

일반논문 (Regular Paper)

방송공학회논문지 제29권 제3호, 2024년 5월 (JBE Vol.29, No.3, May 2024)

<https://doi.org/10.5909/JBE.2024.29.3.321>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

트랜스포머를 이용한 거리함수 기반 포인트 클라우드 개선

김도훈, 김철현^{a)†}

Distance Fields based Point Cloud Enhancing using Transformer

Dohoon Kim and Chulhyun Kim^{a)†}

요약

3D 센서의 놀라운 발전에도 불구하고, 취득된 3차원 포인트 클라우드는 잡음, 낮은 밀집도, 그리고 비균일성과 같은 문제점을 나타내는 경향이 있다. 이러한 데이터를 이용해 3D 콘텐츠의 다양한 산업군에 적용했을 때, 성능이 저하되기 때문에 포인트 클라우드의 품질을 향상시키는 것은 중요하다. 그러나 기존 네트워크 기반의 포인트 클라우드 품질 향상 기법은 신경망의 한계로 인하여 몇 가지의 문제가 있었다. 본 논문에서는 거리함수 기반의 트랜스포머 모델을 제안한다. 제안하는 모델은 실제 포인트 클라우드와 입력 포인트 클라우드 간 거리를 예측한다. 제안하는 방법은 기존 방법들보다 뛰어난 성능을 보였다. Chamfer distance (CD), Hausdorff distance (HD), 그리고 ground truth mesh 데이터와 예측한 포인트 클라우드 간 거리를 나타내는 point-to-surface distance (P2F) 3가지 측정에서 모두 기존 Grad-PU보다 우수한 값을 보였다.

Abstract

Despite of the superior advancement of 3D sensors, raw 3D point clouds are poor quality including noisy, sparse, and irregularity. Because these point clouds cause performance degradation in various 3D applications, enhancing raw point clouds is important. However, existing learning-based point cloud super-resolution methods have some limitations in the architecture of neural networks. In this paper, we propose transformer-based model applying distance fields. Our network predicts the distance between input point cloud and ground truth point cloud. Experiments demonstrate our method outperforms existing methods. It exhibited superior values in all three measurements: Chamfer Distance (CD), Hausdorff Distance (HD), and point-to-surface distance (P2F) when compared to the existing Grad-PU, indicating a closer distance between the predicted point cloud and the ground truth mesh data.

Keyword : Point cloud, Implicit neural representation, Vision transformer

a) 나사렛대학교 방송영상콘텐츠학부(Department of Broadcasting, Visual Design and Media Contests, Korea Nazarene University)

† Corresponding Author : 김철현(Chulhyun Kim)

E-mail: acts@kornu.ac.kr

Tel: +82-41-570-4817

ORCID: <https://orcid.org/0000-0001-9421-8622>

· Manuscript April 3, 2024; Revised May 21, 2024; Accepted May 22, 2024.

1. 서론

오늘날 3차원 영상은 여러 산업군에서 사용하고 있다. 의료, 국방, 자율 주행, 게임, 영화제작까지 많은 분야에 필수적인 요소이다. 3차원 영상은 3차원 위치 값과 속성 값을 갖는 점들의 집합이다. 이러한 점들의 구성과 집합을 포인트 클라우드라고 부른다^[1].

포인트 클라우드 3차원 영상 획득은 개인용 스마트폰부터, 스캐너, 자율 주행 차량 등 다양한 분야에서 사용하고 있다. 적용 분야가 늘고, 수요가 올라가면서 Three-dimensional(3D) 센서 기술도 성장하였다. 최근 많이 응용되는 분야는 자율주행 분야이며, 이 분야를 중심으로 포인트 클라우드 기술이 경쟁적으로 개발되고 있다. 또한 일반 소비자용 제품인 스마트폰에도 Lidar(라이다) 센서가 장착되어, 이를 이용해 3차원 객체를 획득하고 포인트 클라우드로 처리할 수 있다. 유명한 게임엔진 언리얼은 포인트 클라우드 데이터를 다른 처리 없이 게임엔진에 즉시 사용할 수 있도록 플러그인을 제공한다^[2]. 이를 응용해 사용자가 직접 생산해야만 하는 3차원 객체를 저비용으로 구현할 수 있게 했다. 그림 1에 언리얼 게임엔진에 포인트 클라우드 데이터를 처리해 사용하는 장면이 보인다.

다양한 산업분야의 요구와 자율주행의 발전으로 3D 센서 기술은 크게 발전하였지만, 포인트 클라우드 데이터는 몇 가지 약점이 있다. 그중 큰 약점은 영상의 해상도 영역

내에 필연적으로 빈공간이 존재한다는 점이다^[3]. 그림 1에도 유리창, 건물의 모서리 등 라이다 센스가 획득하기 어려운 영역은 노이즈가 심한 것을 볼 수 있다. 2차원 영상은 해상도와 각 점들의 값으로 완벽한 재현이 가능하지만 3D 포인트 클라우드는 완벽한 재현이 어렵다.

