadores.

Outros trabalhos que também se destacam na literatura são os *datasets nuScenes* [Caesar et al. 2020] e *Waymo Open Dataset* [Sun et al. 2020], que ampliam o escopo de aplicações em direção autônoma ao combinar sensores complementares como LiDAR, radares e câmeras. O *nuScenes* possui um alcance de 360 graus, extensa cobertura temporal, além de mais de 1,4 milhões de *bounding boxes* 3D anotadas, incluindo condições variadas de iluminação e clima, o que fortalece a generalização dos modelos para contextos urbanos diversos. Já o *Waymo Open Dataset*, por sua vez, oferece uma escala ainda maior, abrangendo múltiplas cidades dos Estados Unidos, com mais de 12 milhões de anotações em caixas 3D e alta densidade de pontos. Ambos, no entanto, concentram-se em tarefas de detecção e rastreamento de objetos, diferentemente dos *datasets* citados anteriormente, que abordam especificamente a segmentação semântica ponto a ponto.

Focando no contexto nacional, o *CaRINA Dataset* [Shinzato et al. 2016], desenvolvido pelo Laboratório de Robótica Móvel da USP, utiliza sensores avançados como LiDAR, RADAR e câmeras estéreo para representar cenários urbanos brasileiros complexos, incluindo estradas mal conservadas, diferentes tipos de pavimento, rotatórias e ruas estreitas. Sua abordagem enfatiza a definição das bordas das ruas para treinamento de modelos. Por outro lado, o *RTK Dataset* [Rateke et al. 2019], coletado com uma câmera de baixo custo, visa classificar superfícies das vias, abrangendo pavimentos variados e danos estruturais como buracos e rachaduras, oferecendo um cenário mais acessível e próximo da realidade brasileira.

Embora datasets internacionais como o *SemanticKITTI*, *SemanticPOSS*, *nuScenes* e *Waymo Open Dataset* sejam amplamente utilizados e demonstrem grande valor para tarefas de percepção veicular, eles não abordam as particularidades dos ambientes brasileiros. No contexto nacional, datasets como o *CaRINA* e o *RTK*, apesar de contribuírem significativamente, focam prioritariamente na definição de limites de ruas ou na classificação de pavimentos, sem oferecer segmentação semântica detalhada ponto a ponto. Dessa forma, o *AIRCloud* busca preencher essa lacuna, oferecendo um conjunto de dados brasileiro segmentado em cinco classes principais (*Car*, *Person*, *Road*, *Terrain* e *Vegetation*) usando um sensor LiDAR de 16 linhas, mais acessível financeiramente em relação a sensores de maior resolução, ampliando as possibilidades para pesquisas específicas no contexto nacional.

3. O *AIRCloud* Dataset

O *AIRCloud* foi construído com base em uma metodologia semelhante à do *SemanticKITTI* [Behley et al. 2019], dataset amplamente reconhecido na literatura pela sua abrangência e qualidade, contendo 20.201 frames para treinamento e 20.351 frames para teste, além de representar um esforço expressivo de mais de 1.700 horas dedicadas ao processo detalhado de anotação.

3.1. Coleta dos Dados

O veículo utilizado na coleta dos dados foi o *SD Twizy* (Figura 1), desenvolvido pela empresa britânica *StreetDrone*, equipado com um sensor *LiDAR Velodyne PUCK VLP-16*. Este modelo possui controle *drive-by-wire*, permitindo automação de direção, frenagem e aceleração, além de integrar interfaces *CANbus* para comunicação eficiente entre sensores e controle. Os dados capturados são armazenados em *SSDs* de 2TB e incluem informações geoespaciais obtidas por um sistema GPS/IMU.



Figura 1. SD Twizy utilizado na coleta dos dados.

As coletas foram realizadas no estacionamento em frente ao Instituto de Informática (INF) da Universidade Federal de Goiás (UFG), conforme ilustrado na figura 2. Ao todo, ocorreram quatro sessões em dias distintos, com duração média de 30 minutos cada.

O ambiente escolhido representa uma abstração simplificada de cenários urbanos reais, reproduzindo elementos comuns, porém sem contemplar situações complexas como objetos em movimento ou grande diversidade de elementos na mesma cena.



Figura 2. Imagens do estacionamento da UFG onde os dados foram coletados.

