

2차원 영상으로부터 3차원 영상을 모델링하는 기술 동향

□ 조형래, 박구만 / 서울과학기술대학교

요약

2차원 영상을 3차원 모델 영상으로 변환하는 방식이 다양하게 발전해오고 있다. 딥러닝의 발전 중 특히 GAN의 다양한 연구는 2차원 영상의 생성뿐만 아니라 다양한 3차원 영상의 생성에도 진전을 보였다. 본 고에서는 2차원 영상을 3차원 영상으로 변환하는 연구의 필요성을 바탕으로 관련 연구의 내용과 동향을 분석하였다. 주요 내용으로는 딥러닝 기반의 3차원 객체인식, 2D로부터 3D 변환을 위한 신경망에 대한 연구, 생성적 기법을 적용한 연구, 3D 모델링 도구 등이 포함된다. 관련 연구의 전반적인 흐름을 고려했을 때 향후 3D 모델링의 정교한 표현력 향상, 고속의 고해상도 렌더링, 편리한 온라인 접근성 등을 예상하게 된다. 관련 산업 종사자들에게는 생성시간의 단축을 가져올 수 있고 일반인은 전문적인 3D 기술이 없어도 우수한 3D 모델을 생성하고 활용할 수 있을 것으로 기대한다.

1. 서론

3D 형태의 콘텐츠는 사람들에게 실제와 같은 사실감, 입체감과 몰입감을 더해주어 영화, 광고, 게임, 건축, 디자인, 증강현실, 로봇공학 등 여러 분야에서 다양하게 적용 및 활용되고 있다. 최근 딥러닝의 발달로 인해 영상 인식 및 생성 기술은 2D 영상을 3D로 변환하는 연구를 더욱 활발하게 만들었다. 하지만 일반 사용자가 3D 콘텐츠를 제작하는 데는 많은 어려움과 시간이 필요하고 3D 저작 소프트웨어의 사용 방법을 익혀서 작업하는 것도 쉬운 일은 아니다. 본문에서 본격적인 3차원 생성 기술을 소개하기에 앞서서 서론에서 2D 영상을 3D 영상으로 변환하는 도구를 소개하고자 한다.

※ 이 글은 2021년도 과학기술정보통신부의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2021-0-00751, 0.5mm급 이하 초정밀 가시-비가시 정보 표출을 위한 다차원 시각화 디지털 트윈 프레임워크 기술개발)

1. ConvertImage

ConvertImage는 무료 온라인 사진 편집기와 온라인 그래픽 파일 변환기로 웹사이트에서 쉽게 이용할 수 있다. 온라인으로 사진을 변환한 후, 소프트웨어 없이 파일 형식을 변경할 수 있는데 파일형식은 BMP, WBMP, DIB, CUR, GIF, JPG, JPEG, JPE, PCX RLE, PDF, PICT, PCT, PIC, PNG, PSB, PSD, TIF, XCF 등이 가능하다. 세부 사항이나 텍스트가 많은 영상에서는 완벽하게 작동하지 않는 것이 단점이다.



<그림 1> 2D 영상의 색상 차이를 이용해 입체 효과인 애너글리프 (Anaglyph 3D)로 바꿔주는 ConvertImage

2. 3D 음영(3D Shade)

<그림 2>에서 좌측의 모델링은 곡선표면 모델링 기법에 의해 정교한 표현을 하면서도 기존 폴리곤 모델링



<그림 2> 윈도우 기반의 플랫폼 3D 음영

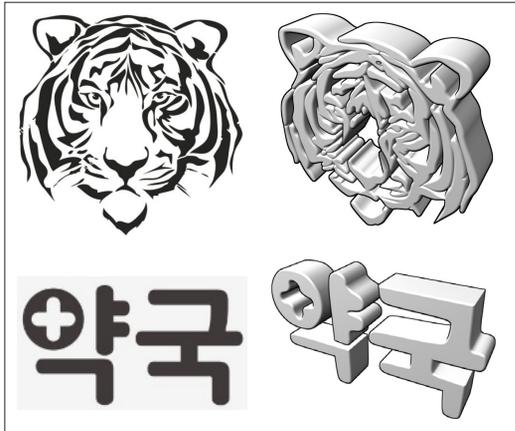
기술보다 더 쉽게 모양을 만들 수 있는 특징을 보여준다. 중앙의 애니메이션 출력은 다양한 관절(예: 뼈대) 및 스킨 설정을 결합하여 애니메이션을 만든 다음 MOV 또는 AVI와 같은 파일 형식으로 출력하는 것을 보여준다. 우측의 3D 프린팅 어시스턴트는 새로운 3D 인쇄 모델을 확인 및 수정하고 3D 인쇄(STL 및 OBJ 형식)를 준비하는 것을 보여준다. 3D 음영은 3D 모델링, 렌더링, 애니메이션, 3D 프린팅이 가능하고 베지어 곡선 모델링을 지원한다. 이는 어도비 일러스트와 같은 베지어 기반 벡터 편집 도구와 유사한 고유한 모델링 방법이며 일부 다른 3D 모델링 응용 프로그램에서는 사용할 수 없다. 3D 음영은 NURBS 기반 모델링도 지원한다. Shade 14는 어도비 형식(ai, psd, swf), 영화 형식(avi, mov, QuickTime VR 파노라마, QuickTime VR 큐빅, QuickTime VR 개체), 3D형식(obj, lwo, 3ds), 비디오 게임 디지털 형식(bvh, direct x, COLLADA, collada animation, Blue Mars, Second Life), 그림 형식(bmp, targa, tiff, png, epix, hdr, openEXR)을 포함한다.

