EasyChair Preprint

№ 15231

Improving Low-Visibility 3D Sensing Performance for Autonomous Vehicles with Camera-Radar Fusion

Ruan Bispo, Bruno Borges de Oliveira and Valdir Grassi Jr

October 18, 2024

Melhorando o desempenho da detecção 3D de baixa visibilidade para veículos autônomos com fusão câmera-radar ⋆

Ruan Bispo * **Bruno Borges** * **Valdir Grassi Jr.** *

* *Departamento de Engenharia Elétrica e de Computação, Escola de Engenharia de São Carlos, Universidade de São Paulo (e-mails: {ruanrobert, bborges.oliveira, vgrassi}@usp.br).*

Abstract: Since the emergence of autonomous vehicles, certain tasks such as object detection have become more necessary, and the adoption or rejection of the technology depends on accurately locating and identifying vehicles and pedestrians on the streets. Considering the current conditions, where human drivers are capable of efficiently recognizing and estimating the distance between these obstacles on roads under any weather and lighting conditions, it is expected that, as feasibility requirements for the adoption of autonomous vehicles on the streets, the vehicle is capable of performing the same function equally or superiorly, considering the same precision and task execution time. Thus, this work presents the modification of a 3D object detection architecture using camera-radar sensor fusion to reduce processing time, data volume, and memory required by the base paper. Results demonstrated a significant reduction in computational cost while maintaining metrics at the same level as the base work.

Resumo: Com o surgimento de veículos autônomos, algumas tarefas como a detecção de objetos no ambiente têm se tornado cada vez mais necessárias, e a adoção ou rejeição da tecnologia depende da acurácia na localização e identificação de veículos e pedestres nas ruas. Considerando as condições atuais, em que motoristas humanos são capazes de reconhecer e estimar, de maneira eficiente, a distância entre esses obstáculos nas vias sob quaisquer condições de clima e iluminação, espera-se que, como requisitos de viabilidade para a adoção de veículos autônomos nas ruas, o veículo seja capaz de desempenhar a mesma função de forma igual ou superior, considerando a mesma precisão e tempo de execução da tarefa. Assim, neste trabalho é apresentada a modificação de uma arquitetura de detecção 3D de objetos utilizando fusão sensorial câmera-radar com o objetivo de reduzir o tempo de processamento, volume de dados e memória requerida pelo artigo base. Resultados demonstraram uma redução significativa no custo computacional mantendo as métricas no mesmo nível do trabalho base *.

* *https://github.com/RuanBispo/radarnet-smalldet3d*

Keywords: Self-driving cars, sensor fusion, nuScenes, camera-radar, adverse weather conditions.
Palavras-chaves: Veículos autônomos, fusão sensorial, nuScenes, câmera-radar, climas adversos.

1. INTRODUÇÃO

Com a busca crescente de evolução na área de veículos autônomos, alguns campos têm se desenvolvido e recebido maior destaque nos últimos tempos, sendo a detecção 3D de objetos um deles. Utilizando imagens de câmeras, nuvem de pontos do LiDAR ou dados de distância do radar como sensoriamento, a detecção 3D tem como objetivo principal, reconhecer e localizar objetos como pedestres, ciclistas e outros veículos no ambiente (Wang et al., 2020a). Esta detecção é realizada através da criação de *bounding boxes* tridimensionais que substituem o objeto detectado por um modelo simplificado de dimensões similares ao objeto em questão.

⋆ Este trabalho foi parcialmente financiado pelo CNPq (465755/2014-3 e 309532/2023-0), CAPES (Código de Financiamento 001), e Fundep - Mover (Linha V, Acordo 27192.02.04/2021.01-00, projeto 27192-54).

A simplificação destes objetos através de *bounding boxes* pode auxiliar na predição de movimentação de objetos dinâmicos no ambiente, pois, uma vez detectado o tipo e localização do objeto, pode-se predizer seu comportamento e evitar colisões com o veículo autônomo. Como exemplo, é possível comparar pedestres e veículos detectados, os quais possuem espaços de configuração diferentes, com velocidades e comportamentos distintos em variados segmentos do caminho percorrido por um veículo autônomo. Por conta disso, a área de detecção 3D tem se desenvolvido consideravelmente nos últimos anos (Wang et al., 2020b).