Durante a coleta, os dados foram inicialmente armazenados em formato *rosbag*, utilizado pelo ROS, e convertidos posteriormente para o formato *PCD* (*Point Cloud Data*)

para facilitar a manipulação e segmentação. Por fim, os arquivos foram convertidos para o formato *BIN*, compatível com as ferramentas do *SemanticKITTI* [Behley et al. 2019]. No total, foram obtidos 2374 *frames*, organizados em quatro sequências, conforme evidenciado na Tabela 1.

Tabela 1. Quantidade de frames por sequência

Número da Sequência	Quantidade de Frames
Sequência 00	479
Sequência 01	750
Sequência 02	709
Sequência 03	436

3.2. Processo de Anotação

Assim como descrito pelos autores do *SemanticPOSS* [Pan et al. 2020] em seu processo de anotação, é mais complexo anotar todos os pontos em uma nuvem de pontos 3D do que anotar cada pixel de imagens 2D, e essa é uma das razões de existir poucos conjuntos de dados 3D. Com isso em mente, diferentes estratégias de anotação são adotadas. O *SemanticKITTI* [Behley et al. 2019], por exemplo, combina múltiplos *frames* em um único mapa 3D utilizando o sistema SLAM baseado em laser, denominado *SuMa* [Behley and Stachniss 2018], permitindo uma anotação mais eficiente.

Inspirado nessa abordagem, este trabalho também adotou o método de construção de mapas 3D utilizando o *SuMa* para obter as poses precisas dos *frames* do LiDAR. Posteriormente, foi utilizada a ferramenta *point labeler*, desenvolvida pelos autores do *SemanticKITTI*, para realizar as anotações. O processo inicial de rotulação durou cerca de duas semanas, seguido por revisões e ajustes contínuos até o resultado final. Devido à limitação da resolução do sensor e ao trajeto do veículo, as anotações concentraram-se principalmente no anel central do estacionamento (Figura 3), onde as leituras apresentavam qualidade suficiente para rotulação detalhada.

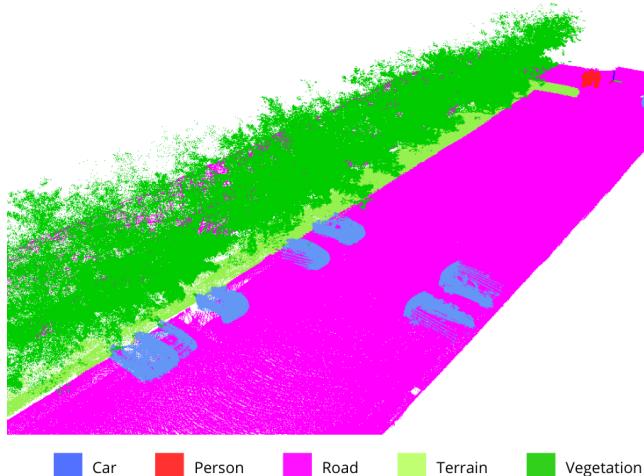


Figura 3. Mapa 3D construído a partir de múltiplos *frames* da sequência 01. Os pontos externos ao estacionamento foram omitidos na visualização.

3.3. Estatísticas do Dataset

Foram definidas cinco classes para a anotação dos objetos no AIRCloud. A Figura 4(a) apresenta a distribuição do número de pontos anotados por classe, evidenciando um desbalanceamento característico entre elas, uma característica comum em conjuntos de dados obtidos em ambientes reais [Behley et al. 2019]. Já a Figura 4(b) mostra a frequência de ocorrência de cada classe em relação ao número total de frames. Embora as classes *car* e *person*, apresentem uma ordem de magnitude menor em relação às outras classes, elas mantêm uma presença consistente no conjunto de dados, o que é um aspecto relevante para sua análise.