3. 메이크스위트 로고(MakeSweet Logo)

메이크스위트 로고는 스텐실을 3D 로고로 변환할 수 있는 무료 온라인 변환기로 최상의 결과를 얻으려면 흑백만 사용해야 하며 사진이 jpeg 또는 png와 같이 일반적으로 사용되는 형식이어야 한다.

사진 크기를 1MB 미만, 그림의 크기를 1000x1000 미만으로 유지해야 한다. 상세한 부분까지 3D변환이 어려운 것과 흑백의 형태로만 되는 것이 단점이다. 오픈소스 2D 벡터그래픽 프로그램인 잉크스케이프(inkscape) 및 3D 프로그램인 블렌더(Blender)를 사용하여 3D 모양을 직접 만들 수 있다.

이외에도 2D 영상을 3D 영상으로 변환하는 도구에는 Clip Studio Paint, Image to Lithophane, ZW3D,



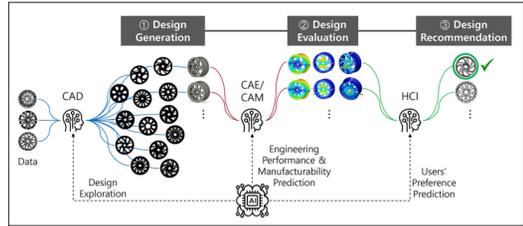
<그림 3> 좌측 상단: 스텐실, 좌측 하단: 로고 영상, 우측: 3D 변환 결과

Embossify, Reliefmod, Smoothie 3D, SculptGL, 3D Face, Reconstruction, Insight3d, 3D-Tool, Selva 3D, Vectary 등이 있다.

4. 제너레이티브 디자인

인공지능 기반의 제너레이티브 디자인은 인공지능에 의해 설계안의 생성, 평가, 추천에 이르는 전 과정을 수행하는데 핵심 인공지능 기술은 다음과 같다.

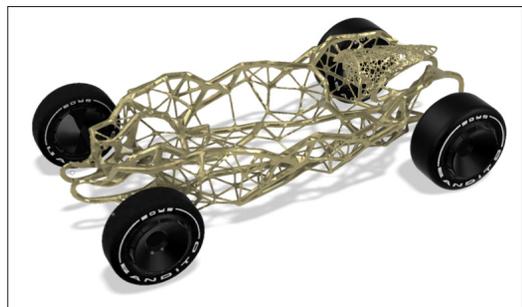
- 설계 생성 단계: 과거설계 데이터를 기반으로 심미적이면서 동시에 공학적으로 최적화된 설계안을 생성
- 설계 평가 단계: 생성된 설계안들의 독창성, 성능, 제조성 및 제조비용 등을 평가/예측/비교할 수 있는 인공지능 기술
- 설계 추천 단계: 평가된 설계안들에 대한 설계자 혹은 고객의 선호도를 예측하여 우수한 설계안 또는 그룹을 추천하고, 생산으로 자동 연결하는 인공지능 기술



<그림 4> 인공지능 기반 제너레이티브 디자인[1]

5. 오토데스크의 프로젝트 드림캐처 (Dreamcatcher)

드림캐처는 디자이너가 목표와 제약 조건을 통해 디자인 문제를 정의할 수 있도록 하는 제너레이티브 디자인 시스템이다. 설계자는 많은 대안적 접근 방식 간의 균형을 탐색하고 제조를 위한 설계 솔루션을 선택할 수 있다. 드림캐처 시스템을 통해 디자이너는 기능 요구 사항, 재료 유형, 제조 방법, 성능 기준 및 비용 제한을 비롯한 특정 디자인 목표를 입력할 수 있다. 설계 요구 사항이 로드되면 시스템은 철차적으로 합성된 설계 공간을 검색하여 설계 요구 사항을 충족하기 위해 생성된 수많은 설계를 평가한다.



<그림 5> 생성 알고리즘으로 설계된 자동차 프레임[2]

본 고는 서론에서 2D 영상을 3D 영상으로 변환하는 도구를 소개하고, 2장에서 3D 데이터셋의 종류, 3장에

서 딥러닝 기반 3D 객체 인식 연구, 4장에서 강화학습을 이용한 3D모델의 자동 생성, 5장에서 3D 변환을 위한 신경망들을 소개한다. 또한 6장에서 다양한 생성적 기법의 3D 모델링, 7장에서 3D 스타일 트랜스퍼를 소개하고, 최종적으로 8장에서 결론을 맺는다.

II. 3D 데이터셋

2D 영상을 자동으로 3D 콘텐츠로 변환하는 기술을 적용할 수 있다면 일반 사용자들도 쉽게 3D 콘텐츠를 제작할 수 있을 것이다. 2D 데이터는 사진기, 녹음기, 스캔 등 다양한 방법으로 확보할 수 있지만, 3D의 경우는 3D 스캐너로 획득하거나, 3D 모델링이나 데이터베이스를 통해서 확보할 수 있는데 3D 데이터셋의 예는 다음과 같다.

