



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

3차원 구름 점 데이터를 위한
회전 불변 지역-전역 표현 학습

Rotation Invariant Local-to-Global
Representation Learning for 3D Point Cloud

2020년 6월

서울대학교 대학원

전기·정보공학부

김서현

3차원 구름 점 데이터를 위한 회전 불변 지역-전역 표현 학습

지도 교수 한보형

이 논문을 공학석사 학위논문으로 제출함
2020년 6월

서울대학교 대학원
전기·정보공학부
김서현

김서현의 공학석사 학위논문을 인준함
2020년 6월

위원장 _____ 이경무 _____ (인)

부위원장 _____ 한보형 _____ (인)

위원 _____ 조남익 _____ (인)

초 록

본 논문은 3차원 구름 점을 이용하는 문제를 해결하기 위한 지역-전역 표현 학습 알고리즘을 제안한다. 제안하는 알고리즘은 구름 점에 적용되는 다양한 기하 변환에 대하여 적절히 다룰 수 있도록 고안되었으며, 특히 다루기 까다로운 회전 변환에 대하여, 데이터 증진 없이 효과적으로 해결할 수 있도록 하였다. 우리 모델은 회전 변환에 강인 하면서 구름 점으로 표현된 3차원 물체에 대한 형태 정보를 기술할 수 있는 기술자를 다중 계층구조로부터 상향식으로 부호화하였다. 이를 기반으로, 그래프 합성곱 신경망을 이용해 다중 계층의 기술자들을 통합한 표현으로부터 물체 인식을 효과적으로 수행하게 된다. 각 계층의 기술자는 3차원 구름 점들을 확률적으로 표집하여 신경망을 이용하여 얻는데, 이러한 확률적 학습법은 과적합 문제를 효과적으로 해결하고, 다양한 입력 값에 대하여 강건한 표현을 얻어낼 수 있게 한다. 본 논문은, 3차원 물체 분류 벤치마크 모델인 모델넷40을 임의의 회전변환으로 증진시킨 데이터셋에서 분류기의 정확도를 측정 한 결과, 본 알고리즘이 최고 수준의 성능을 보이고 있고, 종합적인 실험을 통해 그 특성을 분석하였다.

주요어 : 회전 불변, 3차원 구름 점, 기술자, 확률적 학습법, 그래프
학 번 : 2018-20013

목 차

제 1 장 서론.....	1
제 1 절 연구의 배경 및 내용	1
제 2 장 관련 연구.....	5
제 1 절 딥 러닝을 이용한 3차원 구름 점 인식	3
제 2 절 회전 불변 3차원 구름 점 인식.....	7
제 3 절 3차원 구름 점 작업을 위한 그래프 기반 네트워크	8
제 3 장 회전 불변 표현의 확률적 학습.....	11
제 1 절 기술자 추출	11
제 2 절 기술자 확장.....	14
제 3 절 그래프 기반 추출	16
제 4 장 실험	19
제 1 절 구현 세부 정보.....	19
제 2 절 주요 결과.....	20
제 3 절 오류 분석.....	21
제 4 절 Ablation 실험.....	24
제 5 절 노이즈와 아웃라이어에 대한 강건성	27
제 6 절 정성적 결과.....	28
제 5 장 결론	30
참고문헌.....	31
Abstract.....	34

표 목차

[표 1] 모델넷40 데이터 셋에 대한 3차원 물체 분류 정확도 ..	20
[표 2] 확률적 학습에 대한 기여도	24
[표 3] 서로 다른 방법론에 따른 모델넷40 데이터 셋에 대한 3차원 물체 분류 정확도 (%).....	25

그림 목차

[그림 1] 기술자 추출(descriptor extraction) 과정	6
[그림 2] 확률적 확장 k -NNs 의 예시.....	12
[그림 3] 제안하는 회전 불변 3차원 물체 분류 프레임 워크...	15
[그림 4] 그래프의 확률적 생성 예시	17
[그림 5] 일반 혼동 행렬	22
[그림 6] 정규화 혼동 행렬	23
[그림 7] 섭동 노이즈, 아웃라이어 노이즈에 대한 모델넷40 분류 정확도 평가.....	27
[그림 8] 각 계층 마다 생성된 그래프 예시의 정성적 결과.....	28
[그림 9] 임의로 회전된 물체에 대한 모델들의 예측	29

제 1 장 서 론

제 1 절 연구의 배경 및 내용

3차원 데이터 수집을 위한 하드웨어 기술(예: LiDAR, RGB-D 카메라)의 발전으로 최근 3차원 물체 인식에 대한 그 관심이 매우 높아지고 있다. 자율 주행(automatic driving), 로봇 인지(robot perceptron), 증강 현실(augmented reality) 그리고 3차원 메디컬 이미지 분석(3d medical image analysis) 등 매우 다양한 영역에서 3차원 영상이 점차 활발하게 이용이 되고 있고, 앞으로도 이에 대한 수요는 더욱 커질 것으로 전망 된다. 3차원 영상은 2차원 영상에 하나의 차원이 추가된 데이터이기 때문에, 다양한 공간에 위치한 물체를 이해하는 데 더욱 유리하다. 따라서 많은 어플리케이션에서 더 풍부한 정보를 제공할 수 있다.

입력 3차원 데이터는 구름 점(point cloud), 복셀(voxel), 또는 메쉬(mesh) 등으로 표현 될 수 있다. 3차원 구름 점은 하드웨어 센서(sensor)로부터 얻은 가장 가공되지 않은(raw) 형태의 데이터 포맷(format)이므로, 많은 연구에서 이를 직접적으로 이용해 효율적으로 과제를 수행하고자 한다. 3차원 구름 점 데이터는 본질적으로 희소(sparse) 하고 불규칙적(irregular)인 구조를 가지고 있다. 이러한 특징들은 한 동안 구름 점을 다루는 것을 매우 어렵게 만들었다. 불규칙적인 구조는 2차원 이미지 인식의 큰 성능 향상을 이룬 규칙적인 합성 곱 연산(convolution operation)의 적용이 어려웠고, 순서에 따라 달라지는 데이터 표현 방식은 기하학적 특징을 얻는 과정에서 안정적인 특징 추출을 어렵게 만들었다. 때문에 3차원 공간을 양자화하여 일정한 구조로 변환시키는 복셀화(voxelization)이나 다양한 시점으로부터 2차원 이미지를 다수 얻어내어 2차원 합성 곱 연산을 통해 이미지 특징들을 추출 및 병합하여 사용하는 다중 뷰 기법(multi-view method)을 통해 문제를 해결하고자 하는 방법들이 한 동안 제시되었다. 하지만 이러한 방법들은 불필요한 중복(redundancy)에 의한 상당한 리소스 요구로 문제를 해결하는 데 많은 제약이 있었다.

최근 딥 러닝(deep learning) 기술의 발전과 함께, 3차원 데이터에 이 기술을 적용시키고자 하는 노력들이 이루어졌고, 최근 발표된 포인트 넷[1] 을 시작으로 구름 점 데이터를 직접적으로 사용하는 3차원 물체

인식(3D object recognition)은 놀라운 성과를 보이고 있다 [1-10]. 전역적 특징(global feature) 뿐만 아니라, 구름 점 사이의 에지 벡터(edge vector)를 이용한 지역적 특징(local feature)은 높은 정확도를 달성하는데 크게 기여하였다. 이 뿐만 아니라, 구름 점을 그래프의 노드로 간주하고, 점 들간에 거리를 바탕으로 그래프를 구축하여 전역적, 지역적으로 특징을 추출하는 방법도 제시되고 있다. 가장 최근에는 3차원 공간에 분포한 점들에 대하여 연속적 커널을 정의하여 구름 점들에 대한 특징을 추출하고자 하는 방법들도 제시되고 있다.

심층 신경 망(deep neural network)의 적극적인 사용으로, 3차원 물체 인식에 대한 정확도는 크게 증가하였고, 다양한 태스크에 적용하여 놀라운 성과를 이루었다. 그럼에도 불구하고, 입력에 임의로 적용될 수 있는 기하학적 변환(geometric transformation)에 대한 해결은 여전히 까다롭고 어려운 문제로 남아있다. 구름 점으로 표현된 물체는 실제 응용 환경에서 매우 다양한 형태로 입력될 수 있는데, 이를 다시 말하면 3차원 물체 인식은 기하학적변환에 강건(robust)해야 한다는 뜻이 된다. 그러나, 기존 알고리즘은 직립 객체 자세의 가정에 의존하는 경우가 많고, 현재까지 기하 변환 문제를 온전하게 고려한 프레임워크는 거의 제시된 바 없다. 특히, 회전 변환(rotation transformation)은 특히 실무에서 처리되는 것이 매우 어렵다. 일반적으로 구름 점은 3차원 카르테시안 좌표 계(Cartesian coordinate system)로 표현되기에, 임의의 회전 변환은 입력 값을 완전히 바꾸게 되고, 이는 학습된 모델로 하여금 심각하게 변형된 특징을 추출하게 된다. 이러한 문제를 해결하기 위한 간단한 해결 방법 중 하나는 데이터 증진(data augmentation)을 통해 학습 모델이 최대한 많은 형상 입력을 관측 시키는 방법인데, 이러한 방법은 높은 연산 비용(computational cost)을 필요로 하고 예외 경우(corner case)에 대해 처리하기 어렵다. 데이터 증진으로 모든 회전을 포괄하고, 실제 세계의 모든 예들을 일반화하는 것은 결코 간단한 문제가 아니다.