이러한 포인트 클라우드를 3D 어플리케이션에 적용할 경우 성능이 낮아지기 때문에 이를 가공하는 것이 필수적이다. 그림 1에 제시된 것과 같이 비록 고성능 센서를 이용해 색상정보와 위치 정보 모두를 획득해도 낮은 밀도로 인해 사용자가 제작한 3차원 객체에 비해 해상도가 떨어지는 것을 볼 수 있다. 따라서 포인트 클라우드 개선을 위해 원본 데이터의 근본적인 문제들을 해결해야 한다. 이는 3차원 콘텐츠 제작과 자율 주행 등 포인트 클라우드 응용 산업분야에 매우 주요한 문제이다.

딤러닝 사용 이전 포인트 클라우드를 개선하는 방법은 최적화 기법을 중심으로 사용하였다^{[4][5]}. 이러한 연구는 주로 컴퓨터 그래픽 분야에 응용하기 위한 접근방법이다. 이후 포인트 클라우드 개선에 딤러닝 방법이 사용된 후로^{[6][7]}, 큰 발전이 있었으며, 현재 PU-Net이 좋은 평가를 받고 있다^[8]. 해당 방법은 2D 영상에서의 개선 방법들을 기반으로 각각 3단계의 모듈로 나눌 수 있다.

첫 번째로 포인트 클라우드를 특징 데이터로 변환하는 과정, 두 번째로 특징 데이터에서 점의 개수를 늘려서 업샘플링하는 과정, 마지막으로 해당 특징 데이터를 다시 포인



그림 1. 게임엔진 언리얼에서 3D 포인트 클라우드 데이터 재생 화면
Fig. 1. Point cloud 3D data playback screen in game engine Unreal

트 클라우드로 복원하는 과정으로 나뉜다. 해당 방법을 기반으로, 포인트 클라우드를 패치 단위로 나누어 수 차례 반복하여 점의 개수를 늘리며, 잡음을 제거하는 MPU^[9]와 생성 모델 중 하나인 Generative Adversarial Networks를 포인트 클라우드를 개선하는 데 적용한 PU-GAN^[10]이 등장하였다. 그 후로 기존 graph convolution 방법을 적용시킨 PU-GCN^[11], coarse to fine 기반 end-to-end의 딥러닝 모델을 제안한 Dis-PU^[12], natural language processing(NLP)에서 제안되어 딥러닝을 발전시킨 트랜스포머를 초해상도에 적용한 PU-Transformer^[13]가 등장하였다.

앞서 기술한 포인트 클라우드 개선 기법은 딥러닝 방법의 변화를 거듭하며 성능을 크게 향상시켰으나 여전히 취약한 부분이 존재한다. 이는 특징 데이터의 점의 개수를 늘리는 과정에서, 해당 모델의 파라미터를 잘 학습하지 못할 경우 성능이 떨어지는 문제이다. 이를 개선하기 위한 잔차 연결과 같은 상대적으로 안정적인 학습 방법도 있었으나, 점의 개수를 늘리는 과정에서 데이터 크기가 불일치하는 문제점이 있다.

딥러닝이 포인트 클라우드에 적용되어 성능을 크게 향상시킨 후, 포인트 클라우드 표면 재구성 분야 또한 딥러닝 방법들을 사용할 것이 제안되었다. 그 중 딥러닝을 통해서 3차원 데이터를 표현하는 암시적 신경 표현(Implicit Neural Representation; INR)은 표면 재구성의 성능을 크게 향상시켰다. 이는 임의의 3차원 점들에 대해서 주어지지 않은 정답 표면과의 거리를 계산하는 부호화, 비부호화 거리함수(Signed, Unsigned Distance Function; SDF, USDF)^[14], 정답 표면으로부터 임의의 3차원 점들이 내부, 외부에 위치하는지를 판단하는 값인 확률을 예측하는 점유 네트워크(Occupancy Nets)들이 제안되어 향상시킨 것이다^[15]. 특히 포인트 클라우드에 잡음이 있거나 점 간 위치가 불균일하더라도 모델이 강건하게 예측하는 결과가 나왔는데, 이는 기존에 네트워크가 단순히 입력 포인트 클라우드의 정보를 통하여 표면을 생성하는 것에 비해, 임의의 3차원 점들에 대해서 입력 포인트 클라우드를 기반으로 실제 표면과의 값을 계산하기 때문이다. 이때, 충분한 데이터가 주어질 경우, 네트워크가 정답 표면에 대한 암시적인 정보를 보다 안정적으로 학습할 수 있다.

본 논문에서는 포인트 클라우드 표면 재구성 분야에서

뛰어난 성능을 달성한 암시적 신경 표현 기반의 방법을 포인트 클라우드 개선 기법에 적용시켰다. 그 중 비부호화 거리함수를 사용할 것을 제안하였는데, 먼저 포인트 클라우드를 보간하여 3차원 공간 상의 밀집도를 높인 뒤, 잡음이 심한 보간된 포인트 클라우드를 개선하기 위해, 정답 포인트 클라우드로부터의 3차원 거리를 예측한다. 예측하는 모델은 포인트 클라우드 데이터 구조에 적합한 트랜스포머를 사용할 것을 제안한다. 제안하는 트랜스포머 기반 모델은 잔차 연결과 풀링을 통해 각각 학습의 안정성과 비선형성을 부여하여 기존 방법들보다 강건한 예측이 가능하도록 설계하였다. 또한, 학습이 끝난 후 추론 과정에서 비부호화 거리값을 역전파하여 각 점에 대해 3차원 벡터를 예측하고, 보간된 포인트 클라우드는 예측된 벡터에 의해 잡음이 개선된다.