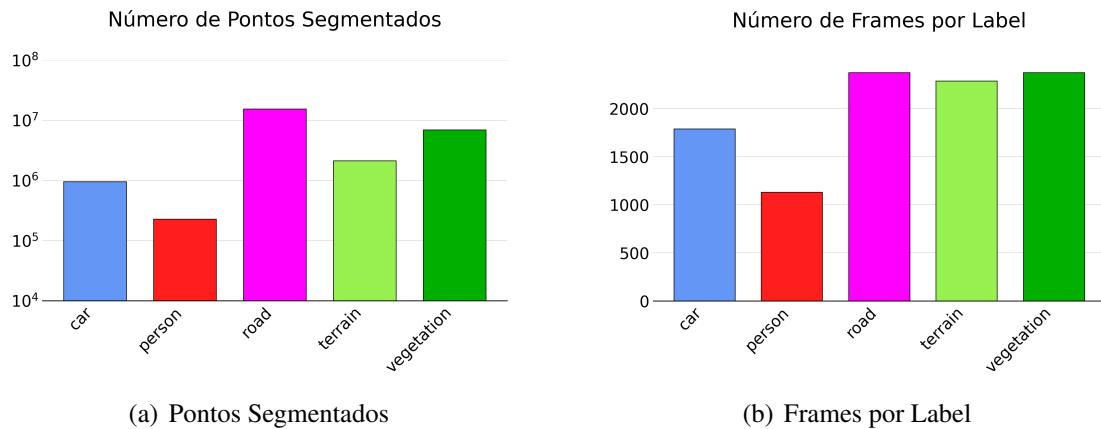


Figura 4. Gráficos de barras contendo análises sobre a frequência e equilíbrio das classes no AIRCloud

4. Resultados

Para avaliar a generalização de modelos treinados com datasets internacionais em cenários nacionais e investigar o impacto da baixa resolução dos sensores LiDAR, é necessário definir um ambiente de avaliação adequado, utilizando metodologias reconhecidas pela comunidade científica.

4.1. Metodologia de Avaliação

Dado o escopo deste trabalho e as limitações computacionais disponíveis, optou-se por utilizar metodologias consolidadas que convertem nuvens de pontos 3D em *range images* 2D [Biasutti et al. 2019, Chen et al. 2020, Wu et al. 2018]. Essa abordagem reduz significativamente a complexidade do processamento, além de ser coerente com o contexto econômico e tecnológico do cenário nacional, por exigir menos recursos de hardware e infraestrutura.

Nesse contexto, este trabalho adota como metodologia de avaliação a arquitetura *Range-Image U-Net (RIU-Net)* [Biasutti et al. 2019], uma adaptação da popular U-Net projetada especificamente para segmentação semântica em dados LiDAR já convertidos em *range images*. A *RIU-Net* aplica convoluções 2D diretamente nessas imagens de profundidade bidimensionais, aproveitando a topologia do sensor para garantir processamento eficiente em tempo real (até 90 *frames* por segundo) com baixa demanda computacional. O modelo foi inicialmente treinado com o dataset *SemanticKITTI* [Behley et al. 2019] e posteriormente avaliado no dataset proposto, o *AIRCloud*, permitindo verificar a adequação dessa abordagem ao contexto brasileiro.

4.2. Ambiente de Avaliação

A RIU-Net [Biasutti et al. 2019] foi treinada conforme descrito no artigo original, utilizando o otimizador Adam (*learning rate* de 0,001) e *batch normalization* com momentum de 0,99. O treinamento ocorreu em 10 épocas, com *batch size* de 8.

A estrutura de pastas do dataset do *SemanticKITTI* [Behley et al. 2019] foi re-organizada conforme a documentação oficial, e um arquivo de configuração (*semantic-kitti.yaml*) foi obtido do repositório oficial, contendo informações essenciais sobre classes e visualização. Ajustes foram necessários para compatibilizar as classes do AIRCloud com o *SemanticKITTI*, e um arquivo de configuração próprio foi criado para o AIRCloud.

Os experimentos foram conduzidos em um ambiente com GPU NVIDIA RTX 2060 SUPER (8GB), processador AMD Ryzen 5 5600 e 32GB de RAM.

4.3. Métricas de Avaliação

Para a tarefa de segmentação semântica utilizando algoritmos de inteligência artificial, os modelos de deep learning recebem a nuvem de pontos do Lidar como entrada e fornecem como saída uma predição semântica para cada um dos pontos. Para avaliar a performance do modelo será utilizado o *mean Jaccard Index* ou *mean intersection over-union*(mIoU), métrica frequentemente utilizada na literatura para avaliar a performance de modelos nas tarefas de segmentação semântica [Behley et al. 2019, Pan et al. 2020]. A fórmula da métrica é descrita como

$$mIoU = \frac{1}{C} \sum_{c=1}^C \frac{TP_c}{TP_c + FP_c + FN_c} \quad (1)$$

onde TP_c , FP_c e FN_c correspondem ao número de predições de verdadeiros positivos, falsos positivos e falsos negativos para a classe c , e C é o número total de classes.

