

3차원 삼각형 메쉬를 정확하고 효율적으로 학습하기 위한 CNN 아키텍처

나홍은^o, 김종현^{*}

^o강남대학교 소프트웨어응용학부,

^{*}강남대학교 소프트웨어응용학부

e-mail: jonghyunkim@kangnam.ac.kr

CNN Architecture for Accurately and Efficiently Learning a 3D Triangular Mesh

Hong Eun Na^o, Jong-Hyun Kim^{*}

^oSchool of Software Application, Kangnam University,

^{*}School of Software Application, Kangnam University

● 요약 ●

본 논문에서는 삼각형 구조로 구성된 3차원 메쉬(Mesh)에서 합성곱 신경망(Convolution Neural Network, CNN)을 응용하여 정확도가 높은 새로운 학습 표현 기법을 제시한다. 우리는 메쉬를 구성하고 있는 폴리곤의 edge와 face의 로컬 특징을 기반으로 학습을 진행한다. 일반적으로 딥러닝은 인공지능망을 수 많은 계층 형태로 연결한 기법을 말하며, 주요 처리 대상은 1, 2차원 데이터 형태인 오디오 파일과 이미지였다. 인공지능에 대한 연구가 지속되면서 3차원 딥러닝이 도입되었지만, 기존의 학습과는 달리 3차원 딥러닝은 데이터의 확보가 쉽지 않다. 혼합현실과 메타버스 시장의 확대에 의해 3차원 모델링 시장이 증가하고, 기술의 발전으로 데이터를 획득할 수 있는 방법이 생겼지만, 3차원 데이터를 직접적으로 학습에 이용하는 방식으로 적용하는 것은 쉽지 않다. 그렇게 때문에 본 논문에서는 산업 현장에서 이용되는 데이터인 메쉬 구조를 폴리곤의 최소 단위인 삼각형 형태로 구성하여 학습 데이터를 구성해 기존의 방법보다 정확도가 높은 학습 기법을 제안한다.

키워드: 딥러닝(Deep learning), 합성곱 신경망(Convolutional neural network), 폴리곤 메쉬(Polygon mesh), 삼각형 메쉬(Triangular mesh), 3차원 모델링(3D modeling)

I. Introduction

최근에 다양한 응용 분야에서 전통적인 컴퓨터 비전 기반의 접근 방법(Hand-crafted feature)보다 월등한 성능을 보여주는 딥러닝은 기존의 1차원 데이터인 오디오 파일이나[1], 2차원 영상에서 3차원으로 영역을 확장하고 있다[2, 3]. 코로나로 인해 이르게 비대면 시대를 맞이한 지금 가상세계에서 사람들과 만날 수 있는 메타버스에 대한 관심이 증가했다. 시장 조사에 따르면 단순히 사람들과 만나는 것을 목적으로 하는 것뿐만이 아닌 소셜미디어, 광고, 게임 등에 활용되고, VR, AR, XR에 대한 관심이 늘어나 점차 시장 규모가 증가할 것으로 예측했다[4]. (Fig. 1 참조)

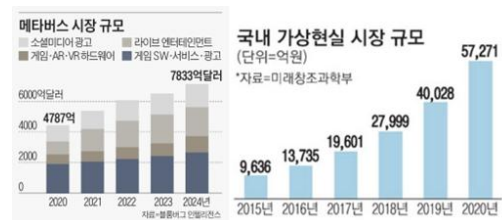


Fig. 1. 메타버스 & 가상현실 시장 규모

혼합현실과 메타버스 시장의 확장으로 3차원 모델링을 활용하는 분야가 점차 늘어나면서 3D 매핑 및 모델링 시장 규모 또한 성장할 것으로 예측된다. 이처럼 최근 3D 모델링의 활용 분야가 게임, 애니메이션, 영상 분야, 건축, 실내 디자인, 메타버스 등으로 늘어났다[5,

6]. 이로 인해 실제 객체를 스캐닝 하여 포인트 클라우드(Point cloud) 데이터를 수집하는 기술이 대중화되기 시작했다. 하지만 물체마다 추출되는 데이터 크기가 다르고, 직접 제작된 모델도 같은 물체를 보고 만들어도 제작자에 따라 데이터 크기가 다를 수 있어 여전히 학습을 위한 3D 데이터 확보는 어려운 난제로 남아있다.