1. Matterport 3D

Matterport 3D가 제공하는 데이터셋은 90개 빌딩 194,400개 RGB+깊이 영상에서 10,800개 정렬된 3D 파노라마 뷰(픽셀 당 RGB+깊이)를 포함한다. 모든 장면은 Matterport Pro 3D 카메라로 취득되었다. 3D모델은 객체를 세분화하여 손으로 레이블링 하였다. Matterport 3D와 관련된 연구에서 딥러닝 기반 기능을 통한 개선된 형상 매칭, 2D 영상 표면 법선 벡터 추정, 복셀(voxel) 기반 모델의 형상 및 객체 식별과 같은 다양한 사용 사례를 보여준다. (<https://matterport.com/ko>)

2. PASCAL3D+

Pascal3D+ 멀티 뷰 데이터 세트는 야생의 영상, 즉

높은 가변성을 나타내는 객체 범주의 영상으로 구성되며, 제어되지 않은 설정, 복잡한 장면 및 다양한 포즈로 캡처된다. PASCAL3D+에는 PASCAL VOC 2012 데이터 집합에서 선택한 12개 범주의 강제 개체가 포함되어 있다. 이러한 개체에는 포즈 정보(각도, 고도 및 카메라까지의 거리)의 주석이 달린다. Pascal3D+는 또한 ImageNet 데이터 세트에서 이러한 12개 범주의 포즈 주석이 달린 영상을 추가한다. (<https://cvgl.stanford.edu/projects/pascal3d.html>)

3. ShapeNet

ShapeNet은 주석이 잘 붙여진 3D 모양의 대규모 데이터 셋으로 컴퓨터 그래픽, 컴퓨터 비전, 로봇 공학 및 기타 관련 분야의 연구에 활용된다. ShapeNet은 ShapeNetCore와 ShapeNetSem의 집합으로 구성되어 있는데 ShapeNetCore는 약 51,300개의 고유한 3D 모델과 함께 55개의 일반적인 개체 범주를 취급한다. 많이 사용되는 컴퓨터 비전 3D 벤치마크 데이터셋인 PASCAL 3D+의 12개 개체 범주는 모두 ShapeNetCore에서 다룬다.

ShapeNetSem은 270개 범주의 광범위한 집합에 걸쳐 12,000개 모델로 구성된 더 상세하게 주석이 달린 하위 집합이다. 수동으로 검증된 범주 레이블 및 일관된 정렬 외에도 이러한 모델에는 실제 치수, 범주 수준에서의 재료 구성 추정치, 총 부피 및 중량 추정치가 주석으로 표시된다. (<https://www.shapenet.org/>)

4. Human3.6M

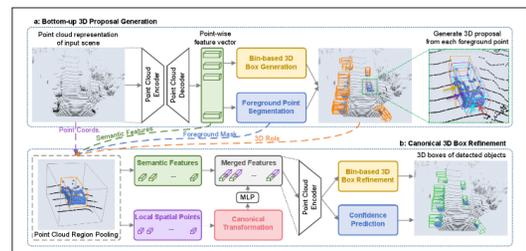
다양성과 규모로는 360만 개의 3D 인간 포즈 및 해당 영상, 전문 배우 11명(남자 6명, 여자 5명), 17가지 시나리오(토론, 흡연, 사진 찍기, 전화 통화 등)가 있고 정

확한 캡처 및 동기화를 통해 4개의 보정된 카메라의 고 해상도 50Hz 비디오, 고속모션 캡처 시스템의 정확한 3D 관절 위치 및 관절 각도, 각 구성에 대한 픽셀 수준의 24가지 신체 부위 레이블, 비행시간 범위 데이터, 배우의 3D 레이저 스캔, 정확한 배경 분리, 사람 경계 상자를 포함한다. (<http://vision.imar.ro/human3.6m/description.php>)

이외에도 가상 합성을 통해 생성된 5만 개 이상의 실내 장면 DB인 SceneNet(<https://robotvault.bitbucket.io/>), 인체 포함하는 다양한 객체를 CAD기반 3D 모델 제공하는 ModelNet(<https://modelnet.cs.princeton.edu/>), 가구 조립품의 371개 샘플과 실측 주석이 포함되어 있는 DB인 IKEA ASM dataset(<https://ikeaasm.github.io/>) 등이 있다.

워크에서는 지역공간을 학습하기 위해 각 제안요소의 합동 포인트를 표준 좌표로 변환하며, 박스 정교화와 예측을 위해 1단계에서 학습한 각 포인트의 전역 의미를 가지는 특성과 결합된다. 포인트 RCNN의 특징은 다음과 같다.

- 포인트 클라우드를 전경객체와 배경으로 분할하여 적은 수치만 고품질의 3D 제안요소를 생성한다.
- 강력한 bin(bin) 기반의 손실 값을 사용한다.



<그림 6> 포인트 RCNN 아키텍처[3]

III. 딥러닝 기반 3차원 객체 인식

딥러닝 기반 3차원 객체를 인식하는 방법은 다음과 같다.

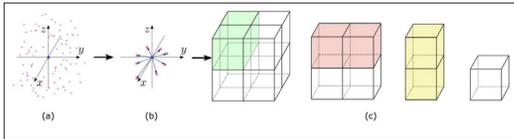
1. 포인트 RCNN(Point RCNN)

포인트 RCNN은 포인트 클라우드를 사용한 3D 물체감지 모델로 전체 프레임워크는 상향식 3D를 생성하기 위한 1단계와 제안요소를 정제하고 최종 검출 결과를 내는 2단계로 구성되어 있다. 선행 연구 방법들처럼 RGB 영상에서 제안요소를 생성하거나 포인트 클라우드를 조감도나 복셀에 투영하는 대신 1단계 서브 네트워크에서는 전체 장면의 포인트 클라우드를 객체와 배경으로 세분화한 뒤 상향식으로 포인트 클라우드에서 고품질 3D 제안요소를 생성한다. 2단계 서브 네트

