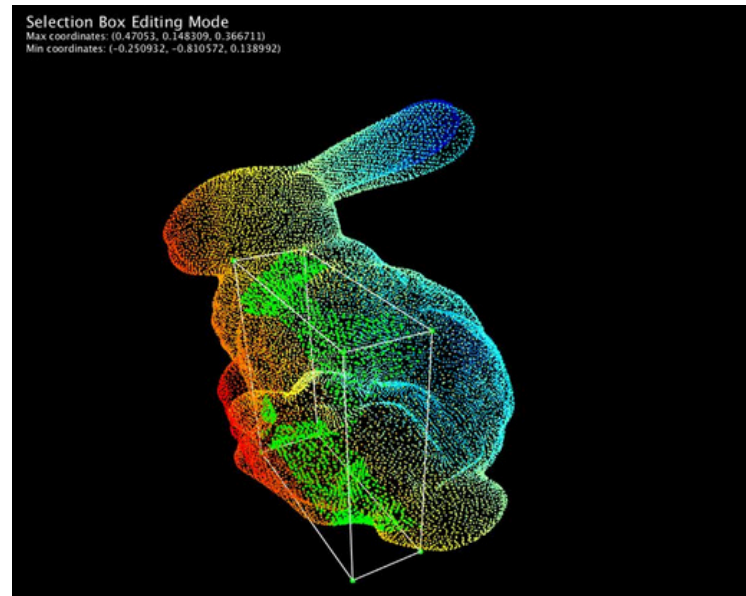


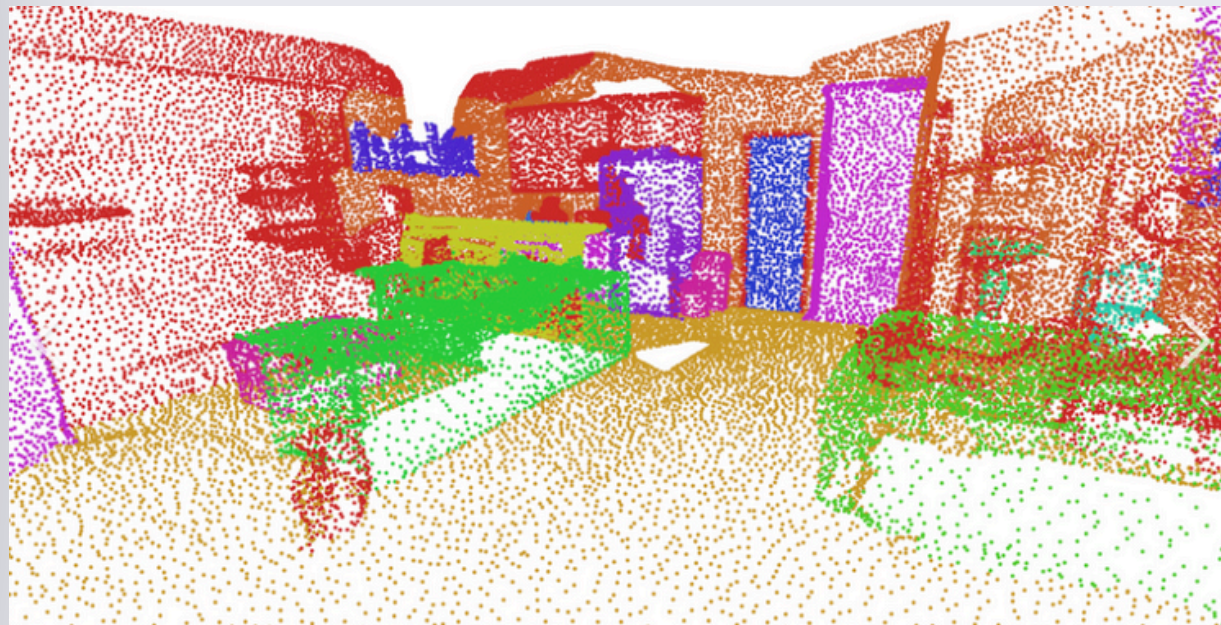


PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation

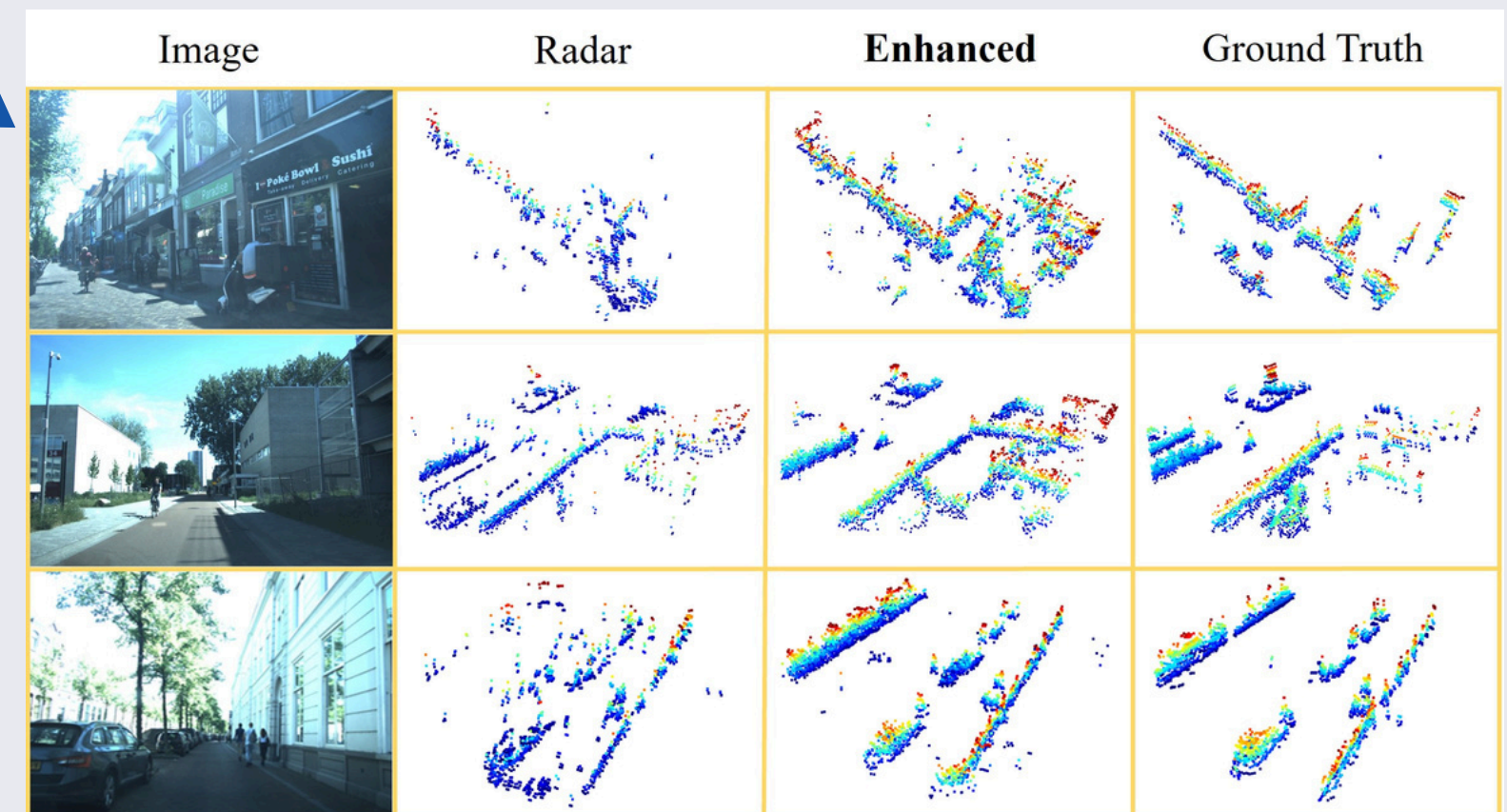
김민성

CVPR 2017

Point Cloud Data



Voxel 3D



Many Image

초록

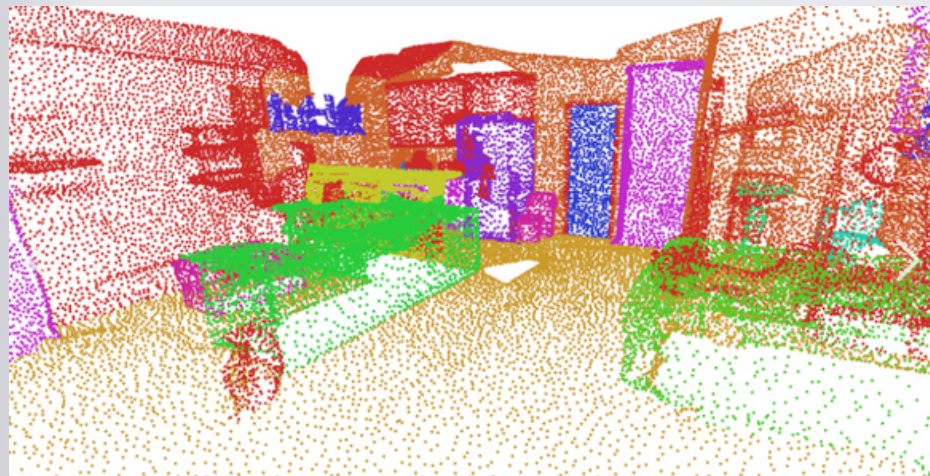
Abstract

CVPR 2017

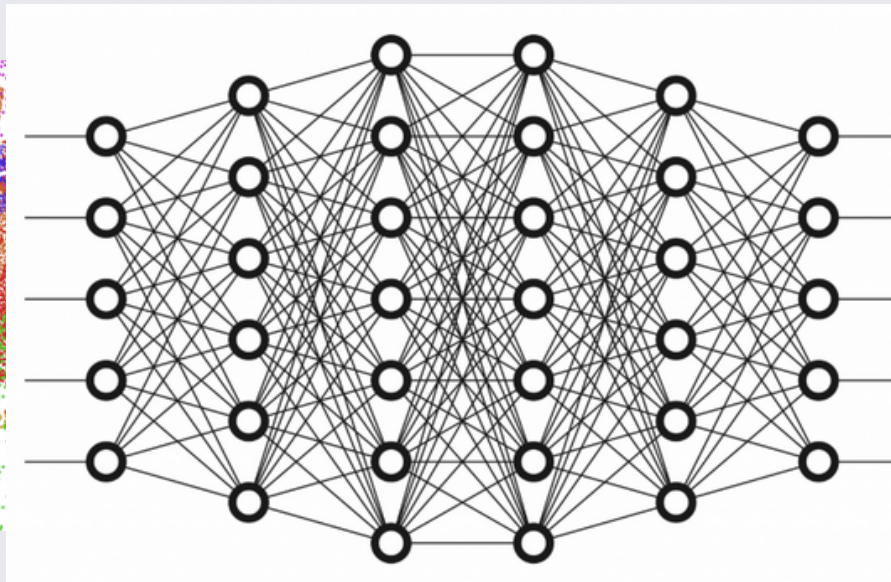
초록

Abstract

Point Cloud Data



Neural Network



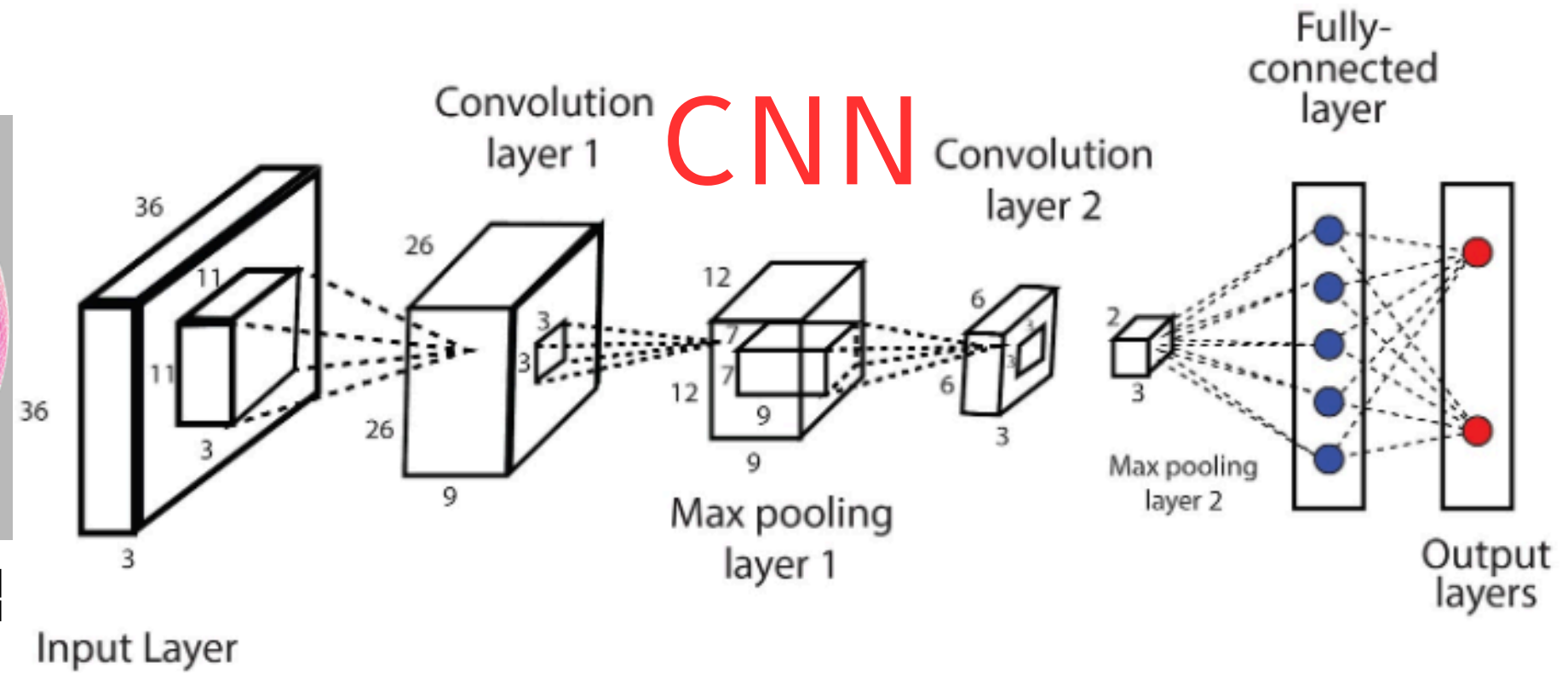
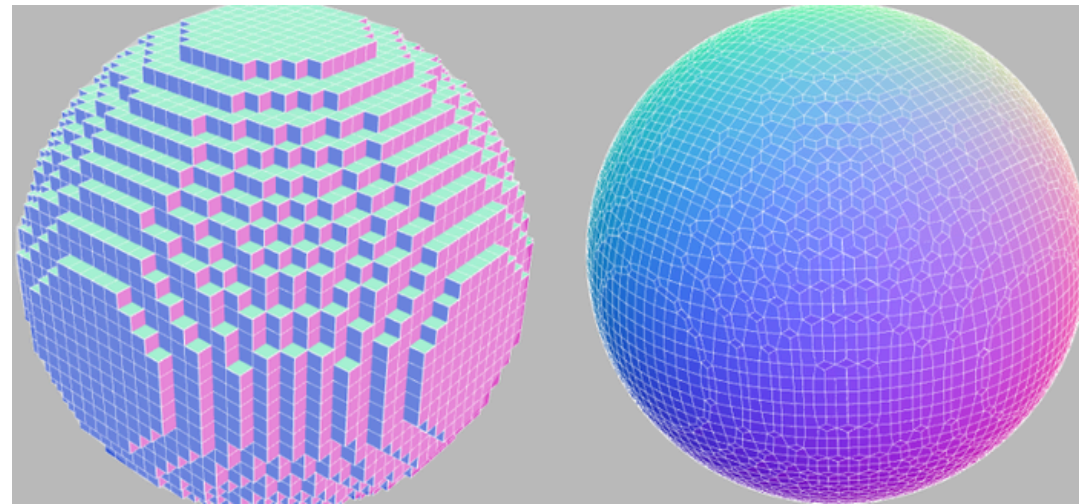
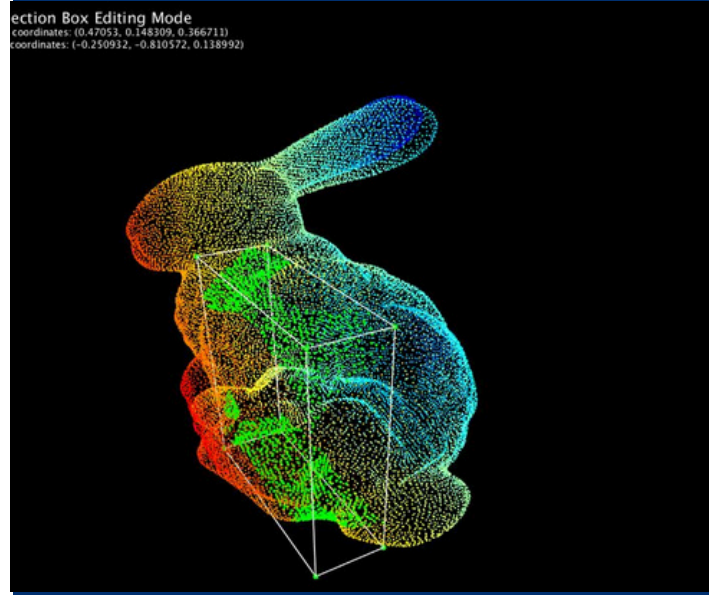
Training