3차원 물체 인식의 맥락에서, 3차원 구름 점의 기하학적 변환에 대처하는 기법이 몇 가지 존재한다. 예를 들어 공간 변환 네트워크 [1, 2]를 사용하여 입력 지점 좌표를 표준화(canonicalize)하는 방법인데, 이 방법은 기본적으로 3차원 좌표 계 입력을 기반으로 학습이 이루어지고, 이에 따라, 회전 변환에 근본적으로 취약하게 되어, 입력 예제의 가변 변환(variable transformation)에 대해 일관성 있게 인식을 수행하기

위해서는 데이터 증진이 필수적이다. 보다 최근의 논문 [11-14]에서는 거리 및 각도와 같은 수작업으로 제작된 변형 불변 특성을 사용하여 강건한(robust) 물체 인식을 수행한다. 이러한 접근방식은 실질적인 성능 향상을 나타내지만, 낮은 수준(low level)의 기하학적 특성은 상세 형상 정보를 표현할 수 있는 제한된 능력을 가지고 있다. 한 단계 낮은 특징의 사용은 기하학적 모순(ambiguity)을 발생시키게 되고, 이에 대한 모순을 해결하기 위해 다수의 점을 공동으로 고려해야 하기 때문에 계산 복잡도의 점진적 증가는 어려움 전반적인 인식 과정을 어렵게 만들 수 있게 된다.

이러한 중요한 문제를 해결하기 위해서, 지역 좌표계를 기반으로 형성된 지역 참조 프레임(local reference frame) 내 구름 점이 나타내는 형상 정보를 기술하는 기술자(descriptor)를 도입하여 새로운 회전 불변 3차원 객체 인식 프레임워크를 도입했다. 기술자의 수용 범위(receptive field)는 확률적(stochastically)으로 확대되며, 생성된 특징(feature)은 보다 정규화(regularize)가 잘 되어있게 되고, 물체의 다양한 형상 변화에도 대표적(representative)이다. 이에 더해, 제안된 방법은 대상 객체의 전역 회전 불변 정보를 효과적으로 표현할 수 있는 기능을 제공하는 기술자를 계층적으로(hierarchically) 설계하였다. 마지막으로, 점진적으로 부호화(encoding)된 지역-전역(local-to-global)에 걸친 형상 정보에 기초하여, 확률적으로 생성된 그래프를 기반으로 그래프 합성곱 신경망(graph convolutional neural network)을 활용하여 입력 구름 점에 적용될 수 있는 노이즈와 아웃라이어(outlier)에 대하여 강건한 인식을 가능하게 하였다.

본 논문의 기여는 아래와 같이 요약될 수 있다.

- 우리는 각도나 거리와 같은 기하학적 특징을 계산하지 않고도 점진적으로 회전 불변 특징을 얻어내기 위해 그래프 신경망 기반의 3차원 구름 점 데이터에 대한 지역-전역 표현 학습법을 제시한다.
- 제안된 알고리즘은 확률적으로 수용성 필드(receptive field)의 크기를 확대시켜, 학습하여 추출하는 특징을 효과적으로 정규화 할 수 있다. 이러한 학습법은 노이즈와 아웃라이어에 대하여 강건성(robustness)을 증가시키는 데 효과적이다.

- 우리의 접근법은 3차원 구름 점 데이터를 기반으로 하는 3차원 물체 인식 벤치마크에 대하여 회전에 대한 데이터 증진 없이도 뛰어난 성능 향상을 보였다.

본 논문의 나머지는 다음과 같은 순서로 정리되었다. 제 2장에서는 3차원 구름 점 데이터에 기반한 물체 인식에 관한 기존 연구들에 대해 논의한다. 제 3장에서는 회전에 불변한 지역 기술자를 생성하고 형상을 계층적으로 사용하여 그래프 합성곱 신경망을 구성하기 위해 제안된 접근 방식을 설명한다. 표준 벤치마크 데이터 셋에 대한 실험 결과를 제 4장에서 제시하며, 논문의 결론은 제 5장에서 제시한다.

제 2 장 관련 연구

제 1 절 딥 러닝을 이용한 3차원 물체 인식

구름 점은 3차원 유클리디안 공간에 희소하고 불규칙적으로 분포하고 있기 때문에, 한 동안 직접적으로 사용되지 못하고, 주로 복셀(voxel)이나 2차원 영상 이미지로 투영(project)되어 이용되었다. 2차원 이미지 처리에서 큰 성능 향상을 이끌어 내었던 합성곱(convolution) 연산자와 같이 규칙적인 연산자를 사용하는 데 제한이 있고, 하나의 물체를 표현하더라도 구름 점의 순서에 따라 그 데이터가 매우 다양하게 표현이 될 수 있기 때문에, 데이터를 직접적으로 다루는 것에 어려움이 있었기 때문이다. 따라서, 이러한 데이터의 본질적인 특성에 의해 초기에는 복셀화(voxelization)과 다중 뷰 방법(multi-view method) 방법으로부터 3차원 형상에 대해 인식하고자 하는 시도가 이루어졌다.

복셀화(voxelization)는 3차원 공간을 양자화(quantize)하여 데이터를 규칙적인(regular) 공간에 분포시키는 방법이다. 이를 바탕으로 3차원 합성곱 신경망(3D convolutional neural network)을 이용하여 우수한 특징을 얻음으로써 인식 문제를 해결했고, 앞서 언급한 두 가지 문제점에서 자유로워 질 수 있었다. 다중 뷰 방법은 3차원 물체에 대해 다양한 뷰에서 얻은 2차원 영상 이미지에 2차원 합성곱(2D convolution)을 적용하여 얻은 특징들을 얻어내었다 [17]. 매우 다양한 뷰에서 물체에 대한 다양한 특징들을 얻기에 매우 높은 인식 정확도를 보인다.

그러나, 앞서 언급한 두 방법 모두 반복적인 연산 및 상당한 연산량을 필요로 하게 된다. 예를 들어, 복셀화의 경우, 3차원 물체를 양자화(quantize)하는 과정에서 해상도(resolution)를 결정해야 하는데, 높은 해상도일수록 형상에 대한 인식을 세부적으로 할 수 있지만, 연산량이 크게 증가한다. 낮은 해상도일수록 연산량은 줄지만, 형상에 대한 이해도가 낮아진다. 다시 말하면, 복셀화는 정보의 손실에 있어 트레이드오프가 존재한다. 다중 뷰 방법의 경우도 마찬가지로, 형상에 대한 인식률을 높이기 위해서는 매우 다양한 뷰를 사전에 선택하여 다수의 이미지를 얻어내어 이를 종합해야 한다. 다시 말해서, 두 가지

방법 모두 확장성(scalability) 측면에서 매우 제한적이고, 불필요한 중복(redundancy)에 의해 낭비되는 리소스가 상당했다. 이러한 문제를 해결하고자, kd-tree[15] 또는 oc-tree[16]와 같은 방법들이 제시되어 연산량을 크게 줄였으나, 앞선 문제점들은 여전히 유효하였다.

최근 심층 신경 망(deep neural network)의 급격한 발전과 함께, 이러한 구름 점을 직접적으로 사용하여 효율적으로 3차원 물체를 인식하고자 하는 과제들이 많은 관심을 끌고 있다. [1]이 제안한 ‘포인트 넷’ (PointNet)은 다중 퍼셉트론 층(multi-layer perceptron layer)을 기반으로 하여 순서가 정해지지 않은(unordered) 점 집합을 직접 입력으로 받아, 순서에 불변한(permutation invariant) 전역적(global) 특징을 효과적으로 추출하였다. 순서 문제는 대칭식(symmetric function)을 사용하여 다양한 순서에도 불변한 특징 추출을 가능하게 하였다. 맥스 풀링(max pooling) 또는 평균 풀링(average pooling)등이 사용될 수 있다. 매우 간단한 방법으로부터 구름 점을 성공적으로 이용할 수 있게 되면서, 이 후 매우 다양한 방법들이 포인트 넷이 사용한 방법을 기반으로 하여 제시되었다. 포인트 넷은 3차원 사물과 관련된 다양한 작업에 결정적 역할을 하는 물체의 지역적(local) 특징을 이해하는 능력이 결여되어 있었는데, 이후 계층적 모델 [3]을 구축하여 일정 반경 내의 점들을 찾아 해당 점들을 군집화(grouping)하여 지역 특징으로 추출하는 방법 또는 점들의 특징 공간(feature space)으로부터 에지(edge) 정보의 통합과 [2] 같은 지역적 특징을 추출하는 일부 전략들이 제시되었다. 또한, 구름 점 입력에 대해 적용할 수 있는 커널(kernel)을 디자인하여 합성곱을 수행하고자 하는 방법론들도 제시되었는데, [18]은 수공 커널(handcrafted kernel)을 사용하여, 지역 별 반응(response) 값들을 얻어내어, 지역적 특징으로써 사용하였다. 이와 같은 지역적 특징을 추출하고자 하는 다양한 노력들은 3차원 물체 분류 과제에 인상적인 성능 향상을 이끌어 내었다.

3차원 구름 점을 직접적으로 사용하여 전역적 그리고 지역적인 특징을 얻고자 하는 연구들은 수 없이 진행되어 왔지만, 회전된 입력에 대해서는 단순히 데이터 셋에 임의의 회전 변환을 적용하여 입력 공간을 무한한 공간으로 확장하여 해결하고 있다. 이러한 방법은 근본적인 문제를 해결하지 못하면서, 기본적으로 많은 리소스를 요구하는 3차원 인식 모델의 수용도(capacity)를 더 높여 학습 리소스를 더욱 무겁게 할 뿐이다. 따라서 본 연구에서는 회전에

강건한 3차원 객체 인식 프레임워크를 구축하는 것을 목표로 하고 있다.