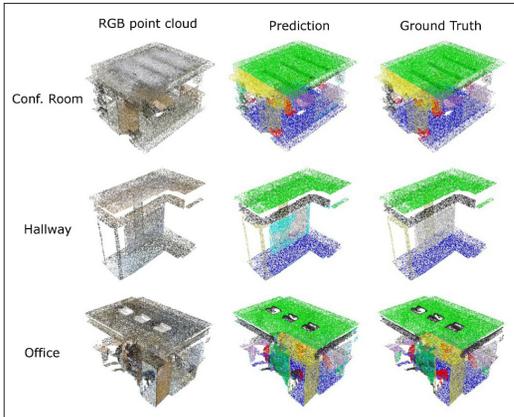
2. 포인트 SIFT(Point SIFT, 포인트 스케일 불변량 기능 변환)

포인트 SIFT는 3D 포인트 클라우드를 위한 시멘틱 분할 프레임워크이다. 이것은 8방향의 인접 포인트에서 특징을 추출하는 간단한 모델을 기반으로 한다. 최근 3차원 이해 연구에서는 포인트 클라우드에서 직접 특징을 추출하는데 더 많은 관심을 기울이고 있다. 따라서 점에서 모양 패턴 설명을 탐색하는 것이 필수다. 뛰어난 2D 모양 표현인 SIFT(Scale Invariant Feature Transform)는 영상의 지역특징을 감지하고 설명하기 위해 다양한 방향의 정보를 인코딩하고 모양의 크기에 적응하는 포인트 SIFT모듈을 설계한다. 특히, 방향 인코딩 단위는 8개의 중요한 방향을 설명하도록 설계되었

다. 따라서 여러 방향 인코딩 단위를 쌓아 다중 스케일 표현을 얻을 수 있다.



<그림 7> 방향 인코딩 포인트 컨볼루션 레이어의 세부 정보 그림. (a) 3D 공간의 포인트 클라우드. (b) 여덟 방향의 이웃. (c) 3단계 컨볼루션은 모든 기능을 결합

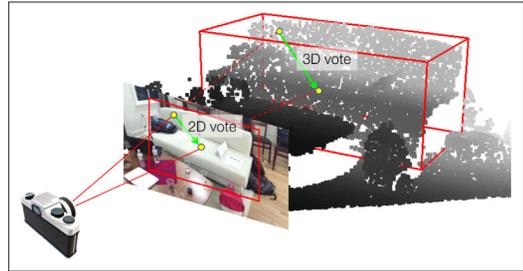


<그림 8> S3DIS 데이터 세트(3차원 포인트 클라우드 데이터에서 시멘틱 기반 객체 인식을 위한 딥러닝 학습 및 예측 모델)에 대한 시각화 결과[4]

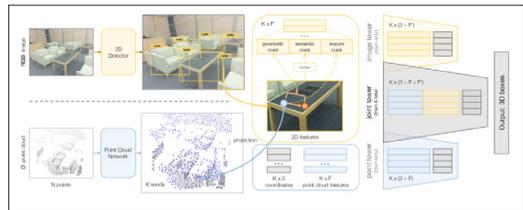
3. ImVoteNet(영상과 보트넷의 합성)

ImVoteNet은 ‘Image’와 ‘VoteNet’의 합성어로 RGB 영상을 참조하고 VoteNet은 3D 객체 감지 구조를 뜻한다. ImVoteNet은 포인트 클라우드 입력만으로 높은 성능을 보인다.

포인트 클라우드 데이터에는 한계가 있는데, 이들의 분포는 희박하고 색상 정보가 부족하며 종종 센서 잡음으로 인해 품질이 떨어질 수 있다. 반면 영상은 해상도가 높고 질감이 풍부하고 영상은 반사 표면으로 인해 종종 발생하는 활성 깊이 센서의 “사각형 영역”을 덮



<그림 9> 영상의 2D 투표와 포인트 클라우드의 3D 투표를 융합



<그림 10> 3D 물체 감지 파이프라인[5]

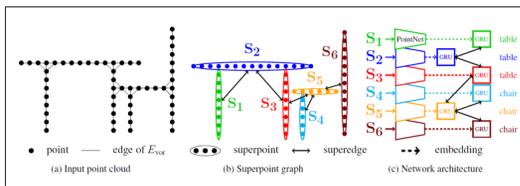
을 수 있다. 따라서 ImVoteNet은 2D 정보가 포인트 클라우드가 제공하는 3D 기하요소를 보완할 수 있는 장점을 가진다.

<그림 10>은 ImVoteNet 구조로서 왼쪽 상단은 Faster RCNN을 사용하여 2D 감지에서 영상 투표값을 추출하고, 왼쪽 하단은 PointNet++을 사용한 딥 허프 보팅(Deep Hough Voting)을 나타낸다. 오른쪽은 기능 융합 및 다중 타워 훈련을 나타낸다.

