



3차원 포인트 클라우드 데이터를 활용한 객체 탐지 기법인 PointNet과 RandLA-Net

이동건¹ · 지승환^{2,3,†} · 박본영²
목포해양대학교 조선해양공학과¹
목포해양대학교 대학원 해양시스템공학과²
현대삼호중공업 정보기술부³

PointNet and RandLA-Net Algorithms for Object Detection using 3D Point Clouds

Dong-Kun Lee¹ · Seung-Hwan Ji^{2,3,†} · Bon-Yeong Park²
Department of Naval Architecture and Ocean Engineering, Mokpo National Maritime University¹
Department of Ocean System Engineering, Graduate School, Mokpo National Maritime University²
Department of Information Technology, Hyundai Samho Heavy Industries Co., Ltd.³

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Research on object detection algorithms using 2D data has already progressed to the level of commercialization and is being applied to various manufacturing industries. Object detection technology using 2D data has an effective advantage, there are technical limitations to accurate data generation and analysis. Since 2D data is two-axis data without a sense of depth, ambiguity arises when approached from a practical point of view. Advanced countries such as the United States are leading 3D data collection and research using 3D laser scanners. Existing processing and detection algorithms such as ICP and RANSAC show high accuracy, but are used as a processing speed problem in the processing of large-scale point cloud data. In this study, PointNet a representative technique for detecting objects using widely used 3D point cloud data is analyzed and described. And RandLA-Net, which overcomes the limitations of PointNet's performance and object prediction accuracy, is described a review of detection technology using point cloud data was conducted.

Keywords : Deep learning(딥러닝), Point clouds(포인트 클라우드), Object detection(객체 탐지), PointNet(포인트넷), RandLA-Net(랜들라넷)

1. 서론

2D 데이터(픽셀 형태의 이미지 & 영상)를 활용한 객체 인식 알고리즘 연구는 여러 제조 산업에서 상용화하여 적용하는 수준 까지 발전했지만, 2D 데이터를 활용한 객체 인식의 한계는 활용 데이터의 근본에서 찾아볼 수 있다. 2D 데이터는 x축과 y축으로만 구성된 평면 데이터로서 깊이감(z축)이 부재하기 때문에 실제적인 관점에서 접근하였을 때 2D 데이터를 활용한 객체 인식은 모호성을 가질 수 있다.

이미 미국 등 IT선진국에서는 자율주행자동차 등의 LiDAR나 3D 레이저 스캐너로 수집한 3D 데이터를 활용한 연구를 선도적으로 수행하고 있었고, 아직 정립되지 않은 알고리즘 학습에 관한 연구를 학문적, 실무적인 접근을 통해 오픈소스 형상관리 공

간(Github)에서 개발자들끼리 기술을 공유해오며 발전해왔다. 최근 조선해양분야에서도 인공지능 기반 균열 탐지를 현장에 적용하기 위한 효율성 연구 (Song et al., 2022a)와 3D 포인트 클라우드 데이터 기반 변형 탐지 알고리즘을 통해 선체 구조 변형 영역의 손상 정도를 확인한 연구 (Song et al., 2022b)가 수행된 바 있다. 이처럼 객체 인식은 자율운항선박, 스마트조선소 공정관리 등 향후 활용 여지가 높은 기술로 판단된다.

본 논문에서는 활용도가 높아지고 있는 포인트 클라우드 데이터 처리를 위한 기법을 설명하고자 한다. 이에 3D 스캔 데이터를 활용한 객체 인식 알고리즘 학습을 획기적인 방법으로 시도하여 관련 분야에서 대표 기법으로 인정받고 있는 PointNet과 함께, 대규모 3D 데이터 활용이 불가능하고, 부족한 예측 정확도를 보이는 PointNet의 한계를 극복한 기법인 RandLA-Net을 분석하였다.

2. PointNet의 딥러닝 아키텍처

2.1 순서 불변 & 변환 불변

PointNet 기법은 스캐닝 포인트 클라우드 데이터만을 활용한 대표적인 객체 탐지 기법으로 기존 연구에서 활용되었던 2D 데이터(이미지, 영상)를 활용하지 않는다는 점에서 차이가 있다. 2D 데이터는 픽셀 단위로 구성되어 있고 픽셀은 가로, 세로 데이터 영역의 크기가 동일하기 때문에 2D 데이터에 대한 컨볼루션 연산이 쉽게 이루어져 CNN에 적용하는 데 문제가 없었지만 입력 데이터가 스캐닝 포인트 클라우드 데이터로 확장되면서 문제가 제기되었다.

3D 스캐너를 통해 데이터를 수집할 때에는 스캐너로부터 객체 데이터까지의 거리가 멀어질수록 수집되는 데이터의 양이 현저하게 줄어들고 데이터를 수집하는 순서가 순차적으로 이루어지지 않다는 특징이 있다. 이러한 특징들로 인해 어떤 위치의 점(point)이 어떤 객체를 나타내는지, 라벨 정보를 몇 번째 점 데이터에 부여하고 이들을 입력 데이터로 활용하여 컨볼루션 연산을 수행해야 하는지에 관한 연구는 곧바로 이루어지지 않았다.