AI
Model
(PointNet)

1. Introduction

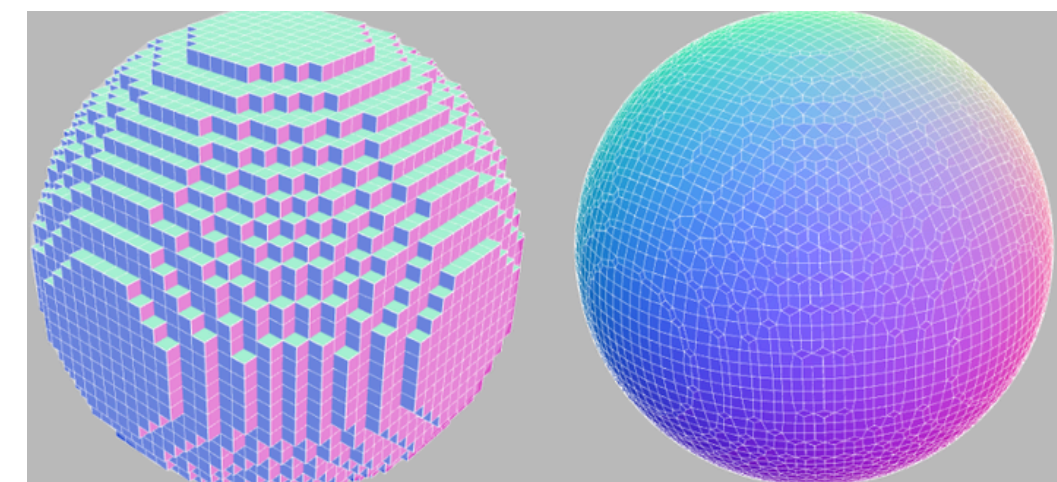
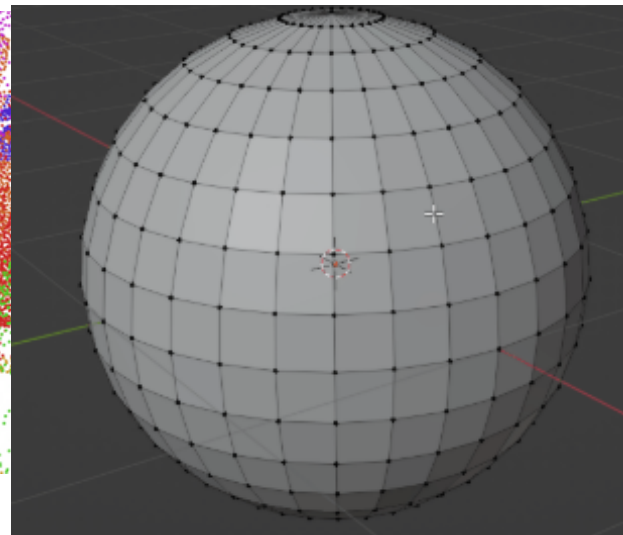
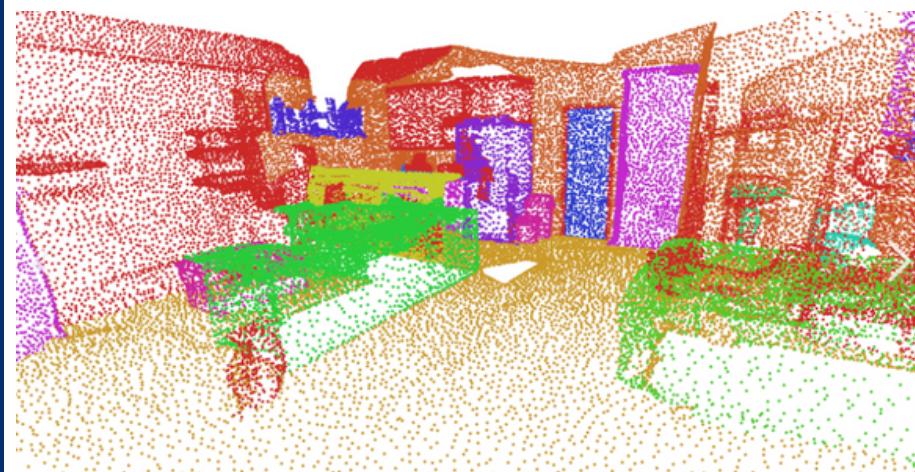
기존 방법



Introduction

voxel이나 grid 방식과 같이 규칙적인 데이터를 통해 학습

문제점



비규칙적인 point cloud, mesh 표현 방식을

voxel이나 많은 이미지와 같이 규칙적인 데이터 표현 방식으로 변환해서 학습

연구 내용 및 예상 목표

연구 방법 및 단계

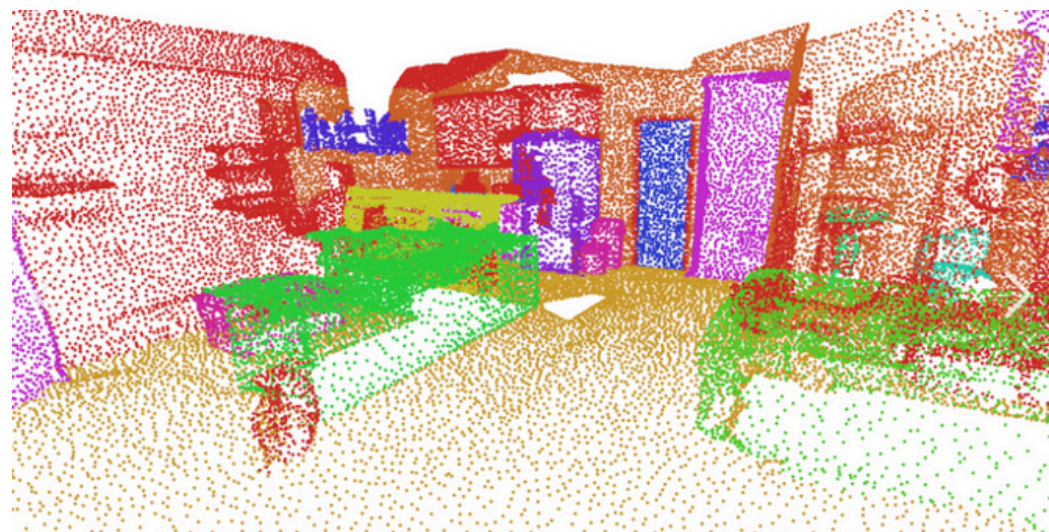
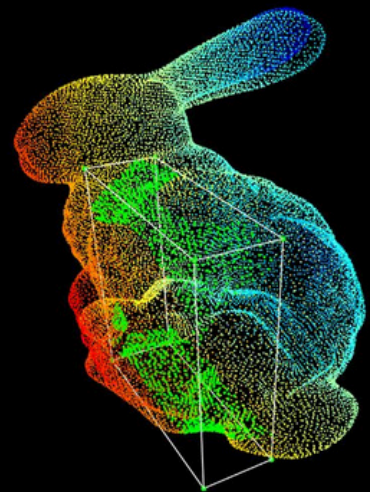
연구 성과 및 사용

종결 및 전망

1. Introduction

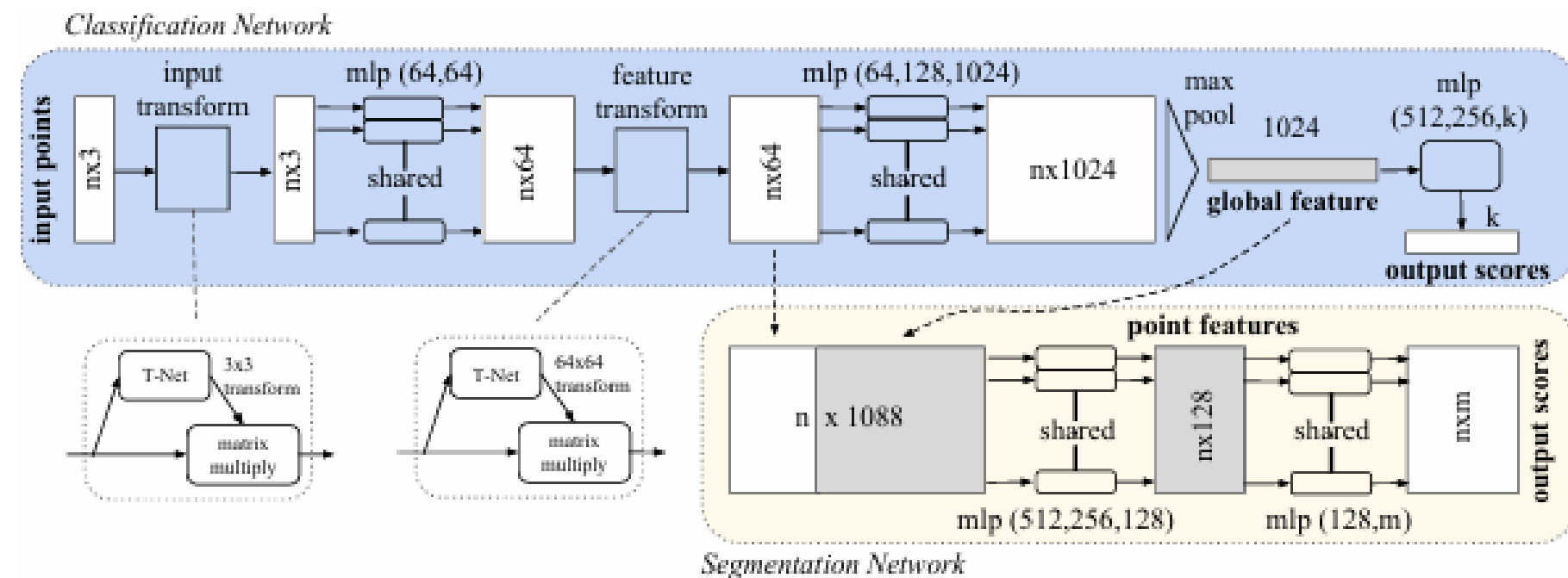
제안

Section Box Editing Mode
Coordinates: (0.47051, 0.148309, 0.366711)
Coordinates: (-0.250932, -0.810572, 0.118992)



포인트 클라우드는 단순하고 통일된 구조로, 메시의 조합적 불규칙성과 복잡성을 피할 수 있어 학습이 더 쉽습니다.

PointNet



Training

Introduction

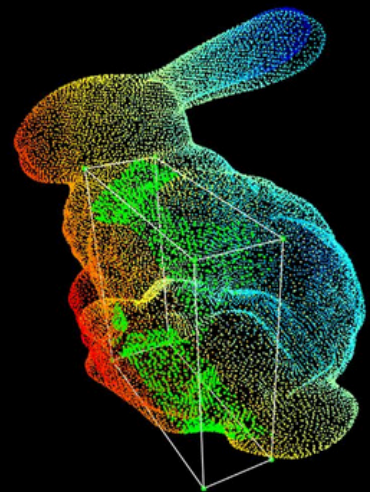
연구 내용 및 예상 목표

연구 방법 및 단계

연구 성과 및 사용

종결 및 전망

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



Introduction

연구 내용 및 예상 목표

연구 방법 및 단계

연구 성과 및 사용

종결 및 전망

1. Introduction

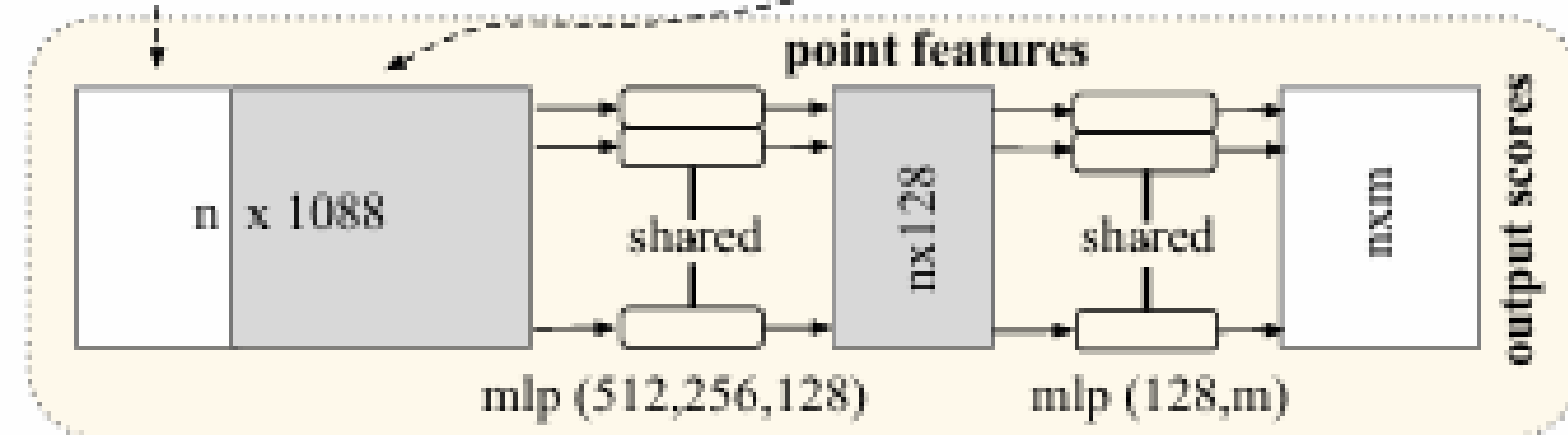
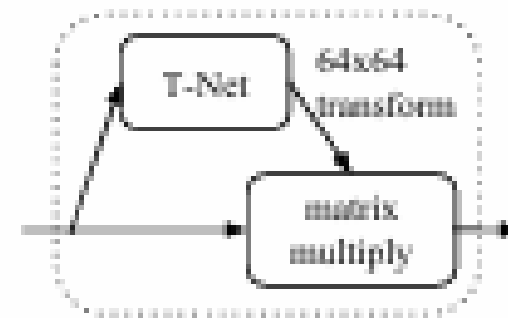
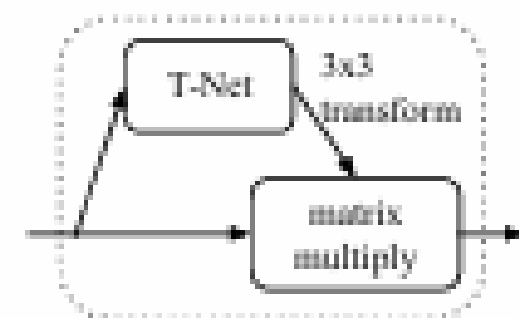
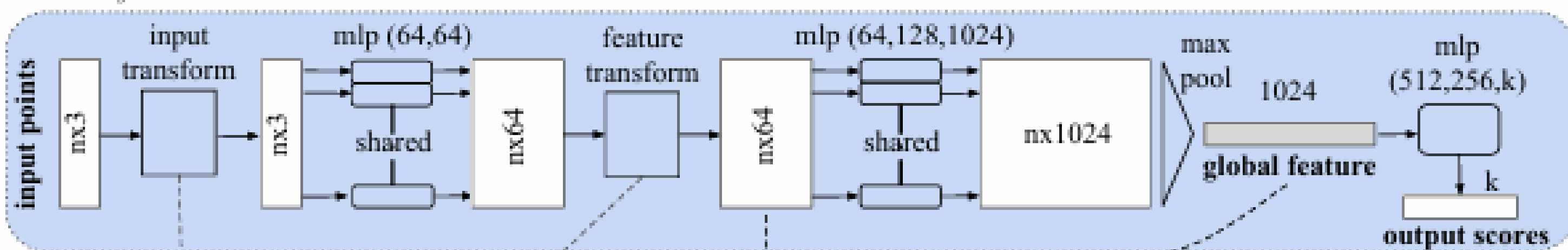
제안

input points
(x, y, z)

모든 점들을 독립적으로 받습니다.

max pool 이미지의 크기를 줄임

Classification Network



Segmentation Network

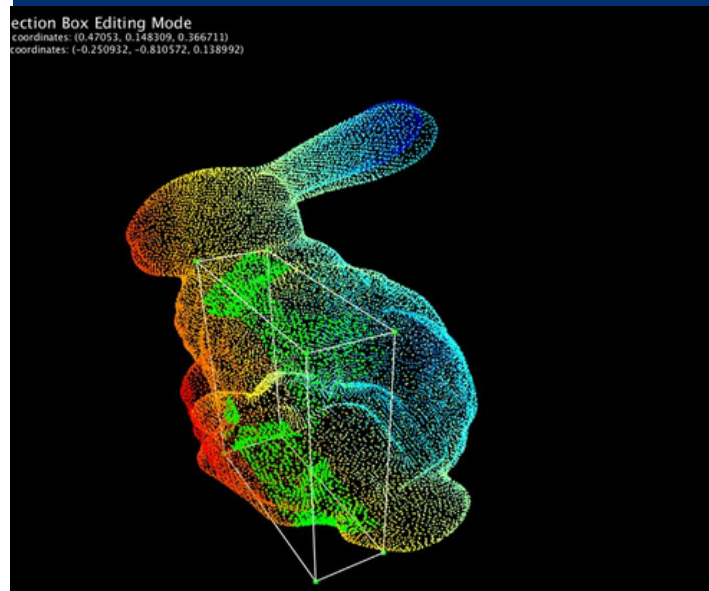
PointNet

1. Introduction

제안

grid transform(그리드 변환)

affine transform(아핀 변환)



