

Network combines 3D LiDAR and 2D image data to enable more robust detection of small objects

January 9 2024

Towards Better Automatic 3D Detection of Small Objects

Multi-modal 3D object detectors in autonomous vehicles (AVs) and robots combine LiDAR and 2D image data to obtain better results

However, accurately aligning data from different sensors is challenging, especially for small objects

- ⚠ Complex calibration
- ⚠ Weather conditions
- ⚠ Occlusion

Dynamic Point-Pixel Feature Alignment Network (DPPFA-Net) for small 3D object detection

Deformable Point-Pixel Fusion (DPPF) modules

- ✓ Establish interactions between key high-resolution features
- ✓ Reduce computational complexity

Memory-based Point-Pixel Fusion (MPPF) modules

- ✓ Intra-modal and cross-modal feature interactions based on a memory bank
- ✓ Easier network learning with promotion of key features
- ✓ Robust against noise in LiDAR data

Semantic Alignment Evaluator (SAE) modules

DPPFA-Net achieves a new state-of-the-art for 3D object detection, paving the way for safer and more capable AVs and robots

The proposed model adopts innovative strategies that enable it to accurately combine 3D LiDAR data with 2D images, leading to a significantly better performance than state-of-the-art models for small target detection, even under adverse weather conditions. Credit: Hiroyuki Tomiyama, Ritsumeikan University

Robotics and autonomous vehicles are among the most rapidly growing domains in the technological landscape, potentially making work and transportation safer and more efficient. Since both robots and self-driving cars need to perceive their surroundings accurately, 3D object detection methods are an active study area.

Most 3D object detection methods employ LiDAR sensors to create 3D point clouds of their environment. Simply put, LiDAR sensors use laser beams to rapidly scan and measure the distances of objects and surfaces around the source. However, using LiDAR data alone can lead to errors due to the high sensitivity of LiDAR to noise, especially in adverse weather conditions like during rainfall.

To tackle this issue, scientists have developed multi-modal 3D object detection methods that combine 3D LiDAR data with 2D RGB images taken by standard cameras. While the fusion of 2D images and 3D LiDAR data leads to more accurate 3D detection results, it still faces its own set of challenges, with accurate detection of small objects remaining difficult.

The problem mainly lies in adequately aligning the semantic information extracted independently from the 2D and 3D datasets, which is hard due to issues such as imprecise calibration or occlusion.

Against this backdrop, a research team led by Professor Hiroyuki Tomiyama from Ritsumeikan University, Japan, has developed an innovative approach to make multi-modal 3D object detection more accurate and robust. The proposed scheme, called "Dynamic Point-Pixel Feature Alignment Network" (DPPFA–Net), is described in their paper [published in *IEEE Internet of Things Journal*](#).

The model comprises an arrangement of multiple instances of three novel modules: the Memory-based Point-Pixel Fusion (MPPF) module, the Deformable Point-Pixel Fusion (DPPF) module, and the Semantic Alignment Evaluator (SAE) module.

The MPPF module is tasked with performing explicit interactions between intra-modal features (2D with 2D and 3D with 3D) and cross-modal features (2D with 3D). The use of the 2D image as a memory bank reduces the difficulty in network learning and makes the system more robust against noise in D point clouds. Moreover, it promotes the use of more comprehensive and discriminative features.

In contrast, the DPPF module performs interactions only at pixels in key positions, which are determined via a smart sampling strategy. This allows for feature fusion in high resolutions at a low computational complexity. Finally, the SAE module helps ensure semantic alignment between both data representations during the fusion process, which mitigates the issue of feature ambiguity.

The researchers tested DPPFA–Net by comparing it to the top performers for the widely used KITTI Vision Benchmark. Notably, the proposed network achieved average precision improvements as high as 7.18% under different noise conditions. To further test their model's capabilities, the team created a new noisy dataset by introducing artificial multi-modal noise in the form of rainfall to the KITTI dataset.

The results show that the proposed network performed better than existing models not only in the face of severe occlusions but also under various levels of adverse weather conditions. "Our extensive experiments on the KITTI dataset and challenging multi-modal noisy cases reveal that DPPFA-Net reaches a new state-of-the-art," says Prof. Tomiyama.

Notably, there are various ways in which accurate 3D object detection methods could improve our lives. Self-driving cars, which rely on such techniques, have the potential to reduce accidents and improve traffic flow and safety. Furthermore, the implications in the field of robotics should not be understated. "Our study could facilitate a better understanding and adaptation of robots to their working environments, allowing a more precise perception of small targets," explains Prof. Tomiyama.

"Such advancements will help improve the capabilities of robots in various applications." Another use for 3D object detection networks is pre-labeling raw data for deep-learning perception systems. This would significantly reduce the cost of manual annotation, accelerating developments in the field.

More information: Juncheng Wang et al, Dynamic Point-Pixel Feature Alignment for Multi-modal 3D Object Detection, *IEEE Internet of Things Journal* (2023). [DOI: 10.1109/JIOT.2023.3329884](https://doi.org/10.1109/JIOT.2023.3329884)

Provided by Ritsumeikan University

Citation: Network combines 3D LiDAR and 2D image data to enable more robust detection of small objects (2024, January 9) retrieved 29 April 2024 from <https://techxplore.com/news/2024-01-network-combines-3d-lidar-2d.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.