이후, 포인트 클라우드 데이터를 3D 복셀 그리드로 변환시킨 뒤 복셀 그리드를 컨볼루션 연산하는 기법들이 개발되었으나 이는 입력 데이터를 가공한다는 점에서 포인트 클라우드 데이터 자체 데이터를 활용한다는 것과 거리감이 있었다.

하지만 딥러닝 기반 3D 객체 탐지 기법인 PointNet(Fig. 1)이 등장하면서 스캐닝 포인트 클라우드 자체를 활용한 컨볼루션 연산이 가능함을 입증하였는데 PointNet은 크게 순서 불변(permutation invariance), 변환 불변(transformation invariance)을 따른다. 이 두 가지 법칙은 포인트 클라우드를 그대로 활용하기 위한 전제조건으로 우선, 3D 스캐너를 통해 수집된 데이터는 Fig. 2와 같은 형태로 수집된다. 좌측 인덱스(index)는 점 각각의 순번을 나타내고 우측에는 그와 대응되는 점 좌표를 나타낸다.

통상 3D 스캐너로 계측된 포인트 클라우드는 무수한 점 데이터의 집단으로, 점 데이터 각각은 번호(순서)를 나타내는 인덱스(index)와 해당 인덱스에 대응하는 좌표 데이터(x, y, z)가 주어진다. 인덱스에 따른 데이터는 그 자체로 고유한 특성을 가지게 되

는데, 이러한 집단 데이터가 심층 신경망(deep neural network)을 통하면서 서로 뒤섞인다면 그 의미는 없어진다. 순서 불변은 심층 신경망 학습 도중 데이터가 뒤섞이더라도 인덱스와 그에 대응하는 좌표 데이터의 고유적 특성을 보존하는 법칙이고 PointNet은 이를 따른다.

이미지의 문자를 식별하는 딥러닝 기법은 어떠한 방향에서 쓰여진 문자도 올바르게 예측하였다. 그것이 가능한 이유는 공간 변형 네트워크(STN, Spatial Transformer Networks) 때문인데 이는, 입력 데이터의 좌표축을 변환시키는 변환 행렬을 곱해 어떤 방향에서 쓰여진 문자들도 한 방향에서 바라보는 효과를 보이는 네트워크이다 (Jaderberg et al., 2015).

PointNet은 공간 변형 네트워크를 변형시켜 3D 입력 데이터인 포인트 클라우드에 적용하였으며 이를 ‘Transformation-Network (T-Net)’이라 칭하였다 (Qi et al., 2017). PointNet의 T-Net은 Fig. 3과 같이 입력 데이터의 차원을 확장시킨 뒤 최대 풀링(max pooling)하고, 차원을 축소시켜 모든 포인트 클라우드를 동일 방향에서 바라보고 식별하기 위한 변환 행렬을 도출한다. 이 과정이 T-Net이며 도출된 변환 행렬을 입력 데이터에 곱해 줌으로써

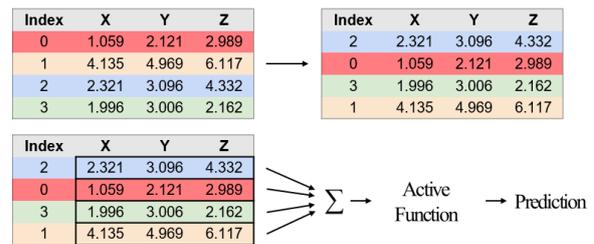


Fig. 2 Permutation invariance

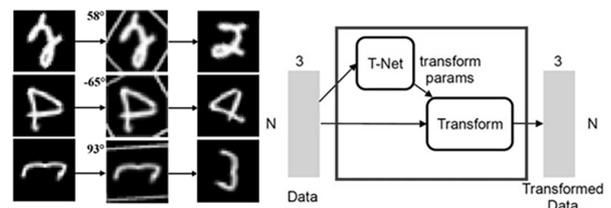


Fig. 3 Data Transformation network about PointNet (Qi et al., 2017)

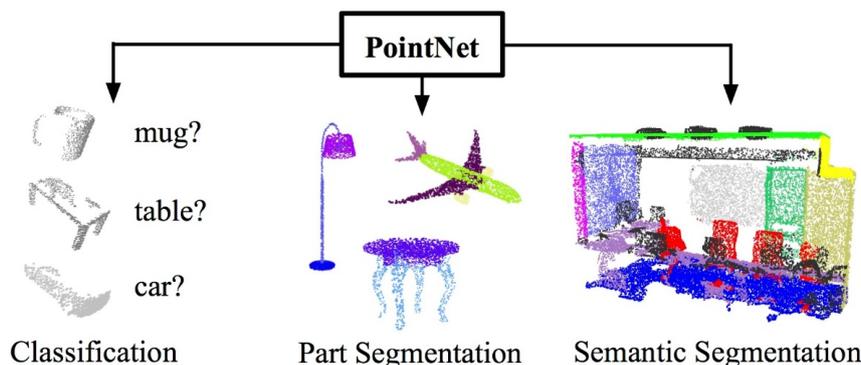


Fig. 1 PointNet architecture for 3D object detection (Qi et al., 2017)

포인트 클라우드가 각기 다른 좌표축을 기준으로 생성된 형상 데이터일지라도 기준 좌표축을 하나로 통일시켜 한 방향에서 바라보도록 변환시켜준다. 따라서 포인트 클라우드가 3축 좌표계를 기준으로 이동(move)되거나 회전(rotation)된 상태일지라도 동일한 형상을 가진 물체나 공간이라면 T-Net에서 그 특징들을 찾아내 동일한 물체나 공간이라는 것을 식별할 수 있게 된다.