본 논문은 1장 서론에서 포인트 클라우드 필요성과 연구 동향 및 본 논문의 연구 방법을 제안하고 2장 기존 연구에서 지금까지 포인트 클라우드 개선 방법 중 2017년 포인트 넷 발표 이후 연구를 정리하였다. 3장 포인트 클라우드에 MLP(Multi Layer Perceptron)를 적용하고 트랜스포머와 거리함수를 이용해 개선하는 방법을 제안, 설명하였다. 4장에서는 3장에 제안한 내용을 토대로 실험결과를 정리하였고 기존 연구보다 시각적으로 우수한 결과를 보였다. 마지막 5장, 결론으로 본 논문을 정리하였다.

II. 관련 연구

2장에서는 2019년 포인트넷을 활용한 포인트 클라우드 개선방법 이후 등장한 주요 기법들을 소개하고 이중 최근 연구이며 효능이 좋은 암시적 신경 표현(Implicit Neural Representation; INR)기법을 설명하고 이를 개선시키고자 한다.

1. 딥러닝 기반 포인트 클라우드 처리 기법

2000년대 초반부터 딥러닝을 2차원 영상처리 분야에 적용하는 연구가 활발하게 수행되었음에도 불구하고, 3차원 좌표값의 집합으로 나타나는 포인트 클라우드에는 딥러닝을 적용하는 것이 어려웠다. 이는 포인트 클라우드가 2차원 비전 영상과 다르게 행의 순서가 바뀌어도 같은 데이터라

는 특성인 순열 불변성(permutation invariance)을 가지기 때문이다. 따라서 순열 가변성(permutation equivariance)의 특성을 가지는 Convolutional Neural Network(CNN)를 포인트 클라우드에 적용하지 못하였고, 이에 2차원 영상과 유사한 구조를 가진 복셀에 딥러닝을 적용하는 연구가 여러 차례 등장하였다.

스탠포드 대학의 Charles R. Qi와 연구팀이 제안한 PointNet은 앞서 기술한 순열 불변성 문제를 해결하기 위하여 포인트 클라우드를 multi-layer perception(MLP)에 적용시키고, 포인트 클라우드의 모든 점들에 대해 max pooling을 적용하였다. 그러나 MLP는 2차원 영상에 적용되는 컨볼루션과 다르게 데이터의 지역적인 정보를 학습할 수 없는 문제가 있다. 이를 해결하기 위해 연구팀은 PointNet++를 추가로 제안하였는데^[7], 제안하는 방법은 k 최근접 이웃(k Nearest Neighbor; kNN), ball query와 같은 방법들을 포인트 클라우드에 적용함으로써 포인트 클라우드의 이웃한 점들을 군집시키고 이를 MLP에 적용시키는 것으로 기존 PointNet 대비 높은 성능을 달성하였다. PointNet++가 제안됨에 따라 본격적으로 포인트 클라우드 분할, 분류의 태스크를 다루는 딥러닝 기반 포인트 클라우드 처리 기법의 연구가 시작되었다. 또한, 현재까지도 이 논문에서 제안한 U-Net 구조의 네트워크와 포인트 클라우드 군집화 방법은 포인트 클라우드에 딥러닝을 사용할 때에 기저가 된다.

한편 최근 연구되고 있는 분야인 NLP(Natural Language Processing)는 트랜스포머가 제안되어 놀라운 발전을 이루었다. 해당 방법 중 셀프 어텐션은 각 데이터의 특성별로 적용되는 weight를 학습하기 때문에 각 데이터의 특징을 잘 학습할 수 있다. 이는 포인트 클라우드에도 적용되고 있다. 해당 방법은 그 우수성이 2차원 비전 영상에 적용되는 것으로 증명되었으며^[16], 3차원 포인트 클라우드에도 적용되었다^{[17][18]}. 본 논문에서도 포인트 클라우드 특성에 적합한 트랜스포머를 네트워크의 모듈로 사용하였다.

2. 포인트 클라우드 개선 기법

원시 포인트 클라우드는 위치값의 잡음, 공간적 밀집도의 불균일성, 객체의 일부 누락 등의 여러 문제가 있다. 포인트 클라우드를 개선하는 방법은 이러한 문제들을 개별

적으로 처리하는 것이다. 먼저, 위치값에 잡음이 있는 경우, 디노이징(denoising) 알고리즘을 사용하여 포인트 클라우드의 위치값을 보정한다. 이를 위해 딥러닝 네트워크는 포인트 클라우드의 각 점이 이동하는 벡터를 예측하도록 설계된다. 이때, 입력 포인트 클라우드와 출력 포인트 클라우드의 점 개수는 동일하다. 포인트 클라우드의 밀집도가 낮은 경우, 점의 개수를 늘려 고해상도의 포인트 클라우드로 변환하는 것이 필요하다. 그러나 점의 개수를 단순히 늘리면 잡음이 증폭되는 트레이드오프 문제가 발생한다. 따라서 밀집도를 높임과 동시에 잡음을 줄이는 초해상도(super-resolution)는 잡음 제거보다 더 어려운 작업이다.