No contexto de veículos autônomos, a tarefa de detecção de objetos 3D possui uma grande importância, uma vez que, com as informações de objetos dinâmicos no ambiente, é possível moldar comportamentos e garantir a segurança em ambientes que possuem robôs na direção. No entanto, apesar dos avanços recentes em tecnologias de sensoriamento aplicadas a esses veículos, os sensores utilizados

ainda possuem limitações devido a ruídos, condições adversas de ambiente, variações de luz, ou mesmo devido à natureza do sensoriamento de cada dispositivo. Com base nisso, o uso de apenas um sensor tem se mostrado uma alternativa menos viável, evidenciando a necessidade de utilização de vários sensores em conjunto, de tal modo que seja possível extrair as vantagens de um sensor de forma complementar aos demais (Fayyad et al., 2020).

Dentre os vários desafios que envolvem o sensoriamento na detecção 3D de objetos, pode-se citar a restrição de sensores sob condições adversas. Em geral, os sistemas de sensoriamento são desenvolvidos para trabalhar sob determinadas condições, as quais, por muitas vezes, tem se tornado muito restritas a ambientes controlados. É possível encontrar vários estudos que abordam o problema de sensoriamento utilizando condições ideais, em que os sensores não são obstruídos, seja por iluminação, poeira, oclusão, neblina, chuva, dentre outros motivos (Pfeuffer and Dietmayer, 2018).

A baixa disponibilidade de *Datasets* com condições adversas tem sido outro motivo da detecção 3D ter sido mais abordada para condições mais controladas de clima e iluminação. Também destacam-se a dificuldade de processamento de múltiplos sensores em tempo real, que pode ser considerado 10 Hz devido a limitações de tempo de amostragem dos sensores (Arnold et al., 2019), e a necessidade de uma variedade maior de sensores nos *Datasets* para realização de estudos de fusão sensorial.

Levando isso em consideração, foi possível observar que a área de detecção 3D sob condições adversas tem sido um campo pouco explorado, e que necessita de mais estudo para se desenvolver de forma que supra as necessidades atuais (Bijelic et al., 2020). Assim, neste artigo é proposta a modificação de uma arquitetura de detecção 3D que aborda a característica de condições adversas com o objetivo de aprimorar sua eficiência, em termos de tempo de execução e custo de memória, sem o comprometimento do desempenho esperado para a sua categoria.

2. REVISÃO DA LITERATURA

Um dos grandes desafios dentro da área de percepção em veículos autônomos, atualmente, é a consideração de condições adversas de clima e luminosidade no sistema de sensoriamento. Como exemplos, pode-se citar a variação de estações do ano com presença de chuva, neve, neblina, variações de iluminação com a mudança de horas durante o dia, presença de luz forte do sol direta sobre os sensores e poeira (Valada et al., 2017).

No que se refere aos *Datasets*, é possível observar um número pequeno de conjuntos de dados relacionados com condições adversas de tempo e iluminação. Isso se deve à característica recente de pesquisas em detecção de objetos, levando a uma restrição do campo de pesquisa a condições mais simples de aplicação, visando cenários mais controlados antes de abordar ambientes com maiores variações. Somente a partir de 2017 deu-se início à criação de *Datasets* com o foco nesse tipo de dados, envolvendo condições adversas no ambiente de percepção.

São exemplos de *Datasets* encontrados que possuem condições variadas de operação o *Dataset* de Oxford (Maddern et al., 2017) e sua posterior extensão (Barnes et al., 2020), os quais foram coletados no decorrer de um ano inteiro, sendo possível observar variações de estação, juntamente com iluminação, neve e neblina, que é atualmente um dos maiores desafios na detecção de objetos neste contexto. Outros *Datasets* com características parecidas surgiram posteriormente (Huang et al., 2018; Pham et al., 2020; Braun et al., 2019; Caesar et al., 2020).

Quanto às técnicas utilizadas para lidar com as condições adversas, destaca-se o trabalho de Mehra et al. (2020), o qual aborda condições climáticas que envolvem ambientes com neblinas. Neste trabalho, é proposta a ReViewNet, que utiliza dados de câmera, abordando o desembaçamento de imagens para detecção de objetos, sendo esta abordagem passível de utilização em tempo real, devido a sua natureza rápida de processamento, ultrapassando as CNNs e arquiteturas baseadas em GAN, tanto em precisão quanto em velocidade de processamento.