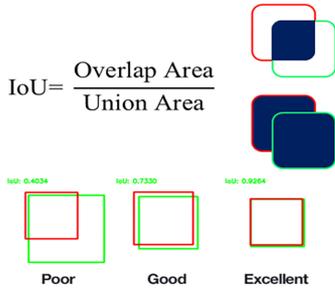


Fig. 4 Performance with intersection over union (Lee et al., 2021)

2.2 미니배치, 에포크 및 IoU

배치 크기(batch size)와 에포크(epoch)는 개발자가 설정할 수 있는 파라미터(parameter)로 두 개념은 컴퓨팅 시스템 하드웨어 한 계로 인해 만들어졌다. 딥러닝은 초기 설정한 배치 크기를 기반으로 가중치와 편향으로 인한 손실을 줄여나가는 경사 하강법 (gradient descent)을 따르고, 업데이트하는 역전파 알고리즘 (back-propagation algorithm)으로 수행되는데, 배치 크기가 너무 작으면 딥러닝 모델은 학습검증 데이터에 과대적합(overfitting)될 수 있고, 배치 크기가 너무 크면 과소적합(underfitting)될 가능성이 크다. 적절한 배치 크기와 에포크 설정은 모델 학습 시 중요한 요소로서 보통 연구 초반에는 해당 연구 도메인에서 자주 사용하는 배치 크기를 기준으로 다양한 실험을 통해 최적의 값을 탐색한다.

IoU(Intersection over union)는 경계 상자가 의미하는 객체를 정확히 판단하기 위한 지표(Fig. 4)로서 기준에 생성한 경계 상자로부터 예측된 상자가 동일하다면 IoU는 100%를 나타낸다. 일반적으로 IoU가 50% 이상이면 ‘동일 객체’라고 판단한다.

2.3 PointNet 기법의 객체 탐지

PointNet 기법의 분류(classification) 아키텍처는 x, y, z개의 차원을 가진 데이터가 입력되고 이들을 라벨링한 뒤, T-Net, MLP(Multi-Layer Perceptron) 반복, 최대 풀링, 다시 MLP, 예측(prediction) 순으로 진행된다(Fig. 5). 이는 입력 데이터에 대해 하나의 라벨만을 예측할 수 있기 때문에 분할(segmentation)을 통해 점 각각에 대한 라벨 예측을 수행할 수 있다.

분할 아키텍처는 분류 아키텍처와 크게 다르지 않지만, 분류 아키텍처에서 최대 풀링 및 글로벌 레지스트레이션이 수행된 데이터셋을 가져옴으로써 차이가 발생한다. 글로벌 레지스트레이션이 수행되면 딥러닝 아키텍처에 입력되는 각각의 점 데이터들을 하나로 합치게 되는데 이때 점 데이터들이란, 포인트 각각에 대한 라벨 정보를 가진 데이터들을 뜻한다. 글로벌 레지스트레이션이 수행된 데이터를 기존의 T-Net까지 수행된 데이터셋에 붙여 줌으로써 순서 불변, 변환 불변이 적용된 스캐닝 데이터의 점 각각에 대한 라벨 정보를 대치시키는 것과 같다. 이로써 학습하고자 하는 모든 점 데이터와 대치된 라벨 정보를 학습하고 객체 탐지를 수행하기 위한 입력 데이터가 완성된다.

포인트 클라우드에 대한 PointNet의 제약된 처리능력을 고려하여 학습검증 데이터를 결정하였는데, 2,048개 이하의 데이터를 학습시키는 것이 하드웨어적 안정성과 예측 정확성을 동시에 가져갈 수 있음을 알 수 있다 (Fig. 6).

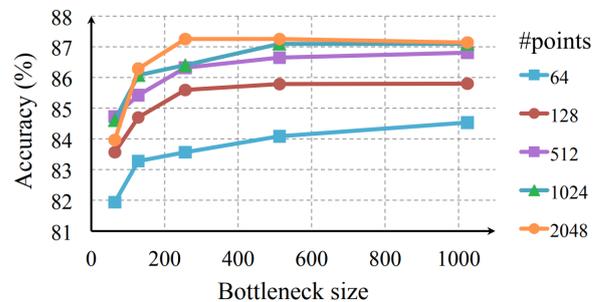


Fig. 6 Effects of bottleneck size and number of input points (Qi et al., 2017)