마지막으로, 객체의 일부가 누락된 경우는 앞서 언급한 두 가지 경우와는 다른 문제이다. 이 경우, 각 객체별로 누락된 부분을 파악하고 점을 생성하여 포인트 클라우드를 완성(completion)시키는 과정이 필요하다. 이 과정은 지도 학습의 일종으로, 포인트 클라우드에서 객체를 명확히 파악하지 못하거나 학습에 사용되지 않은 객체가 입력으로 들어왔을 때는 좋은 결과를 도출하지 못하는 단점이 있다. 본 논문에서는 포인트 클라우드의 밀집도를 높여 고해상도의 포인트 클라우드를 생성하는 동시에, 이로 인해 발생하는 위치값의 잡음을 정제하는 과정을 제안하고 있다.

3. 암시적 신경 표현

딥러닝이 포인트 클라우드에 성공적으로 적용됨에 따라, 최적화 기법이 주를 이루었던 포인트 클라우드 표면 재구성 분야에도 딥러닝이 적용되었다. 그 중, 네트워크를 통해 3차원 표면 정보를 표현하는 암시적 신경 표현은 기존 방법 대비 개선된 성능을 보이며 현재까지도 활발하게 연구되고 있다^{[19][20]}. 이는 기존의 3차원 데이터를 획득하는 과정에서 이산화된 매쉬, 복셀, 포인트 클라우드와 다르게 임의의 토폴로지들에 대한 거리 또는 확률을 출력함으로써 암시적으로 3차원의 연속적인 데이터를 네트워크가 표현하게 하는 방법이다.

이를 제안한 방법 중 하나인 DeepSDF^[14]는 포인트 클라우드를 네트워크에 인코딩하여 특징 데이터로 변환한 후, 이 정보를 통하여 3차원 임의의 점들에 대해서 알지 못하는 정답 표면과의 거리를 예측한다. 그림 2는 입력 포인트 클

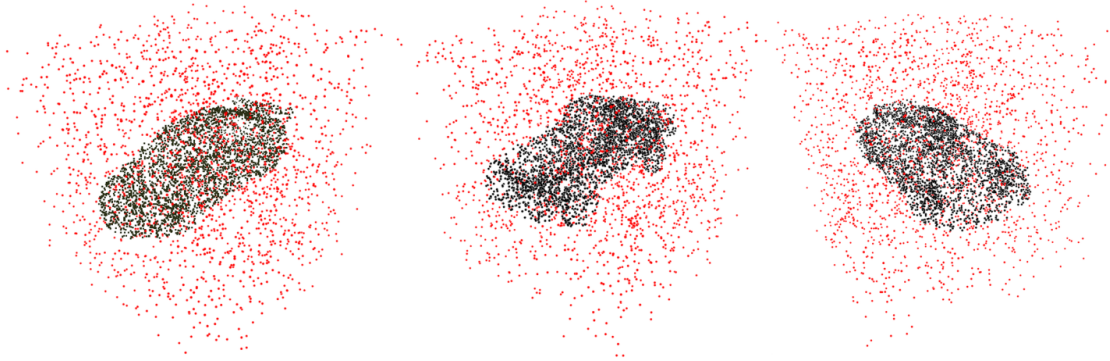


그림 2. 암시적 신경 표현의 입력인 포인트 클라우드(흑색)와 임의의 3차원 점(적색)
Fig. 2. Point cloud(black) and random 3D points(red) as input to implicit neural representation

라우드와 임의의 3차원 점을 시각화한 것으로, 적색 포인트 클라우드가 입력으로 들어오면 정답 표면으로부터의 거리가 출력되며, 이 거리의 오차를 바탕으로 학습된다. 학습이 완료되면, 추론 과정에서는 수만 개에서 수십만 개의 임의의 점을 3차원 공간 상에 위치시킨 후, 네트워크의 결과인 거리값이 0이 되는 점들이 정답 표면 상의 위치한 점이 된다. 이 점들을 사용해 **marching cubes** 알고리즘으로 메쉬로 변환하여 표면을 재구성한다.

본 논문에서는 최근의 방법과 유사하게 이러한 개념을 사용한다^[21]. 먼저 입력 포인트 클라우드를 보간하여 고해상도 포인트 클라우드를 생성하고, 각 점에 대해 주어지지 않은 정답 표면으로부터의 거리를 예측함으로써, 네트워크가 암시적으로 정답 표면에 근접하게 위치한 고해상도 포인트 클라우드를 추출할 수 있도록 한다.

III. 본 론

본 장에서는 먼저 딥러닝 기반 포인트 클라우드 처리 기법에서 사용하는 방법과 거리함수에 대한 기본적인 내용에 대해 다룬다. 그 후, 제안하는 딥러닝 모델과 손실 함수에 대해 다룬다.

1. 배경 이론

딥러닝을 포인트 클라우드에 적용하는 방법이 제안되었

음에도 불구하고, 2D 영상에서의 딥러닝 방법과는 달리 포인트 클라우드 기반 딥러닝 방법은 제한적인 일부 모델의 정보만 학습할 수 있다는 한계가 있었다. **PointNet++**^[7]는 이러한 문제를 효과적으로 해결하기 위해 **k** 최근점 이웃 (**kNN**)을 사용함으로써 포인트 클라우드를 하이퍼 파라미터 **k**의 수만큼 군집시키고, **Multi Layer Perceptron(MLP)**를 적용하는 방법을 제안하였다.