Pfeuffer and Dietmayer (2018) utilizaram a abordagem de transformação de *Datasets*, a partir do *Dataset* KITTI (Geiger et al., 2012), que é um *Dataset* clássico no estudo de veículos autônomos, sendo um dos mais utilizados até então. Neste trabalho foram incorporadas modificações que simulassem condições adversas ao *Dataset* para a etapa de treinamento e validação. Como exemplo dessas modificações, pode-se citar pontos brancos adicionados às imagens e aos pontos do LiDAR, uma vez que, sob luz intensa do sol ou em ambiente com neve é esperado que estes efeitos ocorram, além da exclusão de sensores de alguns trechos para simular sua falha.

Foi observado que, sob essas condições adversas, o desempenho das redes treinadas apenas em condições normais de operação decaiu em cerca de 20%, sendo possível concluir que, para estes exemplos, as fusões sensoriais do tipo*Late Fusion* obtiveram melhores resultados, devido aos aspectos de falha sensorial. Neste tipo de fusão, cada sensor possui camadas de extração de características individuais, de modo que a presença de falha em um sensor não influencia os demais sensores, obtendo resultados melhores para esse tipo de situação (Pfeuffer and Dietmayer, 2018).

Outros trabalhos podem ser encontrados com aplicações a condições adversas, principalmente considerando os *Datasets* que possuem tal característica. Como exemplo, pode-se citar os trabalhos de Nabati and Qi (2021), Xu et al. (2021) e Liu et al. (2022). No entanto, apesar de utilizar dados com condições adversas disponíveis, estes trabalhos não possuem o foco neste tipo de situação, não possuindo, assim, um estudo muito aprofundado sobre como lidar com tais problemas, ou restrições dos algoritmos para estas condições.

No cenário de fusão sensorial, cada sensor atua a complementar a informação do outro para garantir a melhor e mais precisa percepção do ambiente. No entanto, sensores como câmeras e LiDAR sofrem grande interferência de condições adversas, tendo seus dados parcialmente ou completamente inutilizados nesses cenários. Em contrapartida, os radares apresentam maior imunidade à condições adversas, como evidenciado por Sheeny et al. (2021). Em seu trabalho foi desenvolvido o *Dataset Radiate*, contendo dados de radar, câmera estéreo, LiDAR e GPS coletados

simultaneamente em climas diversos. Também foi comparado o desempenho de modelos treinados com dados de radar evidenciando a baixa interferência do clima no radar.

Adicionalmente, alguns trabalhos têm tentado utilizar os dados de radar em conjunto com dados de LiDAR, seja apenas na etapa de treinamento, ou através da geração de mapas de profundidade. Esses métodos atualmente configuram o estado da arte em termos de métricas na página oficial do nuScenes, conseguindo alcançar o nível de detecção obtido utilizando o LiDAR acima de 0.5 de NDS *(nuScenes detection score)*(Lei et al., 2023; Kim et al., 2023). No entanto, essa abordagem possui a desvantagem de ainda utilizar informações de LiDAR, o que aumenta o volume de dados que precisam ser processados, elevando o tempo de execução e o custo geral da implementação, uma vez que um LiDAR tem o custo aproximado ao de um veículo completo.

Considerando a restrição de uso apenas da câmera e radar, alguns trabalhos como REDFormer (Cui et al., 2023), propõe a fusão sensorial câmera-radar com objetivo de melhorar as métricas de detecção gerais e em condições adversas, acrescentando redes de extração de *features* do radar e uma abordagem *multi-task learning* para detectar as condições adversas, ao custo do tempo de processamento. Outros trabalhos focam em melhorar o problema do alinhamento de dados de radar com dados da câmera, como em MVFusion (Wu et al., 2023), mas também ao custo de um maior processamento e volume de dados.

3. METODOLOGIA

Para este trabalho foi utilizado o *Dataset* nuScenes (Caesar et al., 2020), devido a sua variedade de ambientes: noturno, chuvoso e dia, além da sua disponibilidade de dados com câmera e radar, possuindo ainda um volume de dados superior aos demais *Datasets* atuais. Considerando a disposição dos dados, o *Dataset* nuScenes possui um conjunto de 1x LiDAR, 5x radares e 6x câmeras para detecção 3D; 1000 cenas de 20 segundos cada; 1.400.000 imagens de câmera; 390.000 revoluções de LiDAR; duas cidades diferentes: Boston e Singapura; além de 1.4M de 3D *bounding boxes* manualmente anotadas para 23 classes de objetos diferentes.