Introduction

연구 내용 및 예상 목표

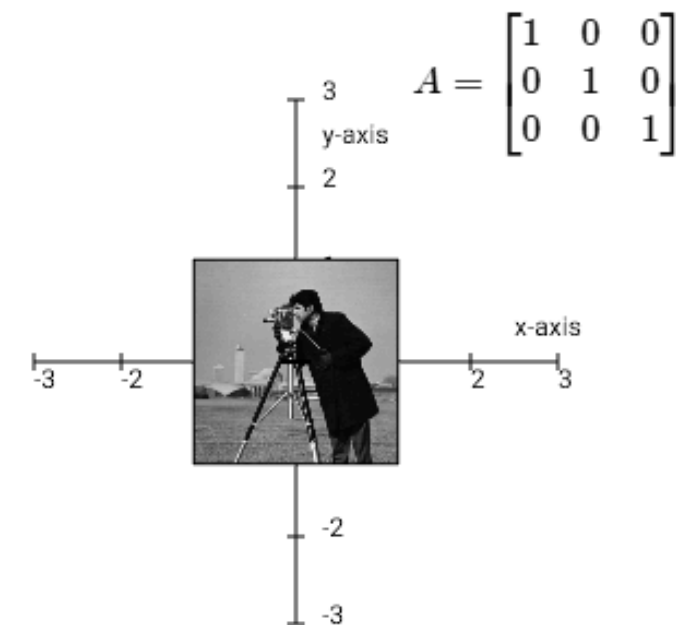
연구 방법 및 단계

연구 성과 및 사용

종결 및 전망

input points
(x, y, z)

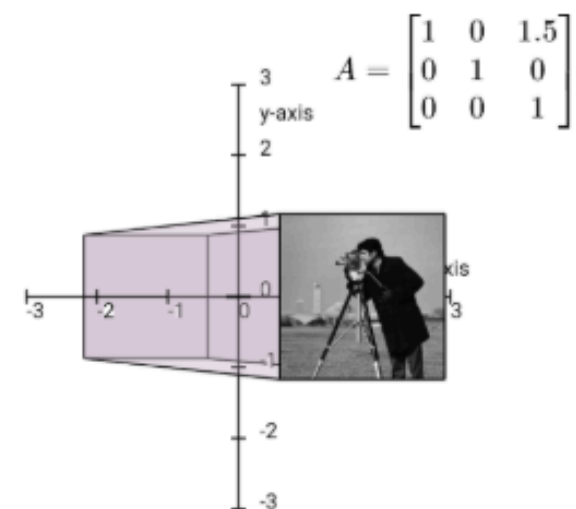
모든 점들을 독립적으로 받는 이유는 그리드 변환과, 아핀 변환을 쉽게 하기 위해서.



(c)공돌이의수학정리노트



1행 3열의 값을 양으로 변화시키면 3차원 공간에서는 정육면체 윗면이 오른쪽으로 밀리게 되지만, 조감도로 투영해보면 사진이 오른쪽으로 이동하는 것과 같은 효과를 가져오게 된다.



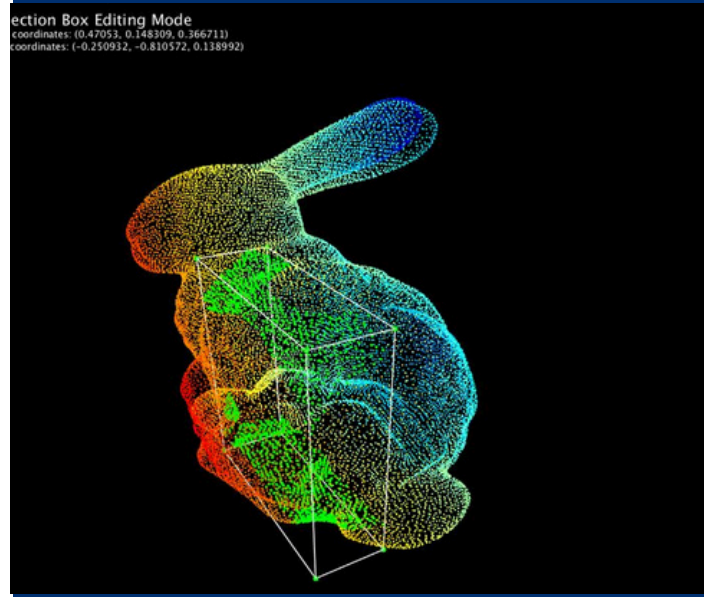
(c)공돌이의수학정리노트



그림 9. 3차원 공간에서 1행 3열의 원소값 변화는 조감도로 보면 x축 상에서 평행이동해준 것과 같은 효과를 보인다.

1. Introduction

프로젝트 결과물(PointNet)



Introduction

연구 내용 및 예상 목표

연구 방법 및 단계

연구 성과 및 사용

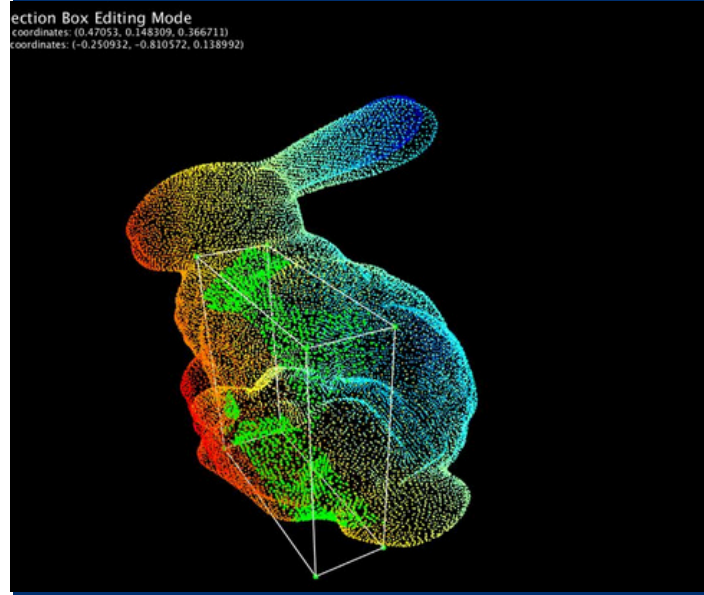
종결 및 전망

1. 이론적 분석과 실험적 평가를 모두 제공
2. 입력 포인트 클라우드를 희소한 핵심점 집합으로 요약하는 법을 학습
3. 이론적 분석은 왜 우리의 PointNet이 입력 포인트의 작은 변화뿐만 아니라 포인트 삽입(이상치)이나 삭제(누락된 데이터)로 인한 손상에도 매우 강인한지 이해할 수 있도록 해줍니다
4. 3D에서 순서가 없는 점 집합을 처리할 수 있는 새로운 딥넷 구조를 설계

1. Related Work

Point Cloud Feature

포인트 특징은 종종 점들의 특징 통계적 속성을 인코딩하며, 일반적으로 고유적(intrinsic) 또는 외적(extrinsic)으로 분류되는 특징 변환에 대해 불변하도록 설계됩니다. 또한 지역적 특징(local features)과 전역적 특징(global features)으로 분류할 수 있습니다.



과제 선택 배경 및 의미

Related Work

연구 방법 및 단계

연구 성과 및 사용

종결 및 전망

1. Related Work

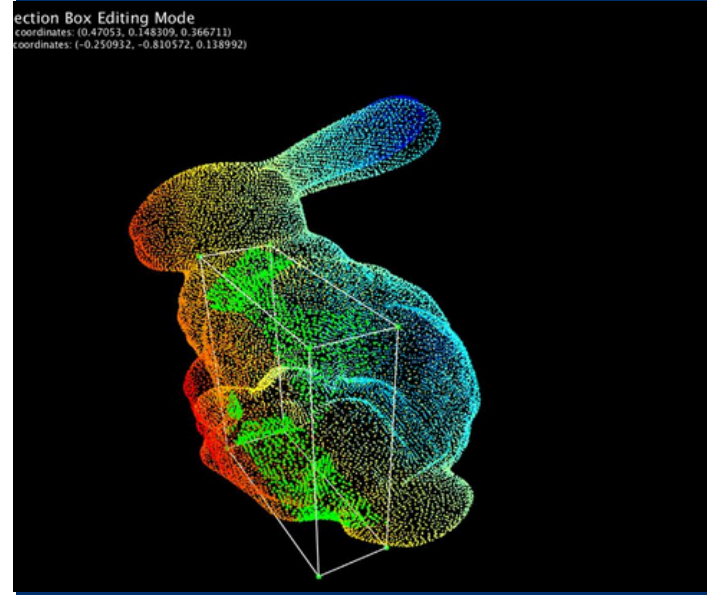
다른 모델들의 문제점

Volumetric CNNs

볼류메트릭 표현은 데이터 희소성과 3D 컨볼루션의 연산 비용 때문에 해상도에 제한이 있다.

FPNN, Vote3D

희소성 문제를 다루기 위해 특별한 방법을 제안했지만, 그들의 연산은 여전히 희소한 볼륨에서 이루어지기 때문에 매우 큰 포인트 클라우드를 처리하는 데에는 어려움이 있다.



과제 선택 배경 및 의미

Related Work

연구 방법 및 단계

연구 성과 및 사용

종결 및 전망

1. Related Work

다른 모델들의 문제점

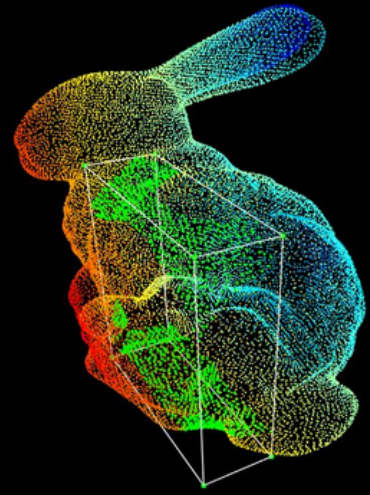
Multiview CNNs

3D 포인트 클라우드나 형태를 2D 이미지로 변환한 후 2D 합성곱 신경망을 적용하여 분류하려고 시도했습니다. 잘 설계된 이미지 CNN을 사용하면, 이 방법은 형태 분류 및 검색 작업에서 우수한 성능을 보여주었습니다. 이러한 방법을 장면 이해나 포인트 분류, 형태 완성 같은 다른 3D 작업으로 확장하는 것은 간단하지 않습니다.

Spectral CNNs

유기체 같은 매니폴드 메시(Manifold Mesh)에 제한되어 있으며, 가구와 같은 비등각 형태로 확장하는 방법은 명확하지 않습니다.

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



과제 선택 배경 및 의미

Related Work

연구 방법 및 단계

연구 성과 및 사용

종결 및 전망

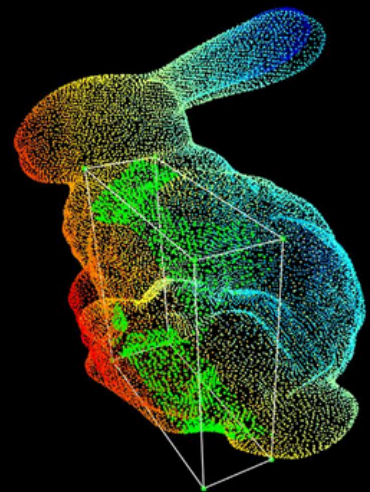
2. Related Work

다른 모델들의 문제점

비정렬 집합에서 최근 연구

포인트 집합에 대한 딥러닝 연구는 많지 않습니다. Oriol Vinyals 등 의 최근 연구는 이 문제를 다루고 있습니다. 그들은 비정렬 입력 집합을 처리하기 위해 어텐션 메커니즘을 사용하는 읽기-처리-쓰기 네트워크를 사용하며, 그들의 네트워크가 숫자를 정렬할 수 있는 능력이 있음을 보여줍니다. 그러나 그들의 연구는 일반 집합과 NLP 응용에 중점을 두고 있기 때문에 집합에서 형상의 역할이 부족합니다.

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



과제 선택 배경 및 의미

Related Work

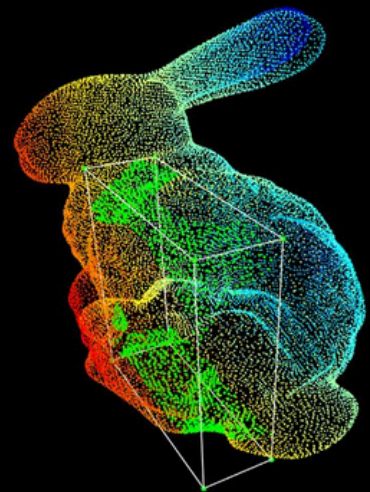
연구 방법 및 단계

연구 성과 및 사용

종결 및 전망

3. Problem Statement

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



과제 선택 배경 및 의미

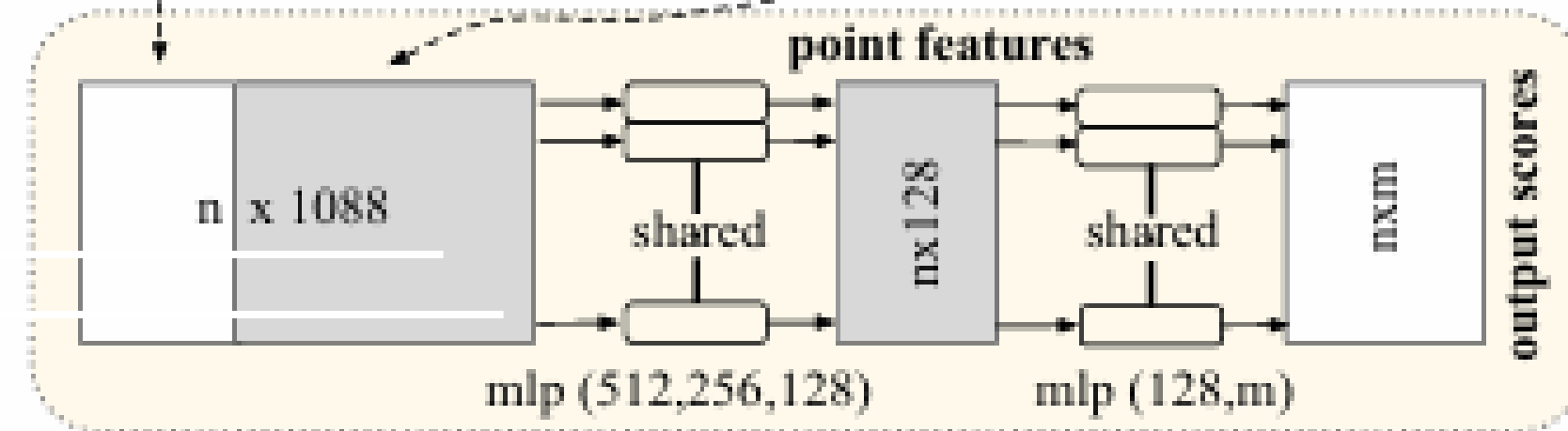
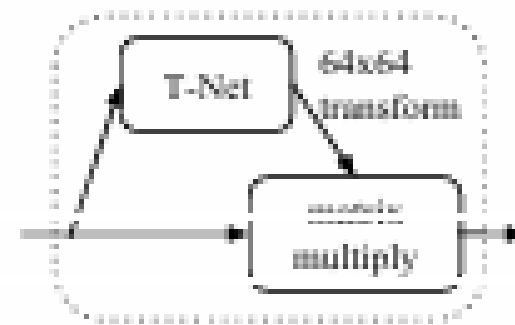
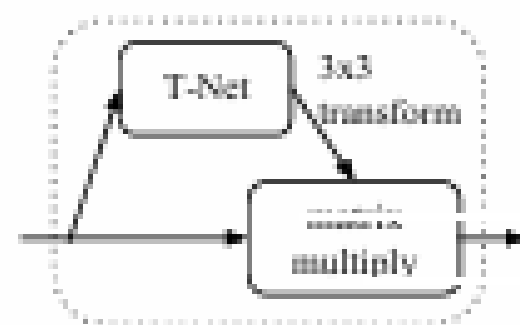
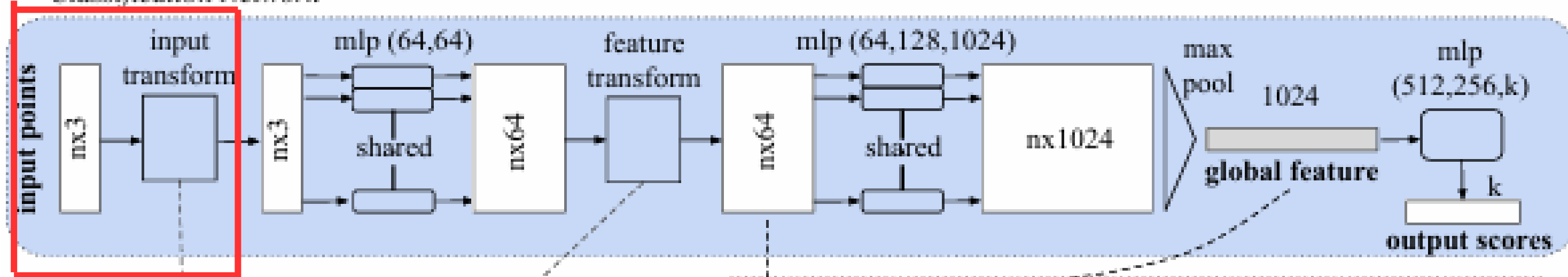
연구 내용 및 예상 목표

Problem Statement

연구 성과 및 사용

종결 및 전망

Classification Network

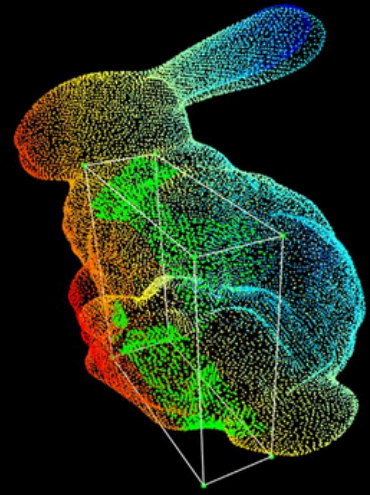


Segmentation Network

3D 점 집합 $\{P_i \mid i = 1, \dots, n\}$ 으로 표현되며, 각 점 P_i 는 (x, y, z) 좌표와 색상, 법선 등과 같은 추가 특징 채널을 가진 벡터입니다. 점의 채널로 (x, y, z) 좌표만 사용합니다.