제 2 절 회전 불변 3차원 구름 점 인식

3차원 물체 인식에서 경쟁력 있는 성능을 달성하기 위해서는, 회전에 강건한 특징 표현(feature representation)을 얻는 것이 매우 중요하다. 초기에는, 공간 변환 네트워크 [1, 2]를 사용하여 기하학적 강체 변환(geometric rigid transformation)에 대한 강건함을 제공했다. 그러나 기본적으로 공간 변환 네트워크는 3차원 좌표 계를 그대로 입력으로 받았기 때문에 회전 변환에 의한 특징 변형으로부터 강건하지 못했다. 이 후, 회전에 대한 강건함을 향상시키기 위해, 구형 함수(spherical function)에 기초한 회전 등가 네트워크(rotation-equivariant network) [19, 20]이 제안되었다. 회전 등가 네트워크는 완전히 회전에 불변한 특징을 얻어내지 못했지만, 회전에 일관성 있는 특징 추출이 가능해지게 되었고, 성능 향상과 함께 학습 모델 수용성의 부담을 크게 줄이게 되었다. SFCNN [21] 은 3차원 구름 점을 정 이십면 체 격자에 매핑 시켜 사용하는 접근법을 제안했다. 이러한 방법은 불규칙적인 구름 점 데이터를 규칙적인 공간에 매핑 시킴으로써 회전 변환에 강건한 입력 체계를 형성하였다. 그러나, 앞서 언급한 회전 등가 네트워크나 규칙적 구조에 매핑 시키는 방법들은, 여전히 관찰되지 못한 회전에 대하여 취약했고, 회전 변환을 이용한 데이터 증진에 크게 의존해야 물체를 문제 없이 인식할 수 있었다.

가장 최근에 제시된 방법은, 점들 간의 거리나 각도와 같은 기하학적 특징을 추출하여 특징으로 사용한다. [13]은 쌍을 이루는 점들을 대상으로 각각의 법선 벡터(normal vector)와 두 점의 차이 벡터(difference vector)간의 상대적 각도와, 차이 벡터의 거리를 함께 사용하여 회전 불변의 특징으로 사용한다. [11]은 두 점 사이의 방위각(azimuthal angle), 극 각(polar angle), 그리고 방사상의 거리(radial distance) 을 통합하여 점의 특징을 생성해낸다. 하나의 점을 기준으로 다수의 주변 이웃 점을 이용해 위의 특징들을 추출하여 이어 붙여 하나의 점을 표현한다. 이와 비슷하게, [12]는 참조 포인트(reference point), 지역 이웃 중심점(local neighborhood centroid) 그리고 지역 이웃 점(local neighbor point)들로부터 형성된 지역 삼각 구조(local triangular structure)에서 추출한 거리와 각도를

활용한다. 그러나 위의 [11-13] 과 같은 방법들은 저 수준(low-level)의 특징으로 변환하는 과정에 의해 개체를 인식하는 데 모호성(ambiguity)을 야기하고 이를 최소화하기 위해 주변 이웃 점들에 대하여 특징들을 다수 모아 사용해야 한다.

이와 달리, 우리는 회전 불변의 지역 기술자를 얻기 위해 지역 참조 프레임(local reference frame)을 이용하였다. 이에, 모델은 형상 정보를 유지하고, 로컬 기술자를 계층 적으로(hierarchically) 통합하여 전역적 특징을 얻을 수 있다.

제 3 절 3차원 구름 점 작업을 위한 그래프 기반 네트워크

그래프 합성 곱 신경 망은 그 동안 많은 발전을 이루었으며 [22-25] 많은 기계 학습과 관련된 과제에서 성공적인 결과를 이끌어 내었다. 그래프 기반 접근 방식은 주로 (1) 공간 구조(spatial structure)를 고려한 접근 방식 (2) 스펙트럼 그래프 이론(spectral graph theory)을 바탕으로 한 접근 방식 두 그룹으로 분류된다. 그래프 합성 곱에 대한 관심이 높아지고, 그 효과와 유효성이 입증되면서, 최근에는 3차원 데이터에 그래프 합성 곱을 적용하여 문제를 해결하고자 하는 시도들이 많이 제시되었다.

일반적으로 공간 그래프를 기반으로 하는 사례 들은 에지 정보에 초점을 맞추게 된다 [2, 26, 27]. DGCNN [2] 은 에지 합성 곱 (edge convolution) 연산자를 제안했는데, 이는 모든 계층에서 동적으로(dynamically) 그래프를 구축하여 지역적으로 기하학적 특징을 추출하였다. 그래프 구축은 일반적으로 KNNs (K-Nearest Neighbors) 알고리즘이 사용된다. ECC [22] 는 에지 별 가중치 행렬(edge-specific weight matrix)을 생성하고 주변 특징을 공간적으로 통합하는 조건부 에지 합성 곱 네트워크(edge-conditioned convolutional network)를 제안했다. DeepGCN [27] 공간 그래프 합성 곱 층에 잔류 연결(residual connection)을 사용함으로써 매우 깊게 쌓아질 수 있고, 이로부터 그래디언트가 사라지는 문제(vanishing gradient problem)를 완화시킬 수 있음과 동시에, 과 적합 문제(overfitting problem)를 해결 시킬 수 있음을 발견했다.

동시에, 스펙트럼 그래프 이론에 기반한 접근 방식 [22-24] 이 함께 발전했다. [7] 은 전체 3차원 구름 점 입력에 대한 그래프 구조를

구성하고 체비셰프 다항식(Chebyshev polynomial)에 의해 근사 된 필터를 사용하여 스펙트럼 그래프 필터링(spectral graph filtering)을 적용했다 [23]. [4]는 전체 구름 점에 그래프 합성 곱을 적용하는 대신, 스펙트럼 그래프 필터링 [24] 을 사용하여 계층 적 구조로 지역 기술자를 계산한다. 이는 다중 퍼셉트론 계층(multi-layer perceptron)을 사용하는 것보다 성능이 우수하다는 것을 밝혀내었다. 그러나, 그래프 신호가 3차원 좌표로 표현되기 때문에, [4, 7]은 여전히 기하 변환에 취약하다. 대조적으로, 우리는 회전 불변 방식으로 생성 된 지역 기술자들을 바탕으로 근사 된 스펙트럼 그래프 필터링을 사용하였다.

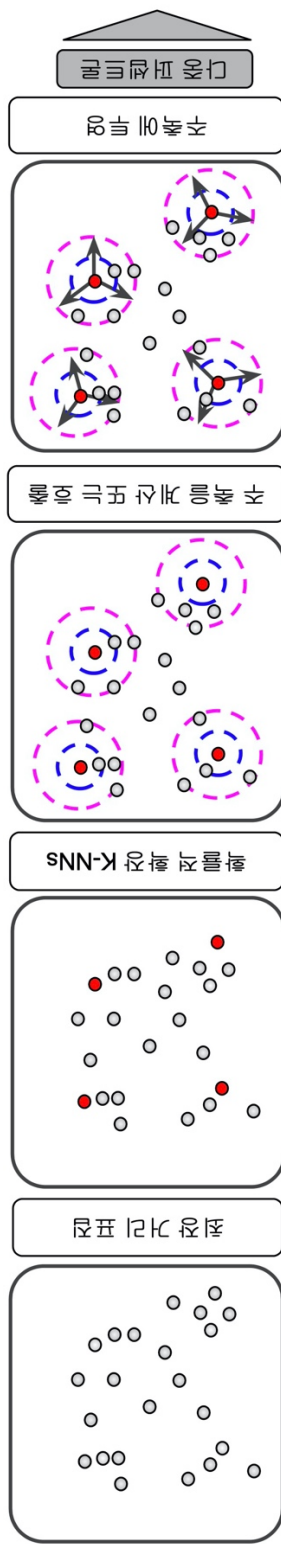


그림 1. 기술자 추출(descriptor extraction) 과정. 주어진 점 집합에 대하여 최장 거리 표집 알고리즘을 이용하여 대표 점들을 선정한다. 이 후, 확률적 확장 k -NNs 알고리즘을 통해 이웃 점들을 탐색한다. 각 지역 영역은 회전 불변한 공간으로 변환되기 위해 주 축(principal axes)을 계산한다. 각 지역 영역은 주 축으로 투영되고 다중 퍼셉트론을 통과하여 고차원의 특징으로 매핑 된다.

제 3 장 회전 불변 표현의 확률적 학습

우리의 목표는 순서가 정해지지 않은 점 구름 집합, $\mathcal{P} = \{p_1, \dots, p_N\}$, 을 사용하여 회전 불변의 3차원 물체 인식 프레임워크를 구축하는 것이다. 여기서 위 집합의 원소는 $p_i \in \mathbb{R}^3$ ($i = 1 \dots N$) 와 같이 표현된다. 다시 말해서, 우리는 임의의 회전 변환 $r: \mathbb{R}^{N \times 3} \rightarrow \mathbb{R}^{N \times 3}$ 에 대하여 다음과 같은 특성을 만족시키는 미분 가능한 함수 $f: \mathbb{R}^{N \times 3} \rightarrow \mathbb{R}^F$ 를 학습시키는 것을 목표로 한다.

$$f(\mathcal{P}) = f(r(\mathcal{P})) \quad (1)$$

우리의 프레임워크는 기술자 추출(descriptor extraction), 기술자 확장(descriptor extension) 그리고 그래프 기반 추출(graph-based abstraction)로 총 3개의 모듈로 구성되어 있다. 첫 번째 과정인 기술자 추출 모듈은 대표 점(representative point)을 확률적으로 표집하고, 신경망 모델을 사용하여 해당 지역 참조 프레임에 해당되는 점들을 인코딩하여 지역 형상 기술자를 추출하게 된다. 다음으로, 기술자 확장 모듈은 각 기술자의 범위(scope)를 점진적으로 확장하고, 지역에서 전역으로 표현 계층을 체계화하게 된다. 마지막으로, 그래프 기반 추출 단계에서는 그래프 합성 곱 신경망 네트워크를 사용하여 인접한 기술자들을 통합함으로써 문맥 인식(context-aware)의 특성 표현을 획득할 수 있다.