O *Dataset* nuScenes utiliza, como métrica de avaliação principal para validação de desempenho dos modelos, a métrica de detecção NDS *(nuScenes detection score)*. Essa métrica usa como parâmetros a precisão média mAP *(mean average precision)*, além de um conjunto de métricas de verdadeiro positivo mTP *(true positive metrics)* como: ATE *(average translation error)*; ASE *(average scale error)*; AOE *(average orientation error)*; AVE *(average velocity error)* e AAE *(average attribute error)*. A seguinte equação ilustra como é realizado o cálculo de avaliação:

$$NDS = 0.5 \times mAP + \sum 0.1 \times max((1 - mTP), 0) \quad (1)$$

3.1 Ambiente de implementação e teste

Para treinamento e validação do modelo foi utilizada a biblioteca Pytorch, MMDetection3D toolbox (Contributors, 2020) e linguagem Python. O modelo foi desenvolvido utilizando o computador disponível no Laboratório de Sistemas Inteligentes (LASI) da USP-SC, utilizando uma GPU NVIDIA GeForce RTX 3090 (24G) e processador AMD® Ryzen™ 9 5950X (3.4GHz). Foi utilizado todo o conjunto de dados de treinamento do nuScenes e a validação foi realizada dividindo o conjunto de validação utilizando sua descrição nos metadados para dados chuvosos e noturnos, de modo a viabilizar o estudo sob condições adversas de clima e baixa visibilidade.

3.2 Modelo base

Como modelo base foi utilizada a arquitetura REDFormer (Cui et al., 2023) que adapta o modelo BEVFormer (Li et al., 2022), que usa apenas câmeras, para uso com radar. Para este trabalho foram utilizadas imagens de tamanho reduzido de forma que fosse possível atuar com modelos mais simples de computador, uma vez que o modelo base é treinado utilizando oito GPUs NVIDIA V100, reduzindo o tempo de treinamento e tornando viável testes para instituições com menos recursos. Adicionalmente são comparadas as arquiteturas MVFusion (Wu et al., 2023) e CenterFusion (Nabati and Qi, 2021), que são arquiteturas que estão na mesma categoria de fusão sensorial que não utilizam a nuvem de pontos do LiDAR ou mapas de profundidade para complementar a fusão câmera-radar.

3.3 Arquitetura proposta

A arquitetura completa utilizada neste trabalho parte da rede REDFormer (Cui et al., 2023), que realiza a fusão sensorial dos dados de radar através da proposta principal da inclusão de uma *Backbone* para o radar, a qual insere as informações de posição do radar nas camadas de agregação de *features* da câmera para o espaço 2D, com o objetivo de aprimorar a regressão da detecção 3D da arquitetura BEVFormer (Li et al., 2022). É possível visualizar a arquitetura completa na Figura 1, em que são recebidos como dados de entrada as imagens da câmera e os pontos dos radares, sendo sua saída a estimativa das *bounding boxes* e detecção de chuva e ambiente noturno.

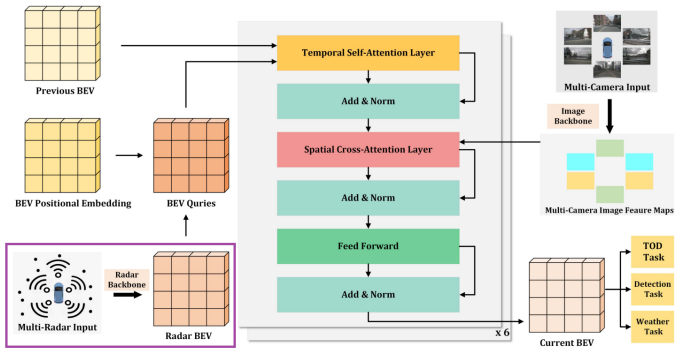

Figura 1. Arquitetura completa (Cui et al., 2023).

Como principal contribuição deste artigo pode-se destacar a modificação da *Backbone* do radar, destacada no canto esquerdo da Figura 1, para atuar aproveitando melhor as informações fornecidas pelo sensor através da adaptação da rede para lidar com uma maior variedade de *features*, inclusão de técnicas de *data augmentation*, juntamente com uma redução no tamanho da imagem original para melhorar o desempenho da rede. Deste modo, podem-se destacar as seguintes modificações no modelo base:

- Refinamento da arquitetura do radar através da implementação de uma arquitetura multicamadas para as *features* do histograma do radar;
- Implementação de duas técnicas de pré-processamento de dados para realização de *data augmentation*. Abordagem multi-ciclos, utilizando dados de instantes de tempo anteriores e realização de operação de fechamento para aumentar a densidade de pontos do radar;
- Aumento do desempenho da rede, reduzindo a quantidade de memória requeria e elevando os FPS através da redução da imagem de entrada.