3. Problem Statement

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



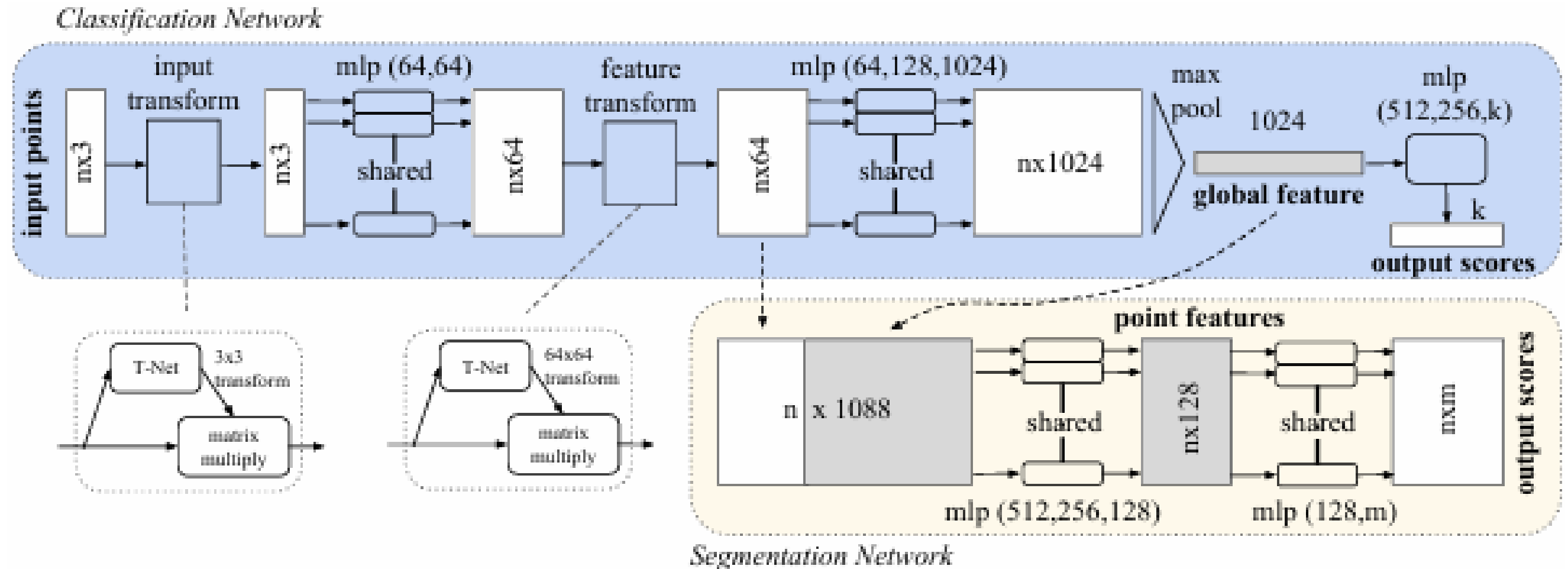
과제 선택 배경 및 의미

연구 내용 및 예상 목표

Problem Statement

연구 성과 및 사용

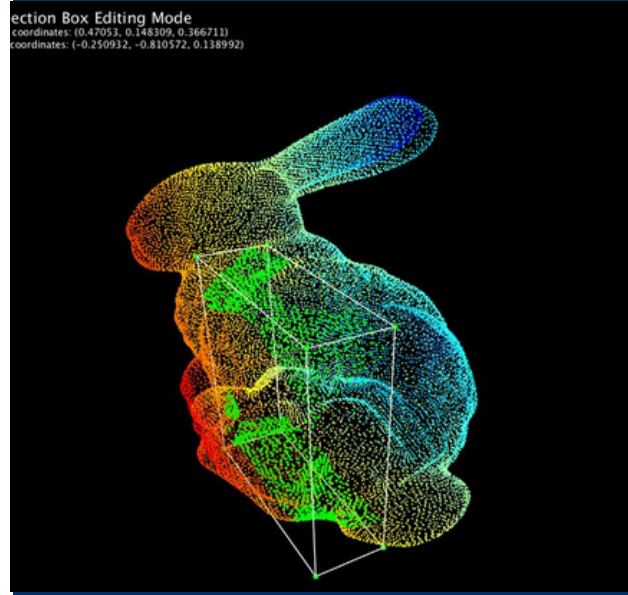
종결 및 전망



물체 분류 작업에서는 입력 점 구름이 형상에서 직접 샘플링되거나 장면 점 구름에서 미리 분할됩니다. 우리가 제안하는 딥 네트워크는 k개의 후보 클래스 각각에 대해 k개의 점수를 출력합니다. 의미론적 분할(semantic segmentation)의 경우, 입력은 부분 영역 분할을 위한 단일 객체이거나, 객체 영역 분할을 위한 3D 장면의 하위 볼륨일 수 있습니다. 우리의 모델은 n개의 점과 m개의 의미론적 하위 카테고리 각각에 대해 $n \times m$ 개의 점수를 출력합니다.

4. Deep Learning on Point Sets

PointNet을 입력받기 위한 3D Point Sets 의 특징



과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

Deep Learning on
Point Sets

종결 및 전망

1. 순서가 없습니다. 이미지의 픽셀 배열이나 그리드의 복셀 배열과 달리, 포인트 클라우드는 특정한 순서가 없는 점들의 집합입니다. 다시 말해, N개의 3D 포인트 세트를 입력으로 받는 네트워크는 데이터 입력 순서에서의 N! 가지 순열에 대해 불변이어야 합니다.
2. 점들은 거리 측정 기준이 있는 공간에서 나옵니다. 이는 점들이 고립되어 있지 않으며, 인접한 점들이 의미 있는 부분집합을 형성함을 의미합니다.
3. 변환에 대한 불변성. 기하학적 객체로서, 학습된 점 집합의 표현은 특정 변환에 대해 불변해야 합니다. 예를 들어, 점들을 함께 회전하거나 이동시키더라도 전체 점 구름의 범주나 점들의 분할은 바뀌지 않아야 합니다.

4. Deep Learning on Point Sets

PointNet Architecture

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

Deep Learning on
Point Sets

종결 및 전망

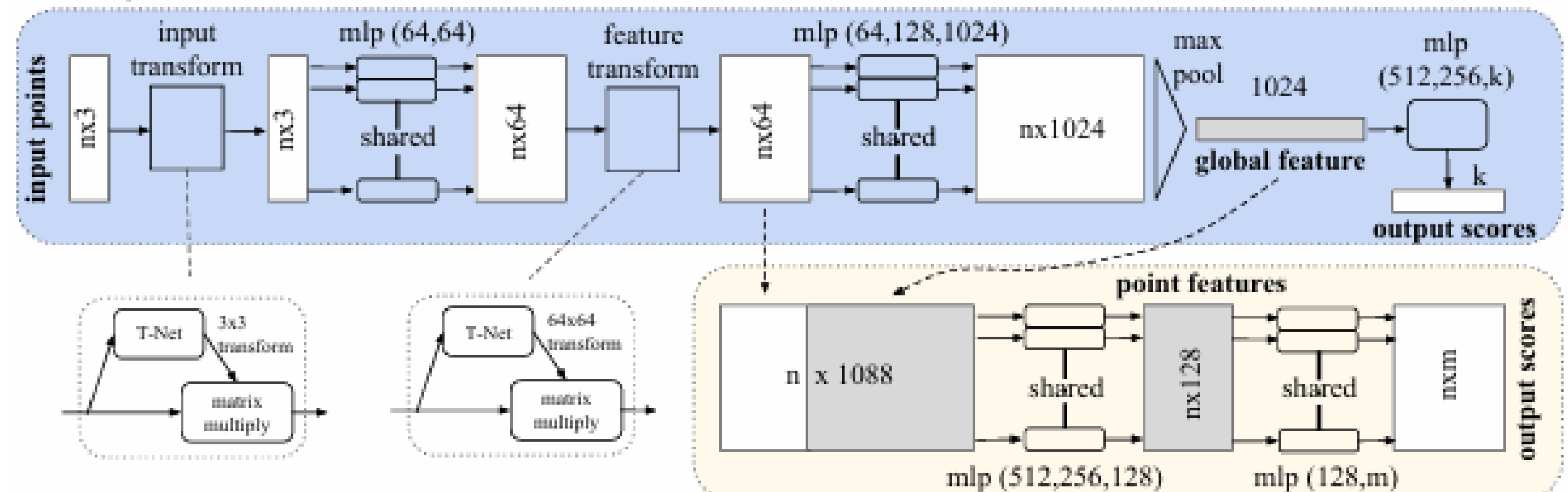
1. 순서가 없습니다. 이미지의 픽셀 배열이나 그리드의 복셀 배열과 달리, 포인트 클라우드는 특정한 순서가 없는 점들의 집합입니다. 다시 말해, N 개의 3D 포인트 세트를 입력으로 받는 네트워크는 데이터 입력 순서에서의 $N!$ 가지 순열에 대해 불변이어야 합니다.
2. 점들은 거리 측정 기준이 있는 공간에서 나옵니다. 이는 점들이 고립되어 있지 않으며, 인접한 점들이 의미 있는 부분집합을 형성함을 의미합니다.
3. 변환에 대한 불변성. 기하학적 객체로서, 학습된 점 집합의 표현은 특정 변환에 대해 불변해야 합니다. 예를 들어, 점들을 함께 회전하거나 이동시키더라도 전체 점 구름의 범주나 점들의 분할은 바뀌지 않아야 합니다.

4. Deep Learning on Point Sets

PointNet Architecture

우리 네트워크에는 세 가지 핵심 모듈이 있습니다

Classification Network



과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

Deep Learning on
Point Sets

종결 및 전망

1. Symmetry Function for Unordered Input

(순서가 없는 입력을 위한 대칭 함수)

2. Local and Global Information Aggregation

(local 및 global 정보 통합)

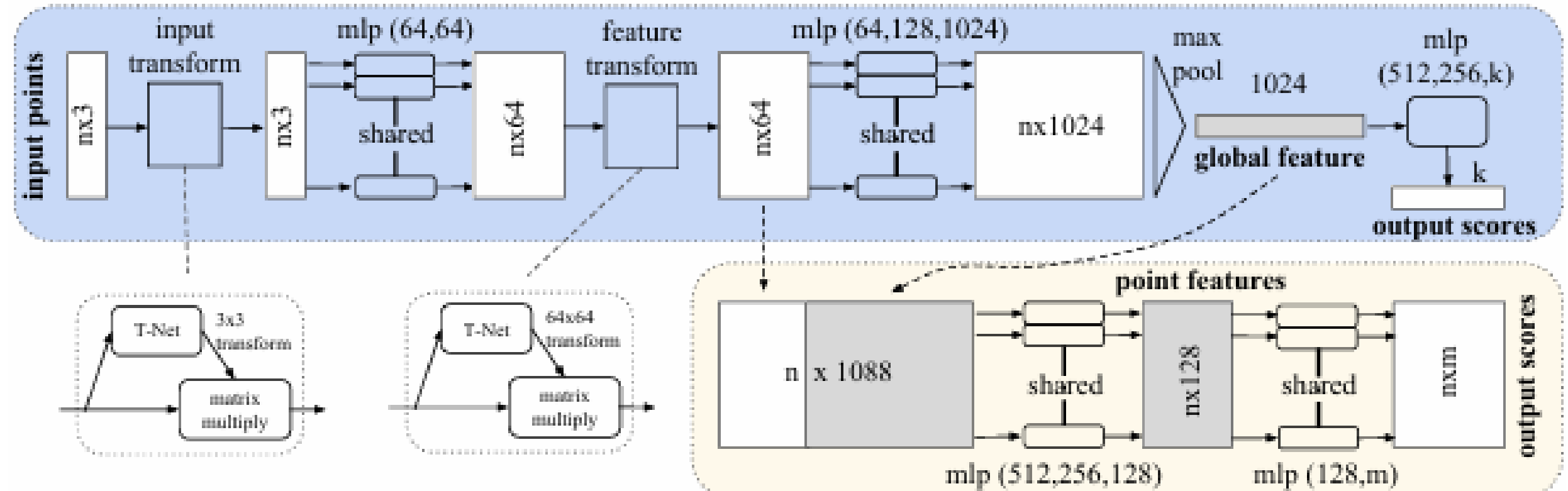
3. Joint Alignment Network

(관절 정렬 네트워크)

4. Deep Learning on Point Sets

PointNet Architecture

Classification Network



Symmetry Function for Unordered Input

입력 순서에 대해 모델을 불변하게 만들기 위해 세 가지 전략이 있습니다

- 1) 입력을 표준 순서로 정렬하기
- 2) 입력을 시퀀스로 처리하여 RNN을 학습하되, 학습 데이터를 모든 종류의 순열로 확장하기
- 3) 각 점에서 정보를 집계하기 위해 단순한 대칭 함수를 사용하기. 여기서 대칭 함수는 n 개의 벡터를 입력으로 받아 입력 순서에 상관없이 새로운 벡터를 출력합니다.

과제 선택 배경 및 의미

연구 내용 및 예상 목표

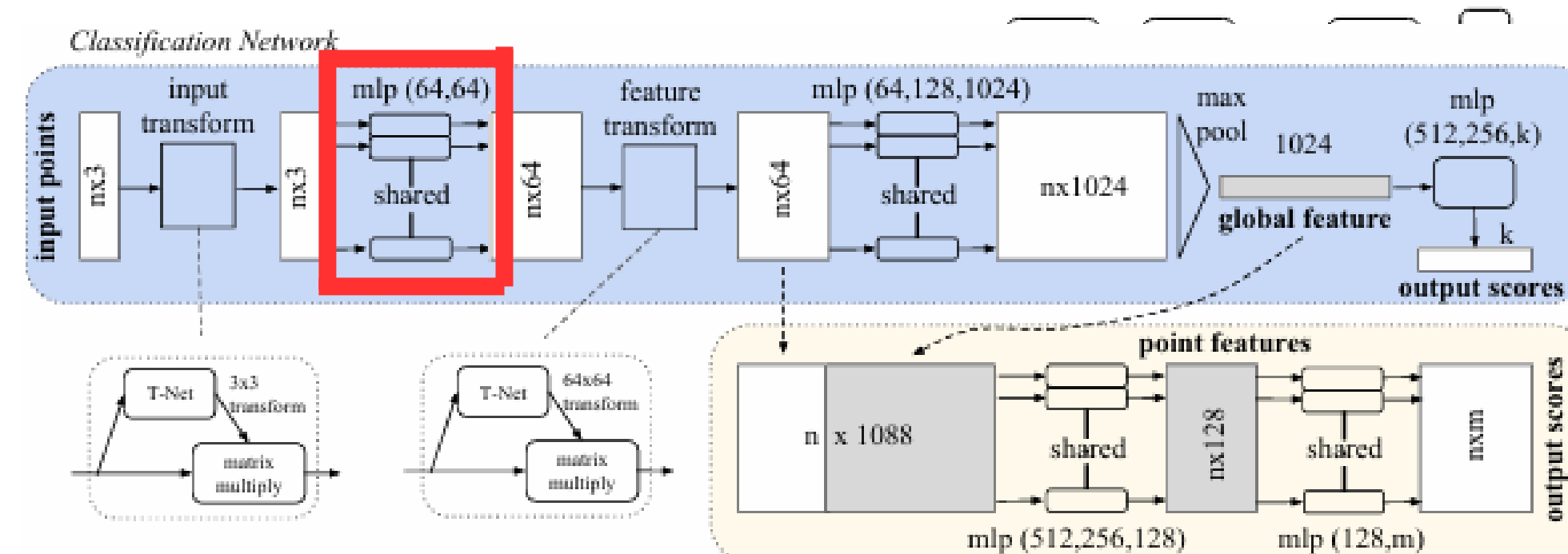
연구 방법 및 단계

Deep Learning on
Point Sets

종결 및 전망

4. Deep Learning on Point Sets

PointNet Architecture



	accuracy
MLP (unsorted input)	24.2
MLP (sorted input)	45.0
LSTM	78.5
Attention sum	83.0
Average pooling	83.8
Max pooling	87.1

과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

Deep Learning on
Point Sets

종결 및 전망

Symmetry Function for Unordered Input

정렬된 점 집합에 MLP를 직접 적용하면 정렬되지 않은 입력을 직접 처리하는 것보다 약간 나은 성능을 보이지만, 전반적으로 성능이 저조합니다.