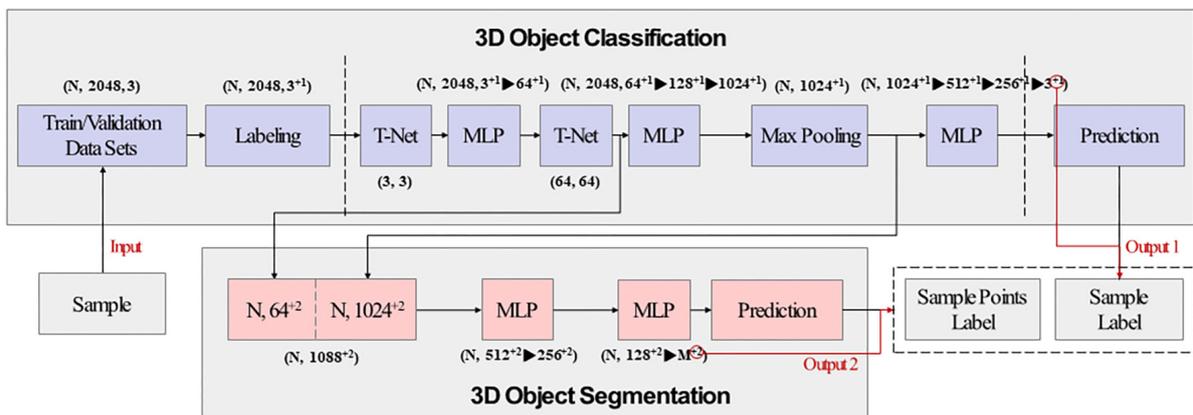


Fig. 5 PointNet architecture for 3D object detection (Lee et al., 2021)

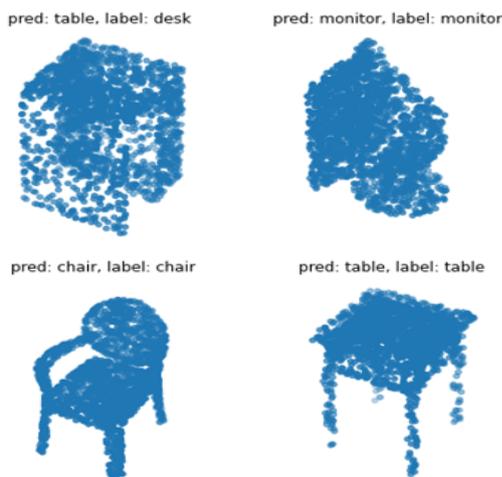
3. PointNet 기법을 활용한 객체 탐지

3.1 분류 예측

PointNet 기법을 적용한 분류 예측 결과는 Table 1과 같이 나타나고, 각 형상을 가진 3,000개 미만의 데이터들에 대해 한 가지 라벨을 예측한다. PointNet의 논문에서 수록된 자료를 살펴보면 타 알고리즘 대비 뛰어난 성능을 나타낸다는 것을 정량적인 지표 비교를 통해 증명하였다.

Table 1 Quantitative results for ModelNet40 of various classification methods (Qi et al., 2017)

Methods	Input	#Views	Accuracy avg. class	Accuracy overall
3DCNN	Point	-	72.6	77.4
PointNet	Point	1	86.2	89.2



Qualitative results for 3D classification of PointNet

3.2 분할 예측

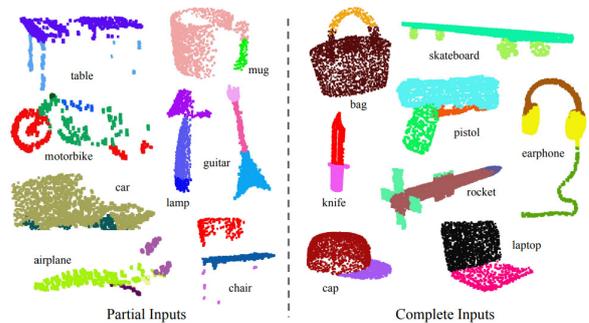
다음은 PointNet 기법을 적용한 요소 분할(part segmentation) 과 의미론적 분할(semantic segmentation) 예측 결과를 나타낸다 (Table 2; Table 3). 요소 분할은 미리 정의된 클래스(테이블, 의자, 머그잔 등)로 사전에 라벨링된 데이터들을 학습하여 예측 정확도를 평가하기 위함이고, 의미론적 분할은 같은 클래스에 속하는 객체를 구별하기 보다는 데이터 자체가 어떤 클래스에 속하는지에 중점을 두고 수행된 예측이다. 의미론적 분할에 활용된 데이터는 스탠퍼드 대학의 모든 강의실을 스캔한 대규모 데이터 (S3DIS)로써, PointNet은 이러한 대량의 데이터셋을 활용해야 하는데 모든 데이터를 3차원 그리드 공간에서 각 그리드마다 포인트 클라우드 데이터를 2,048개씩 균일하게 쪼개어 각각의 소규모 데이터들을 하나씩 학습하고 각각을 예측, 종합하여 예측 모델을 생성하고 예측 정확도를 평가하였다.