4.4. Experimentos

Considerando que o *AIRCloud* possui apenas 16 linhas, enquanto o modelo foi treinado com o *dataset SemanticKITTI*, que conta com 64 linhas, além do fato de que os dados foram coletados em contextos distintos — no Brasil e na Alemanha —, levantou-se a hipótese de que essas diferenças, dentro outros fatores, poderiam impactar negativamente a qualidade dos resultados. Com base nessa premissa, foram propostos diferentes experimentos para avaliar essas questões.

4.4.1. Primeiro Experimento

O primeiro experimento tem como objetivo avaliar diferentes estratégias de pré-processamento das *range images* utilizadas como entrada para o modelo, com o propósito de realizar o *resize* de 16 para 64 linhas e analisar possíveis ganhos de desempenho. As estratégias avaliadas são as seguintes:

- Interpolação Bilinear: Calcula o valor de cada *pixel* baseado nos *pixels* vizinhos.
- Interpolação por Vizinhos Mais Próximos (*Nearest Neighbors*): Copia o valor do *pixel* mais próximo da posição original.

É importante destacar que, ao redimensionar as imagens de entrada para 64x1024, a saída do modelo terá a mesma dimensão da entrada. No entanto, as *labels* utilizadas para o cálculo das métricas permanecem na dimensão original dos dados (16x1024). Por isso, torna-se necessário o uso de técnicas de pós-processamento para ajustar os resultados à dimensão original. Uma abordagem simples para isso consiste em selecionar, da predição do modelo, apenas as linhas correspondentes à imagem original, conforme exemplificado na Figura 5.

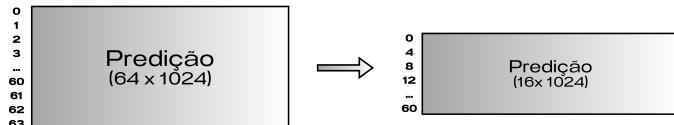


Figura 5. Exemplo da técnica de pós-processamento que consiste em selecionar uma a cada quatro linhas, com o intuito de reduzir de 64 para 16.

Analizando os resultados na Tabela 2, é notório que as estratégias de pré-processamento trouxeram ganhos quantitativos para as classes *Car*, *Road* e *Vegetation*. Esses resultados sugerem que melhorar a resolução dos dados de entrada pode impactar positivamente na performance do modelo. Entretanto, a classe *Terrain* apresentou perdas significativas de desempenho com as técnicas de pré-processamento, levantando questões sobre as causas desse comportamento. Embora os experimentos realizados não sejam suficientes para fornecer uma resposta definitiva, algumas hipóteses, como a possibilidade de mistura com outras classes durante o processo de interpolação, ou diferenças contextuais nos ambientes de coleta dos dados de treinamento e inferência, podem ajudar a esclarecer essa questão.

Tabela 2. Resultados do primeiro experimento.

Inferência RIU-Net [Biasutti et al. 2019]	Car	Person	Road	Terrain	Vegetation	mIoU
Dados Originais	20,6	0.0	76.0	23.8	83.8	40.9
Interpolação Bilinear	26.8	0.1	80.8	5.3	91.1	40.8
Interpolação <i>Nearest Neighbors</i>	34.6	0.0	83.2	7.4	91.7	43.4

4.4.2. Segundo Experimento

A técnica de pós-processamento utilizada nos resultados apresentados anteriormente foi suficiente para alcançar o objetivo de retornar a predição à dimensão original dos dados de entrada. Contudo, é importante observar que, ao selecionar apenas um subconjunto de linhas dos dados, parte da informação é descartada. Para explorar alternativas mais robustas de pós-processamento, um novo experimento foi proposto, mantendo a mesma estrutura de pré-processamento mencionada anteriormente. Nesse experimento, em vez de selecionar apenas a primeira linha a cada grupo de quatro e descartar as demais, optou-se por calcular a média das quatro linhas, resultando em uma única linha agregada, como exemplificado na figura 6. Essa abordagem busca preservar e consolidar a informação, ao invés de descartá-la.