RNN을 사용하는 아이디어는 점 집합을 순차적 신호로 간주하고, 무작위로 순열된 시퀀스로 RNN을 훈련함으로써 RNN이 입력 순서에 대해 불변하게 되기를 기대합니다.

RNN 기반 모델이 우리가 제안한 방법만큼 성능이 좋지 않다는 것을 보여주었습니다.

4. Deep Learning on Point Sets

PointNet Architecture

$$f(\{x_1, \dots, x_n\}) \approx g(h(x_1), \dots, h(x_n)), \quad (1)$$

where $f : 2^{\mathbb{R}^N} \rightarrow \mathbb{R}$, $h : \mathbb{R}^N \rightarrow \mathbb{R}^K$ and $g : \underbrace{\mathbb{R}^K \times \dots \times \mathbb{R}^K}_n \rightarrow \mathbb{R}$ is a symmetric function.

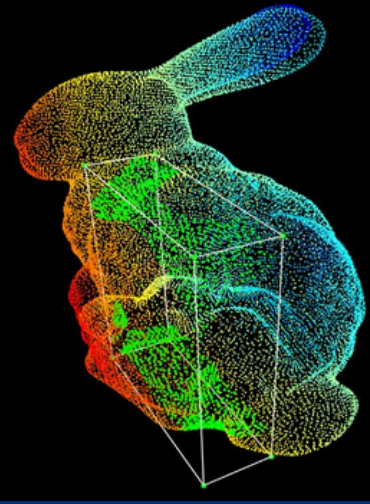
Symmetry Function for Unordered Input

집합의 요소에 대칭 함수를 적용하여 근사하는 것입니다.

h 는 다층 퍼셉트론 네트워크로 근사화하고, g 는 단일 변수 함수와 최대 풀링 함수의 합성으로 근사화합니다.

h 들의 모음을 통해, 우리는 집합의 다양한 속성을 포착하는 여러 f 를 학습할 수 있습니다. 모듈의 단순성 덕분에, 이론적 분석도 제공할 수 있습니다.

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

Deep Learning on
Point Sets

종결 및 전망

4. Deep Learning on Point Sets

PointNet Architecture

$$f(\{x_1, \dots, x_n\}) \approx g(h(x_1), \dots, h(x_n)), \quad (1)$$

where $f : 2^{\mathbb{R}^N} \rightarrow \mathbb{R}$, $h : \mathbb{R}^N \rightarrow \mathbb{R}^K$ and $g : \underbrace{\mathbb{R}^K \times \dots \times \mathbb{R}^K}_n \rightarrow \mathbb{R}$ is a symmetric function.

Local and Global Information Aggregation

위 섹션의 출력은 벡터 $[f_1, \dots, f_K]$ 를 형성하며, 이는 입력 집합의 전역적인 특징입니다. 이 전역적인 특징을 사용하여 SVM 또는 다층 퍼셉트론 분류기를 쉽게 학습시킬 수 있습니다.

하지만 점 분할에는 지역적 지식과 전역적 지식의 조합이 필요합니다.

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

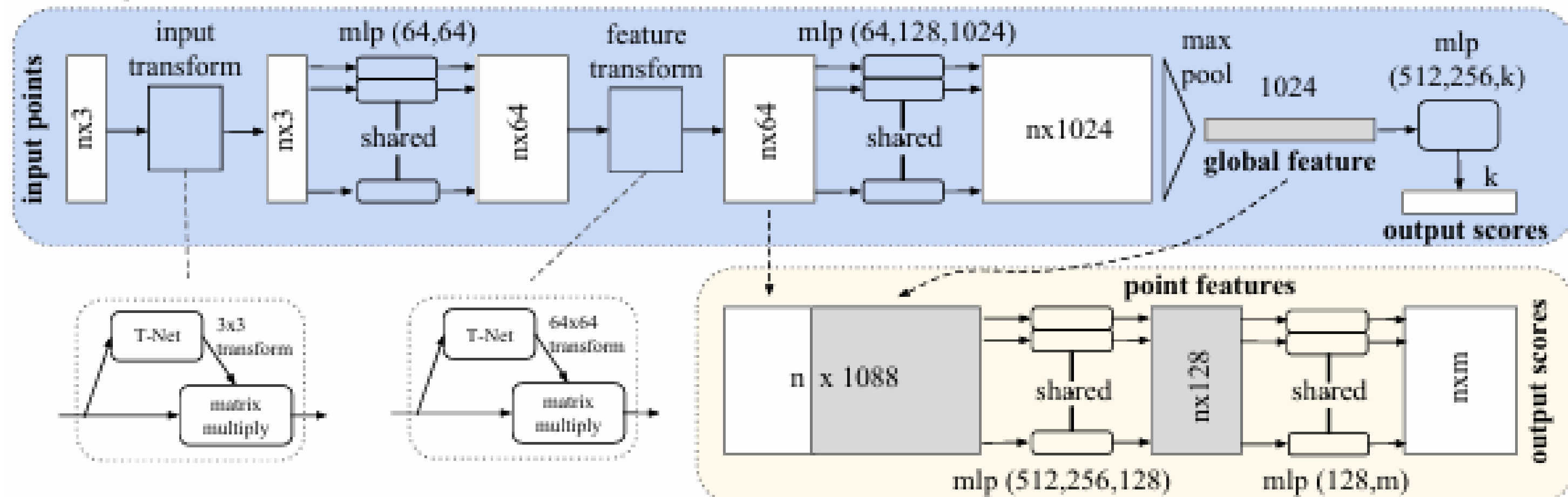
Deep Learning on
Point Sets

종결 및 전망

4. Deep Learning on Point Sets

PointNet Architecture

Classification Network



과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

Deep Learning on
Point Sets

종결 및 전망

Local and Global Information Aggregation

전역 포인트 클라우드 특징 벡터를 계산한 후,

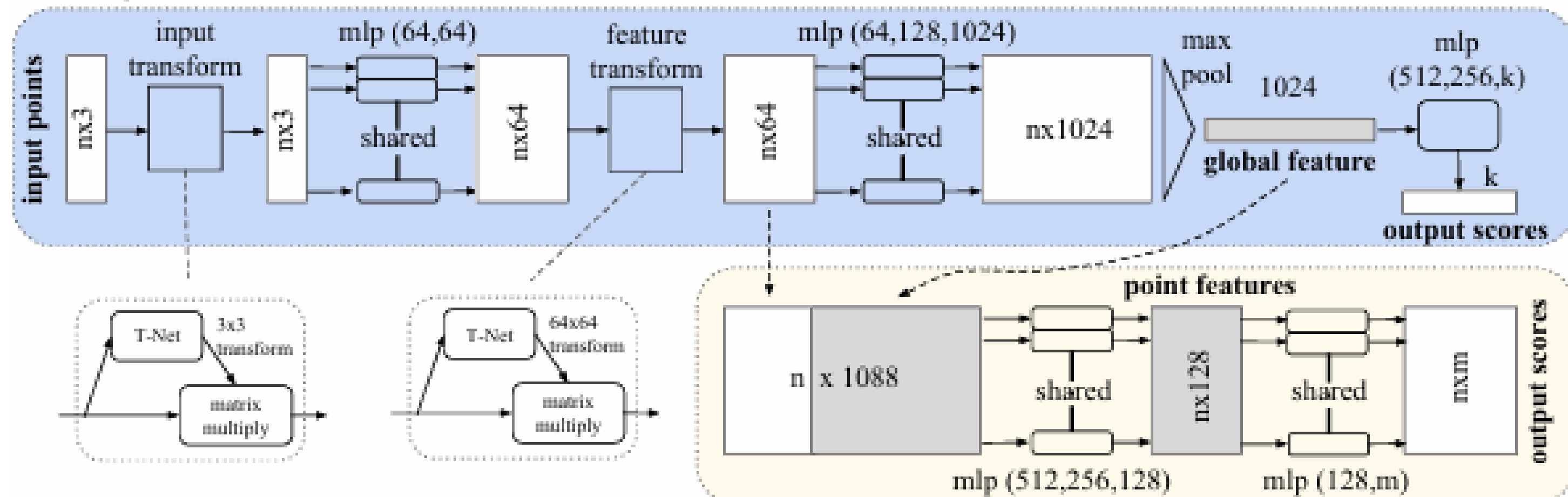
전역 특징 벡터를 각 포인트 특징 벡터와 연결하여 포인트별 특징으로 변환합니다.

그런 다음, 결합된 포인트 특징 벡터를 기반으로 새로운 포인트별 특징을 추출합니다. 이때 포인트별 특징은 지역 정보와 전역 정보를 모두 인식하게 됩니다.

4. Deep Learning on Point Sets

PointNet Architecture

Classification Network



Local and Global Information Aggregation

지역 기하학적 정보와 전역 의미론적 정보를 모두 활용하는 각 지점별 값을 예측할 수 있습니다. 예를 들어, 각 지점의 법선 벡터를 정확하게 예측할 수 있으며 이는 네트워크가 해당 지점의 주변 영역에서 정보를 요약할 수 있음을 입증합니다.

과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

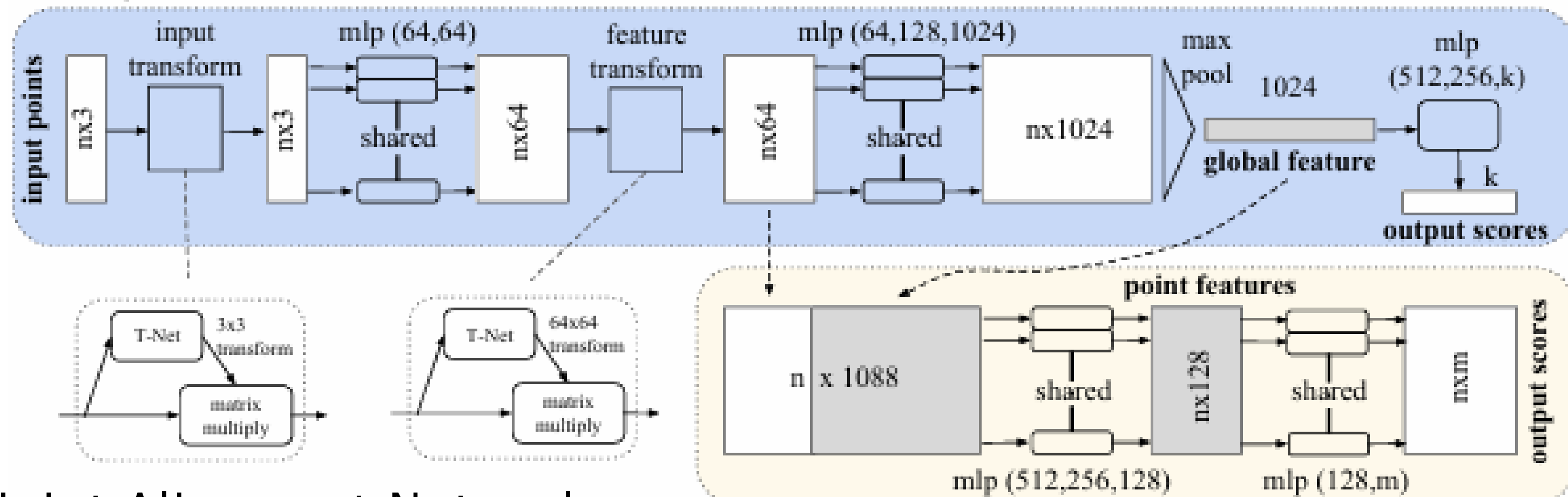
Deep Learning on
Point Sets

종결 및 전망

4. Deep Learning on Point Sets

PointNet Architecture

Classification Network



Joint Alignment Network

포인트 클라우드의 의미 레이블링은 변환을 겪는 경우에도 불변해야 합니다.

포인트 세트에 의해 학습된 표현은 이러한 변환에 대해 불변해야 합니다.

특징 추출 전에 모든 입력 세트를 정규 공간에 정렬하는 것입니다.

기존 방법은 아래와 같은 논문을 통해 해결.

M. Jaderberg, K. Simonyan, A. Zisserman, et al. Spatial transformer networks. In *NIPS 2015*. 4

과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

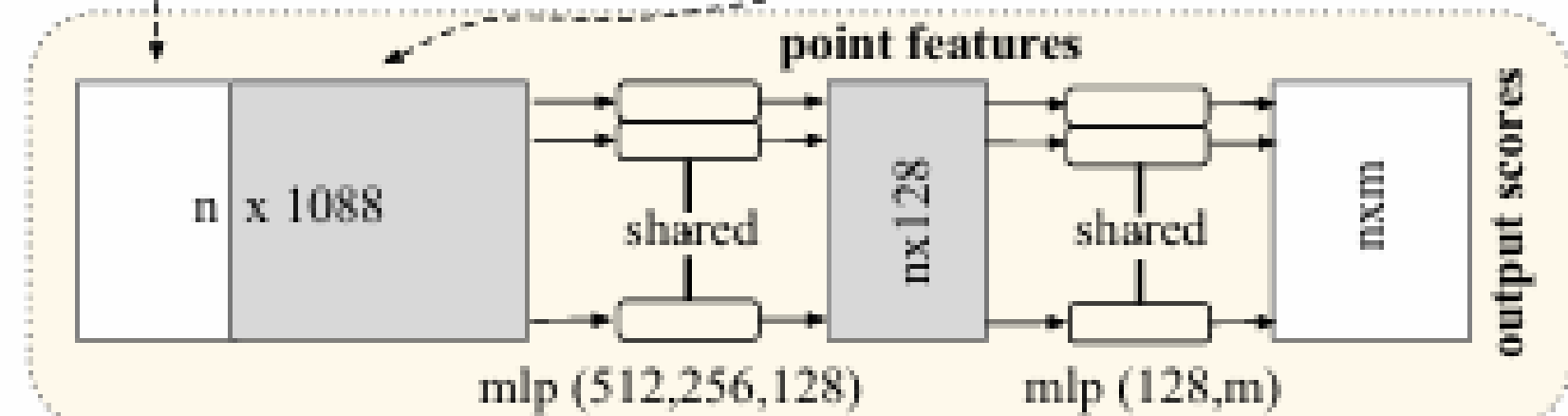
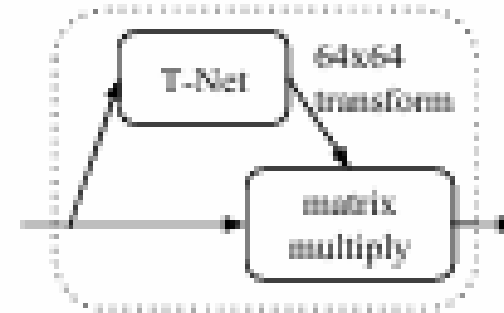
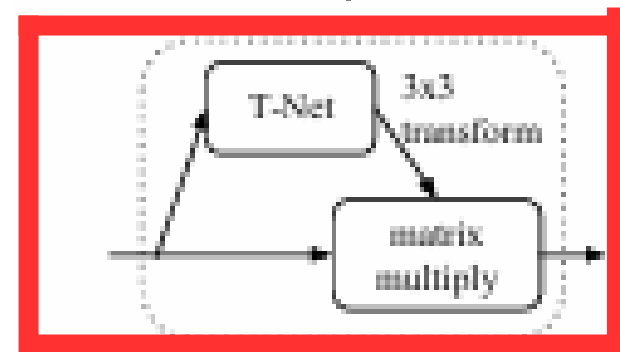
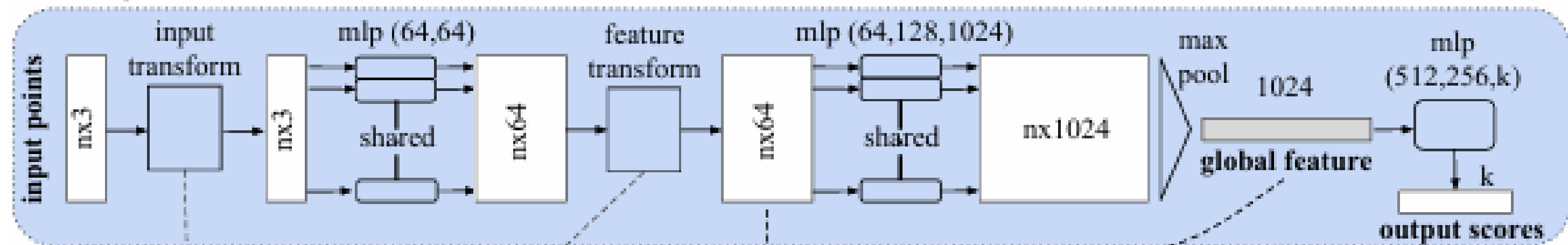
Deep Learning on
Point Sets

종결 및 전망

4. Deep Learning on Point Sets

PointNet Architecture

Classification Network



Joint Alignment Network

새로운 레이어를 만들 필요가 없으며 이미지의 경우처럼 에일리어싱도 도입되지 않습니다.

미니 네트워크(T-net)를 통해 아핀 변환 행렬을 예측하고 이 변환을 입력 포인트의 좌표에 직접 적용합니다.

미니 네트워크 자체는 큰 네트워크와 유사하며 포인트 독립 특징 추출, 최대 풀링 및 완전 연결 레이어의 기본 모듈로 구성됩니다.