제 1 절 기술자 추출

이 모듈은 대표 점과 그 이웃 점(neighbor point)들을 이용하여 일련의 지역 기술자들을 학습하는 것을 목표로 한다. 그러한 강건성을 얻기 위해, 인접 점들은 지역 영역의 주 성분 분석(principal component analysis)을 통해 구축된 지역 참조 프레임(local reference frame)에 투영된다. 이는, 전역 좌표 계에 적용되는 회전에 대하여 지역 영역 단위로 형성된 지역 기술자는 영향을 크게 받지 않는다는 사실에 입각하여 동기 부여 될 수 있다. 그림 1은 이 모듈의 개요를

보여준다. 모듈의 첫 번째 단계는 대표적인 점들을 파악하고 그것들의 이웃 점들을 찾아내는 것이다. 우리는 대표 점 집합 $Q = \{q_1, \dots, q_M\} \in \mathbb{R}^{M \times 3} \subset \mathcal{P}$ ($M \ll N$) 를 최장 거리 표집(Farthest Point Sampling) 알고리즘 [28]을 이용하여 선정하였다. 최장 거리 표집 알고리즘의 목적 함수(objective function)은 다음과 같다.

$$\max_{p \in \mathcal{P}} \min_{q \in Q} \text{dist}(p, q), \quad (2)$$

여기서 $\text{dist}(\cdot, \cdot)$ 는 점 들간의 거리를 계산하는 함수이다. 최적 Q 를 찾아내는 문제는 NP-hard 이지만, 선택된 점들의 수가 M 에 도달할 때까지 탐욕적(greedy) 방법으로 모든 단계에서 가장 먼 지점을 반복적으로 찾아 해결책을 대략적으로 구할 수 있다. 이러한 해결책은 2-approximation 알고리즘으로 알려져 있다.

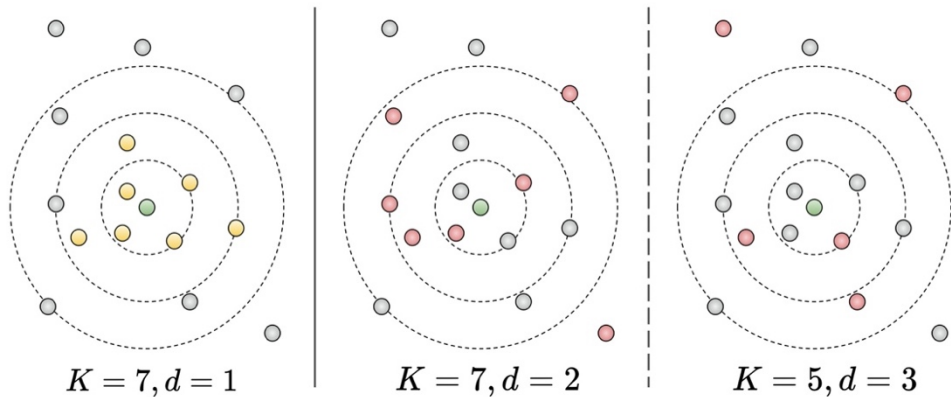


그림 2. 확률적 확장 k -NNs 의 예시

각 대표 점 q_i ($i = 1 \dots M$) 은 기존 점 집합 \mathcal{P} 로부터 이웃 점들을 탐색하여 지역 영역을 정의한다. 여기서 다양한 형상 변화를 다루기 위하여 우리는 최근 제안된 확장 k -NN 탐색 법(dilated k -NN) [27] 을 확률적 방식으로 사용하였다. 앵커(anchor) 점 q_i 와 확장율(dilation rate) d 가 주어지는 경우, 확장된 k -NN 탐색의 출력은 집합 \mathcal{P} 내 k 개 점들의 인덱스 집합인데, 해당 점들은 앵커 점 q_i 로부터 가까운 순서로 정렬된 점들의 매 d 번째에 위치하고 있다. 학습 과정에서 우리는 k, d 를 모두 랜덤 변수(random variable)로 설정하여 확장 k -NN 탐색을 확률적으로 수행함으로써 프레임워크에서 얻어내는

표현에 대한 정규화 효과를 얻어 성능을 향상 시켰다. 그림 2는 확률적 확장 k -NNs에 대한 예시를 나타낸다. 가장 좌측은 일반적인 k -NNs 알고리즘을 나타낸다. 중앙과 우측에 묘사된 그림은, 일반 k -NNs에 확장 율 d 를 설정하여 매 d 번째 점들을 찾아 지역 영역을 형성하는 것을 확인 할 수 있다. 이러한 확장 율을 적용하는 이유는, 보통 다수의 점들에 대해 모든 각 점들에 대하여 지역 영역을 형성해 다중 퍼셉트론을 이용하여 부호화하는 과정은 많은 메모리 리소스 뿐만 아니라, 학습을 위한 연산량을 크게 증가시키게 된다. 따라서 보다 더 넓은 영역을 보기 위해서는 k 값을 크게 설정하여 많은 이웃 점들을 찾아야 하는데, 이는 매우 비효율적인 학습을 진행하게 된다. 따라서, 확장 율을 설정함으로써, 같은 메모리를 사용하면서 보다 더 넓은 기하 구조를 파악하기 위해서 사용된다.

이 후, q_i 와 그 이웃 점들을 기준으로 하여 지역 참조 프레임을 구축하였다. 그림 1에 묘사된 것처럼, 우리는 각각의 앵커 점 q_i 의 두 가지 서로 다른 탐색 영역 $\mathcal{N}_{k,1}(q_i)$ (파란색 원) and $\mathcal{N}_{k,d}(q_i)$ (분홍색 원) 을 사용하였다. $\mathcal{N}_{k,1}(q_i)$ 은 확장 율로 1을 사용한 점 집합이고, $\mathcal{N}_{k,d}(q_i)$ 은 d 을 사용한 점 집합이다. 지역 참조 프레임은 $\mathcal{N}_{k,d}(q_i)$ 의 주 성분 분석을 통해 얻어 내었고, 각각의 점 집합은 예측한 참조 프레임에 투영된다. 이처럼 투영된 점 집합을 $\mathcal{N}'_{k,d}(q_i)$ 로 표현하였다. 이 모듈의 마지막 단계는 지역 참조 프레임에 회전-정규화된(rotation-normalized) 정보를 기반으로 지역 기술자를 생성하는 것이다. q_i 에 대응되는 기술자 $\phi_i \in \mathbb{R}^c$ 는 다음과 같이 구한다.

$$\phi_i = f_{\odot} \left(\left[g_1 \left(\mathcal{N}'_{k,1}(q_i) \right), g_2 \left(\mathcal{N}'_{k,d}(q_i) \right) \right] \right), \quad (3)$$

위 식에서 $f_{\odot}(\cdot)$, $g_1(\cdot)$ 그리고 $g_2(\cdot)$ 은 맥스 풀링(max pooling)이 뒤따라 오는 다중 퍼셉트론을 의미하고, $[\cdot, \cdot]$ 은 채널 방향으로 주어진 인자 텐서(tensor)들을 이어 붙임(concatenation) 하는 것을 나타낸다.

제 2 절 기술자 확장

이전 기술자 추출 모듈로부터 점 집합 $Q = \{q_1, \dots, q_M\}$ 과 각 점들에 대응되는 기술자 집합 $\mathcal{H} = \{\phi_1, \dots, \phi_M\}$ 를 얻어 내었다고 하자. 그 다음, 기술자 확장 첫 번째 단계에서는 두 집합 $Q_0 \equiv Q$ 그리고 $\mathcal{H}_0 \equiv \mathcal{H}$ 를 입력으로 사용하고, 후속 단계에서 사용하게 될 동일 유형의 집합 Q_1 \mathcal{H}_1 을 생성하게 된다. 이 때, $|Q_1| \ll |Q_0|$ 를 만족한다. 기술자 확장 모듈은 미리 정의된 반복 횟수에 따라 이 절차를 반복하게 된다. 기본적으로 기술자 확장 모듈의 절차는 기술자 추출 모듈과 거의 유사하지만, 다음과 같은 차이가 존재한다. 첫 째, 각 단계는 최장 거리 표집을 적용하여 감소된 수의 점들을 다루며, 이는 지역 점 집합을 업데이트 하게 된다. 둘째, 이 모듈의 각 단계에서는 식 3의 $g_2(\mathcal{N}'_{k,d}(q_i))$ 이 이전 모듈에서 얻은 기술자 표현(descriptor representation)으로 대체된다 (예: ϕ_i). 셋 째, 기술자 추출 모듈에서 얻은 주요 축을 재사용한다. 계층이 깊어질 수록 점들은 희소해지고, 이로부터 얻어낸 축은 안정적이지 못할 수 있기 때문이다.

정리하면, 이 모듈은 기술자 추출 모듈에서 얻은 지역 기술자의 기술 범위를 확장하고, 확장된 범위와 함께 기술자의 의미적 수준(semantic level)을 높이기 위해 계층 구조를 구성한다. 이는 주변 형상 정보를 종합함으로써 달성되는데, 이는 보다 넓은 영역에 걸쳐 지역에 대한 기하학적 이해를 용이하게 하고, GCN의 크기를 줄임으로써 계산 비용을 절감해 준다. 지역 형상 특징에 대해 보다 넓은 수용 범위(receptive field)를 부여함으로써 지역 형상 정보를 효과적으로 특징 짓고, 파악할 수 있다는 점에서, 기술자 범위 확장은 그 과정이 매우 바람직하다는 것을 알 수 있다. 그림 3은 기술자 확장 모듈의 파이프라인을 묘사한다.