4. 슈퍼 포인트 그래프(Superpoint graph)[6]

기존에 진행된 많은 AI 기반의 포인트 클라우드 연구는 좋은 성능을 보여주었지만, 입력 용량이 큰 한계 때문에 적은 수의 점들로 구성된 포인트 클라우드에 대해서만 적용할 수 있었다. 신경망은 몇 백만 개 이상의 점들로 구성된 LiDAR 스캔을 직접 다루기 어려웠고, 축소 처리를 거쳐서 네트워크의 입력으로 이용했다. 하지

만 이는 포인트 클라우드의 장점인 물체에 대한 정교한 표현력을 떨어뜨릴 수밖에 없었다. 슈퍼포인트 그래프(superpoint graph)는 기존의 연구와는 달리 몇 백 만 개 단위의 점들로 구성된 포인트 클라우드를 대상으로 의미론적으로 세분화를 수행하였다. 유사한 구조의 점들을 슈퍼포인트라는 하나의 점으로 모아서 새로운 그래프를 구성했고 슈퍼포인트 그래프는 많은 점들로 구성되어 물체 간의 의미론적 관계에 대한 풍부한 정보를 담고 있기 때문에 의미론적 세분화 성능을 끌어올릴 수 있었다. 슈퍼포인트 그래프는 포인트 클라우드로부터 크게 세 단계를 거쳐서 생성된다. 우선, 전체 포인트 클라우드에서 기하학적으로 비슷한 구조를 가진 점들을 묶어서 작은 여러 개의 포인트 클라우드로 분리한다. 이후, 분리된 각각의 포인트 클라우드를 하나의 점으로 변환하는데, 이것을 슈퍼포인트라고 정의한다. 이때 분리된 포인트 클라우드 내에 속한 점들의 특징벡터를 통합해서 각 슈퍼포인트의 임베딩 벡터를 추출한다. 마지막으로 슈퍼포인트 간의 연결 관계를 파악해서 에지를 연결하고, 이를 그래프 컨볼루션 네트워크에 넣어서 세분화 작업을 수행한다.



<그림 11> 슈퍼그래프(SPG)

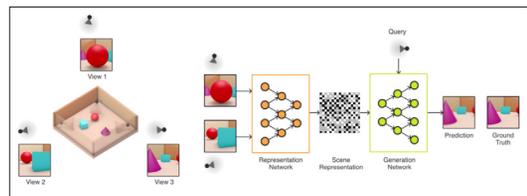
이외에 딥러닝 기반 3차원 객체 인식 방법으로 포인트 클라우드에서 실시간 3D 물체를 감지하는 네트워크 Complex-YOLO[7], 3D 포인트 클라우드 시맨틱 분할을 위한 심층 신경망을 구현하는 Semantic3dnet[8], 대규모 포인트 클라우드에 대한 포인트별 의미를 직접 추론하여 효율적이고 가벼운 신경 아키텍처인 RandLA-

Net[9], 동시에 두 가지 작업을 수행하는 다중 작업 포인트별 네트워크JSIS3D[10], 연속적인 정규화 흐름을 통한 3D 포인트 클라우드를 생성하는 PointFlow[11], 완전 컨볼루션 One-Stage Monocular 3D 물체 감지 프레임워크 FCOS3D[12] 등이 있다.

IV. 강화학습을 이용한 2D로부터 3D 모델 자동 생성

1. GQN(Generative Query Network)

2D 영상을 보고 3D 모델을 자동 생성할 수 있는 비전 알고리즘 GQN은 사람의 지시나 훈련 없이 여러 각도의 2D 영상을 보고 3D 모델을 렌더링 할 수 있다. 예를 들면 바닥에 빨간색 상자가 놓여있는 그림이 몇 장 제시되면 조명 등을 계산해 그림에 나와 있지 않은 빨간 상자 부분까지 렌더링해서 3D 모델을 만들어 낼 수 있다.



<그림 12> 생성 쿼리 네트워크

(출처 : torch-gqn, <https://github.com/iShohei220/torch-gqn>, (Accessed September 26, 2021))

현재 대다수 시각 인지 시스템은 사람이 제작한 방대한 데이터 세트를 이용해 훈련된다. 데이터셋의 각 장면들은 사람들이 작성한 주석이 달려있는데 이같은 데이터를 확보하려면 많은 비용과 시간이 들어간다. GQN은 주변에서 접할 수 있는 데이터들로도 주변 환경을 인식할 수 있게 하려고 사람의 뇌처럼 주변 환경과 물체

간 물리적 상호 작용을 인지하는 것이 목표다. 사람이 다리가 4개인 의자를 봤을 때 눈앞에는 다리가 3개만 보여도 눈에 보이지 않는 4번째 다리가 있음을 추론하거나 그림자 방향을 보고 광원의 위치를 예상할 수 있는 것처럼 GQN도 주변 환경을 인식할 수 있게 하는 것이다. GQN은 표현 네트워크(representation network)와 생성 네트워크(generation network) 두 부분으로 구성된다. 표현 네트워크는 입력 데이터를 기본 장면을 설명하는 수학적 표현, 즉 벡터로 변환한다. 생성 네트워크는 시각 시스템이 보지 못한 관점에서 장면을 예측해 영상화한다. 표현 네트워크는 생성 네트워크가 예측할 수 있는 관점의 장면을 알 수 없기 때문에 관찰한 장면의 실제 레이아웃을 가능한 한 정확하게 표현하는 방법을 찾게 된다. 객체 위치, 색상, 공간 레이아웃 등 가장 중요한 요소들을 캡처하는 것이다. 생성 네트워크는 장면 표현과 새로운 카메라 관점이 제공되면 원근법, 교합, 조명 등이 사전에 지정되지 않아도 선명한 영상을 생성한다.

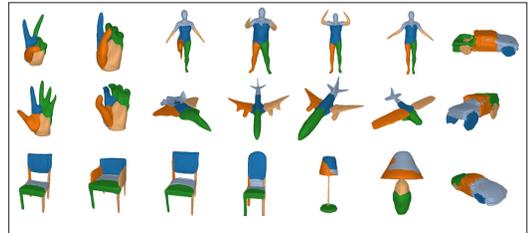
V. 2D로부터 3D 변환을 위한 신경망

2D 영상에서 3D 영상으로 변환하는 연구 중 신경망 방식을 적용한 예를 다음과 같이 소개한다.