포인트 클라우드 표면 재구성은 촬영되어 이산화된 포인트 클라우드가 입력받아, 이를 바탕으로 연속적인 표면을 생성하는 것을 의미한다. 기존 연구는 포인트 클라우드의 점의 개수를 늘리거나 잡음을 제거한 후에 표면을 재구성하였으나, **DeepSDF**^[14]는 임의의 토폴로지를 사용하는 것을 제안하였다. 해당 방법은 3차원에 임의의 랜덤한 점들을 생성하고, 각 점들에 대해서 알지 못하는 표면과의 거리를 예측하는 방식으로, 최근 연구들도 이러한 거리함수 기반의 방법들을 사용하고 있다. 본 논문에서는 이러한 방법을 기반으로 입력 포인트 클라우드에 대해 보간을 수행한 후, 각 점들에 대해서 거리를 예측하는 것을 제안한다.

2. 제안된 방법

저해상도 포인트 클라우드 $X^{LR} = \{x_i\}_{i=1}^N$ 가 입력으로 주어진다고 가정하면, 여기서 N 은 점의 개수이다. 주어진 X^{LR} 은 잡음이 있으며, 3차원 공간을 충분히 채울 정도의 정보를 가지고 있지 않다. 이를 바탕으로 잡음을 줄인 고해

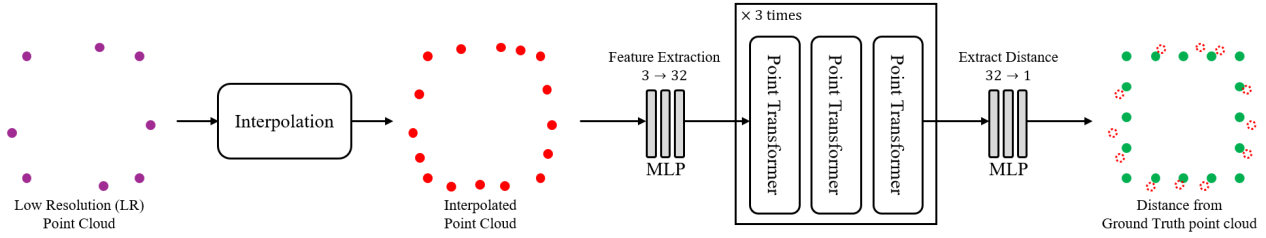


그림 3. 제안하는 네트워크의 파이프라인
 Fig. 3. Proposed Network Module Pipeline

상도 포인트 클라우드 $Y^{HR} = \{y_i\}_{i=1}^{rN}$ 를 생성하는 것이 포인트 클라우드 초해상도이다. 여기서 r 은 초해상도 배율을 나타내며 y_i 는 실제 데이터의 표면 위에 위치하는 점이다. 앞서 기술한 바와 같이 본 논문에서는 먼저 X^{LR} 을 보간하여 점의 개수를 늘린 후, 각 점들에 대한 거리 $D_X \in \mathbb{R}^{rN}$ 을 예측하는 거리함수 f_θ 를 사용한다.

$$D_X = f_\theta(I_r(X^{LR})) \tag{1}$$

여기서 $I_r(\cdot)$ 는 X^{LR} 에 대해 r 배 만큼 보간하는 함수를 의미한다.

그림 3에는 제안한 네트워크의 파이프라인이 나타나 있다. 여기서 보간 방법으로는 kNN을 사용하여 이웃점 x_k 를 생성하고, x 와 평균값에 해당하는 위치에 점을 생성한다. 그런 다음 가장 먼 점 샘플링(Farthest Point Sampling)을 사용하여 rN 개의 점이 남도록 샘플링한다. 이렇게 보간된 점들은 잡음이 심하기 때문에 이를 정답 표면에 맞춰 위치 값을 조정해야 한다. 이를 위해 초기 MLP를 통해 포인트 클라우드를 특징으로 변환한 후, 제안한 트랜스포머 모듈을 3번 반복하여 통과시키고, 마지막 MLP를 통해 ground truth(GT) point cloud로부터의 거리 값 D_x 를 예측한다. 제안한 네트워크에서 각 레이어의 수는 실험적으로 결정된다.

그림 4는 그림 3에서 사용된 제안한 포인트 트랜스포머 모듈이다. 먼저 kNN을 사용하여 계산된 오프셋 $\phi_x = x - x_k$ 과 kNN 인덱스를 통해 복제된 특징 $F_{x,k}$ 를 입력받는다. 기존 트랜스포머의 decoder에서 사용되는 attention 방법과 마찬가지로, cross attention을 통하여 입력 포인트 클라우드와 특징 간의 유사도를 계산하고 이를 특