Assim, para a etapa de implementação da abordagem multicamadas foram testadas mais duas camadas de *features* disponíveis nos dados do radar: velocidade radial e a *flag* dynProp. Sendo a velocidade radial obtida pelo efeito doppler, que mede a velocidade do ponto detectado pelo radar para o veículo utilizando o sensor e a *flag* caracterizada pela informação de propriedades dinâmicas do objeto atingido pelo radar, para indicar se está se movendo.

Considerando que os radares possuem uma densidade pequena de pontos, cerca de 625 pontos por ciclo, uma das técnicas de *data augmentation* utilizadas foi amostrar pontos em instantes de tempo anteriores $t - i$ propagando as posições futuras considerando a velocidade radial encontrada através do efeito doppler do radar, em que, para esta técnica foi utilizado o valor empírico de $i = 3$. Para valores maiores de i foi observado uma tendencia alta de desvio na propagação de posição para objetos em movimento, o que era esperado devido a aceleração dos objetos.

Em seguida, foi realizada a operação morfológica de fechamento para aumentar a densidade de pontos do radar. Considerando que o objetivo da operação era preencher os espaços vazios entre os pontos próximos de radar optou-se por esta técnica de processamento de imagem. Para essa operação foram testados alguns valores de erosão e dilatação, sendo o melhor resultado obtido através de um ciclo de dilatação seguido de um ciclo de erosão com um *kernel* de valor unitário e tamanho (3×3). A Figura 2 ilustra as modificações citadas na *Backbone* do radar, sendo f as *features* utilizadas e x e y coordenadas espacias.

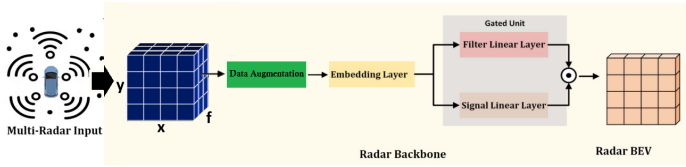

Figura 2. *Backbone* do radar adaptado (Cui et al., 2023).

Por fim, foi escolhido para este trabalho o tamanho de imagem (960x540), equivalente a 60% do tamanho total da imagem disponível no *Dataset* nuScenes (1600x800). Este ajuste teve por base trabalhos anteriores que reduziram sua imagem para (1280x720), cerca de 80% do valor original, sem perda considerável do desempenho. No final da sessão de resultados é apresentada o desempenho para uma única época de treinamento considerando a arquitetura desenvolvida de forma completa para fins de comparação.

4. RESULTADOS

Para verificar o desempenho da arquitetura proposta na detecção 3D de objetos, foi realizada sua validação considerando o conjunto de dados de validação no nuScenes e os resultados obtidos nos modelos base, sendo extraídas suas duas principais métricas para este trabalho. Na Tabela 1 são ilustrados os resultados completos da arquitetura implementada, em que é ilustrado em cinza o desempenho da arquitetura base com o tamanho de imagem padrão e reduzido, seguida da arquitetura com modificações propostas neste trabalho.

Tabela 1. Métricas do conjunto de validação completo do nuScenes.

Método	Imagem	NDS ↑	mAP ↑
CenterFusion	800×450	0.453	0.332
MVFusion	1600×640	0.455	0.380
BEVFormer	1280×720	0.479	0.370
REDFormer	1280×720	**0.486**	**0.385**
REDFormer	960×540	0.437	0.316
(Proposta)	960×540	0.475 (+9%)	0.366 (+16%)

Como etapa adicional de validação, foi realizada a divisão dos dados de validação em dois subconjuntos exemplificados na Tabela 2, sendo um com dados noturnos e outro com dados de chuva, de modo que fosse possível comparar o desempenho da arquitetura atual sob condições variadas de clima e tempo. Foi possível observar que mesmo com o tamanho de imagem reduzido, a arquitetura proposta consegue manter o mesmo nível de desempenho ou mesmo superar o trabalho base, como é o caso da abordagem noturna que teve o resultado mais promissor.

Tabela 2. Métricas dos subconjuntos extraídos dos dados de validação. "N" representa os dados à noite e "C" os dados com chuva.