3.3 RandLA-Net의 대규모 포인트 클라우드 데이터를 활용하기 위한 방법

RandLA-Net 기법은 대량의 로우 데이터를 활용하면서 높은 정확도까지 보유할 수 있도록 연구되어 개발된 알고리즘이다. PointNet을 기점으로, 3D 객체 탐지를 위한 딥러닝 연구는 활발히 진행됐으나 대부분 기법은 하드웨어적인 성능 한계(GPU 성능, 메모리 한계 등)가 있었다. 대규모 데이터 처리를 위해 1×1 세제곱미터의 3차원 그리드로 쪼개는 복셀화 과정(voxelization)을

Table 2 PointNet and 3DCNN set as baseline (Metric is mIoU[%] on points) (Qi et al., 2017)

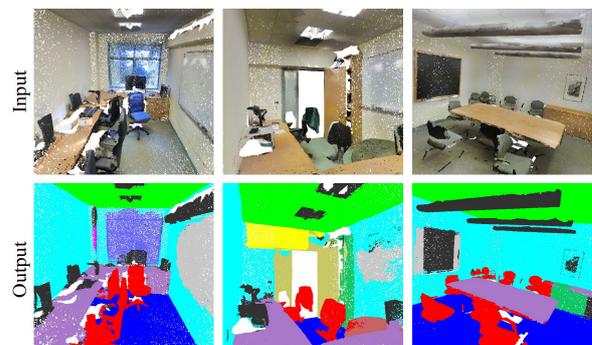
Methods	mean	laptop	mug	chair	knife	aero
3DCNN	79.4	93.9	91.8	87.2	79.6	75.1
PointNet	83.7	95.3	93.0	89.6	85.9	83.4



Qualitative results for 3D classification

Table 3 Comparison between PointNet and 3DCNN set as baseline (Metric is mIoU[%] on points) (Qi et al., 2017)

Methods	mean IoU (%)	Overall accuracy
3DCNN	20.12	53.19
PointNet	47.71	78.62



Qualitative results for semantic segmentation

거치고, 복셀 그리드 각각에 내재한 데이터들을 학습하여 예측 모델을 생성하여 그리드들을 종합하는 과정이 거의 필수적이였다. 이 방법은 하드웨어적 성능을 고려하면서 대규모 데이터를 학습하기 위한 차선책으로 많이 활용되었지만, 서브샘플링(sub sampling) 과정에서 데이터의 기하학적 손실이 발생하였고 학습에 활용되는 데이터 공간을 좁게였기 때문에 공간 전체의 구조를 파악하는 결과를 산출하기 어렵다는 단점이 있었다. RandLA-Net은 이러한 단점을 극복하기 위해 랜덤 샘플링을 적용하였는데, 이는 학습에 활용되는 입력 데이터 양에 구애받지 않아 계산 효율이 뛰어나고, 계산을 위한 GPU 메모리 한계치가 낮다는 장점이 있다.

4. RandLA-Net의 딥러닝 아키텍처

4.1 3D 객체 탐지 기법을 위한 샘플링 메소드

최대 수백 미터의 공간에서 수집된 대규모 포인트 클라우드 데이터는 높은 예측 성능을 위해 기하학적으로 의미 있는 포인트 간의 특성을 유지하면서 각 신경 계층마다 점진적, 효율적으로 다운샘플링 되어야 한다. 다운샘플링 방법은 크게 휴리스틱 샘플링(heuristic sampling), 학습 기반 샘플링(learning-based sampling)으로 구분할 수 있고, 휴리스틱 샘플링에는 최대 지점 샘플링(FPS, Farthest Point Sampling), 역밀도 중요도 샘플링(IDIS, Inverse Density Importance Sampling), 랜덤 샘플링(RS, Random Sampling), 학습 기반 샘플링에는 생성기 기반 샘플링(GS, Generator-based Sampling), 연속 이완 샘플링(CRS, Continuous Relaxation based Sampling), 정책 기울기 기반 샘플링(PGS, Policy Gradient based Sampling)이 있다.

2~3장에 설명된 PointNet의 경우 휴리스틱 샘플링의 최대 지점 샘플링(FPS, Farthest Point Sampling)을 적용하였다.

RandLA-Net 논문에서 수록된 자료(Fig. 7)에 따르면 FPS는 대규모 데이터(N ~ 106)를 처리할 때 단일 GPU에서 최대 200초 소요되고, IDIS는 경험적으로 100만 포인트 클라우드 데이터를

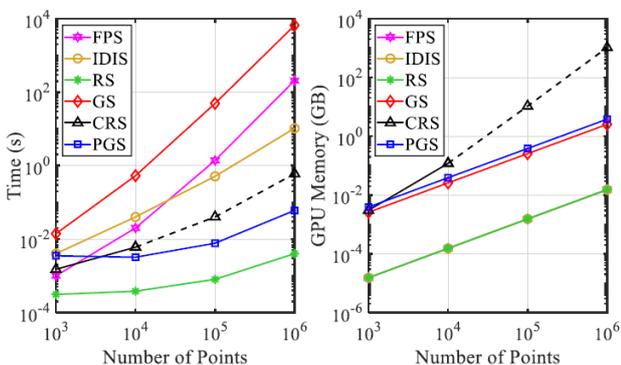


Fig. 7 Time and memory consumption of different sampling approaches (Hu et al., 2019)

처리할 때 10초 소요되어 FPS에 비해 효율적이지만 이상치(outlier)에 민감하고 실시간 시스템에 사용하기엔 여전히 미흡하다는 단점이 있다. RS는 가장 뛰어난 계산 효율을 나타내는데, 이는 약 100만개의 데이터를 처리할 때 0.04초가 소요된다. 하지만 RS는 데이터를 무작위로 샘플링하기 때문에 그 과정에서 기하학적으로 중요한 위치의 데이터를 가져오지 않을 수 있는데 이러한 단점을 보완하기 위해 로컬 특징 집계 모듈(LFAM, Local Feature Aggregation Module)을 도입하였고, 이는 4.2절에서 설명할 예정이다.