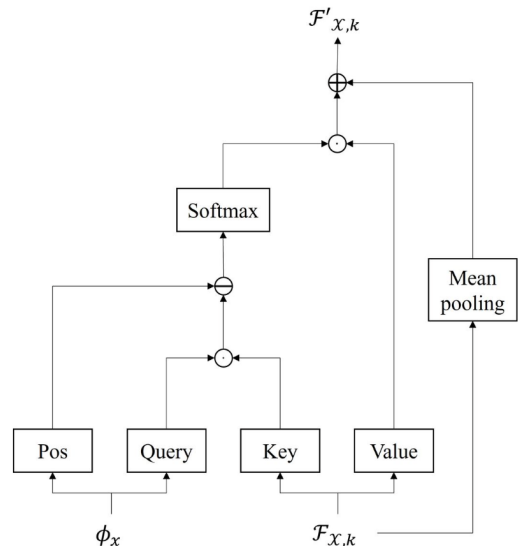


그림 4. 제안하는 포인트 트랜스포머 모듈
 Fig. 4. Proposed Point Transformer Module

징에 적용하는 방법을 사용한다. 기존 포인트 클라우드 트랜스포머 방법과 유사하게, 제안한 방법으로는 positional encoding을 위해 오프셋에 MLP를 적용하고 이를 attention 행렬에 빼줌으로써 포인트 트랜스포머 모듈이 학습하는 과정에서 입력 포인트 클라우드의 위치 정보를 유지하였다. 마지막으로 attention 행렬과 value에 해당하는 특징과 곱하여 정보를 업데이트하고, 잔차 연결로 기울기 손실 문제를 해결했으며 학습에 안정성을 높였다.

3. 훈련 전략(Training Strategy)

손실 함수로는 네트워크를 통하여 예측된 거리 D_X 와 실제 정답 포인트 클라우드 Y^{HR} 및 보간된 저해상도 포인트

클라우드 X^{LR} 간 거리의 유사도를 사용한다. 유사도를 측정하는 매트릭은 l_1 loss를 사용하였다.

$$\ell = \frac{1}{|Y^{HR}|} \left| \left\| Y^{HR} - I_r(X^{LR}) \right\|_2^2 - D_X \right| \quad (2)$$

학습을 마치고 추론하는 과정에서 거리값은 스칼라 값이고, 입력 포인트 클라우드는 3차원 값을 가지므로 포인트 클라우드의 위치값을 변경할 수 없다. 이 문제는 추론 과정에서 D_X 를 backward하여 각 좌표 값의 기울기를 계산함으로써 해결할 수 있다.

$$\nabla D_X = \frac{\partial D_X}{\partial x} + \frac{\partial D_X}{\partial y} + \frac{\partial D_X}{\partial z} \quad (3)$$

여기서 각 편미분 값은 x, y, z 로 나뉘지며 이를 보간된 입력 포인트 클라우드의 각 좌표에 대해서 빼줌으로써 고해상도 포인트 클라우드를 예측한다.

IV. 실험 결과

본 논문의 실험 환경으로는 Intel i7-8700K CPU, 64GB RAM, 그리고 NVIDIA GTX 2080ti를 탑재한 하드웨어를

사용하였다. 딥러닝 모델의 학습에 필요한 epoch은 200으로 설정하였으며, batch size는 32, 학습률은 0.001로 설정된 Adam 최적화 기법을 사용하였다. 학습에 필요한 데이터셋으로는 PU1K 데이터셋을 사용하였다. 해당 데이터셋은 Qian et al.^[11]이 제안한 것으로, ShapeNet 데이터셋의 GT mesh에서 포인트들을 샘플링하여 저해상도 포인트 클라우드와 고해상도 포인트 클라우드 데이터 쌍을 생성하였다.

1. 정성적 실험 결과

그림 5는 제안한 딥러닝 모델의 정성적 실험 결과를 기존 방법들과 비교한 것이다. (a) 입력된 저해상도 포인트 클라우드는 2,048개의 점을 가지며, (b)-(g) 고해상도 포인트 클라우드는 8,192개의 점을 포함하여 3D 공간에서 데이터를 표현하기에 충분한 밀도를 제공한다. 딥러닝 기반 포인트 클라우드 개선 기법의 결과인 (c)-(g)는 (b)와 유사할수록 더 좋은 결과로 평가된다. 제안한 방법은 기존 방법보다 국소적으로 점의 개수가 적은 지역에서도 포인트 클라우드의 표면을 더욱 정확하게 예측함을 보여준다.