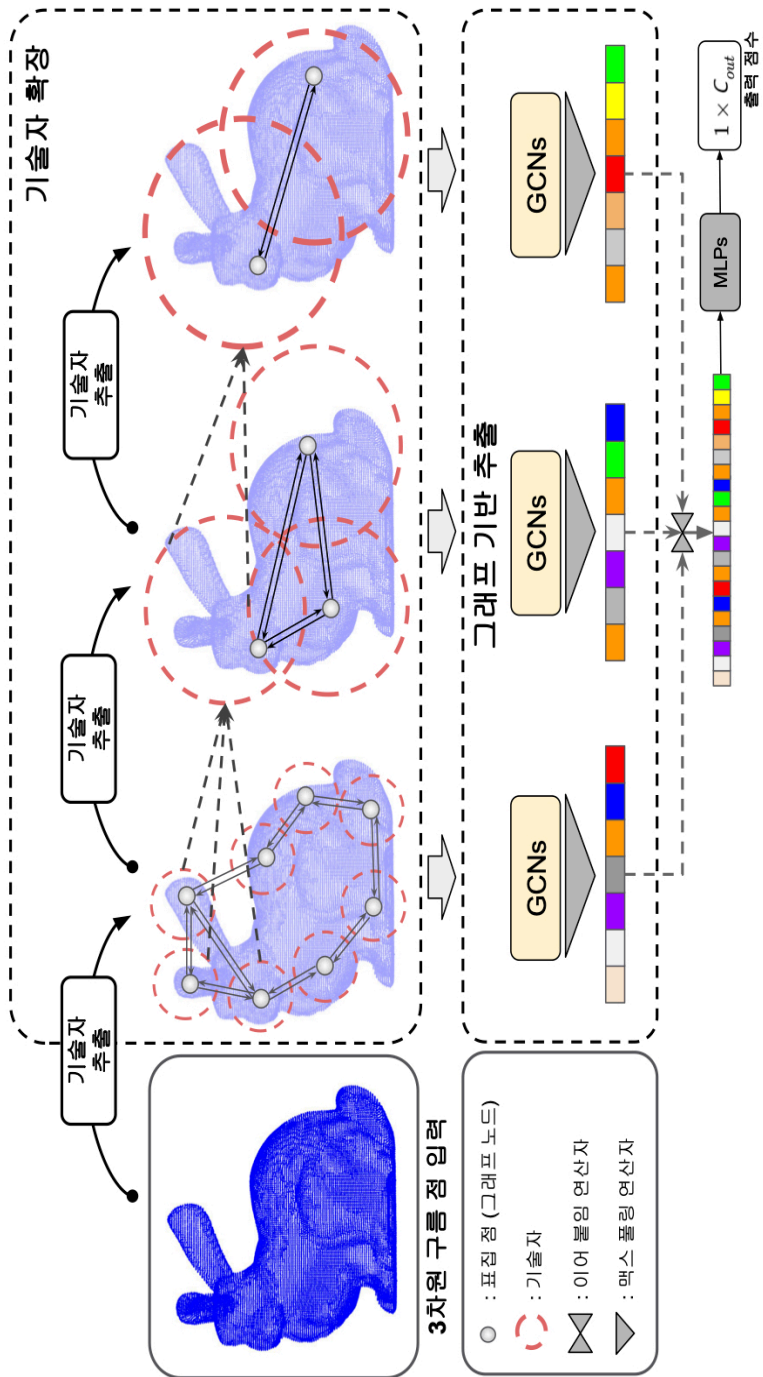


그림 3. 회전 불변 3차원 물체 분류 프레임워크. 기술자 확장은 기술자 추출 모듈을 여러 개 쌓아 올린 형태를 하고 있으며, 기술자의 범위를 회전 불변성을 유지하며 지역 특징들을 그룹화 하며 확장시킨다. 그래프 기반 추출 모듈은 그래프 합성곱 신경망에서 각 계층에서 위상 구조를 생성함으로써, 지역 기술자들을 통합한다.

제 3 절 그래프 기반 추출

그래프 기반 추출 모듈의 목적은 기술자 간의 연결을 참조하여 문맥을 인식하는 (context-aware) 기술자를 얻는 것이다. 지금까지 각 계층에서 학습한 정보는 물체의 부분적인 형상에 불과하므로, 각 기술자가 그 기하학적 맥락을 잘 파악하여 전체적인 형상을 더 잘 이해하는 것은 바람직하다. 기술자의 정보를 종합하기 위해, 우리는 각 계층 레벨의 마지막 단계에서 스펙트럼 기반의 GCN을 적용하였다. GCN은 그래프에 존재하는 정점의 신호 값들을 그래프 푸리에 기저에 투영하여 스펙트럼 필터링(spectral filtering) 수행하게 되는데, 이에 대한 연산을 위해 그래프 라플라시안(graph Laplacian)을 필요로 하게 된다. 그래프 라플라시안 $\mathcal{L} \in \mathbb{R}^{N \times N}$ 은 일반적으로 다음과 같이 계산된다.

$$\mathcal{L} = D - A \quad (4)$$

여기서 D 는 차수 행렬(degree matrix), A 는 인접 행렬(adjacency matrix)을 의미한다. 그래프 라플라시안은 본질적으로 양의 준 정부호(positive semi-definite) 성질을 가지고 있기 때문에, 고유 값 분해(eigen-decomposition)을 통해 $\mathcal{L} = U\Lambda U^T$ 와 같이 분해될 수 있다. 여기서 U 는 직교 함수 계(orthonormal)한 고유벡터(eigenvector)를 의미하고, Λ 는 각 고유 벡터에 대응되는 고유 값(eigenvalue)를 의미한다. 즉, 그래프 합성 곱은, 임의의 그래프 신호 $x \in \mathbb{R}^n$ 에 대해 다음과 같은 식으로 표현 될 수 있다.

$$x * g = U(U^T x \circ U^T g) \quad (5)$$

여기서 g 는 필터이고, \circ 는 아다마르 곱(Hadamard product) 연산자이다. 위 식에서, $diag(U^T g)$ 를 g_θ 로 설정하면, 식 5는 다시 다음과 같은 식으로 정리될 수 있다.

$$y = x * g_\theta = U g_\theta(\Lambda) U^T x \quad (6)$$

즉, 위 식 6으로부터 그래프 합성 곱은 입력 그래프 신호를 직교를 이루는 그래프 푸리에 기저에 투영하여, 필터링을 거친 후, 다시 그래프 푸리에 역 변환을 통해 원래 도메인으로 되 돌아온 값을 출력으로 사용한다는 것을 알 수 있다.

모든 스펙트럼 이론에 기반한 그래프 합성 곱은 위의 식 (6)을 기반으로 발전되어 왔는데, 우리는 가장 최근에 스펙트럼 그래프 합성 곱을 1차 근사 시킨 [24]의 그래프 합성 곱 방법을 사용하였다. 이에 더해, 다양한 범위와 맥락에 대하여 학습하기 위해 그래프를 형성할 때 확률적으로 생성하였다. 구체적으로, 각 단계의 기술자들은 먼저 그래프 신호(graph signal) $X_l \in \mathbb{R}^{n_l \times c_l}$ 로 전환된다. 이 때 n_l 노드의 개수, 그리고 c_l 는 l 번째 계층의 특징 차원(feature dimension)을 의미한다. 단순성(simplicity) 을 위해 이하 부분에 대해서는 인덱스 l 은 생략하도록 한다. 인접 행렬 A 를 위해 우리는 k -NN 그래프를 구축하였는데, 이 때 에지 개수 k 는 랜덤 변수로 설정해 주었다.

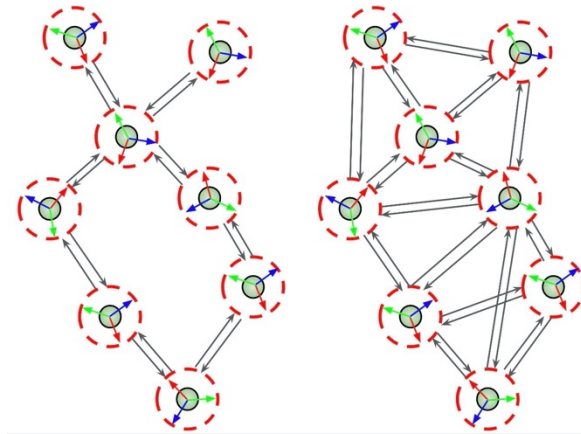


그림 4. 그래프의 확률적 생성 예시. 연두색 원은 정점(점)을 의미하고, 원들 사이를 잇는 실선은 에지를 의미한다. 정점을 둘러싼 붉은 실선의 원은 기술자를 표현한 것이며, 원 내부의 축은 기술자 추출 단계에서 계산한 주축을 의미한다.

그림 4는 생성될 수 있는 그래프에 대한 예시를 나타낸다. 같은 기하 구조를 표현하고 있는 3차원 점들에 대하여, 에지 개수를 확률적으로 설정함으로써 보다 일반화된 학습을 진행할 수 있다. 우리는 그래프의 에지에 가우시안 커널(Gaussian kernel)에 의해

스무딩(smoothing)된 유클리디안 거리(Euclidean distance)를 가중치로서 부여하였다.