3D 데이터는 표현 방법이 다양하고, 각 방법마다 여러 파일 형식을 가지고 있어 이에 따라 학습 사용할 수 있는 데이터가 더 줄어들게 되고, 데이터 표현 방법에 따라 1) 2D 네트워크 구조를 사용할 수 있는 방법이 있고, 2) 표현 방법에 적합한 네트워크 구조를 설계하거나, 3) 3D 데이터를 기존의 적합한 형태로 가공해야 하는 방법이 있다. 그에 따라 다양한 네트워크 구조들이 제시되었으며, 본 논문에서는 2, 3번째 방법을 응용하여 비정규형 3차원 메쉬를 직접적으로 학습 및 테스트할 수 있는 CNN 기반의 딥러닝 네트워크 구조를 제안한다. 또한 학습에 사용할 데이터는 산업 현장에서 가장 많이 사용하는 3D 데이터 구조인 메쉬 구조에 안정적인 포맷인 OBJ 파일로 학습을 진행한다.

II. Related Works

본 논문에서 다루고자 하는 일반적인 3차원 데이터 표현 방식은 총 6가지이다. 3D 데이터에 대한 다시점 영상인 다중 뷰 이미지(Multi-view images), 복셀(Voxel: 부피를 가지는 픽셀 정보)로 정의되는 볼륨 데이터(Volume data), 메쉬의 그 아류인 그래프(Graph), 다차원 데이터에서 실질적으로 의미를 가지는 특징을 모아둔 매니폴드(Manifolds), 3D 공간 위치 및 색상 정보를 포함하는 포인트 정보인 포인트 클라우드, 3D 정점(Vertex), 다각형 면(삼각 혹은 사각형의 면) 등으로 구성된 메쉬이다.

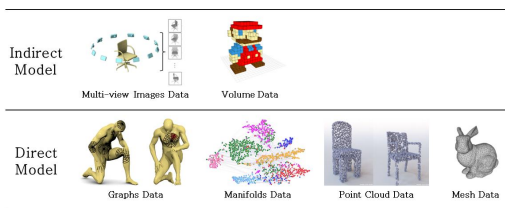


Fig. 2. 3D data representation in network model.

앞에서 언급한 3D 데이터 표현 방식을 2D 데이터 학습 구조로 사용할 수 있는지 없는지로 나누면 Fig. 2와 같이 나눌 수 있다. Indirect Model은 2D 데이터 학습 구조를 사용할 수 있는 데이터로 MVCNN[7]과 Shapnets[8]가 있다. MVCNN같은 경우에는 이미지 CNN을 그대로 사용하였고, Shapnets는 3D CNN을 사용했다. 다만 이 두 방법은 많은 메모리를 낭비하거나 CNN 계산을 중복해서 사용하기 때문에 계산 시간이 데이터의 양에 비해 오래 걸린다. 포인트 클라우드 데이터 형식을 사용한 네트워크 구조로는 PointNet[9]과 PointCNN[10]이 있다. 포인트 클라우드는 위상 정보와 표면 구조를 포함하지 않기 때문에 비슷한 구조를 가진 다른 데이터로 인식될 가능성이 있다. 메쉬를 사용한 네트워크 구조는 페이스(Face) 정보를 바탕으로 네트워크를 구성한 MeshNe[11]과 에지(Edge)를 바탕으

로 네트워크를 구성한 MeshCNN[12]이 있다. 본 논문에서는 정확도가 높은 MeshCNN을 바탕으로 네트워크 구조를 구성한다.

III. The Proposed Scheme

1. Mesh structure

네트워크 구조를 만들기에 앞서 학습에 사용할 데이터 구조를 제대로 파악할 필요가 있다. 메쉬는 폴리곤이 모여서 만들어진 3차원 객체로 본 논문에서는 폴리곤의 최소 단위인 삼각형인 면(face)으로 이루어진 3D 데이터를 사용한다. 간단하게 정점(vertex), 선(edge), 면(face)로 표현할 수 있다. (Fig. 3, 4 참조).

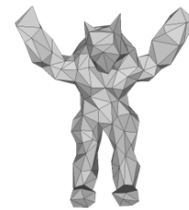


Fig. 3. Mesh Object

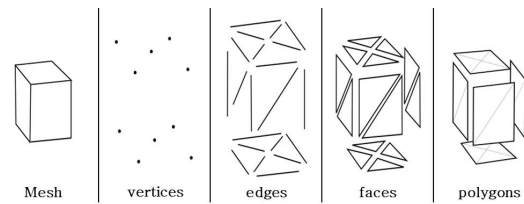


Fig. 4. Mesh structure

정점은 공간상에서 위치(position)을 나타내며, 2개의 정점을 모으면 선이 되고 3개의 정점을 모으면 면이 된다. 본 논문에서 사용하는 OBJ 파일은 Fig. 5와 같이 구성되어 있다. v는 정점으로 x, y, z 좌표 값으로 v 는 면으로 삼각형을 이루는 정점의 번호로 구성되어있다. (Fig. 5 참조)

v 0.023840 0.278918 -0.390713	f 103 81 142
v -0.277316 0.422782 -0.227440	f 145 81 204
...	...
v 0.668955 0.018051 0.466874	f 171 138 144
v 0.620319 -0.371994 0.251170	f 171 144 172
v 0.616258 -0.340800 0.303655	f 156 138 171
Vertices	Faces