2. 정량적 실험 결과

3D 상의 거리를 계산하기 위해 CD, HD 그리고 P2F 방법을 사용하였다. 이 세 가지 방법은 최근 포인트 클라우드

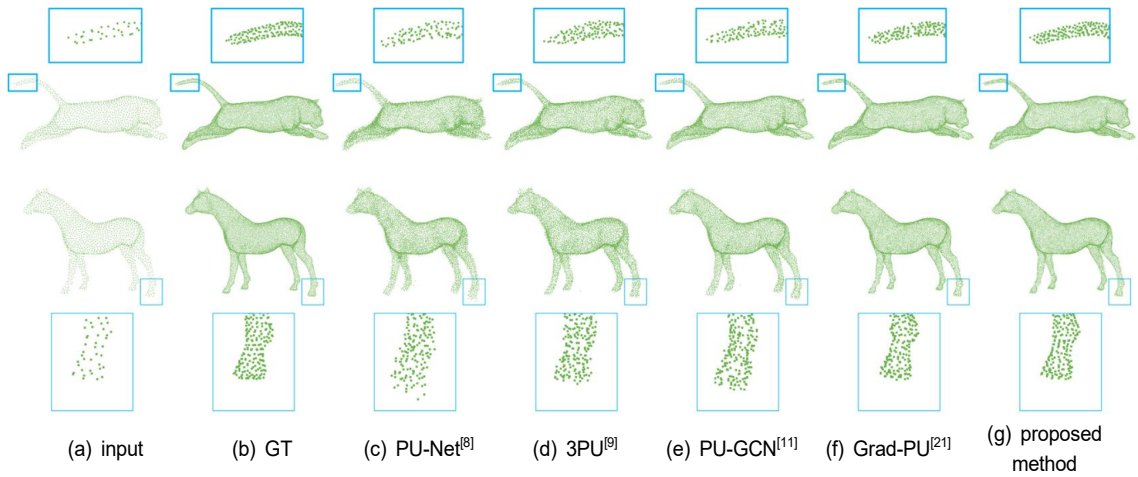


그림 5. 정성적 실험 결과
Fig. 5. Qualitative results

개선 기법 연구에서 정량적 평가에 주로 사용된다. 일반적인 2차원 이미지와 달리, 포인트 클라우드는 순열 불변성과 격자 구조가 없는 특징을 가진다. CD는 두 포인트 클라우드 셋에서 가장 가까운 점 간의 거리를 계산하여 유사도를 측정한다. HD는 두 포인트 클라우드 데이터 간 가장 먼 거리를 계산하여 집합의 밀집도를 측정하는 일반적인 지표다. 마지막으로, P2F는 GT 메쉬 데이터와 예측한 포인트 클라우드 간의 거리를 나타낸다. 세 지표 모두 낮은 값일수록 높은 성능을 나타낸다.

표 1은 기존 방법과 제안한 방법을 비교한 정량적 실험 결과를 보여준다. 제안한 방법은 CD와 HD 값에서 가장 우수한 성능을 보인다. 그러나 P2F에서는 낮은 성능을 보이는데, 이는 포인트 클라우드 보간 방법에서 샘플링된 저해상도 포인트 클라우드에 따라 성능이 크게 변동하는 문제 때문이다.

표 1. 정량적 실험 결과

Table 1. Quantitative results

Methods	CD	HD	P2F
PU-Net ^[8]	1.155	15.170	4.834
3PU ^[9]	0.935	13.327	3.551
PU-GCN ^[11]	0.585	7.577	2.499
PU-Transformer ^[13]	0.451	3.843	1.277
Grad-PU ^[21]	0.404	3.732	1.474
proposed method	0.392	3.521	1.293

V. 결론

포인트 클라우드를 활용한 3D 재현·분석은 AR, VR 그리고 자율주행과 같이 3차원 비전 콘텐츠에서 다양하게 사용된다. 따라서 3차원 실사 데이터 획득에 주로 사용되는 포인트 클라우드를 복원하여, 품질을 향상시키는 것은 중요하다. 본 논문에서는 거리함수 기반의 포인트 클라우드 개선 기법에 대해 제안했다. 제안하는 방법은 포인트 클라우드를 보간하여 3차원 공간 상의 밀집도를 늘리고, 딥러닝 네트워크를 통해 정답 표면과 보간한 포인트 클라우드 간 거리를 예측했다. 딥러닝 네트워크는 포인트 클라우드 데이터 구조에 적합한 트랜스포머와 kNN 방법을 적용하였다. 이를 통해 제안된 모델은 kNN을 사용하여 포인트 클라우드의 지역 정보를, 트랜스포머를 사용하여 전역 정보를

예측할 수 있다. 제안하는 모델을 통하여 정성적 및 정량적 실험결과에서 기존 포인트 클라우드 개선 방법보다 다소 우수한 성능을 달성하였다.

참고 문헌 (References)

- [1] Y. Hwang, J. Kim and K. Kim, "2D Interpolation of 3D Points using Video-based Point Cloud Compression," *Journal of Broadcast Engineering*, Vol.26, No.6, pp.692-703, November 2021. doi: <https://doi.org/10.5909/JBE.2021.26.6.692>.
- [2] Use the LiDAR Point Cloud Plugin to import LiDAR point clouds into Unreal Engine., <https://docs.unrealengine.com/5.0/ko/lidar-point-cloud-plugin-quick-start-guide-in-unreal-engine/> (accessed Jan. 14, 2024).
- [3] J. Kim, M. Lee and K. Kim, "Advanced Point Cloud Density Evaluation Method using Surface Estimation and Outlier Removal," *Journal of Broadcast Engineering*, Vol.26, No.6, pp.771-788. November 2023. doi: <https://doi.org/10.5909/JBE.2023.28.6.771>
- [4] M. Alexa, J. Behr, D. Cohen-Or, S. Fleishman, D. Levin and C.T. Silva, "Computing and rendering point set surfaces," *IEEE Transactions on visualization and computer graphics*, Vol.9, No.1, pp.3-15, February 2003.
- [5] Y. Lipman, D. Cohen-Or, D. Levin and H. Tal-Ezer, "Parameterization-free projection for geometry reconstruction," *ACM Transactions on Graphics(TOG)*, Vol.26, No.3, pp.22 - es, July 2007. doi: <https://doi.org/10.1145/1276377.1276405>
- [6] C. R. Qi, H. Su, K. Mo and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, pp.652-660, 2017. doi: <https://doi.org/10.48550/arXiv.1612.00593>
- [7] 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*, Long Beach, CA, pp.5105-5114, December 2017. doi: <https://doi.org/10.48550/arXiv.1706.02413>
- [8] L. Yu, X. Li, C. Fu, D. Cohen-Or and P. Heng, "Pu-net: Point cloud upsampling network," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, Salt Lake City, UT, pp.2790-2799, June 2018. doi: <https://doi.org/10.48550/arXiv.1801.06761>
- [9] W. Yifan, S. Wu, H. Huang, D. Cohen-Or and O. Sorkine-Hornung, "Patch-based progressive 3d point set upsampling," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, pp.5958-5967, June 2019. doi: <https://doi.org/10.48550/arXiv.1811.11286>
- [10] R. Li, X. Li, C. Fu, D. Cohen-Or and P. Heng, "Pu-gan: a point cloud upsampling adversarial network," In *Proceedings of the IEEE/CVF international conference on computer vision*, Seoul, South Korea, pp. 7203-7212. October 2019.