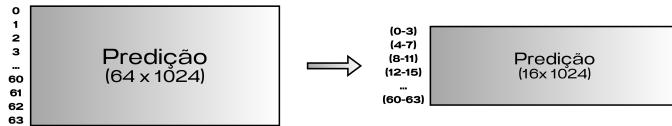


Figura 6. Exemplo da técnica de pós-processamento que consiste em calcular a média de blocos de quatro linhas, com o intuito de reduzir de 64 para 16 sem descartar informação.

Os resultados apresentados na Tabela 3 refletem o desempenho do modelo ao utilizar a nova estratégia de pós-processamento proposta, ao passo de que mantém as estratégias de pré-processamento utilizadas no primeiro experimento. Nota-se uma melhora geral na métrica *mIoU* em comparação com a tabela 2, cujos detalhes e implicações serão discutidos mais profundamente na próxima seção.

Tabela 3. Resultados do segundo experimento.

Inferência RIU-Net [Biasutti et al. 2019]	Car	Person	Road	Terrain	Vegetation	mIoU
Dados Originais	20,6	0.0	76.0	23.8	83.8	40.9
Interpolação Bilinear	30.7	0.1	84.1	7.0	92.1	42.8
Interpolação Nearest Neighbors	35.1	0.0	84.2	11.0	92.6	44.6

4.5. Discussão dos Resultados

Observa-se que o modelo identifica com maior facilidade as classes *Road* e *Vegetation*, provavelmente devido ao maior número de pontos dessas classes nos dados de treinamento. Por outro lado, as classes *Terrain* e *Car* apresentam maiores dificuldades na detecção, enquanto a classe *Person* praticamente não é detectada.

Para a classe *Terrain*, predominante no canteiro central do estacionamento do *AIR-Cloud*, o modelo frequentemente confunde seus pontos com a classe *Road*, como exemplificado na Figura 7, possivelmente devido à proximidade e semelhança entre ambas. Essa limitação pode decorrer da baixa resolução dos dados e das diferenças estruturais do ambiente em relação ao dataset de treinamento (*SemanticKITTI*), dificultando a distinção entre as classes. Quanto à classe *Car*, sua detecção é desafiadora devido à diversidade de formas, tamanhos, movimentação e possíveis oclusões, tornando-a mais suscetível aos efeitos negativos da baixa resolução.

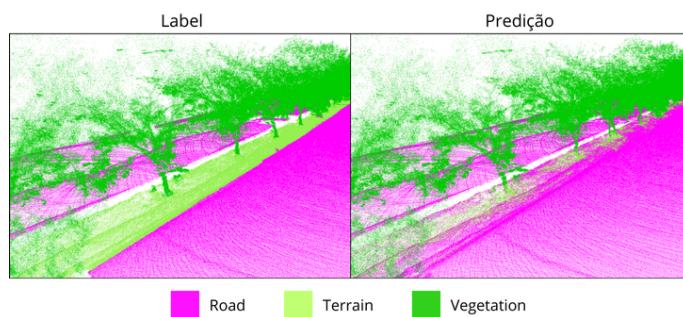


Figura 7. Exemplo da predição do modelo destacando a confusão ocorrida na ilha central entre as classes *Road* e *Terrain*.

Para a classe *Person*, modelos mais simples, como o *RIU-Net*, geralmente apresentam dificuldades significativas na detecção, fato corroborado por estudos anteriores [Behley et al. 2019, Pan et al. 2020]. Isso ocorre devido à alta complexidade intrínseca dessa classe e à escassez de pontos anotados nos conjuntos de dados utilizados para treinamento. O *SemanticPOSS* [Pan et al. 2020] adota uma abordagem que busca melhorar o desempenho da classe *person* ao incluir uma quantidade maior de pontos anotados dessa classe em seu conjunto de dados. No entanto, como esse não é o foco principal deste trabalho, as discussões subsequentes sobre os experimentos realizados não incluirão detalhes específicos relacionados à classe *person*.

