

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG
KHOA CÔNG NGHỆ THÔNG TIN I



BÁO CÁO ĐỀ CƯƠNG GIỮA KỲ
HỌC PHẦN: THỰC TẬP CƠ SỞ
ĐỀ TÀI: 3D PAGNOTIC SEGMENTATION
CHO PHÂN TÍCH CẢNH TỪ DỮ LIỆU LiDAR

Giảng viên hướng dẫn	: TS. Kim Ngọc Bach
Mã sinh viên	: B23DCCN397
Tên sinh viên	: Nguyễn Nhật Huy
Lớp hành chính	: D23CQCN05-B
Nhóm thực tập	: 11

Hà Nội – 2026

I. GIỚI THIỆU DỰ ÁN

1.1. Lý do chọn đề tài

Trong những năm gần đây, cùng với sự phát triển mạnh mẽ của trí tuệ nhân tạo và hệ thống xe tự hành, việc phân tích và hiểu môi trường xung quanh từ dữ liệu cảm biến đã trở thành một bài toán quan trọng trong lĩnh vực thị giác máy tính và robot học.

Trong các loại cảm biến, LiDAR (Light Detection and Ranging) được sử dụng rộng rãi nhờ khả năng thu thập dữ liệu không gian 3D chính xác trong điều kiện ánh sáng khác nhau. Dữ liệu LiDAR được biểu diễn dưới dạng đám mây điểm (point cloud), chứa thông tin hình học của môi trường.

Tuy nhiên, để hệ thống có thể hiểu cảnh một cách đầy đủ, không chỉ cần phân loại ngữ nghĩa (semantic segmentation), mà còn cần phân biệt từng thực thể riêng biệt trong cùng một lớp (instance-level). Ví dụ:

- Không chỉ biết đó là “Car”
- Mà phải biết có bao nhiêu xe và mỗi xe là một thực thể riêng

Vì vậy, bài toán **3D Panoptic Segmentation** ra đời nhằm kết hợp cả hai nhiệm vụ trên thành một mô hình thống nhất.

Đề tài này có ý nghĩa thực tiễn và học thuật cao, phù hợp với định hướng nghiên cứu trong lĩnh vực:

- Thị giác máy tính 3D
- Xe tự hành
- Robot tự hành

1.2. Ý nghĩa của đề tài

Đề tài giúp:

- Hiểu cấu trúc cảnh 3D từ dữ liệu LiDAR
- Phân loại chính xác các lớp ngữ nghĩa
- Phân biệt các thực thể riêng biệt trong cùng một lớp
- Tạo nền tảng cho các bài toán cao hơn như tracking hoặc dự đoán quỹ đạo

1.3. Tính ứng dụng

Hệ thống có thể áp dụng trong:

- Xe tự hành

- Robot tự hành trong nhà máy
- Giám sát giao thông thông minh
- Hệ thống ADAS(Advanced Driver Assistance Systems)

1.4. Giá trị học thuật

Đề tài tích hợp nhiều lĩnh vực:

- 3D Deep Learning
- Xử lý point cloud
- Semantic Segmentation
- Instance Segmentation
- Scene Understanding

Đây là một bài toán nâng cao, yêu cầu hiểu rõ cả mô hình học sâu và đặc trưng hình học 3D.

II. Cơ sở lý thuyết và công nghệ sử dụng

2.1. 3D Segmantic Segmentation

2.1.1. Khái niệm

3D Semantic Segmentation là bài toán gán nhãn ngữ nghĩa cho từng điểm (point-wise classification) trong đám mây điểm (point cloud) thu được từ cảm biến LiDAR.

Mỗi điểm $p_i=(x,y,z,intensity)$ sẽ được gán một nhãn lớp:

- Car
- Pedestrian
- Cyclist
- Road
- Building
- Vegetation
- Pole
- Traffic Sign

Khác với ảnh RGB 2D có cấu trúc lưới đều, point cloud là dữ liệu:

- Không có cấu trúc cố định

- Phân bố không đồng đều
- Mật độ giảm theo khoảng cách
- Có hiện tượng che khuất (occlusion)

Do đó, 3D semantic segmentation phức tạp hơn so với 2D segmentation.

2.1.2. Các cách biểu diễn dữ liệu

* Point-based method

Xử lý trực tiếp từng điểm (ví dụ: PointNet, PointNet++)

Ưu điểm:

- Giữ nguyên cấu trúc 3D
- Không mất thông tin

Nhược điểm:

- Tính toán lớn
- Khó tối ưu cho real-time

* Voxel-based method

Chuyển point cloud thành lưới voxel 3D.