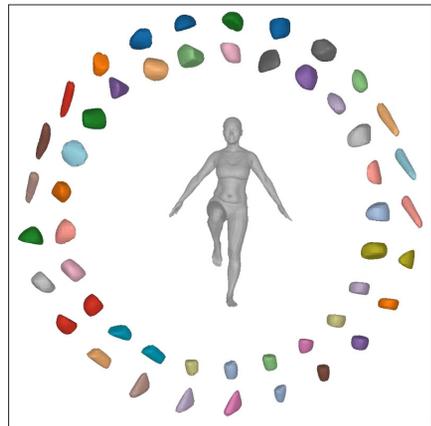
1. 신경망 부분

가역 신경망으로부터 표현적인 3D 모양 추상화 학습을 하는 것으로 Learning Expressive 3D Shape Abstractions with Invertible Neural Networks[13]이 있다. 이 연구에서는 암시적 표면 기능과 메쉬의 효율적인 계산을 위해 표현적인 흐름 기반 3D 표현을 사용하여 단순성과 의미를 결합하려고 시도한다. 표현력이

뛰어난 3D 표현을 통해 비지도 학습으로 3D 개체를 기하학적으로 정확하고 의미가 일관된 배열로 구문적 분석을 해낸다. 모델은 비지도 학습으로 5개의 원시특징(Primitive)만을 사용하여 3D 개체를 분석하는 방법을 배우게 된다. 그리고 ShapeNet 데이터셋, 4D 데이터 셋인 D-FAUST, 손 포즈 데이터셋인 FreiHAND에서 원시특징이 복잡한 기하학을 캡처함으로써, 적은 수의 원시특징만으로도 기하학적으로 정확하고 해석 가능한 3D 모델을 구성한다.

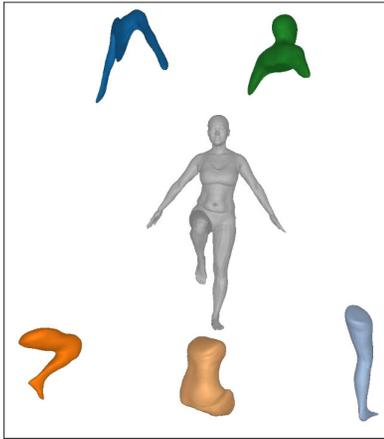


<그림 13> ShapeNet 데이터셋 객체



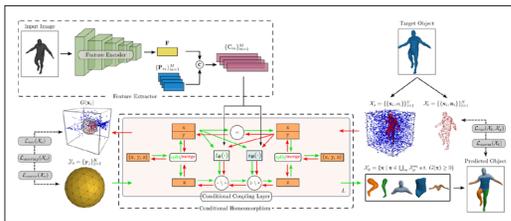
<그림 14> 기존의 원시 기반 방법(Primitive-based Methods)

<그림 14>에서처럼 기존의 방법은 복잡한 3D 모양을 부품으로 분해하기 위해 단순한 모양에 의존한다. 결과적으로 정확한 재구성을 추출하기 위해 많은 수의 원시 특징이 필요하다.

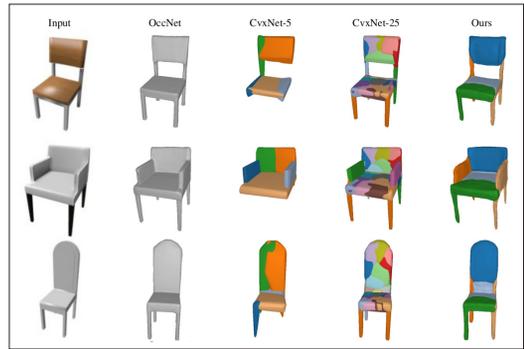


<그림 15> Neural Parts의 새로운 3D 프리미티브 표현

<그림16>에서 보는 바와 같이 신경망 파트는 곡면의 종수가 0인 구체로부터 기하학적으로 더 정확하고 의미있는 모양을 생성한다. 가역 신경 네트워크(INN, Invertible Neural Network)를 사용하면 예측된 모양의 암시적, 명시적 표현을 효율적으로 계산하고 예측된 부분에 다양한 제약 조건을 부과할 수 있고 기본 매개 변수(직육면체의 중심 및 크기)를 직접 예측한다. <그림 15>는 각 원시특징을 완전히 정의하기 위해 가역 신경 네트워크를 사용하고 이를 통해 복잡한 기하학을 캡처하는 원시특징을 가질 수 있으므로 간단한 볼록 모양 원시특징에 의존하는 기존 접근 방식(그림 14)에 비해 훨씬 적은 원시특징을 사용하여 기하학적으로 더 정확한 표현이 가능하게 된다.