GS는 10만개의 포인트 데이터를 처리하는데 최대 1,200초가량 소요되고, CRS는 10만개의 포인트 데이터를 처리하는데 300GB 이상의 메모리 공간이 필요한 것으로 추정하였다. PGS는 대규모 데이터에 활용될 경우 네트워크를 수행하기 어렵다는 것을 경험적으로 발견하였기 때문에 학습 기반 샘플링 방법들은 대체로 3D 객체 탐지 딥러닝 알고리즘에 적용하는데 제한적이라는 것을 알 수 있다.

4.2 RandLA-Net의 로컬 특징 집계 모듈

RandLA-Net은 로컬 기능 수집기(LFA, Local Feature Aggregator)와 랜덤 샘플링(RS)을 적용했는데 알고리즘의 순서는 크게 4개의 인코딩 레이어, 디코딩 레이어 이후 공유된 완전연결 다층 퍼셉트론(shared fully-connected multilayer perceptron) 및 드롭아웃(DP, Drop-out)을 통해 예측이 수행된다. 인코딩 레이어에서는 입력된 데이터를 무작위로 다운샘플링 하지만 각 점의 특징들은 차원을 역으로 확장하여 더 많은 정보를 보존하고자 했다. 디코딩 레이어에서는 업샘플링(up-sampling)을 통해 축소된 데이터를 다시 확장하고, 확장되었던 특징들을 다시 축소함과 동시에 인코딩 레이어에서 중간중간 생성된 피쳐 맵(feature map)들을 잔차 연결(skip connection)하여 데이터의 연결성을 강화한다. 이후 공유된 다층 퍼셉트론과 드롭아웃을 통해 의미론적 분할 예측이 수행된다.

RandLA-Net의 로컬 특징 집계 모듈(LFAM, Fig. 8)은 크게 3가지로 구분할 수 있는데, 이 과정들을 통해 랜덤 샘플링이 적용되더라도 예측할 때 기하학적으로 중요한 데이터를 보존할 수 있다. 일반적인 포인트 클라우드 데이터는 X, Y, Z, R, G, B 속성을 가진 데이터로 구성되는데, 이 집계 모듈에서 N은 입력된 점 데이터의 수를 의미하고, 3은 X, Y, Z를 뜻하며, d는 R, G, B를 뜻한다. LFAM의 상단 패널은 점 데이터 간의 특징을 추출하는 로컬 공간 인코딩(LocSE, Local Spatial Encoding)과 데이터 관계 및 지오메트리에 기반하여 가장 중요한 인접 특징에 가중치를 부여하는 주의 풀링(AP, Attentive Pooling) 메커니즘을, 하단 패널은 상단 패널을 두 번 연결하여 수용되는 필드 크기를 확장하는 방법을 따른다.

RandLA-Net의 LFAM은 크게 LocSE, AP, DRB로 구분할 수 있다. 첫 번째 LocSE에서는 스캔이 포인트 클라우드 데이터의 형상적으로 구분하기 위해 중요한 특징을 가진 점 데이터를 찾기

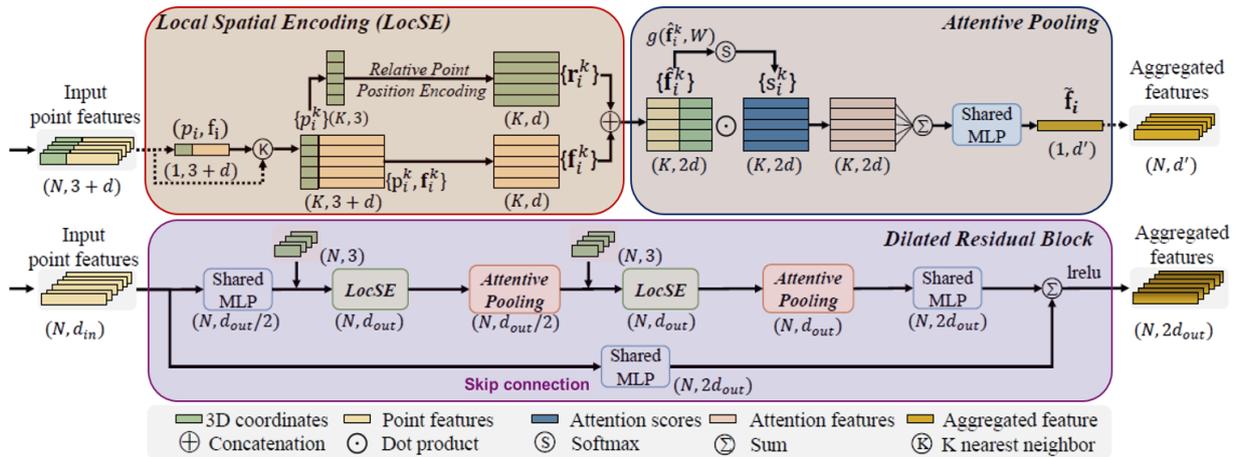


Fig. 8 Local feature aggregation module for RandLA-Net (Hu et al., 2019)