	Método	Imagem	NDS ↑	mAP ↑
N	BEVFormer	1280x720	0.191	0.182
N	REDFormer	1280x720	0.281	**0.203**
N	REDFormer	960x540	0.256	0.179
N	**(Proposta)**	960x540	**0.284** (+11%)	**0.203** (+13%)
C	BEVFormer	1280x720	0.388	0.352
C	REDFormer	1280x720	**0.509**	**0.404**
C	REDFormer	960x540	0.454	0.334
C	**(Proposta)**	960x540	0.508 (+12%)	0.381 (+14%)

Considerando o tamanho base da arquitetura e seu tempo de execução, é possível observar a redução significativa obtida através das modificações propostas, em que, comparado ao modelo base REDFormer, foi possível alcançar uma melhoria de aproximadamente +66% no tempo de execução total, sendo este um parâmetro essencial para a implementação futura em veículos autônomos. A Tabela 3 ilustra o desempenho do BEVFormer mostrado no artigo original com hardware similar e os resultados para a variação da arquitetura REDFormer proposta.

Tabela 3. Comparativo da arquitetura proposta. "C" indica o uso de câmera e "R" radar

Método	Sensor	Imagem	VRAM ↓	FPS ↑
BEVFormer	C	1280x720	10.5G	2.3
REDFormer	C+R	1280x720	9.2G	4.7
(Proposta)	C+R	960x540	7.6G (-17%)	7.8 (+66%)

4.1 Variações na arquitetura proposta

Para a etapa de testes e validação foi realizada uma variação nos parâmetros adicionados à arquitetura treinada com menos épocas, com o intuito de avaliar o impacto das mudança sugeridas. Os efeitos de cada modificação nas camadas do radar são apresentados qualitativamente na Figura 3. Nesta figura é possível observar o histograma de entrada da camada de extração de *features* do radar considerando cada uma das mudanças propostas: apenas um ciclo de coleta de dados; adição da abordagem multi-ciclos; operação de fechamento; e o uso da abordagem *multi-feature* para destacar objetos móveis. Os veículos na cena são ilustrados pelos retângulos em branco.

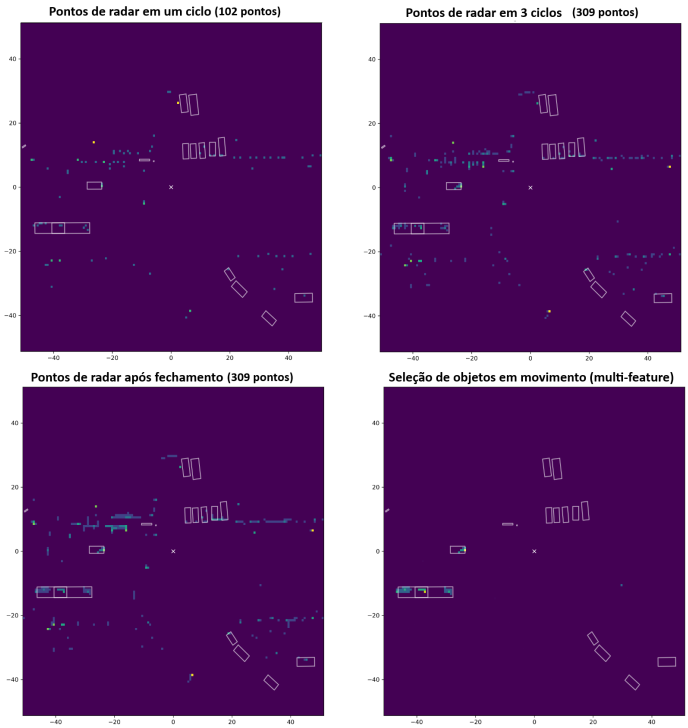

Figura 3. Histograma de pontos do radar em metros. Cores mais destacadas ilustram maior densidade de pontos.

As *features* do radar utilizadas incluem a posição em coordenadas x e y, a velocidade radial em cada coordenada, e um parâmetro que indica as propriedades dinâmicas do objeto. Esses parâmetros são ilustrados na Tabela 4, abreviados como *Pos*, *Vr* e *dynProp*, respectivamente. Para a abordagem multi-ciclos e a operação de fechamento, foram testados diversos valores, obtendo os melhores resultados empiricamente com 3 ciclos e uma operação de dilatação seguida de erosão. Os impactos da adição dessas abordagens são mostrados nas Tabelas 5 e 6.