<그림 16> 가역 신경 네트워크 INN(Invertible Neural Network)



<그림 17> 가역 신경망으로 표현적인 3D 모양 학습 결과

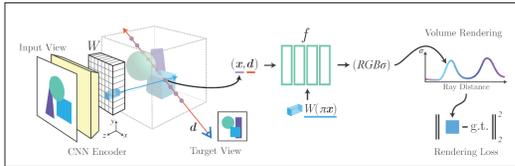
2. pixelNeRF

pixelNeRF[14]는 하나 또는 소수의 입력 영상을 조건으로 연속적인 신경 장면 표현을 예측하는 프레임워크이다. 픽셀 NeRF는 완전한 컨볼루션 방식으로 영상 입력에 대한 NeRF(NeRF의 궁극적인 목표는 뷰 합성으로 하나의 물체를 여러 각도에서 동시에 촬영한 영상 n장을 입력 데이터로 사용하고 그 영상들을 이용해서 영상에 주어지지 않은 새로운 각도에서의 동일 물체의 영상을 만들어내는 것)를 조절하는 아키텍처를 도입하여 독립적으로 모든 뷰에 대한 표현 최적화 작업 시 발생하는 많은 뷰에 대한 계산 시간을 단축하고자 한다. 이



<그림 18> 하나 또는 소수의 입력 영상으로부터 연속 신경 장면 표현을 예측하는 학습 프레임워크

를 통해 네트워크는 여러 장면에서 걸쳐 훈련되어 사진에 장면을 학습할 수 있으므로 최소 뷰 세트(최소한 하나)에서 피드포워드 방식으로 새로운 뷰 합성을 수행할 수 있게 된다.



<그림 19> pixelNeRF 아키텍처

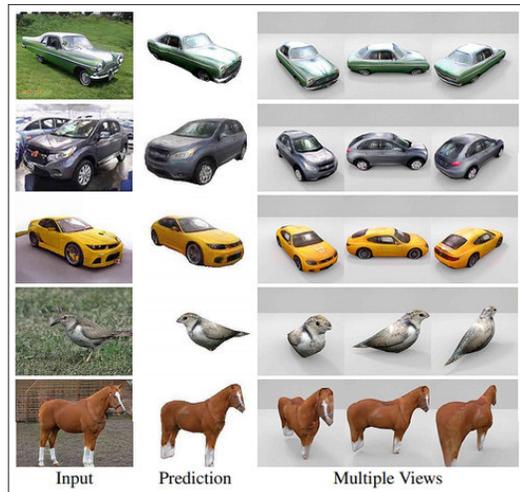
<그림 19>는 pixelNeRF 아키텍처로 뷰 방향이 d 인 대상 카메라 광선을 따라 쿼리 포인트 x 에 대해 a 해당 영상 특징을 투영 및 보간을 통해 특징 볼륨을 W 에서 추출한다. 공간 좌표와 함께 NeRF 네트워크 f 의 출력 RGB 및 농도 값은 볼륨 렌더링되어 대상과 비교된다. 좌표 x 와 d 는 입력 뷰의 카메라 좌표계에 있다.

이외에 2D로부터 3D로 변환하는 네트워크로는 메타 학습을 활용하여 적은 양의 다중 뷰와 일관된 라이트 필드 재구성을 하게 되고 시간과 메모리 복잡성이 크게 감소하여 실시간 렌더링이 가능한 라이트 필드 네트워크(Light Field Networks)[15]와 기하학적 정보와 모양을 모두 부호화하는 연속적인 3D 구조 인식 장면 표현인 SRN(Scene Representation Networks)[16] 장면 네트워크가 있으며, 실제 물체에서 사진을 찍고 이 사진을 3D 장면으로 변환하는 ShaRF[17]에 사용된 모양 네트워크(shape network)가 있다. 모양 네트워크를 통해 영상의 잠재 코드를 복셀로 구성된 3D 모양으로 매핑하여 영상 개체로부터 유용한 정보를 얻게 된다.

VI. 생성적 기법에 의한 2D로부터 3D 모델링

1. GANverse3D

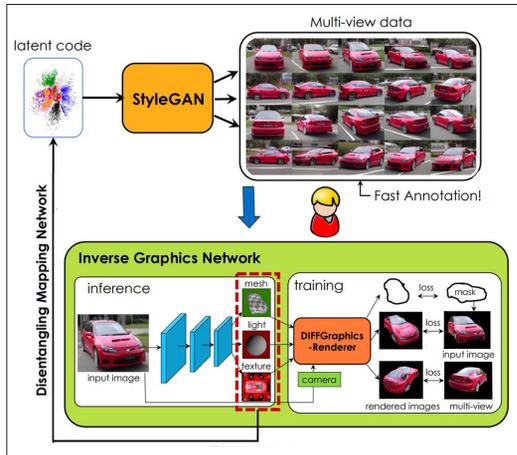
GANverse3D[18]는 2D 영상을 3D 모델로 변환해주고, 가상 환경에서 시각화 및 제어가 가능한 애플리케이션이다. 이는 특정 자동차 사진 1장만으로 사실적인 전 조등, 후미등, 점멸등까지 완벽한 3D 모델을 구축해 가상의 장면에서 주행하게 만들 수 있다. 동일한 객체를 서로 다른 각도에서 캡처하는 데이터셋은 현실적으로 드물다. 따라서 웹상에서 공개적으로 얻을 수 있는 자동차 영상과 같은 실세계 데이터에서 다시 점 영상을 얻기 위해 GAN 모델을 활용했다.



<그림 20> 왼쪽: GAN의 입력 영상과 다양한 시점에서의 렌더링, 오른쪽: 역 그래픽스 네트워크의 입력 영상과 다양한 시점에서의 렌더링

<그림 20>을 보면 입력 영상(첫 번째 열)이 주어졌을 때 3D 모양, 질감 및 동일한 시점(두 번째 열)이 렌더링한다. 나머지 3개의 다른 뷰에서도 렌더링 된다. 모델은 다양한 모양, 질감 및 관점, 단단한(관절 된) 물체, 즉 새

와 말에 대해서도 동일한 접근 방식을 보여준다.