Fig. 5. Vertex and face information

2. Feature extraction

메쉬는 3D 공간상에서 좌표(position), 회전(rotation), 크기(scale)의 변화에 따라 정점의 값이 달라진다. 다만 2d 데이터와는 달리 3d 데이터는 정점의 값이 달라지면 선과 면도 영향을 받아 길이와 넓이가 달라진다. 선을 기준으로 이러한 변화에 변하지 않는 특징을 찾아 feature 값으로 추출을 하였다.

- 선을 공유하는 두 면들의 각도
- 각 면이 공유하는 선의 맞은편에 있는 정점의 각도
- 각 면이 공유하는 선인 e 를 삼각형의 밑면이라 할 때, 각 면의 높이를 e 로 나눈 값

다만 이렇게 기하학적인 특징만 추출할 경우 Fig. 6.과 닮음이고 크기만 다른 선에서 같은 값이 추출 된다. 추출되는 feature 값에 차이를 주려면 길이나 면적에 대한 값이 들어가야 하는데 이럴 경우 메쉬의 변화에 따라 값이 달라지기 때문에 특징을 추출하기 전에 앞서 메쉬의 크기를 조절했다.

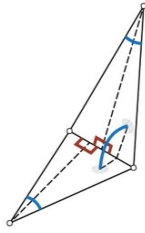


Fig. 6. Geometric features.

3. Mesh resizing

메쉬의 크기 조절은 처음에는 일정한 크기 이하이면 확대, 이상이면 축소를 하는 방향으로 진행했다. 다만 자세를 마음대로 바꿀 수 있는 3D 데이터 특징상 같은 물체여도 최대 길이가 달라 이 방법은 적합하지 않아 새로운 방법을 제안했다. (Fig. 7 참조)

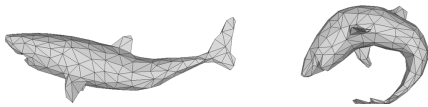


Fig. 7. different maximum length

사람이 자세가 바뀔 때 근육이 줄어들어 있는 곳이 있으면 늘어나는 곳이 있듯이 메쉬 또한 Object의 자세를 바꿀 때 면적이 줄어들어 있는 곳이 있으면 늘어나는 곳도 있다는 것에 착안하여 한 메쉬에서 나오는 면들의 넓이의 평균값을 1이 되도록 정점의 좌표를 수정하였다. 공식은 다음과 같다.

1. $mesh$ 의 중심좌표를 원점으로 이동
2. $v = va$
3. $\alpha = \sqrt{\frac{N}{\sum_{i=0}^{N-1} F_i}}$
(N : face의 개수, $F = face$ 의 넓이)

이에 맞추어 새로운 feature 값을 추가했다.

- 선을 공유하는 두 면의 넓이

4. Mesh convolution

mesh convolution은 이미지 convolution처럼 선 e 를 학습할 때 주변의 선들도 같이 학습한다. 같이 학습하는 선의 구성은 e 를 공유하는 두 면의 나머지 선으로 구성했다. (a, b, c, d) 와 (c, d, a, b) . 유효한

순서가 2가지이기 때문에 본 논문에서는 $(a+c, |a-c|, b+d, |b-d|)$ 로 값이 대칭될 수 있도록 구성하여 학습을 진행한다. (Fig. 8 참조).

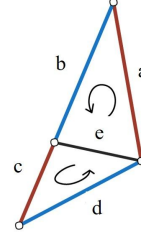


Fig. 8. Edge order in convolution

5. Mesh pooling

CNN에서 Pooling은 첫 번째로 pooling의 영역을 정의하는 것이고 다음으로 각 pooling의 특징을 병합하고, 마지막으로 병합된 특징에 대한 인접성을 재정의 하는 것이다. MeshCNN에서는 이와 유사한 메쉬 간략화(Mesh simplification)를 사용해 특징이 가장 작은 선을 선택해 제거하는 방식으로 진행했다.(Fig. 9 참조).

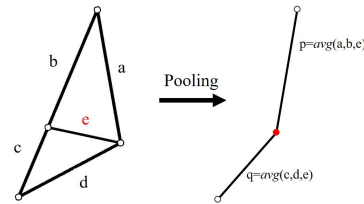


Fig. 9. Mesh pooling.