위에 적용된다. 데이터를 먼저 클러스터링(clustering)하고, 그 영역의 중심점을 찾아 중심점과 인접점(relative points)들의 상관성을 찾아 연결 연산(concatenation operation)을 수행한다. 이를 통해 새로운 인접 특징 집합인 \hat{F} 을 구할 수 있다. Table 4는 연결 연산에 포함되는 파라미터 수에 따른 평균 IoU를 나타내고, 모든 파라미터를 적용함에 따라 최고의 성능을 보이는 것을 알 수 있다.

$$r_i^k = MLP(p_i \oplus p_i^k \oplus (p_i - p_i^k) \oplus \|p_i - p_i^k\|) \quad (1)$$

$$r_i^k \oplus f_i^k = \hat{f}_i^k \quad (2)$$

$$\hat{F}_i = (\hat{f}_i^1 \dots \hat{f}_i^k \dots \hat{f}_i^K) \quad (3)$$

Table 4 The mIoU result of RandLA-Net by encoding different kinds of spatial information (Hu et al., 2019)

LocSE	mIoU(%)
(p_i)	45.5
(p_i^k)	47.7
(p_i, p_i^k)	49.1
$(p_i, p_i^k, \ p_i - p_i^k\)$	50.5
$(p_i, p_i^k, p_i - p_i^k)$	53.6
$(p_i, p_i^k, p_i - p_i^k, \ p_i - p_i^k\)$	54.3

다음은 인접 특징 집합 \hat{F} 으로부터 중요한 특징을 가지는 데이터를 찾아내기 위해 주의 점수(attention scores)를 계산한다. 계산은 공유 함수 $g()$ 에서 이루어지고, 소프트맥스 활성화 함수 \textcircled{S} 를 통해 데이터마다 특정 가중치를 부여해준다. g 를 통해 기하학적으로 예측에 중요한 특징을 가지는 데이터에 높은 가중치를 주고, 중요하지 않은 데이터에는 낮은 가중치를 부여한다. 그리고, 공유 함수 $g()$ 로 계산된 s 와 이전에 구했던 \hat{f} 를 내적하

여 총합한 \tilde{f} 를 구한다.

$$s_i^k = g(\hat{f}_i^k, W) \quad (4)$$

$$\tilde{f}_i = \sum_{k=1}^K (\hat{f}_i^k \cdot s_i^k) \quad (5)$$

RandLA-Net은 LocSE와 AP가 두 번 적용되었는데, 이로써 데이터의 기하학적인 세부사항들이 더 잘 보존시켜 예측 정확도를 증가시켰다. Table 5에서 LocSE와 AP의 적용 횟수에 따른 mIoU를 확인할 수 있고, Table 6에서 RandLA-Net의 주의 풀링과 여타 일반적인 풀링 방법들 및 단순화시킨 DRB와의 비교를 통해 RandLA-Net의 프레임워크가 더 뛰어나다는 것을 증명하였다.

Table 5 The mIoU scores of RandLA-Net regarding different number of aggregation units in a residual block (Hu et al., 2019)

Dilated residual block	mIoU(%)
one aggregation unit	49.8
three aggregation units	51.1
two aggregation units (The Standard Block)	54.3

Table 6 The mean IoU scores of RandLA-Net (Hu et al., 2019)

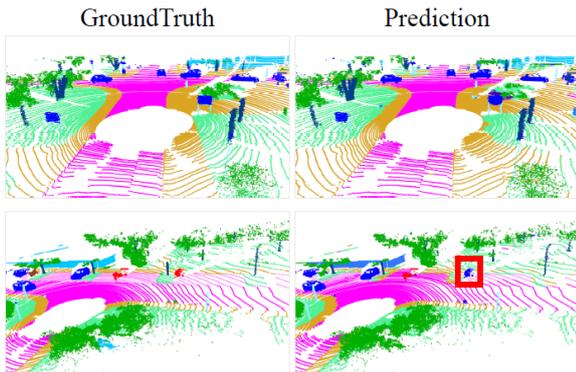
	mIoU(%)
Remove local spatial encoding	49.8
Replace with max-pooling	55.2
Replace with mean-pooling	53.4
Replace with sum-pooling	54.3
Simplify dilated residual block	48.8
The Full framework (RandLA-Net)	57.1

5. RandLA-Net 기법을 활용한 3D 객체 탐지

RandLA-Net의 논문에 수록된 자료를 살펴보면 3D 객체 탐지에 활용되는 공용 데이터셋(S3DIS, SemantickITTI 등)을 활용한 예측 결과를 알 수 있고, 이는 Table 7, Table 8과 같다.