[24]에서는 GCN을 재 정규화 트릭 (renormalization trick) 방법을 이용하여 그래프를 $\hat{A} = \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2}$ 로 근사 시킨다. 이 때, $\tilde{A} = A + I_n$ 는 자가 연결(self-connection)이 고려된 인접 행렬을 의미하고, $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ 는 차수 행렬(degree matrix)를 의미한다. 따라서, GCN은 다음과 같이 적용되어 X 의 표현력을 강화시키게 된다.

$$Y = \text{ReLU}(\hat{A}XW), \quad (4)$$

이 때, W 은 학습 가능한 파라미터 행렬(parameter matrix)를 의미한다. 그림 3에 묘사된 바와 같이, 분류를 위한 최종 표현을 얻기 위해 모든 계층의 출력은 맥스 풀링이 적용되어 이어 붙여져 얻어진다

제 4 장 실험

이 절에서는 기존 접근방식들과 비교한 우리의 알고리즘 실험 결과를 제시한다. 우리는 다양한 비교 실험을 통해, 제안하는 프레임워크의 효과를 설명하였다. 본 논문에서 사용하는 평가 프로토콜(evaluation protocol)은 이전 연구에서 사용한 평가 프로토콜과 동일하다 [11, 12, 16, 17]. 구체적으로, 학습과 평가 시 다음과 같은 세 가지 방식으로 진행했다: 학습과 평가 모두 방위각 회전(azimuthal rotation) 으로 만 데이터를 증진 시킨 방식 (z/z), 학습은 방위각 회전으로, 평가 시에는 임의 회전(arbitrary rotation) 으로 데이터를 증진 시킨 방식($z/SO(3)$), 그리고 학습과 평가 모두 임의 회전으로 데이터를 증진시킨 방식($SO(3)/SO(3)$).

제 1 절 구현 세부 정보

우리 프레임워크는 텐서플로우(Tensorflow)를 이용해 구현되었다. 그리고 모든 실험은 단일 NVIDIA Titan XP GPU 를 이용해 진행하였다. 학습 율(learning rate)는 0.001로 세팅해주었고, 배치 사이즈(batch size)는 32이다. 우리는 각 계층마다 256개, 128개, 그리고 64개의 점들을 기술자 추출 과정에서 표집 하였다. 각 계층마다 사용되는 확률적 확장 k -NNs에서 사용되는 k 값은 모두 각기 다른 단일 분포(uniform distribution) 로부터 표집 하였다. 계층 별 단일 분포는 다음과 같다: $\mathcal{U}(32, 96)$, $\mathcal{U}(16, 48)$, 그리고 $\mathcal{U}(8, 24)$. 첫 번째 기술자 추출과정에서 사용되는 확장 율은 $\mathcal{U}(2, 4)$ 에서 표집 된다. 한 편, 각 계층에서 그래프 합성 곱 연산에 필요한 그래프 구축을 위해서 사용되는 에지의 개수 (\hat{k})도 우리는 확률적으로 표집 했다: $\mathcal{U}(8, 24)$, $\mathcal{U}(4, 12)$, 그리고 $\mathcal{U}(2, 8)$. 평가 시에는, 모든 파라미터를 고정 값으로 세팅하였는데, k 값은 각 계층 순서대로 64, 32, 그리고 16으로, \hat{k} 는 16, 8, 그리고 4로 세팅하였다. 그리고 확장 율은 3으로 설정하였다.

제 2 절 주요 결과

회전에 대한 강건성을 평가하기 위해 3차원 객체 분류 작업에서 가장 많이 이용되는 모델넷40 (ModelNet40) [30] 데이터 셋을 이용하여 최근 3차원 객체 분류 접근법과 비교하였다. 모델넷40은 40개 카테고리의 CAD 모델로 구성되며, 9,843개의 학습 데이터(training data)와 2,468개의 유효 데이터(validation data)를 포함하고 있다. 구름 점은 CAD 모델의 정점(vertex)와 표면에서 균일하게 샘플링 된다. 모든 점들은 단위 구(unit sphere) 내부로 이동되고, 정규화 된다.

표 1. 모델넷40 데이터 셋에 대한 3차원 물체 분류 정확도 (%)

Method	Input	z/z	z/SO(3)	SO(3)/SO(3)
PointNet(w/o STN) [1]	pc (1024 x 3)	88.5	16.4	70.5
Pointnet++(MSG w/o STN) [3]	pc + normal(5000 x 6)	91.9	18.4	74.7
SO-Net(w/o STN) [29]	pc + normal(5000 x 6)	93.4	19.6	78.1
DGCNN(w/o STN) [2]	pc (1024 x 3)	91.2	16.2	75.3
PointNet [1]	pc (1024 x 3)	89.2	16.2	75.5
PointNet++ [3]	pc + normal(5000 x 6)	91.8	18.4	77.4
SO-Net [29]	pc + normal(5000 x6)	91.2	21.1	80.2
DGCNN [2]	pc(1024 x 3)	92.2	20.6	81.1
SpecGCN [4]	pc (1024 x 3)	91.5	28.8	75.3
Spherical CNN [20]	voxel (2 x 64 x 64)	88.9	76.9	86.9
SFCNN [21]	pc(1024 x 3)	91.4	84.8	90.1
SFCNN [21]	pc + normal(1024 x 6)	92.3	85.3	91.0
RICConv [12]	pc (1024 x 3)	86.5	86.4	86.4
ClusterNet [11]	pc(1024 x 3)	87.1	87.1	87.1
Ours	pc(1024 x 3)	89.5	89.5	89.5
Ours (with normal)	pc + normal (1024 x 6)	91.0	91.0	91.0

표 1은 가장 최근 발표된 두 개의 회전 불변 알고리즘 [11, 12]을 포함한 전반적인 3차원 객체 분류의 전반적인 성능을 보여주고 있다. 대부분의 접근법 [1-3, 20, 29]은 z/z 에서 우수한 성능을 보이고 있다. 그러나, 학습 시 관측하지 못한 회전들에 대하여 추론 능력을 평가하였을 때, 매우 심각한 성능 저하가 발생하는 것을 확인할 수

있다. 더욱이, 비록 학습 단계에서 임의의 회전으로 데이터를 증진시켜도, 대부분의 알고리즘은 일반화 성능을 온전하게 회복시키지 못했다. PointNet [1], PointNet++ [3], SO-Net [29], 그리고 DGCNN [2] 은 공간 변환 네트워크를 사용함에도 불구하고, 데이터 증진과 관계없이, 임의의 회전 변환을 처리하는 데 효과적이지 않은 것으로 나타났다. 스펙트럼 그래프 필터링을 활용하여 지역 기술자를 구성하는 SpecGCN [4] 은 그래프 신호가 3차원 좌표와 동일하기 때문에, 회전에 강건한 모습을 보이지 못했다.

한편, SFCNN [21]은 데이터 증진을 통해 훈련 과정 중 관측된 회전에 대해 경쟁력 있는 정확도를 보이지만, 관측되지 못한 회전 유형인 $z/SO(3)$ 에 대해서는 처리하는 데 어려움을 겪으며, $SO(3)/SO(3)$ 의 정확도와 관련하여 큰 격차를 보인다. 최근 회전 불변을 명시적으로 다루는 방법 [11, 12] 는 관측되지 않은 회전 유형에 대한 인식 정확도를 유지하지만, 전반적으로 만족스럽지 못한 성능을 보인다. 우리의 알고리즘은 모든 평가 방법에 대해 일관적으로 매우 높은 성능을 보이고 있으며, 관측되지 못한 회전 변환에 대한 강건함을 보이고 있음을 알 수 있다.

제 2 절 오류 분석

3차원 분류 벤치마크인 데이터 셋 모델넷40은 40개 카테고리의 클래스가 존재한다. 이번 절에서는 우리 프레임워크가 각 클래스 별로 어떤 성능을 보이고 있는 지 혼동 행렬(confusion matrix)을 기반으로 분석한다.

그림 4는 우리 프레임워크의 예측을 기반으로 하는 혼동 행렬을 나타낸 것이다. 그림 5는, 그림 4를 정규화된 표현으로 나타낸 혼동 행렬이다. 벤치(bench)와 테이블(table), 화분(flower pot)과 꽃병(vase), 그리고 화분과 식물(plant) 에서 상대적으로 높은 오류를 보이고 있다. 이는, 각 클래스 별로, 비슷한 형상을 띠고 있는 오브젝트들이 있기 때문인데, 예를 들어, 화분과 꽃병은 그 형상이 매우 비슷하고, 하나의 클래스로 단정 지어 카테고리 분류하기에 그 애매함(ambiguity)이 존재하는 오브젝트들이 클래스 간 존재하게 되는데, 그러한 클래스에 대하여 높은 오류를 보이고 있다.

제 3 절 Ablation 실험

우리는 제안하는 접근법의 효과를 분석하기 위해 다양한 ablation 실험을 수행하였다.

표 2. 확률적 학습에 대한 기여도. 체크 마크 표시는 각 파라미터가 확률성을 부여 받았음을 의미한다.

	Deterministic	Partially Stochastic						Fully Stochastic
d		V			V	V		V
k			V		V		V	V
\hat{k}				V		V	V	V
z/SO(3) (%)	87.7	88.6	88.9	88.0	89.2	89.0	88.9	89.5

1. 확률적 학습의 효과

제시한 확률적 학습의 효과를 보이기 위해, 우리는 지역과 전역 기술자를 학습하는 과정에서 사용하는 세 가지 확률적 요인의 영향을 분석하여, 결정론적 접근법(deterministic approach)과 비교하였다. 이 때, 사용되는 확률적 요인은, 확장 비율(dilation rate), 이웃 포인트의 탐색 개수, 그리고 그래프 구축 단계에서 에지 개수가 된다. 표 2는 순수 결정론적 학습의 모델이 정확도가 낮으며, 확률적 학습 구성요소를 추가하면, 결과가 일관되게 개선된다는 것을 보여주고 있다.