Ưu điểm:

- Áp dụng được 3D CNN

Nhược điểm:

- Tốn bộ nhớ
- Mất chi tiết nếu voxel lớn

* Range-view projection (phổ biến trong xe tự hành)

Chiếu point cloud lên ảnh 2D theo góc quét LiDAR.

Biểu diễn thành ma trận:

$\text{RangeImage} \in \mathbb{R}(H \times W \times C)$

Trong đó:

- H: height
- W: width

- C: channel (range, intensity, x, y, z...)

Phương pháp này cho phép dùng CNN 2D hiệu quả và đạt tốc độ cao.

2.2. Instance Segmentation

2.2.1. Khái niệm

Instance Segmentation là bài toán:

Phân biệt các thực thể riêng biệt trong cùng một lớp ngữ nghĩa.

Ví dụ:

Semantic segmentation chỉ cho biết:

- Car

Instance segmentation phải cho biết:

- Car ID 1
- Car ID 2
- Car ID 3

Phân biệt các thực thể riêng biệt trong cùng một lớp.

Ví dụ:

- Car 1
- Car 2
- Car 3

Các phương pháp:

- Clustering theo embedding
- Center-based detection

2.2.2. Các hướng tiếp cận

* Clustering-based

Mô hình học embedding vector cho mỗi điểm:

$$e_i \in \mathbb{R}^d$$

Các điểm thuộc cùng instance sẽ có embedding gần nhau trong không gian đặc trưng.

Sau đó áp dụng:

- DBSCAN
- MeanShift
- Connected Components

Ưu điểm:

- Linh hoạt
- Không cần xác định trước số instance

Nhược điểm:

- Phụ thuộc tham số clustering
- Tốn thời gian xử lý hậu kỳ

* Center-based / Proposal-based

Dự đoán:

- Tâm của object
- Vector offset từ điểm về tâm

Ý tưởng tương tự:

- Mask R-CNN (trong 2D)

Ưu điểm:

- Ổn định hơn clustering
- Dễ mở rộng sang tracking

2.3. Panoptic Segmentation

2.3.1. Khái niệm

Panoptic Segmentation là phương pháp kết hợp giữa:

- **Semantic Segmentation** (phân loại ngữ nghĩa từng điểm)
- **Instance Segmentation** (phân biệt từng thực thể riêng biệt)

Khái niệm Panoptic Segmentation được đề xuất năm 2019 bởi nhóm nghiên cứu tại Facebook AI Research và University of California, Berkeley.

Mục tiêu của Panoptic Segmentation là tạo ra một đầu ra thống nhất, trong đó **mỗi điểm trong point cloud có đầy đủ cả nhãn ngữ nghĩa và ID thực thể**.

2.3.2. Các giải pháp

- Kết hợp semantic segmentation + instance segmentation để giải quyết vấn đề

- Tạo model để có thể học luôn pagnotic segmentation

2.4 Công nghệ sử dụng

2.4.1. Thư viện Deep Learning: PyTorch

PyTorch là framework chính được sử dụng để xây dựng và huấn luyện mô hình.

Vai trò trong đề tài:

- Xây dựng kiến trúc mạng Semantic Segmentation và Instance Segmentation.
- Huấn luyện mô hình trên GPU.
- Tối ưu bằng backpropagation.
- Triển khai inference trên dữ liệu LiDAR.

Ưu điểm:

- Linh hoạt, dễ tùy biến kiến trúc.
- Cộng đồng mạnh.
- Phù hợp nghiên cứu học thuật.

2.4.2. Thư viện xử lý dữ liệu số: NumPy

NumPy được sử dụng để:

- Xử lý mảng điểm (x, y, z, intensity).
- Tính toán khoảng cách Euclidean trong clustering.
- Thao tác tensor trước khi đưa vào mạng.

NumPy giúp tăng tốc các phép toán tuyến tính và xử lý dữ liệu lớn.

2.4.3. Thư viện xử lý và hiển thị Point Cloud: Open3D

Open3D được sử dụng để:

- Đọc file .bin hoặc .pcd từ LiDAR.
- Hiển thị point cloud 3D.
- Gán màu theo semantic label hoặc instance ID.
- Kiểm tra trực quan kết quả segmentation.

Open3D hỗ trợ trực quan hóa 3D hiệu quả và thuận tiện cho phân tích kết quả.

2.4.4. Thư viện trực quan hóa: Matplotlib

Matplotlib được sử dụng để:

- Vẽ biểu đồ loss theo epoch.
- Vẽ biểu đồ so sánh mIoU, PQ.
- Phân tích kết quả huấn luyện.