Tabela 4. Teste com variação de *features* para conjunto de validação com multiciclos.

Pos	Vrad	dynProp	NDS ↑	mAP ↑
x			0.466	0.356
x	x		0.464	0.356
x		x	0.465	0.358
x	x	x	0.469	0.361

Por fim, foram realizados testes variando o tamanho da imagem para identificar seus impactos na nova arquitetura em termos de eficiência e métricas atingidas. Nesse teste foi utilizada a arquitetura desenvolvida, treinada para apenas uma época, com o objetivo de visualizar o impacto da variação no tamanho da imagem no tempo de execução, balizando assim trabalhos futuros. Foi escolhido apenas uma época para fins de comparação, pois foi observado que, após a primeira época, há variações apenas na terceira e quarta casa decimal por época, não ultrapassando uma mudança na escala de dezenas de centésimos nas métricas, mesmo para um número grande de épocas.

Tabela 5. Teste com variação de multiciclos para conjunto de validação com *multi-features*.

Técnica de pré-processamento	NDS ↑	mAP ↑
sem abordagem multi-ciclos	0.468	0.358
com abordagem multi-ciclos	0.469	0.361

Tabela 6. Teste de validação do impacto do uso da operação de fechamento sobre as demais.

Técnica de pré-processamento	NDS ↑	mAP ↑
sem operação de fechamento	0.469	0.361
com operação de fechamento	0.471	0.361

Tabela 7. Métricas do conjunto de validação do nuScenes para diferentes tamanhos de imagem para a arquitetura proposta.

Imagem	NDS ↑	mAP ↑	VRAM ↓	FPS ↑
1280x720	0.486	0.382	9.2G	5.1
1120x630	0.483	0.377	8.6G	6.0
960x540	0.468	0.352	7.6G	7.8
800x450	0.455	0.337	7.1G	9.4

5. CONCLUSÃO

Com o objetivo de possibilitar mais trabalhos na área de detecção 3D de objetos levando em consideração variações climáticas e de iluminações nos cenários, foi desenvolvida uma variação da arquitetura REDFormer, atacando pontos propostos pelo artigo base, o qual cita sua alta taxa de processamento e tamanho. Com uma redução no tamanho da imagem e realização de modificações na arquitetura do radar, foi implementado um histograma multicamadas e duas técnicas de *data augmentation* aplicando a adição de multi-ciclos de dados e operações de fechamento.

Assim foi possível reduzir significativamente o tempo de execução, memória requerida, e aumentar os FPS sem comprometer seu desempenho. Além disso, foram apresentados testes de variação da arquitetura, que podem balizar futuros trabalhos que possuem limitação computacional, como alguns sistemas embarcados ou instituições com menos recursos, ou que tenham o objetivo de aprimorar o uso dos dados do radar, que se mostra um sensor promissor para lidar com condições adversas de iluminação e clima.

Para trabalhos futuros foi possível observar que apesar do tempo de execução ser significativamente reduzido ainda são necessários ajustes para a arquitetura alcançar a execução em tempo real. Outro ponto observado foi a redução nas métricas para dados noturnos, pois apesar da arquitetura desenvolvida conseguir os resultados mais promissores nesse quesito, ainda há uma variação grande em relação aos dados coletados durante o dia.

AGRADECIMENTOS

Este trabalho foi realizado com o apoio do CNPQ, CAPES, Fundep - Mover e InSAC, através da concessão de bolsa parcial de estudos e infraestrutura para pesquisa.

REFERÊNCIAS

Arnold, E., Al-Jarrah, O.Y., Dianati, M., Fallah, S., Oxtoby, D., and Mouzakitis, A. (2019). A survey on 3d object detection methods for autonomous driving applications. *IEEE Transactions on Intelligent Transportation Systems*, 20(10), 3782–3795.

Barnes, D., Gadd, M., Murcutt, P., Newman, P., and Posner, I. (2020). The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 6433–6438. IEEE.

Bijelic, M., Gruber, T., Mannan, F., Kraus, F., Ritter, W., Dietmayer, K., and Heide, F. (2020). Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 11679–11689. doi:10.1109/CVPR42600.2020.01170.

Braun, M., Krebs, S., Flohr, F., and Gavrila, D.M. (2019). Eurocity persons: A novel benchmark for person detection in traffic scenes. *IEEE transactions on pattern analysis and machine intelligence*, 41(8), 1844–1861.