확률적 학습의 도입은 과 적합 문제를 효과적으로 해결할 수 있고, 다양한 형태로 입력될 수 있는 구름 점 데이터에 대하여 그 기하 형상 정보에 대한 특징 추출 과정에서 일반화된 특징을 얻을 수 있게 된다. 특히, 우리 프레임워크의 경우, 확률적으로 형성되는 지역 영역에 대해 적용 될 수 있는 정규화 과정은 지역 특징에 긍정적인 영향을 주게 되어, 이는 단일 전역 특징이 더욱 대표적(representative)으로 추출될 수 있음을 유추할 수 있다.

표 3. 서로 다른 방법론에 따른 모델넷40 데이터 셋에 대한 3차원 물체 분류 정확도 (%)

(a)			(b)		
Architecture variation	z/SO(3)	SO(3)/SO(3)	Transformation scope	z/SO(3)	SO(3)/SO(3)
MLP	89.1	89.2	Global transformation	87.2	87.2
GCN (ours)	89.5	89.5	Local transformation (ours)	89.5	89.5

(c)			(d)		
Levels of models	z/SO(3)	SO(3)/SO(3)	Stochastic dilation	z/SO(3)	SO(3)/SO(3)
Single level	88.7	88.9	Deterministic		
Two levels	89.1	89.3	Dilated k-NN [27]	88.3	88.3
Three levels (ours)	89.5	89.5	Stochastic		
Four levels	89.3	89.4	Dilated k-NN (ours)	89.5	89.5

2. 그래프 신경망과 다중 퍼셉트론에 대한 비교

표 3 (a)는 그래프 합성 곱 신경 망을 사용한 모델과, 다중 퍼셉트론을 사용한 모델과의 비교 결과를 제시하고 있다. 비교를 위해 우리는 모든 계층에 존재하는 단일 그래프 신경 망 레이어를 단일 완전 연결 레이어(fully-connected layer)로 대체하였고, 생성되는 특징의 차원은 동일하게 맞추었다. 그래프 신경 망의 사용은 다중 퍼셉트론을 사용한 것과 비교하였을 때 성능 향상을 보였다. 이는, 그래프 합성 곱 신경 망을 이용해 단일 노드가 이웃 노드에 대한 특징을 함께 고려하여 기하학적 문맥을 파악하는데 도움을 주기 때문이다.

3. 지역 변환과 전역 변환에 대한 비교

우리는 지역 별로 주 성분 축을 계산하여 이를 기반으로 특징 학습을 하는 것과, 전역에 대한 주 성분 축을 기반으로 특징 학습을 하는 것에 대한 비교를 진행하였다. 이를 위해서, 우리의 방법과 동일한 새로운 모델을 만들었고, 지역 별로 주 성분 축을 계산하지 않고, 전체 입력 구름 점을 대상으로 주 성분 분석(principal component analysis)을 통해 전역 회전 행렬을 얻어내어 입력 구름 점들을 회전 시켜 학습에 이용하였다. 네트워크의 하이퍼 파라미터와 학습과정은 우리의 모델과 동일하게 설정하였다. 표 3

(b) 는 지역 변환을 통한 좌표 계를 기반으로 하는 학습 방법이 전역 변환을 사용한 것과 비교하였을 때, 더 우수한 성능을 보이는 것을 확인 할 수 있다.

4. 계층 적 모델링의 효과

우리가 제안하는 계층 적 표현에 대한 효과를 입증하기 위해, 우리는 신경 네트워크 아키텍처의 계층 수와 관련하여 분류 정확도를 평가하였다. 표 3 (c)은 계층 수를 증가시켜 기술자가 수용할 수 있는 영역을 확대시킴으로써 분류 정확도가 증가함을 알 수 있으며, 반면, 계층 수를 계속적으로 추가하는 것은 성능 향상에 특별한 도움을 주지않는다는 것에 대해 확인했다.

5. 확률적 확장 k -NN 탐색의 효과

[27]에서 사용되는 탐색 방법과 달리, 우리가 제안하는 알고리즘은 k 와 d 를 랜덤 변수로 설정하여 확률적으로 확장된 k -NN 탐색을 사용하게 된다. 이와 같이, 탐색을 위한 이웃 점들의 범위는 $k*d$ 로 확률적으로 결정되게 되는데 이는 지역 기술자에 대해 정규화 효과를 얻어낼 수 있고, 노이즈나 형상 변형에 강건한 모델이 되도록 할 수 있다. 표 3 (d)는 우리의 확률적 이웃 탐색 전략의 우수성을 명백하게 보여주고 있다.

제 4 절 노이즈와 아웃라이어에 대한 강건성

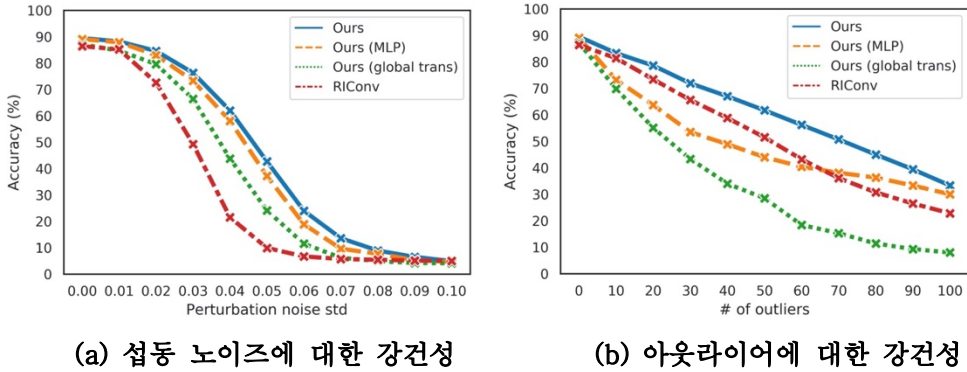


그림 7. 섭동 노이즈, 아웃라이어에 대한 모델넷40 분류 정확도 평가

우리는 또한 데이터의 노이즈에 대한 강건성을 평가하기 위한 실험을 수행하였다. 위 실험을 수행하기 위해서 섭동(perturbation)과 아웃라이어 두 가지 형태의 노이즈를 적용하였다. 섭동 노이즈를 이용한 실험은 평균이 0인 정규분포에 대하여 표준편차 값을 변화시켰고, 아웃라이어는 단위 구 내부에서 샘플된 점들을 각 객체에 대한 입력 포인트에 추가하여 각 실험에 대한 분류 정확도를 측정하였다. 모든 비교 대상 알고리즘에 입력되는 3차원 물체는 1,024개의 점들로 구성된다.

그림 3 (a) 는 GCN에 기초하여 제안된 알고리즘의 섭동 노이즈에 대한 강건성에 관한 곡선을 나타낸다. 그래프 컨벌루션 신경 망 대신 다중 퍼셉트론을 사용하면 섭동 노이즈가 적용된 물체 분류 정확도가 소폭 저하된다. 한편, 전역 변환을 기반으로 학습된 모델은 지역 변환 기반의 방법보다 효과적이지 못했다. 그리고 RIConv [12]는 섭동 노이즈에 대하여 매우 현저한 성능 저하를 겪음을 확인할 수 있었다.

한편, 그림 3 (b)는 아웃라이어가 존재하는 상황에서의 성능을 나타내며, 여기서 우리의 알고리즘은 다른 모든 방법과 비교하여 우수한 성능을 나타낸다. 특히 그래프 합성 곱 신경 망은 다중 퍼셉트론에 비하여 상당한 정확도 상승을 달성했다. 이 실험에서, 지역 변환은 전역 변환보다 우수한 성능을 나타냈는데, 이는 전역 변환은, 추가된 아웃라이어 점에 대하여 전역에 대한 축 정보가 지역적으로 형성되는 축들에 비해 방향 정보에 대한 왜곡이 훨씬 심하다는 것을

유추할 수 있다. 마지막으로, RConv [12] 의 결과는 낮은 레벨의 특징을 가진 표현 학습이 섭동 노이즈 뿐만 아니라 아웃라이어에 대하여 충분히 강하지 않음을 보여준다.

제 4 절 정성적 결과

그림 8은 계층 구조를 따라 생성된 그래프의 예를 보여준다. 첫 번째 계층 그래프는 에지들이 밀집하게 얽혀 있으나, 계층이 증가할수록 점점 희소해지게 된다. 계층 구조가 깊어질수록 전체적인 골격이 포착되어 전역적 표현을 추출하는 데 도움이 된다. 따라서 여러 계층에서 그래프 합성곱 신경 네트워크에서 학습한 표현은 의미적 수준에서 서로 보완된다는 것을 예측할 수 있다.