과제 선택 배경 및 의미

연구 내용 및 예상 목표

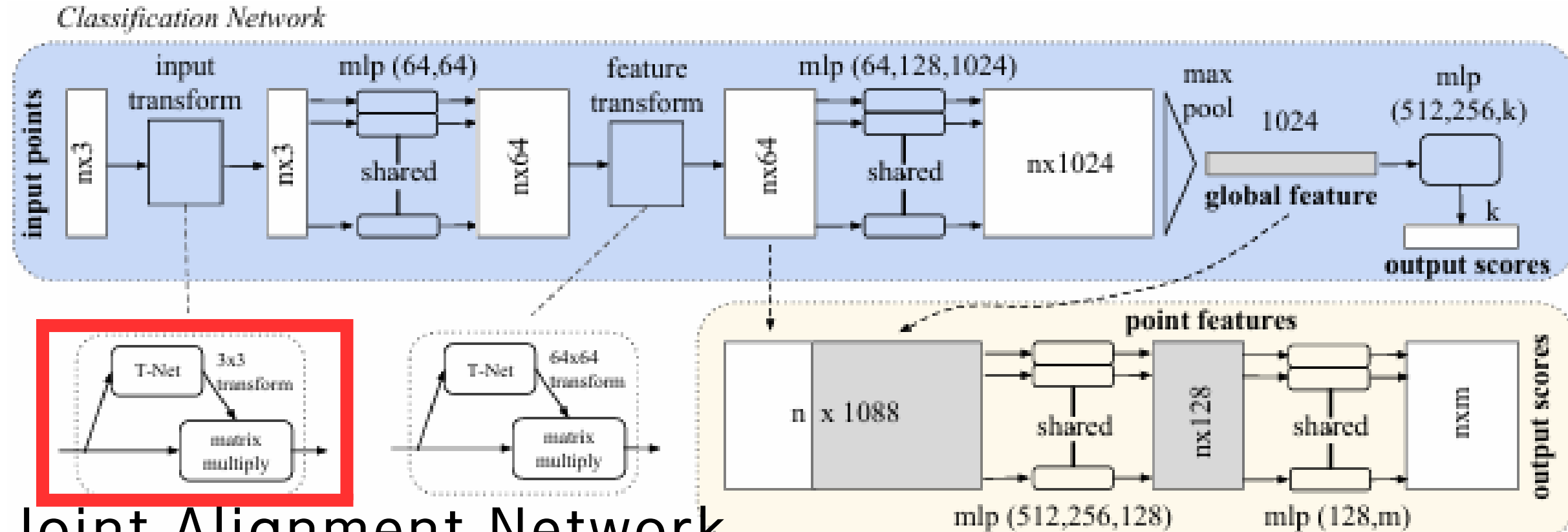
연구 방법 및 단계

Deep Learning on
Point Sets

종결 및 전망

4. Deep Learning on Point Sets

PointNet Architecture



Joint Alignment Network

이 아이디어는 특징 공간(feature space)의 정렬에도 확장될 수 있습니다. 점 특징에 대한 또 다른 정렬 네트워크를 삽입하고 특징 변환 행렬을 예측하여 서로 다른 입력 점 구름의 특징을 정렬할 수 있습니다. 그러나 특징 공간(feature space)의 변환 행렬은 공간 변환 행렬보다 차원이 훨씬 높기 때문에 최적화 난이도가 크게 증가합니다. 따라서 소프트맥스 학습 손실에 정규화 항을 추가합니다. 특징 변환 행렬이 직교 행렬에 가깝도록 제약 조건을 적용합니다.

과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

Deep Learning on
Point Sets

종결 및 전망

4. Deep Learning on Point Sets

PointNet Architecture

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



$$L_{reg} = \|I - AA^T\|_F^2,$$

Joint Alignment Network

A는 미니 네트워크가 예측한 특징 정렬 행렬입니다.

직교 변환은 입력의 정보를 손실하지 않으므로 바람직합니다.

정규화 항을 추가함으로써 최적화가 더욱 안정되고 모델의 성능이 향상됨을 확인했습니다.

과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

Deep Learning on
Point Sets

종결 및 전망

4. Deep Learning on Point Sets

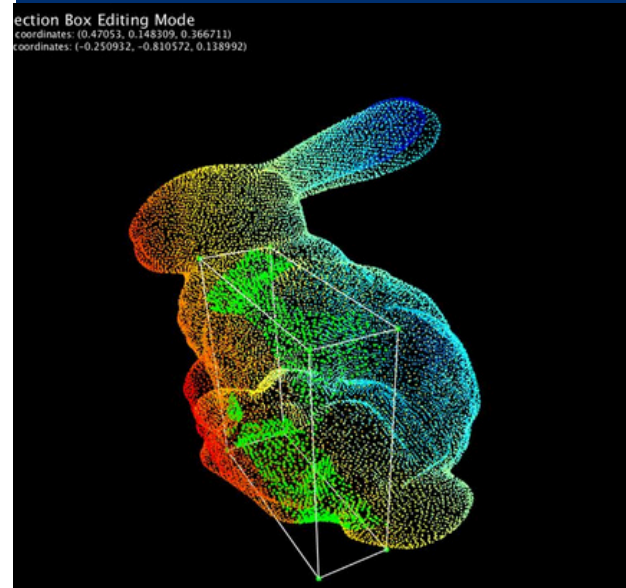
PointNet Architecture

We define $u = \text{MAX}_{x_i \in S} \{h(x_i)\}$ to be the sub-network of f

병목 현상의 규모 및 안정성

실험적으로 우리는 우리 네트워크의 표현력이 최대 풀링 레이어의 차원, 즉 K 에 의해 크게 영향을 받는다는 것을 발견했습니다. 여기서 우리는 분석을 제공하며, 이는 또한 우리 모델의 안정성과 관련된 속성을 드러냅니다.

$[0, 1]^m$ 범위의 점 집합을 K 차원 벡터로 매핑하는 함수입니다.



과제 선택 배경 및 의미

연구 내용 및 예상 목표

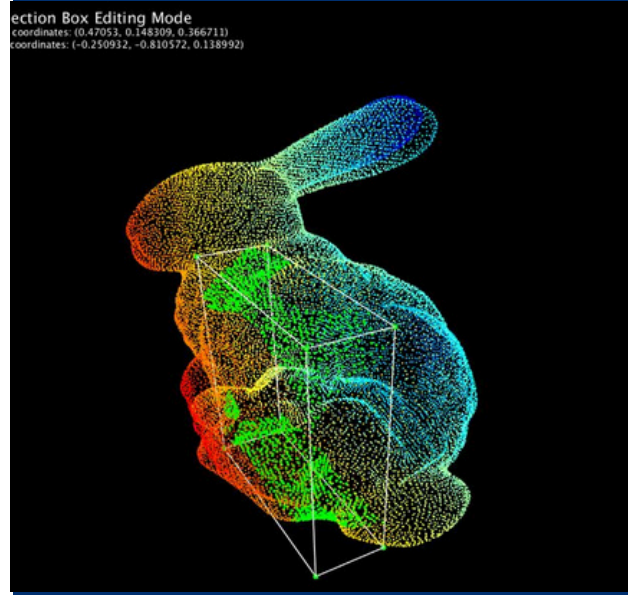
연구 방법 및 단계

Deep Learning on
Point Sets

종결 및 전망

4. Deep Learning on Point Sets

PointNet Architecture



	input	#views	accuracy avg. class	accuracy overall
SPH [11]	mesh	-	68.2	-
3DShapeNets [28]	volume	1	77.3	84.7
VoxNet [17]	volume	12	83.0	85.9
Subvolume [18]	volume	20	86.0	89.2
LFD [28]	image	10	75.5	-
MVCNN [23]	image	80	90.1	-
Ours baseline	point	-	72.6	77.4
Ours PointNet	point	1	86.2	89.2

Theorem 2. Suppose $\mathbf{u} : \mathcal{X} \rightarrow \mathbb{R}^K$ such that $\mathbf{u} = \text{MAX}_{x_i \in S} \{h(x_i)\}$ and $f = \gamma \circ \mathbf{u}$. Then,

(a) $\forall S, \exists \mathcal{C}_S, \mathcal{N}_S \subseteq \mathcal{X}, f(T) = f(S)$ if $\mathcal{C}_S \subseteq T \subseteq \mathcal{N}_S$;

(b) $|\mathcal{C}_S| \leq K$

과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

Deep Learning on
Point Sets

종결 및 전망

$f(S)$ 는 입력 손상에 대해 $\mathcal{C}_S$ 의 모든 점이 보존되는 경우 변경되지 않습니다. 또한 추가 노이즈 점이 $\mathcal{N}_S$ 까지 있는 경우에도 변경되지 않습니다. (b)는 $\mathcal{C}_S$ 가 (1)의 K 에 의해 결정되는 유한한 수의 점만 포함한다는 것을 의미합니다. 즉,

$f(S)$ 는 실제로 K 개 이하의 요소를 가진 유한 부분집합 $\mathcal{C}_S \subseteq S$ 에 의해 완전히 결정됩니다.

따라서 $\mathcal{C}_S$ 를 S 의 임계점 집합이라고 하고 K 를 f 의 병목 차원이라고 합니다.

h 의 연속성과 결합하여, 이는 점 교란, 손상 및 추가 노이즈 점에 대한 우리 모델의 견고성을 설명합니다.

네트워크는 희소한 키 포인트 집합을 통해 형태를 요약하는 방법을 학습합니다.

5. Experience

PointNet Architecture

	mean	aero	bag	cap	car	chair	ear phone	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate board	table
# shapes		2690	76	55	898	3758	69	787	392	1547	451	202	184	283	66	152	5271
Wu [27]	-	63.2	-	-	-	73.5	-	-	-	74.4	-	-	-	-	-	-	74.8
Yi [29]	81.4	81.0	78.4	77.7	75.7	87.6	61.9	92.0	85.4	82.5	95.7	70.6	91.9	85.9	53.1	69.8	75.3
3DCNN	79.4	75.1	72.8	73.3	70.0	87.2	63.5	88.4	79.6	74.4	93.9	58.7	91.8	76.4	51.2	65.3	77.1
Ours	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6

Table 2. Segmentation results on ShapeNet part dataset. Metric is mIoU(%) on points. We compare with two traditional methods [27] and [29] and a 3D fully convolutional network baseline proposed by us. Our PointNet method achieved the state-of-the-art in mIoU.

PointNet이 다양한 3D 인식 작업에 적용될 수 있음

네트워크가 3D 객체 분류, 객체 부분 분할, 그리고 의미론적 장면 분할을 수행하도록 학습되는 방법을 보여줍니다.

40개의 인공 객체 범주에서 가져온 12,311개의 CAD 모델이 있으며, 이 중 9,843개는 훈련용으로, 2,468개는 테스트용으로 나뉩니다.

메쉬 면에서 면적에 따라 1024개의 점을 균일하게 샘플링하고 이를 단위 구로 정규화합니다.

학습 과정에서 점 구름을 실시간으로 증강하기 위해 객체를 위쪽 축을 따라 무작위로 회전시키고 각 점의 위치를 평균이 0이고 표준 편차가 0.02인 가우시안 노이즈로 필터링합니다.

위 표는 기존 방법과 우리의 PointNet을 비교합니다.

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

연구 성과 및 사용

Experience

5. Experience

PointNet Architecture

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



	mean	aero	bag	cap	car	chair	ear phone	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate board	table
# shapes		2690	76	55	898	3758	69	787	392	1547	451	202	184	283	66	152	5271
Wu [27]	-	63.2	-	-	-	73.5	-	-	-	74.4	-	-	-	-	-	-	74.8
Yi [29]	81.4	81.0	78.4	77.7	75.7	87.6	61.9	92.0	85.4	82.5	95.7	70.6	91.9	85.9	53.1	69.8	75.3
3DCNN	79.4	75.1	72.8	73.3	70.0	87.2	63.5	88.4	79.6	74.4	93.9	58.7	91.8	76.4	51.2	65.3	77.1
Ours	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6

Table 2. Segmentation results on ShapeNet part dataset. Metric is mIoU(%) on points. We compare with two traditional methods [27] and [29] and a 3D fully convolutional network baseline proposed by us. Our PointNet method achieved the state-of-the-art in mIoU.

PointNet이 다양한 3D 인식 작업에 적용될 수 있음

범주별 및 평균 IoU(%) 점수를 보고합니다.

평균 IoU가 2.3% 향상되었으며,

본 연구 결과는 대부분의 범주에서 기존 방법보다 우수한 성능을 보였습니다.

과제 선택 배경 및 의미

연구 내용 및 예상 목표

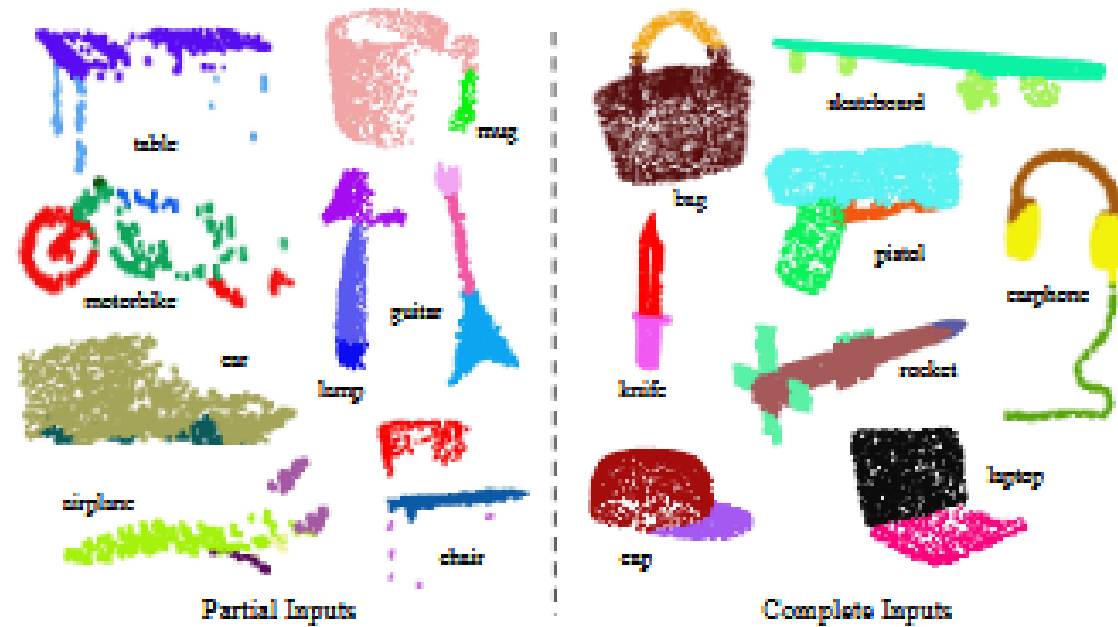
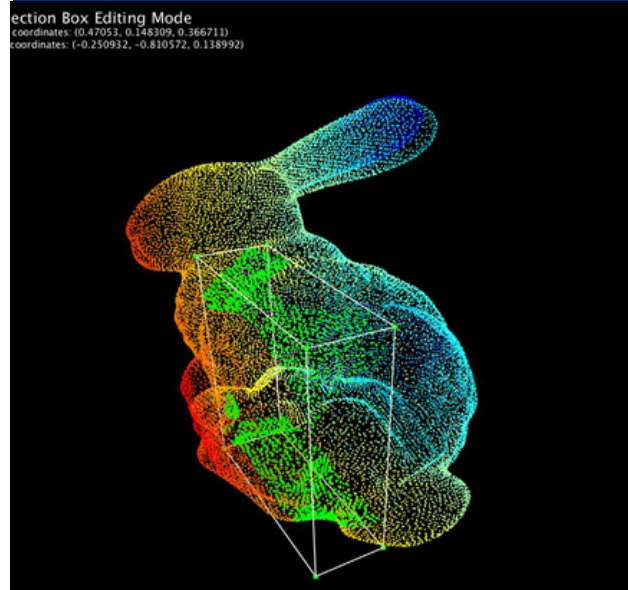
연구 방법 및 단계

연구 성과 및 사용

Experience

5. Experience

PointNet Architecture



과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

연구 성과 및 사용

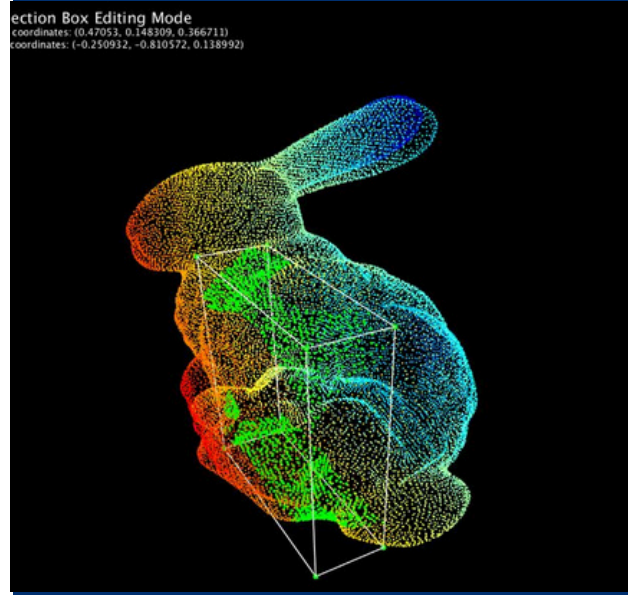
PointNet이 다양한 3D 인식 작업에 적용될 수 있음

ShapeNet 부품 데이터 세트의 모든 CAD 모델에 대해 Blensor Kinect 시뮬레이터[7]를 사용하여 6개의 임의 시점에서 불완전한 포인트 클라우드를 생성합니다. 동일한 네트워크 아키텍처와 훈련 설정을 사용하여 완전한 형상과 부분 스캔에 대해 PointNet을 훈련합니다. 결과는 평균 IoU가 5.3%만 손실됨을 보여줍니다. 그림 3에서는 완전한 데이터와 부분 데이터 모두에 대한 정성적 결과를 제시합니다. 부분 데이터는 상당히 어렵지만 예측 결과가 합리적임을 알 수 있습니다.