4.5.1. Sobre os Experimentos Propostos

A Figura 8 compara as estratégias de pré e pós-processamento do primeiro experimento com os resultados obtidos sem processamento adicional (baseline). Valores à direita indicam ganhos e, à esquerda, perdas na métrica *mIoU*. A estratégia *Nearest Neighbors* superou a *Bilinear*, provavelmente devido à sensibilidade dos dados *LiDAR*. Isso ocorre porque a interpolação *Bilinear* gera novos valores, enquanto a *Nearest Neighbors* replica diretamente os dados originais, preservando a integridade das informações.

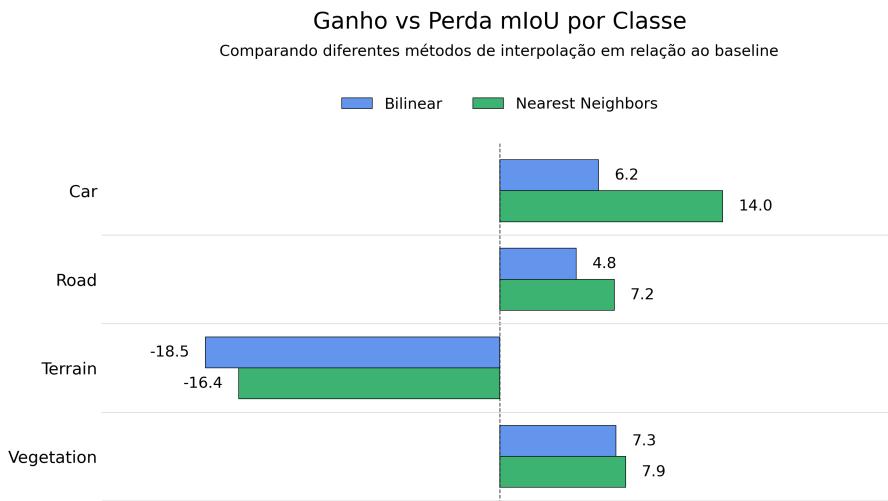


Figura 8. Resultados referentes a tabela 2

Ao analisar a Figura 9, que compara as estratégias de pré e pós-processamento do segundo experimento com o *baseline*, observa-se novamente superioridade da interpolação *Nearest Neighbors*. Contudo, a interpolação *Bilinear* apresentou melhora significativa comparada ao primeiro experimento, indicando que o pós-processamento por média das linhas pode mitigar limitações dessa abordagem, melhorando o desempenho geral.

5. Conclusão

Este trabalho abordou a construção de um *dataset* segmentado com dados coletados por um sensor *LiDAR* de baixo custo, propondo estratégias para mitigar as limitações decorrentes da baixa resolução do sensor. A criação do *dataset* envolveu desafios significativos, incluindo coleta de dados, cálculo das *poses*, anotação detalhada e revisão extensa.

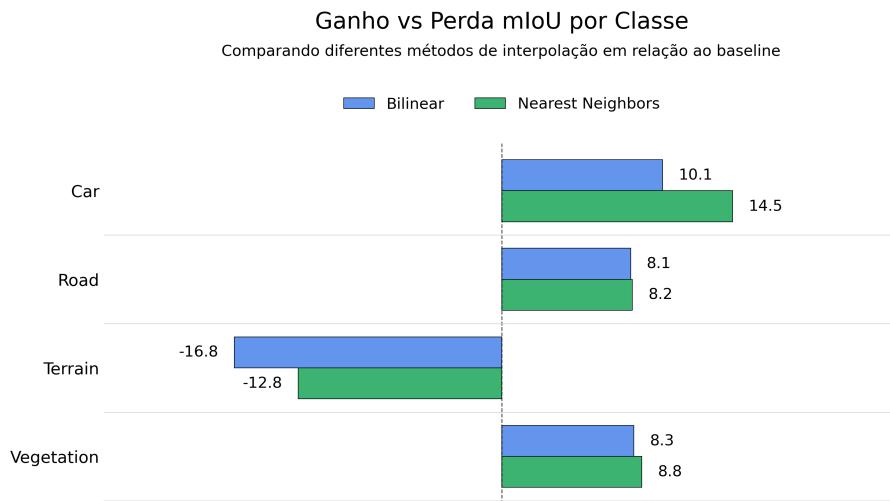


Figura 9. Resultados referentes a tabela 3

Os resultados sugeriram diferenças contextuais importantes entre dados coletados nacionalmente e os disponíveis na literatura internacional, evidenciadas pela dificuldade do modelo em identificar estruturas específicas do ambiente local.