- doi: <https://doi.org/10.48550/arXiv.1907.10844>
- [11] G. Qian, A. Abualshour, G. Li and A. Thabet, "Pu-gcn: Point cloud upsampling using graph convolutional networks," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, pp.11683-11692. June 2021.
doi: <https://doi.org/10.48550/arXiv.1912.03264>
- [12] R. Li, X. Li, P. Heng and C. Fu, "Point cloud upsampling via disentangled refinement," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Virtual, pp.344-353. June 2021.
doi: <https://doi.org/10.48550/arXiv.2106.04779>
- [13] S. Qiu, S. Anwar and N. Barnes, "Pu-transformer: Point cloud upsampling transformer," In Proceedings of the Asian Conference on Computer Vision, Macau SAR, China, pp.2475-2493. December 2022.
doi: <https://doi.org/10.48550/arXiv.2111.12242>
- [14] J. Park, P. Florence, J. Straub, R. Newcombe and S. Lovegrove, "Deepsdf: Learning continuous signed distance functions for shape representation," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Long Beach, CA, pp.165-174. June 2019.
doi: <https://doi.org/10.48550/arXiv.1901.05103>
- [15] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin and A. Geiger, "Occupancy networks: Learning 3d reconstruction in function space," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Long Beach, CA, pp.4460-4470. June 2019.
doi: <https://doi.org/10.48550/arXiv.1812.03828>
- [16] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit and Neil Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 7th International Conference on Learning Representations, Virtual, May 2021.
doi: <https://doi.org/10.48550/arXiv.2010.11929>
- [17] H. Zhao, L. Jiang, J. Jia, P. Torr and V. Koltun, "Point transformer," In Proceedings of the IEEE/CVF international conference on computer vision, Montreal, QC, Canada, pp.16259-16268. October 2021.
doi: [10.1109/ICCV48922.2021.01595](https://doi.org/10.1109/ICCV48922.2021.01595)
- [18] X. Wu, Y. Lao, L. Jiang, X. Liu and H. Zhao, "Point transformer v2: Grouped vector attention and partition-based pooling," Advances in Neural Information Processing Systems, November 2022, pp.33330-33342.
doi: <https://doi.org/10.48550/arXiv.2210.05666>
- [19] Z. Wang, S. Zhou, J. Park, D. Paschalidou, S. You, G. Wetzstein, L. Guibas and A. Kadambi, "ALTO: Alternating Latent Topologies for Implicit 3D Reconstruction," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, Louisiana, pp.259-270, June 2022.
doi: <https://doi.org/10.48550/arXiv.2212.04096>
- [20] C. Koneputugodage, Y. Ben-Shabat and S. Gould, "Octree Guided Unoriented Surface Reconstruction," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver Canada, pp. 16717-16726. June 2023.
doi: [10.1109/CVPR52729.2023.01604](https://doi.org/10.1109/CVPR52729.2023.01604)
- [21] Y. He, D. Tang, Y. Zhang, X. Xue and Y. Fu, "Grad-PU: Arbitrary-Scale Point Cloud Upsampling via Gradient Descent with Learned Distance Functions," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver Canada, pp.5354-5363. June 2023.
doi: <https://doi.org/10.48550/arXiv.2304.11846>

저 자 소 개



김도훈

- 2022년 2월 : 호서대학교 컴퓨터공학과
- 2024년 2월 : 중앙대학교 첨단영상대학원 공학석사
- 주관심분야 : Point Cloud, 3D vision, Image Processing, Artificial Intelligence.



김철현

- 2008년 8월 : 중앙대학교 첨단영상대학원 공학석사
- 2011년 11월 : 중앙대학교 첨단영상대학원 공학박사
- 2013년 3월 ~ 현재 : 나사렛대학교 방송시각영상디자인학부 교수
- ORCID : <https://orcid.org/0000-0001-9421-8622>
- 주관심분야 : 3D Content, VR content, Cinema Technology, Entertainment Technology