Experience

5. Experience

PointNet Architecture



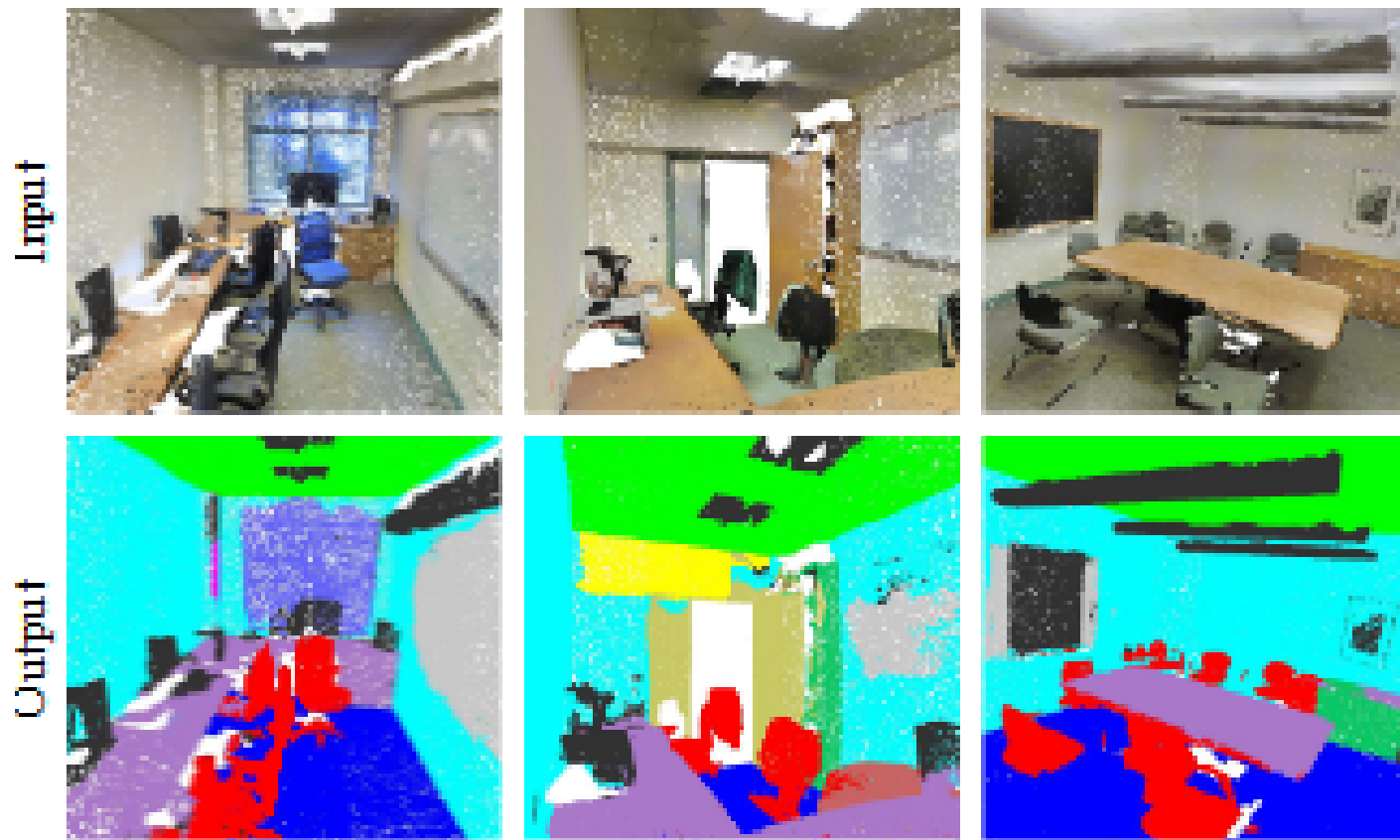
과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

연구 성과 및 사용

Experience



I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese. 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2016. 6, 7

PointNet이 다양한 3D 인식 작업에 적용될 수 있음

점들을 방별로 분할하고,

그다음 방들을 1m x 1m 크기의 블록으로 샘플링합니다.

각 블록에서 각 점의 클래스를 예측하도록 PointNet의 분할 버전을 훈련합니다. 각 점은 XYZ, RGB 및 방에 대한 정규화된 위치(0에서 1까지)를 포함하는 9차원 벡터로 표현됩니다.

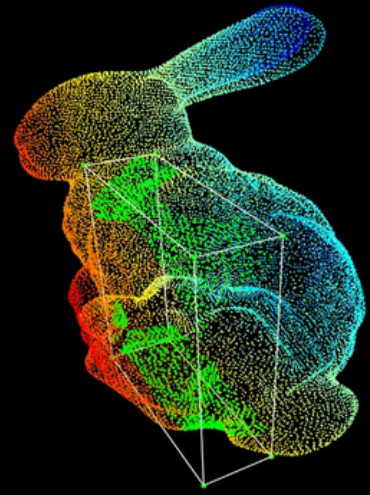
훈련 시에는 각 블록에서 4096개의 점을 무작위로 샘플링합니다.

테스트 시에는 모든 점을 대상으로 테스트합니다. 위 논문과 동일한 프로토콜을 따르며, 훈련과 테스트에 k-겹 교차 검증 전략을 사용합니다.

5. Experience

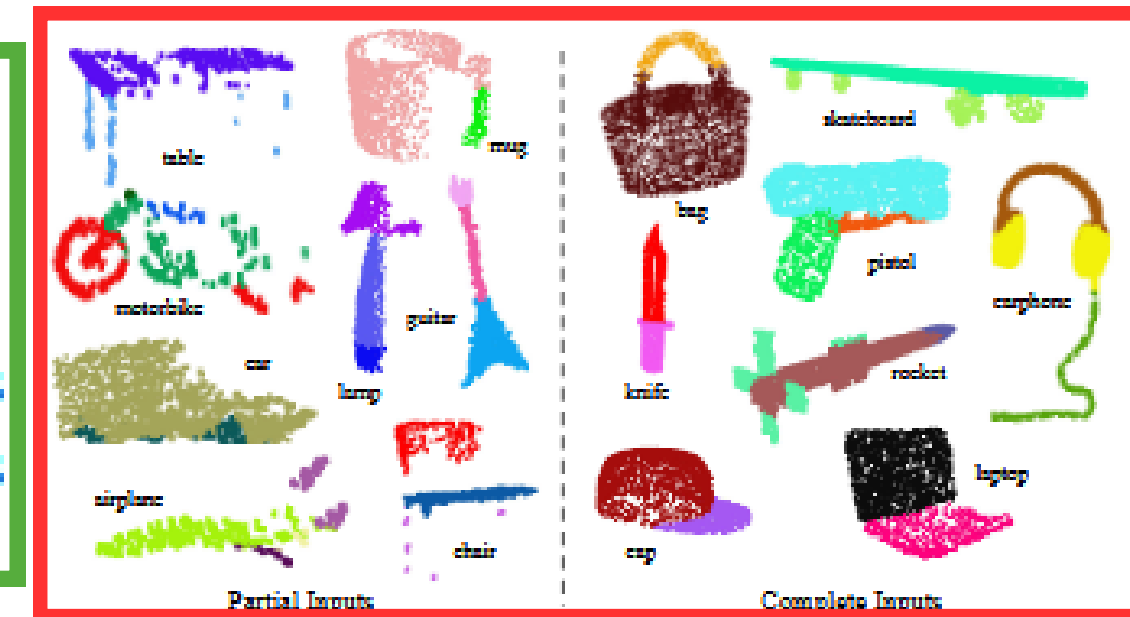
PointNet Architecture

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



	mean IoU	overall accuracy
Ours baseline	20.12	53.19
Ours PointNet	47.71	78.62

Table 3. Results on semantic segmentation in scenes. Metric is average IoU over 13 classes (structural and furniture elements plus clutter) and classification accuracy calculated on points.



과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

연구 성과 및 사용

PointNet이 다양한 3D 인식 작업에 적용될 수 있음

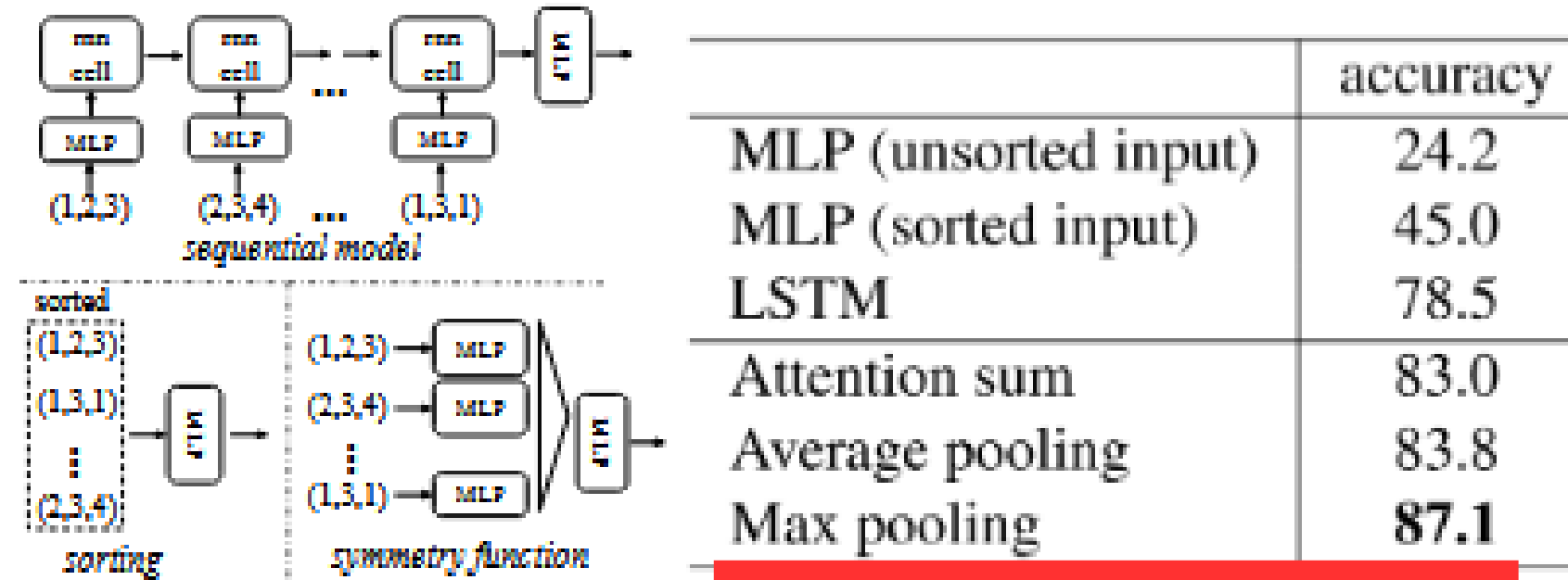
기존 모델은 동일한 9차원 로컬 특징과 세 가지 추가 특징(로컬 포인트 밀도, 로컬 곡률, 법선 벡터)을 추출합니다. 분류기로는 표준 MLP를 사용합니다. 결과는 왼쪽 이미지와 같습니다.

PointNet 방법이 기존 모델보다 훨씬 우수한 성능을 보입니다. 오른쪽 이미지에서는 정성적 분할 결과를 보여줍니다.

Experience

5. Experience

PointNet Architecture



다른 순서 불변 방법과의 비교

Figure 5. Three approaches to achieve order invariance. Multi-layer perceptron (MLP) applied on points consists of 5 hidden layers with neuron sizes 64,64,64,128,1024, all points share a single copy of MLP. The MLP close to the output consists of two layers with sizes 512,256.

정렬되지 않은 점과 정렬된 점을 $n \times 3$ 배열로 처리하는 다층 퍼셉트론, 입력 점을 시퀀스로 고려하는 RNN 모델, 그리고 대칭 함수 기반 모델이 포함됩니다.

실험에 사용한 대칭 연산에는 최대 풀링, 평균 풀링, 그리고 어텐션 기반 가중 합이 있습니다. 각 점 특징에서 스칼라 점수를 예측한 다음, 소프트맥스를 계산하여 점들 간에 점수를 정규화합니다. 그런 다음 정규화된 점수와 점 특징에 대해 가중 합을 계산합니다. 위에서 볼 수 있듯이, 최대 풀링 연산이 큰 차이로 최고의 성능을 달성하여 우리의 선택이 옳았음을 입증합니다.

과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

연구 성과 및 사용

Experience

5. Experience

PointNet Architecture

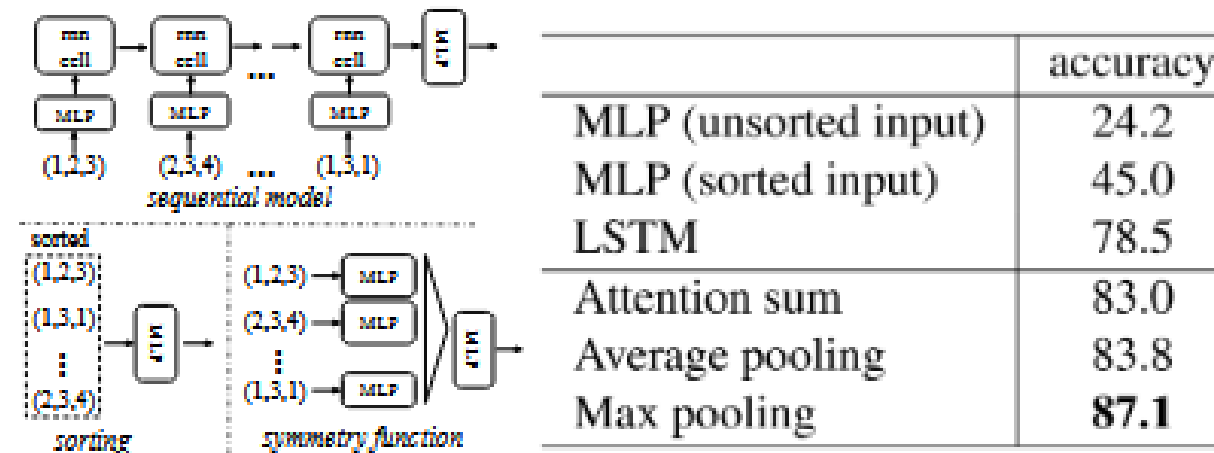
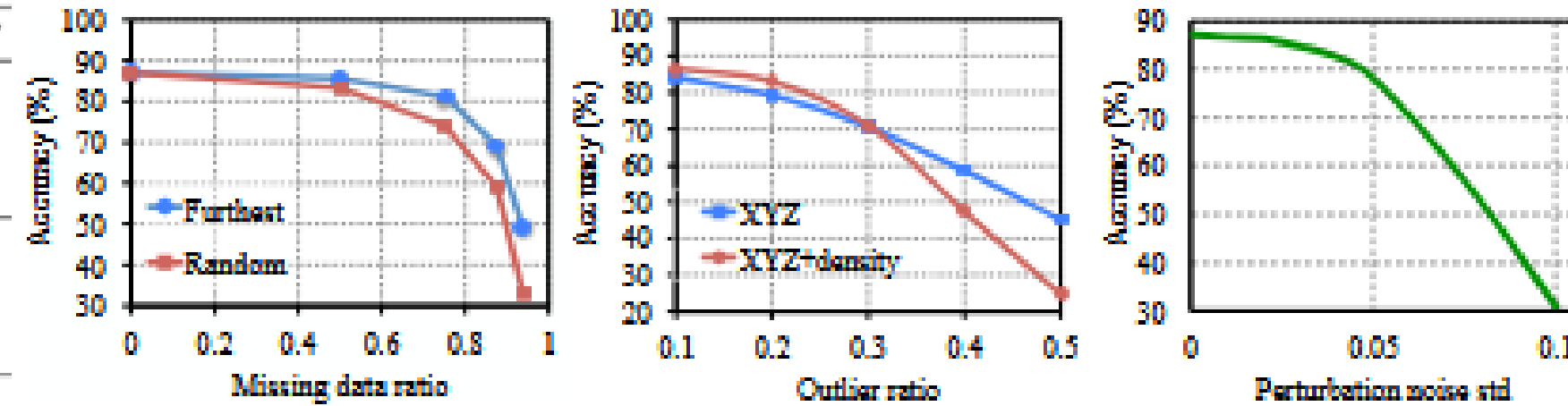


Figure 5. Three approaches to achieve order invariance. Multi-layer perceptron (MLP) applied on points consists of 5 hidden layers with neuron sizes 64,64,64,128,1024, all points share a single copy of MLP. The MLP close to the output consists of two layers with sizes 512,256.



Robustness Test

맥스 풀링 네트워크와 동일한 아키텍처를 사용합니다. 입력 포인트는 단위 구로 정규화됩니다. 결과는 오른쪽 이미지에 나와 있습니다.

결측 포인트의 경우, 50%의 포인트가 결측되었을 때, 정확도는 가장 먼 지점 샘플링과 무작위 입력 샘플링에 비해 각각 2.4%와 3.8%만 감소합니다. 또한 본 네트워크는 이상치에도 강건합니다.