2.4.5. Bộ dữ liệu sử dụng

Đề tài sử dụng bộ dữ liệu:

SemanticKITTI

Đặc điểm:

- Thu thập từ xe tự hành.
- Dữ liệu LiDAR 3D dạng point cloud.
- Có nhãn semantic và instance.
- Phù hợp cho bài toán 3D Panoptic Segmentation trên single frame.

III. Phân tích yêu cầu dự án

3.1. Yêu cầu chức năng

Hệ thống 3D Panoptic Segmentation phải:

- Đọc và xử lý dữ liệu point cloud từ file LiDAR (.bin hoặc .pcd).
- Thực hiện 3D Semantic Segmentation (gán nhãn ngữ nghĩa cho từng điểm).
- Thực hiện 3D Instance Segmentation (phân biệt các thực thể riêng biệt).
- Kết hợp hai đầu ra thành Panoptic Output:
 - Semantic label
 - Instance ID
- Trực quan hóa kết quả segmentation trên point cloud.
- Tính toán các chỉ số đánh giá:
 - mIoU (semantic)
 - PQ (Panoptic Quality)
- Lưu kết quả dự đoán ra file để đánh giá hoặc so sánh.

3.2. Yêu cầu phi chức năng

- Độ chính xác đạt mức chấp nhận được trên bộ dữ liệu chuẩn.
- Hệ thống ổn định khi xử lý point cloud kích thước lớn.

- Có khả năng mở rộng sang 4D Panoptic trong tương lai.
- Tổ chức mã nguồn rõ ràng, dễ bảo trì.
- Thời gian suy luận (inference) hợp lý cho một frame LiDAR.

3.3. Thách thức

- Dữ liệu LiDAR thưa và không đều.
- Mật độ điểm giảm theo khoảng cách.
- Hiện tượng occlusion (che khuất).
- Khó phân biệt các instance gần nhau.
- Class imbalance (ví dụ: road rất nhiều điểm, pedestrian rất ít).
- Tối ưu bộ nhớ khi xử lý point cloud lớn.

IV. Kế hoạch thực hiện dự án

Tuần 1: Nghiên cứu lý thuyết

- 3D Semantic Segmentation
- Instance Segmentation
- Panoptic Segmentation
- Cấu trúc point cloud

Tuần 2: Tìm hiểu bộ dữ liệu

- Cấu trúc nhãn của SemanticKITTI
- Phân tích semantic label và instance ID
- Viết code đọc file .bin và label

Tuần 3: Xây dựng module Semantic Segmentation

- Chọn kiến trúc (Range-view hoặc point-based)
- Triển khai mạng cơ bản
- Huấn luyện thử nghiệm

Tuần 4: Xây dựng module Instance Segmentation

- Thiết kế embedding head hoặc center-based head
- Triển khai clustering

Tuần 5: Kết hợp thành Panoptic

- Ghép semantic + instance
- Xử lý conflict

- Sinh panoptic ID

Tuần 6: Huấn luyện và tối ưu

- Điều chỉnh hyperparameter
- Giảm overfitting
- Tăng mIoU và PQ

Tuần 7: Xây dựng module đánh giá

- Tính mIoU
- Tính PQ
- So sánh với baseline

Tuần 8: Kiểm thử hệ thống

- Kiểm tra độ ổn định
- Visualize kết quả
- Phân tích lỗi

Tuần 9: Tìm hiểu, mở rộng thêm các chức năng cho đề tài

Tuần 10: Hoàn thiện báo cáo và chuẩn bị demo

- Hoàn thiện tài liệu
- Chuẩn bị slide
- Demo mô hình

Tài liệu tham khảo

- [1] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [2] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- [3] B. Wu, A. Wan, X. Yue, and K. Keutzer, "SqueezeSeg: Convolutional Neural Nets with Recurrent CRF for Real-Time Road-Object Segmentation from 3D LiDAR Point Cloud," *IEEE Int. Conf. Robotics and Automation (ICRA)*, 2018.
- [4] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "RangeNet++: Fast and Accurate LiDAR Semantic Segmentation," *IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, 2019.

- [5] T. Cortinhal, G. Tzelepis, and E. Erdal Aksoy, “SalsaNext: Fast, Uncertainty-Aware Semantic Segmentation of LiDAR Point Clouds,” *Proc. European Conf. Computer Vision (ECCV)*, 2020.
- [6] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017.
- [7] J. Kirillov et al., “Panoptic Segmentation,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [8] J. Behley et al., “SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences,” *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2019.