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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11621–11631.

Contributors, M. (2020). MMDetection3D: OpenMMLab next-generation platform for general 3D object detection. `https://github.com/open-mmlab/mmdetection3d`.

Cui, C., Ma, Y., Lu, J., and Wang, Z. (2023). Redformer: Radar enlightens the darkness of camera perception with transformers. *IEEE Transactions on Intelligent Vehicles*.

Fayyad, J., Jaradat, M.A., Gruyer, D., and Najjaran, H. (2020). Deep learning sensor fusion for autonomous vehicle perception and localization: A review. *Sensors*, 20(15), 4220.

Geiger, A., Lenz, P., and Urtasun, R. (2012). Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, 3354–3361. IEEE.

Huang, X., Cheng, X., Geng, Q., Cao, B., Zhou, D., Wang, P., Lin, Y., and Yang, R. (2018). The apolloscape dataset for autonomous driving. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 954–960.

Kim, Y., Shin, J., Kim, S., Lee, I.J., Choi, J.W., and Kum, D. (2023). Crn: Camera radar net for accurate, robust, efficient 3d perception. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 17615–17626.

Lei, K., Chen, Z., Jia, S., and Zhang, X. (2023). Hvdetfusion: A simple and robust camera-radar fusion framework. *arXiv preprint arXiv:2307.11323*.

Li, Z., Wang, W., Li, H., Xie, E., Sima, C., Lu, T., Qiao, Y., and Dai, J. (2022). Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers. In *European conference on computer vision*, 1–18. Springer.

Liu, Z., Tang, H., Amini, A., Yang, X., Mao, H., Rus, D., and Han, S. (2022). Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. *arXiv preprint arXiv:2205.13542*.

Maddern, W., Pascoe, G., Linegar, C., and Newman, P. (2017). 1 year, 1000 km: The oxford robotcar dataset. *The International Journal of Robotics Research*, 36(1), 3–15.

Mehra, A., Mandal, M., Narang, P., and Chamola, V. (2020). Reviewnet: A fast and resource optimized network for enabling safe autonomous driving in hazy weather conditions. *IEEE Transactions on Intelligent Transportation Systems*, 22(7), 4256–4266.

Nabati, R. and Qi, H. (2021). Centerfusion: Center-based radar and camera fusion for 3d object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1527–1536.

Pfeuffer, A. and Dietmayer, K. (2018). Optimal sensor data fusion architecture for object detection in adverse weather conditions. In *2018 21st International Conference on Information Fusion (FUSION)*, 1–8. IEEE.

Pham, Q.H., Sevestre, P., Pahwa, R.S., Zhan, H., Pang, C.H., Chen, Y., Mustafa, A., Chandrasekhar, V., and Lin, J. (2020). A* 3d dataset: Towards autonomous driving in challenging environments. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2267–2273. IEEE.

Sheeny, M., de Pellegrin, E., Mukherjee, S., Ahrabian, A., Wang, S., and Wallace, A. (2021). Radiate: A radar dataset for automotive perception in bad weather. *Proceedings - IEEE International Conference on Robotics and Automation*, 2021-May, 5617–5623. doi:10.1109/ICRA48506.2021.9562089.

Valada, A., Vertens, J., Dhall, A., and Burgard, W. (2017). Adapnet: Adaptive semantic segmentation in adverse environmental conditions. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 4644–4651. doi:10.1109/ICRA.2017.7989540.

Wang, L., Chen, T., Anklam, C., and Goldluecke, B. (2020a). High Dimensional Frustum PointNet for 3D Object Detection from Camera, LiDAR, and Radar. *IEEE Intelligent Vehicles Symposium, Proceedings*, (Iv), 1621–1628. doi:10.1109/IV47402.2020.9304655.

Wang, L., Chen, T., Anklam, C., and Goldluecke, B. (2020b). High dimensional frustum pointnet for 3d object detection from camera, lidar, and radar. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, 1621–1628. IEEE.

Wu, Z., Chen, G., Gan, Y., Wang, L., and Pu, J. (2023). Mvfusion: Multi-view 3d object detection with semantic-aligned radar and camera fusion. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2766–2773. IEEE.

Xu, S., Zhou, D., Fang, J., Yin, J., Bin, Z., and Zhang, L. (2021). Fusionpainting: Multimodal fusion with adaptive attention for 3d object detection. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 3047–3054. IEEE.