과제 선택 배경 및 의미

연구 내용 및 예상 목표

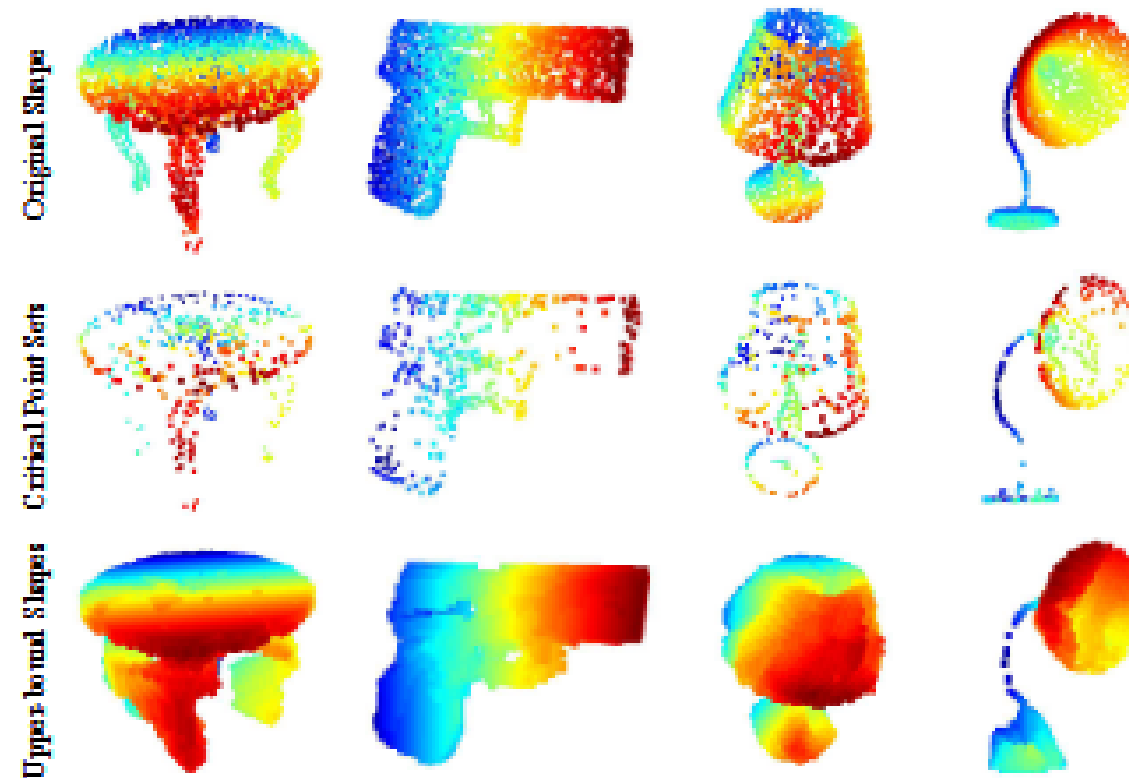
연구 방법 및 단계

연구 성과 및 사용

Experience

5. Experience

PointNet Architecture



	#params	FLOPs/sample
PointNet (vanilla)	0.8M	148M
PointNet	3.5M	440M
Subvolume [18]	16.6M	3633M
MVCNN [23]	60.0M	62057M

임계점 집합 CS(최대 풀링된 특징에 기여한 점들)가 형상의 골격을 요약한다는 것을 명확하게 알 수 있습니다. 상한 형상 NS는 입력 점 구름 S와 동일한 전역 형상 특징 $f(S)$ 를 제공하는 가능한 가장 큰 점 구름을 나타냅니다. CS와 NS는 PointNet의 견고성을 반영합니다. 즉, 일부 중요하지 않은 점을 잃더라도 전역 형상 특징 $f(S)$ 는 전혀 변하지 않습니다.

MVCNN 및 Subvolume(3D CNN)은 높은 성능을 달성하지만, PointNet은 계산 비용 측면에서 훨씬 효율적입니다

또한 PointNet은 네트워크의 매개변수 개수 측면에서 MVCNN보다 공간 효율성이 훨씬 높습니다 더 나아가 PointNet은 확장성이 훨씬 뛰어납니다. 공간 및 시간 복잡도는 $O(N)$ 으로 입력 포인트 수에 비례합니다.

과제 선택 배경 및 의미

연구 내용 및 예상 목표

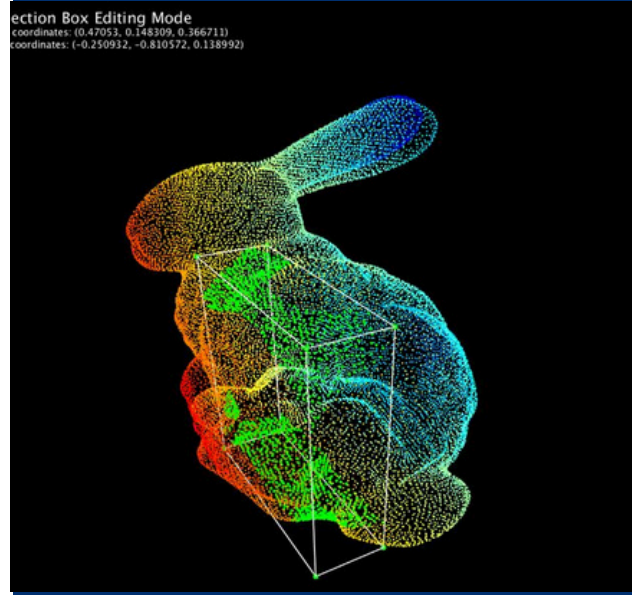
연구 방법 및 단계

연구 성과 및 사용

Experience

5. Experience

PointNet Architecture

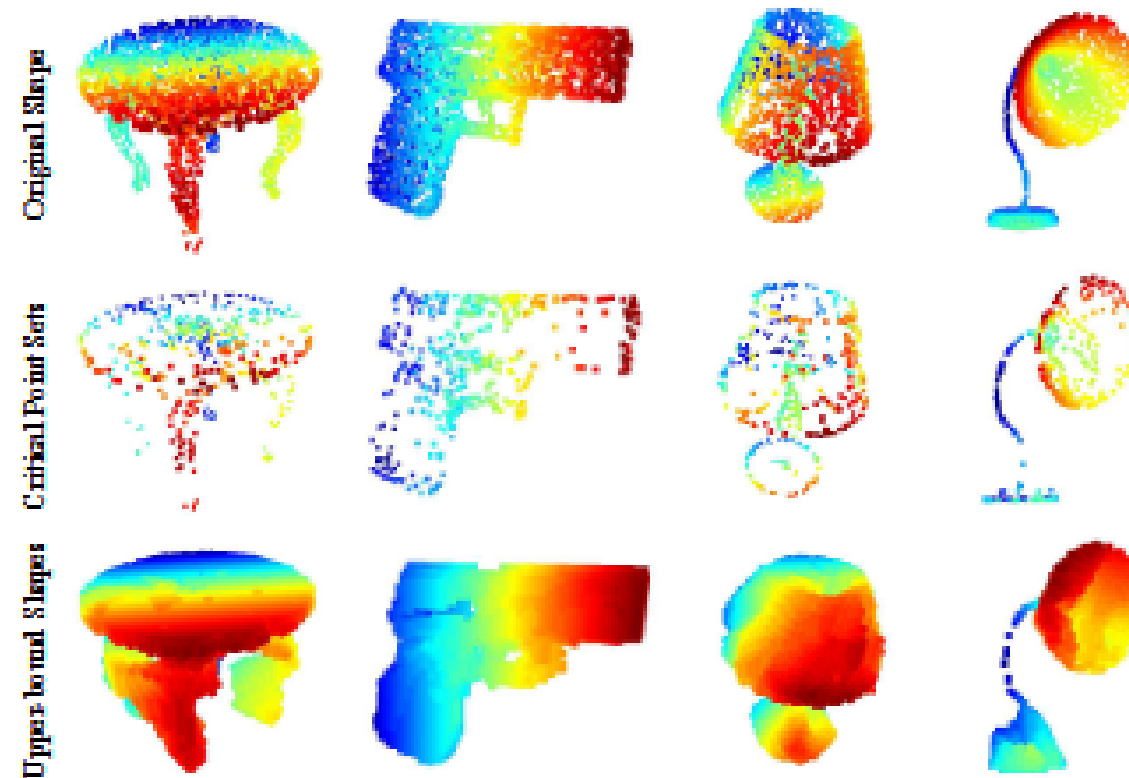


과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

연구 성과 및 사용



	#params	FLOPs/sample
PointNet (vanilla)	0.8M	148M
PointNet	3.5M	440M
Subvolume [18]	16.6M	3633M
MVCNN [23]	60.0M	62057M

PointNet은 TensorFlow 기반의 1080X GPU를 사용하여 포인트 클라우드 분류(초당 약 1,000개 객체) 또는 의미론적 분할(초당 약 2개 방)을 위해 초당 백만 개 이상의 포인트를 처리할 수 있습니다. 이는 실시간 애플리케이션에 대한 큰 잠재력을 보여줍니다.

Experience

5. Experience

PointNet Architecture

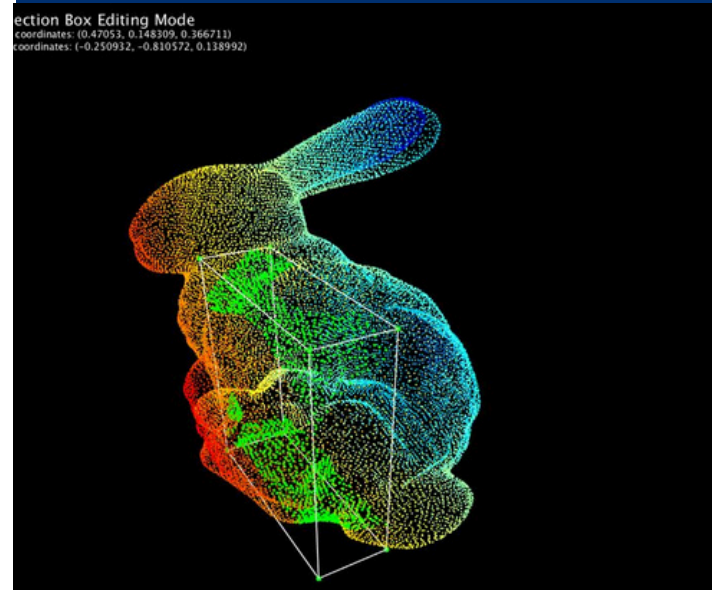
Transform	accuracy
none	87.1
input (3x3)	87.9
feature (64x64)	86.9
feature (64x64) + reg.	87.4
both	89.2

입력 및 기능의 효율성

입력 및 특징 변환(정렬용)의 긍정적인 효과를 확인할 수 있습니다.

가장 기본적인 아키텍처에서도 이미 상당히 만족스러운 결과를 얻을 수 있다는 점이 흥미롭습니다. 입력 변환을 사용하면 성능이 0.8% 향상됩니다. 정규화 손실은 고차원 변환이 제대로 작동하기 위해 필요합니다.

두 변환과 정규화 항을 결합하면 최상의 성능을 얻을 수 있습니다.



과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

연구 성과 및 사용

Experience

5. Experience

PointNet Architecture

	mean	aero	bag	cap	car	chair	ear phone	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate board	table
# shapes		2690	76	55	898	3758	69	787	392	1547	451	202	184	283	66	152	5271
Wu [27]	-	63.2	-	-	-	73.5	-	-	-	74.4	-	-	-	-	-	-	74.8
Yi [29]	81.4	81.0	78.4	77.7	75.7	87.6	61.9	92.0	85.4	82.5	95.7	70.6	91.9	85.9	53.1	69.8	75.3
3DCNN	79.4	75.1	72.8	73.3	70.0	87.2	63.5	88.4	79.6	74.4	93.9	58.7	91.8	76.4	51.2	65.3	77.1
Ours	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6

Table 2. Segmentation results on ShapeNet part dataset. Metric is mIoU(%) on points. We compare with two traditional methods [27] and [29] and a 3D fully convolutional network baseline proposed by us. Our PointNet method achieved the state-of-the-art in mIoU.

네트워크가 학습한 내용을 시각화

$f(S)$ 는 입력 손상에 대해 CS 의 모든 점이 보존되는 경우 변경되지 않습니다. 또한 추가 노이즈 점이 NS 까지 있는 경우에도 변경되지 않습니다. (b)는 CS 가 (1)의 K 에 의해 결정되는 유한한 수의 점만 포함한다는 것을 의미합니다. 즉,

$f(S)$ 는 실제로 K 개 이하의 요소를 가진 유한 부분집합 $CS \subseteq S$ 에 의해 완전히 결정됩니다.

따라서 CS 를 S 의 임계점 집합이라고 하고 K 를 f 의 병목 차원이라고 합니다.

h 의 연속성과 결합하여, 이는 점 교란, 손상 및 추가 노이즈 점에 대한 우리 모델의 견고성을 설명합니다.

네트워크는 희소한 키 포인트 집합을 통해 형태를 요약하는 방법을 학습합니다.

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

연구 성과 및 사용

Experience

5. Experience

PointNet Architecture

Section Box Editing Mode
coordinates: (0.47051, 0.148309, 0.366711)
coordinates: (-0.250932, -0.810572, 0.118992)



	mean	aero	bag	cap	car	chair	ear phone	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate board	table
# shapes		2690	76	55	898	3758	69	787	392	1547	451	202	184	283	66	152	5271
Wu [27]	-	63.2	-	-	-	73.5	-	-	-	74.4	-	-	-	-	-	-	74.8
Yi [29]	81.4	81.0	78.4	77.7	75.7	87.6	61.9	92.0	85.4	82.5	95.7	70.6	91.9	85.9	53.1	69.8	75.3
3DCNN	79.4	75.1	72.8	73.3	70.0	87.2	63.5	88.4	79.6	74.4	93.9	58.7	91.8	76.4	51.2	65.3	77.1
Ours	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6

Table 2. Segmentation results on ShapeNet part dataset. Metric is mIoU(%) on points. We compare with two traditional methods [27] and [29] and a 3D fully convolutional network baseline proposed by us. Our PointNet method achieved the state-of-the-art in mIoU.

시간 및 공간 복잡도를 분석

$f(S)$ 는 입력 손상에 대해 CS 의 모든 점이 보존되는 경우 변경되지 않습니다. 또한 추가 노이즈 점이 NS 까지 있는 경우에도 변경되지 않습니다. (b)는 CS 가 (1)의 K 에 의해 결정되는 유한한 수의 점만 포함한다는 것을 의미합니다. 즉,

$f(S)$ 는 실제로 K 개 이하의 요소를 가진 유한 부분집합 $CS \subseteq S$ 에 의해 완전히 결정됩니다.

따라서 CS 를 S 의 임계점 집합이라고 하고 K 를 f 의 병목 차원이라고 합니다.

h 의 연속성과 결합하여, 이는 점 교란, 손상 및 추가 노이즈 점에 대한 우리 모델의 견고성을 설명합니다.

네트워크는 희소한 키 포인트 집합을 통해 형태를 요약하는 방법을 학습합니다.

과제 선택 배경 및 의미

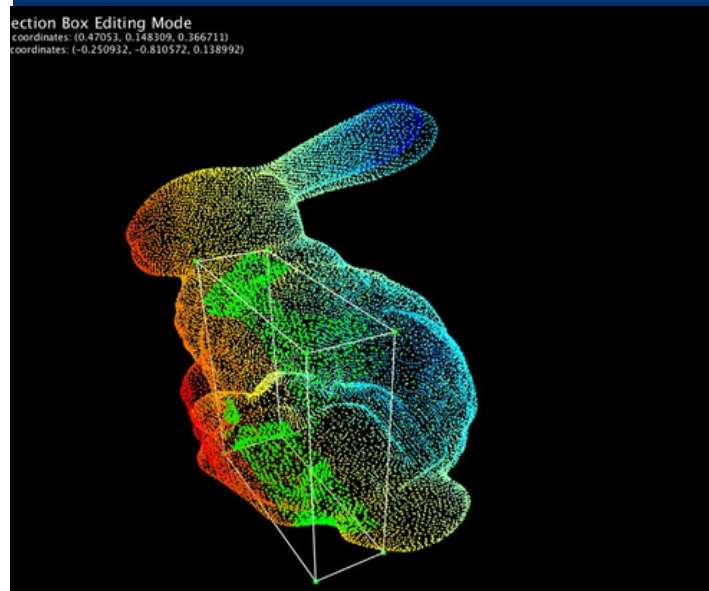
연구 내용 및 예상 목표

연구 방법 및 단계

연구 성과 및 사용

Experience

5. 결론



본 연구에서는 포인트 클라우드를 직접 입력으로 사용하는 새로운 심층 신경망 PointNet을 제안합니다. 제안된 네트워크는 객체 분류, 부분 분할, 의미론적 분할 등 다양한 3D 인식 작업에 대한 통합적인 접근 방식을 제공하며, 표준 벤치마크에서 최첨단 기술과 동등하거나 그 이상의 성능을 달성합니다. 또한, 본 네트워크의 작동 원리를 이해하기 위한 이론적 분석 및 시각화 자료를 제공합니다.

과제 선택 배경 및 의미

연구 내용 및 예상 목표

연구 방법 및 단계

연구 성과 및 사용

결론

감 사 합 니 다

Thank you