Testes mostraram que técnicas de pré e pós-processamento melhoraram o desempenho geral do modelo, apesar de algumas limitações em classes específicas, como *Terrain*. Esses resultados indicam o potencial de uso de modelos treinados com sensores de maior resolução para auxiliar análises e anotações em sensores mais simples.

Como trabalhos futuros, recomenda-se ampliar a coleta para ambientes mais diversos — incorporando pedestres e veículos em movimento e aumentando o número de classes —, investigar estratégias de aprendizagem por transferência a partir de bases de maior escala para adaptar representações ao contexto nacional, submeter o AIRCloud a diferentes arquiteturas de segmentação como *KPConv*, *PointNet++* e *SalsaNext*, realizar estudos comparativos com outros conjuntos de dados nacionais e explorar técnicas de pré-processamento avançadas, como *upsampling* via inteligência artificial, a fim de mitigar as limitações do sensor LiDAR de 16 linhas.

Entre as limitações deste estudo, destacam-se o ambiente controlado das coletas, que não reflete totalmente cenários dinâmicos reais, e as dificuldades técnicas enfrentadas na etapa de anotação. Adicionalmente, o uso de um único modelo restringe generalizações dos resultados, indicando a importância de explorar múltiplos modelos em pesquisas futuras.

Referências

- Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., and Gall, J. (2019). SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences. In *Proc. of the IEEE/CVF International Conf. on Computer Vision (ICCV)*.
- Behley, J. and Stachniss, C. (2018). Efficient surfel-based slam using 3d laser range data in urban environments.

- Biasutti, P., Bugeau, A., Aujol, J.-F., and Brédif, M. (2019). Riu-net: Embarrassingly simple semantic segmentation of 3d lidar point cloud. *arXiv preprint arXiv:1905.08748*.
- Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., and Beijbom, O. (2020). nuscenes: A multimodal dataset for autonomous driving. In *CVPR*.
- Chen, K., Oldja, R., Smolyanskiy, N., Birchfield, S., Popov, A., Wehr, D., Eden, I., and Pehserl, J. (2020). Mvlidarnet: Real-time multi-class scene understanding for autonomous driving using multiple views. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2288–2294.
- Chen, S., Liu, B., Feng, C., Vallespi-Gonzalez, C., and Wellington, C. (2021). 3d point cloud processing and learning for autonomous driving: Impacting map creation, localization, and perception. *IEEE Signal Processing Magazine*, 38(1):68–86.
- Elhousni, M. and Huang, X. (2020). A survey on 3d lidar localization for autonomous vehicles. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 1879–1884.
- Geiger, A., Lenz, P., and Urtasun, R. (2012). Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 3354–3361.
- Pan, Y., Gao, B., Mei, J., Geng, S., Li, C., and Zhao, H. (2020). Semanticposs: A point cloud dataset with large quantity of dynamic instances.
- Pendleton, S. D., Andersen, H., Du, X., Shen, X., Meghjani, M., Eng, Y. H., Rus, D., and Ang, M. H. (2017). Perception, planning, control, and coordination for autonomous vehicles. *Machines*, 5(1).
- Rateke, T., Justen, K., and Von Wangenheim, A. (2019). Road surface classification with images captured from low-cost camera - road traversing knowledge (rtk) dataset. *Revista de Informática Teórica e Aplicada*, 26.
- Shinzato, P. Y., dos Santos, T. C., Rosero, L. A., Ridel, D. A., Massera, C. M., Alencar, F., Batista, M. P., Hata, A. Y., Osório, F. S., and Wolf, D. F. (2016). Carina dataset: An emerging-country urban scenario benchmark for road detection systems. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 41–46.
- Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., Vasudevan, V., Han, W., Ngiam, J., Zhao, H., Timofeev, A., Ettinger, S., Krivokon, M., Gao, A., Joshi, A., Zhang, Y., Shlens, J., Chen, Z., and Anguelov, D. (2020). Scalability in perception for autonomous driving: Waymo open dataset. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2443–2451.
- Wu, B., Wan, A., Yue, X., and Keutzer, K. (2018). Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1887–1893.