IV. Results

본 논문에서는 30개의 클래스에 각 20개의 데이터를 준비해 총 600개의 Training data set을 준비했고 이중 480개를 Training data, 120개를 Test data로 사용했다.

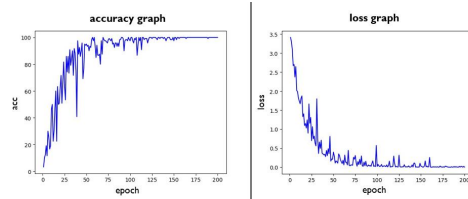


Fig. 10. Training accuracy & loss

학습 결과는 앞선 연구결과인 98.3%에서 약 0.9% 증가한 99.2%가 나왔다.

정확한 비교를 위해 학습에 사용된 모델과는 다른 모델을 테스트 한 결과 기존 기법보다 정확한 결과가 나왔다.(Table 1, 2. 참조)

Table 1. lamp and rabbit models.





	lamp	rabbit
(a) train data		
(b) new test data		

Table 2. Comparison results

	(b) lamp	(b) rabbit
MeshCNN[12]	snake	rabbit
Previous our method	laptop	two_balls
Our method	lamp	rabbit

V. Conclusions

본 논문에서는 현재 산업현장에서 많이 사용하는 메쉬 데이터를 사용한 3D 데이터 학습을 위한 네트워크를 구성했다. 기존의 MeshCNN는 오직 예지를 기준으로 위치와 방향, 크기 조절에 영향을 받지 않는 값으로 학습을 진행했기 때문에 데이터 모델과 전혀 연관이 없는 학습 결과가 나오기도 했다. 향후 메쉬의 방향성을 비등방성 (Anisotropic)하게 학습할 수 있도록 특징을 추가하고 분류할 수 있는 3d 모델을 늘리기 위해 클래스를 추가할 계획이다.

REFERENCES

[1] Pons, Jordi, Joan Serra, and Xavier Serra. "Training neural audio classifiers with few data." In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 16-20. IEEE, 2019.

[2] Li, Ruihui, Xianzhi Li, Ka-Hei Hui, and Chi-Wing Fu. "SP-GAN: Sphere-guided 3D shape generation and manipulation." ACM Transactions on Graphics (TOG) 40, no. 4 (2021): 1-12.

[3] Barsoum, Emad, John Kender, and Zicheng Liu. "Hp-gan: Probabilistic 3d human motion prediction via gan." In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 1418-1427. 2018.

[4] Chemerys, Hanna, Kateryna Osadcha, Viacheslav Osadchyi, and Vladyslav Kruhlyk. "Increase of the

Level of Graphic Competence Future Bachelor in Computer Sciences in the Process of Studying 3D Modeling." In ICTERI Workshops, pp. 17-28. 2019.

[5] Seok, Hyeonseon, and Seong do Yoon. "Companies Entering the Metabus Industry-Major Big Data Protection with Remote-based Hard Disk Memory Analysis Audit (AUDIT) System." Journal of the Chosun Natural Science 14, no. 4 (2021): 189-196.

[6] Zhou, S. Kevin, Hayit Greenspan, Christos Davatzikos, James S. Duncan, Bram Van Ginneken, Anant Madabhushi, Jerry L. Prince, Daniel Rueckert, and Ronald M. Summers. "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises." Proceedings of the IEEE 109, no. 5 (2021): 820-838.

[7] Yu, Qian, Chengzhan Yang, Honghui Fan, and Hui Wei. "Latent-MVCNN: 3D shape recognition using multiple views from pre-defined or random viewpoints." Neural Processing Letters 52, no. 1 (2020): 581-602.

[8] Wang, Rui, Xiaoqian Wang, and David I. Inouye. "Shapley explanation networks." arXiv preprint arXiv:2104.02297 (2021).

[9] Qi, Charles R., Hao Su, Kaichun Mo, and Leonidas 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, pp. 652-660. 2017.

[10] Li, Yangyan, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. "Pointnet: Convolution on x-transformed points." Advances in neural information processing systems 31 (2018).

[11] Feng, Yutong, Yifan Feng, Haoxuan You, Xibin Zhao, and Yue Gao. "Meshnet: Mesh neural network for 3d shape representation." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 01, pp. 8279-8286. 2019.

[12] Hanocka, Rana, Amir Hertz, Noa Fish, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. "Meshcnn: a network with an edge." ACM Transactions on Graphics (TOG) 38, no. 4 (2019): 1-12.