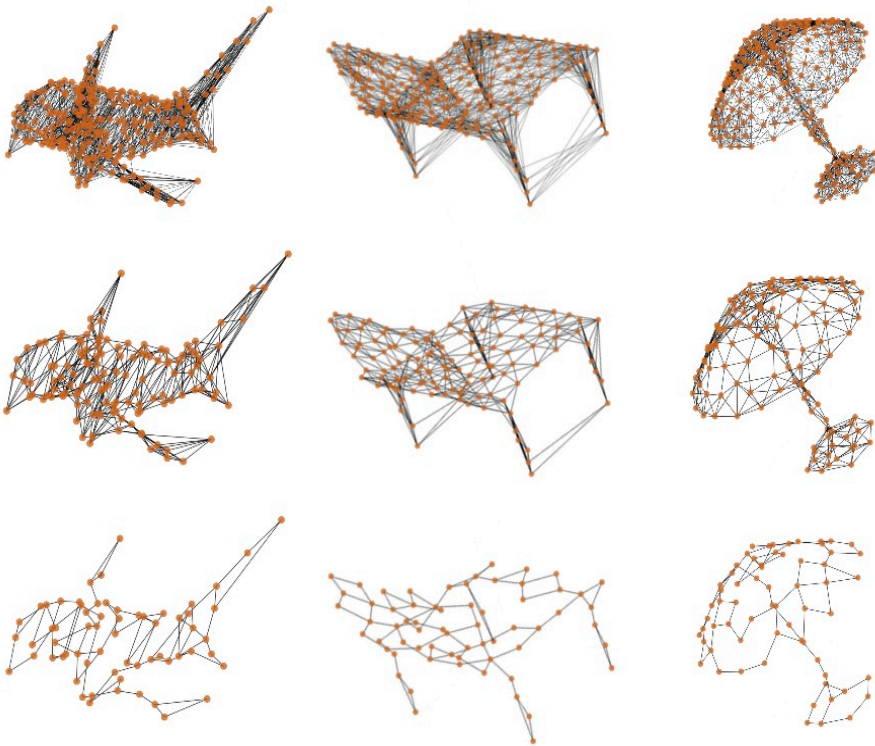


그림 8. 각 계층마다 생성된 그래프 예시의 정성적 결과. 갈색 원은 그래프 노드를 나타내고, 실선은 노드 사이의 에지를 나타낸다.






	GT	PointNet	PointNet++	DGCNN	SpecGCN	RIConv	Ours
(a) 	airplane	plant	plant	plant	plant	airplane	airplane
(b) 	table	stairs	door	stool	chair	table	table
(c) 	bowl	plant	vase	vase	plant	tent	bowl
(d) 	car	stairs	bottle	plant	bottle	stairs	car
(e) 	xbox	tv-stand	tent	table	tent	range-hood	radio

그림 9. 임의로 회전된 물체에 대한 모델들의 예측

그림 9는 모델넷40 내 몇 가지 카테고리에 존재하는 물체를 임의로 회전한 것에 대하여 다양한 프레임워크로부터 예측된 결과를 나타낸 그림이다. 포인트넷, 포인트넷++, DGCNN, 그리고 SpecGCN은 회전 변환을 고려한 요소가 없기 때문에, 임의의 회전에 대해 정확한 예측을 하지 못한다는 것을 확인할 수 있다. 언급한 프레임워크는 모두 임의의 회전이 입력되지 않을 경우, 매우 우수한 인식 성능을 보이는 프레임워크라는 사실에 비추었을 때, 회전 변환에 대한 요소 유무가 실제 응용 환경에서 매우 중요한 이슈가 될 것이라는 것을 확인할 수 있다. 반면, RIConv는, 점 들간의 거리와 각도를 계산하여 회전에 불변한 특징들을 사용했기 때문에 임의의 회전에도 정확한 예측을 하고 있다. (a), (b)의 경우, RIConv와 우리 프레임워크는 모두 정확한 예측을 하였다. 그러나, (a)와 (b)와 달리, (c), (d)와 같이 특징적인 부분이라도드러지는 형상이 많지 않거나, 거의 없는 물체의 경우 RIConv는 형상을 제대로 파악하지 못했고, 우리의 프레임워크만이 정확한 예측을 하였다. (e)는 분별하기 매우 어려운 형태를 가졌기에, 모든 프레임워크가 예측에 실패한 것을 확인할 수 있다.

제 5 장 결론

우리는 회전 변환에 강건한 3차원 물체 분류 프레임워크를 제안하였다. 본 연구에서 제안된 프레임워크는 회전에 불변한 지역 기술자를 학습시키고, 이를 위상학적 구조에서 그래프 합성 곱을 수행하여 지역적 특징을 집적하는 전역 기술자를 추출하여 3차원 물체 분류를 수행한다. 본 프레임워크에서 사용되는 확률적 학습 전략은 3차원 물체의 기하학적 특징을 효과적으로 정규화 하였고, 이는 물체 분류 성능을 크게 증가시켰다. 거리, 각도와 같은 낮은 수준의 특징으로 변환하여 사용하던 기존의 방법들과 달리, 우리의 방법은 그러한 변환 없이 직접적으로 구름 점을 사용한다는 점에서 유일하다. 또한, 계층 적 구조 기반의 그래프 합성 곱 신경 망은 계층이 깊어짐에 따라 넓은 영역에 대한 기하학적 구조를 기술하는 지역 기술자를 성공적으로 통합함으로써 임의의 회전으로 증진된 3차원 물체 분류 벤치마크 데이터 셋에서 최고 성능을 보이고 있다. 마지막으로, 다양한 실험환경에서 수행된 광범위한 실험은 우리 프레임워크의 효과를 검증하였다.

참고 문헌

- [1] Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3D classification and segmentation. In CVPR. (2017)
- [2] Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph cnn for learning on point clouds. ACM Transactions on Graphics (TOG) 38(5) (2019) 146
- [3] Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In NIPS. (2017)
- [4] Wang, C., Samari, B., Siddiqi, K.: Local spectral graph convolution for point set feature learning. In ECCV. (2018)
- [5] Atzmon, M., Maron, H., Lipman, Y.: Point convolutional neural networks by extension operators. In SIGGRAPH. (2018)
- [6] Xu, Y., Fan, T., Xu, M., Zeng, L., Qiao, Y.: SpiderCNN: Deep learning on point sets with parameterized convolutional filters. In ECCV. (2018)
- [7] Zhang, Y., Rabbat, M.: A graph-cnn for 3D point cloud classification. In ICASSP. (2018)
- [8] Xiong, Y., Ren, M., Liao, R., Wong, K., Urtasun, R.: Deformable filter convolution for point cloud reasoning. arXiv preprint arXiv:1907.13079 (2019)
- [9] Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J.: KPConv: Flexible and deformable convolution for point clouds. In ICCV. (2019)
- [10] Wu, W., Qi, Z., Fuxin, L.: PointConv: Deep convolutional networks on 3D point clouds. In CVPR. (2019)
- [11] Chen, C., Li, G., Xu, R., Chen, T., Wang, M., Lin, L.: ClusterNet: Deep hierarchical cluster network with rigorously rotation-invariant representation for point cloud analysis. In CVPR. (2019)

- [12] Zhang, Z., Hua, B.S., Rosen, D.W., Yeung, S.K.: Rotation invariant convolutions for 3D point clouds deep learning. In International Conference on 3D Vision (3DV). (2019)
- [13] Deng, H., Birdal, T., Ilic, S.: PPF-FoldNet: Unsupervised learning of rotation invariant 3D local descriptors. In ECCV. (2018)
- [14] Deng, H., Birdal, T., Ilic, S.: PPFNet: Global context aware local features for robust 3D point matching. In CVPR. (2018)
- [15] Klokov, R., Lempitsky, V.: Escape from cells: Deep kd-networks for the recognition of 3d point cloud models. In The IEEE International Conference on Computer Vision (ICCV). (2017)
- [16] Wang, P.S., Liu, Y., Guo, Y.X., Sun, C.Y., Tong, X.: O-cnn: Octree-based convolutional neural networks for 3d shape analysis. ACM Transactions on Graphics (TOG) 36(4) (2017) 1-11
- [17] Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E.: Multi-view convolutional neural networks for 3d shape recognition. In The IEEE International Conference on Computer Vision (ICCV). (2015)
- [18] Shen, Y., Feng, C., Yang, Y., Tian, D.: Mining point cloud local structures by kernel correlation and graph pooling. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018)
- [19] Cohen, T.S., Geiger, M., Kohler, J., Welling, M.: Spherical CNNs. In ICLR. (2018)
- [20] Esteves, C., Allen-Blanchette, C., Makadia, A., Daniilidis, K.: Learning SO (3) equivariant representations with spherical cnns. In ECCV. (2018)
- [21] Rao, Y., Lu, J., Zhou, J.: Spherical fractal convolutional neural networks for point cloud recognition. In CVPR. (2019)
- [22] Bruna, J., Zaremba, W., Szlam, A., LeCun, Y.: Spectral networks and locally connected networks on graphs. In ICLR. (2014)
- [23] Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In NIPS. (2016)

- [24] Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In ICLR. (2017)
- [25] Hammond, D.K., Vandergheynst, P., Gribonval, R.: Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis* 30(2) (2011) 129-150
- [26] Simonovsky, M., Komodakis, N.: Dynamic edge-conditioned filters in convolutional neural networks on graphs. In CVPR. (2017)
- [27] Li, G., Müller, M., Thabet, A., Ghanem, B.: DeepGCNs: Can GCNs Go as Deep as CNNs? In ICCV. (2019)
- [28] Gonzalez, T.F.: Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science* 38 (1985) 293-306
- [29] Li, J., Chen, B.M., Hee Lee, G.: SO-Net: Self-organizing network for point cloud analysis. In CVPR. (2018)
- [30] Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3D shapenets: A deep representation for volumetric shapes. In CVPR. (2015)

Abstract

Rotation-Invariant Local-to-Global Representation Learning for 3D Point Cloud

Seohyun Kim

Electrical and Computer Engineering

The Graduate School

Seoul National University

We propose a local-to-global representation learning algorithm for 3D point cloud data, which is appropriate to handle various geometric transformations, especially rotation, without explicit data augmentation with respect to the transformations. Our model takes advantage of multi-level abstraction based on graph convolutional neural networks, which constructs a descriptor hierarchy to encode rotation-invariant shape information of an input object in a bottom-up manner. The descriptors in each level are obtained from neural networks based on graphs via stochastic sampling of 3D points, which is effective to make the learned representations robust to the variation of the input data.

The proposed algorithm presents the state-of-the-art performance on the rotation-augmented 3D object recognition benchmarks and we further analyze its characteristics through comprehensive ablation experiments.

Keywords : Rotation-invariant, 3D Point Cloud, Stochastic Learning, Graph

Student Number : 2018-20013