Table 7 Quantitative results of RandLA-Net of different approaches on SemantickITTI (Hu et al., 2019)

Methods	Size	mIoU(%)
PointNet	50K pts	14.6
SPG		17.4
SPLATNet		18.4
PointNet++		20.1
TangentConv		40.9
RandLA-Net		53.9



Qualitative results of RandLA-Net on the validation split of SemantickITTI. Red box shows the failure case

Table 8 Quantitative results of RandLA-Net of different approaches on S3DIS (6-fold cross-validation). Overall Accuracy (OA, %), mean class Accuracy (mAcc, %), mean IoU (mIoU, %) are reported (Hu et al., 2019)

Methods	OA(%)	mAcc(%)	mIoU(%)
PointNet	78.6	66.2	47.6
RSNet	-	66.5	56.5
3P-RNN	86.9	-	56.3
SPG	86.4	73.0	62.1
PointCNN	88.1	75.6	66.7
RandLA-Net	88.0	82.0	70.0



Semantic segmentation results of RandLA-Net on the complete point clouds of Areas 1-2 in S3DIS. Left: predicted labels, right: ground truth

6. 결론

본 논문에서는 LiDAR, 3차원 스캐너 등에서 생성되는 3차원 포인트 클라우드를 활용한 인공지능 기반 탐색기법에 대한 고찰을 수행하였다. 관련 계측장비가 발달함에 따라 계측되는 정보의 양과 용량이 빠르게 확대되고 있는 상황에서 기존 알고리즘 기반의 탐색은 처리속도에 있어서 한계점이 있다. 이에 사진, 동영상 객체인식에서 착안된 딥러닝 기반 3차원 데이터 탐색기법 연구가 지속되고 있다. 기존에 활발히 연구되었던 2D 객체 탐지 기법은 산업에서 활용할 정도로 이미 고도화돼버린 시점이고, 더욱 정밀한 객체 예측을 위해 3D 데이터를 활용한 객체 탐지 기법 (Qi et al. 2020; Chen et al. 2021; Jing et al. 2022)들이 화두 되고 있다. 하지만, 아직 3D 객체 탐지 기법은 활용 데이터의 특성 (unordered and sparse)으로 인해 예측 정확도의 고도화가 필요한 시점이다. 또한 점차 고해상도 및 고정밀 데이터가 확보되고 누적되는 시점에 기존 알고리즘 기반의 탐색으로는 처리시간의 한계가 있다. 이에, 현재 가장 널리 활용되고 있는 기법 중 하나인 PointNet에 대한 분석과 이후 파생된 기법 중 하나인 RandLA-Net을 분석하여 기술하였다. 아직까지 현장에 적용할 만큼 충분한 탐색성능을 보여주고 있지는 못하지만 자율주행차 영역에서 핵심적인 기술로 연구되고 있으므로, 가까운 시일 내에 더욱 정확하고 신속한 탐지 및 분석예측이 가능할 것으로 판단되며, 이에 관련 조선해양 연구자들에게 기반 기법에 대한 내용을 공유하고자 하였다.

후기

이 논문은 2022년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원(P0017006, 2022년 산업혁신인재성장지원사업)과 한국해양교통안전공단의 2022년 자체연구개발사업 [현장검사 지

원을 위한 고정밀 형상계측 기반 디지털선박 원격검사 및 이력관리 기술개발]의 지원을 받아 수행된 연구임

References

- Chen, X., Jiang, K., Zhu, Y., Wang X. and Yun, T., 2021. Individual tree crown segmentation directly from UAV-Borne LiDAR data using the PointNet of deep learning. *Journal of the Forests*, 12(2), pp.131.
- Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N. and Markham, A., 2019. RandLA-Net: Efficient semantic segmentation of large-scale point clouds. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, arXiv: 1911.11236v3.
- Jaderberg, M., Simonvan, K., Zisserman, A. and Kavukcuoglu, K., 2015. Spatial transformer networks. *Part of Advances in Neural Information Processing Systems 28*, paper ID:1213. Montreal, Canada, 7 - 12th December 2015.
- Jing, L., Yu, R., Kretschmar, H., Li, K., Qi, C.R., Zhao, H., Ayvaci, A., Chen, X., Cower, D., Li, Y., You, Y., Deng, H., Li, C. and Anguelov, D., 2022. Depth estimation matters most: improving per-object depth estimation for monocular 3D detection and tracking. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, arXiv: 2206.03666.
- Lee, D.K., Ji, S.H. and Park, B.Y., 2021. Object detection and post-processing of LNGC CCS scaffolding system using 3D point cloud based on deep learning. *Journal of the Society of Naval Architects of Korea*, 58(5), pp.303-313.
- Song, S.H., Lee, G.H., Han, K.M. and Jang, H.S., 2022a. Field Applicability Study of Hull Crack Detection Based on Artificial Intelligence. *Journal of the Society of Naval Architects of Korea*, 59(4), pp.192-199.
- Song, S.H., Lee, G.H., Han, K.M. and Jang, H.S., 2022b. Application of point cloud based hull structure deformation detection algorithm. *Journal of the Society of Naval Architects of Korea*, 59(4), pp.235-242.
- Qi, C.R., Chen, X., Litany, O. and Guibas, L.J., 2020. Imvotenet: boosting 3d object detection in point clouds with image votes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, arXiv: 2001.10692v1.
- Qi, C.R., Su, H., Mo, K. and Guibas, L.J., 2017. PointNet: Deep learning on point sets for 3D classification and segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, arXiv: 1612.00593v2.