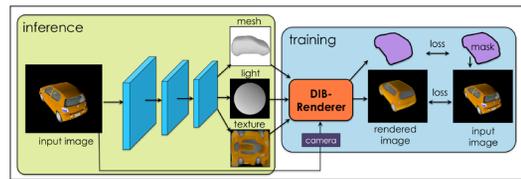
<그림 21> GANverse3D의 두 가지 렌더러(Renderers)

<그림 21>는 두 가지 렌더러(Renderers)를 보여주는데 GAN(작업에서는 StyleGAN)과 차별 가능한 그래픽 렌더러(DIB-R) 스타일을 사용한다. GAN은 합성 데이터 생성기로서, 이 데이터에 레이블을 붙이고, 데이터셋은 영상에서 3D 속성을 예측하는 역 그래픽 네트워크를 훈련시키는데 사용된다.

2. DIB-R

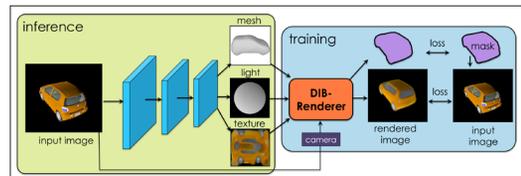
2D 사진은 3D 장면의 투영이다. 3D 장면은 카메라나 관점에서 본 3D 메쉬, 점점, 면, 텍스처 맵 및 광원의 모음이다. 3D 개체를 재구성하기 위한 대입 방식은 점점, 면, 광원 및 텍스처의 가능한 모든 조합을 계산하는 것이다. 이 조합은 2D로 투영될 때 입력으로 제공된 것과 동일한 2D 영상을 생성해야 한다. DIB-R[19]은 2D 사진 한 장으로 3D 모델을 만드는 것으로 신경망을 훈련시켜 완벽한 질감을 갖춘 3D 모델을 생성한다. 물론 이전에도 3D 모델을 자동 생성하는 유사 기술은 있었

다. 하지만 정확한 결과를 얻으려면 다양한 각도에서 촬영한 사진이 많이 필요하고 소프트웨어 영상의 특정 개체 크기와 모양을 이해시키려면 인간이 숫자를 입력할 필요가 있었다. DIB-R은 이런 문제를 고급 기술이 없는 사람이라도 간단하게 3D 모델링 할 수 있게 만들었다.



<그림 22> DIB-R 구조

<그림 22>에서 첫 번째 열은 실측 영상이고 두 번째와 세 번째 열은 DIB-R 모델의 예측, 네 번째와 다섯 번째 열은 SoftRas-Mesh의 결과, 마지막 두 열은 N3MR의 결과이다.

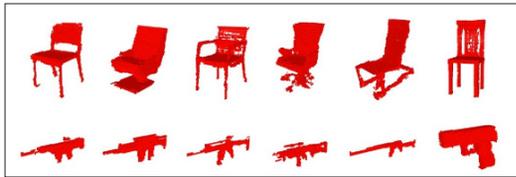


<그림 23> 결과: 단일 영상에서 3D 개체 예측, 기하학 및 색상

3. 3D-GAN(3D Generative Adversarial Network)

3D-GAN[20]은 생성적 적대 네트워크와 함께 체적형(3D) 컨볼루션 네트워크(Volumetric Convolutional Neural Network)를 활용한다. 3D 객체 생성은 영상 생성보다 더 어려운데 이것은 2D 영상의 공간에 비해 3D 모양의 공간을 모델링하는 것이 더 어렵기 때문이다.

3D-GAN은 체적형 컨볼루션 네트워크와 GAN을 활용하여 확률적 공간에서 3D 객체를 생성한다. 이 모델의 장점은 세 가지이다. 첫째, 생성기가 객체 구조를 암시적으로 캡처하고 고품질 3D 객체를 합성할 수 있다. 둘째, 생성기는 저차원 확률 공간에서 3D 개체의 공간으로 매핑을 설정하여 참조 영상 또는 CAD 모델 없이 개체를 샘플링하고 3D 개체를 다양하게 탐색할 수 있다. 셋째, 판별자는 비지도 학습된 강력한 3D 모양을 제공한다. <그림 24>는 고품질 3D 객체를 생성한 결과이다.

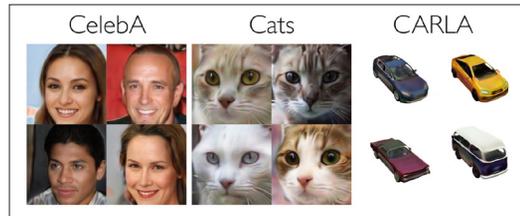


<그림 24> 3D-GAN으로 합성한 모양

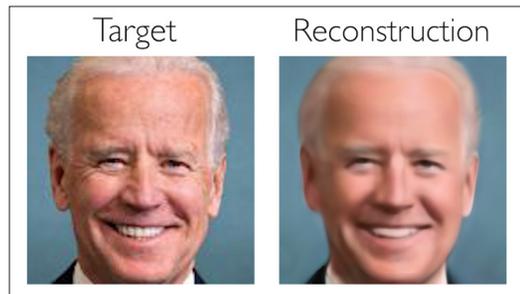
4. pi-GAN

생성적 시각 모델 및 신경 렌더링의 최근 발전을 통해 3D 인식 영상 합성은 빠른 발전을 했다. 그러나 기존 접근 방식은 두 가지 면에서 부족하다. 첫째, 기본 3D 표현이 부족하거나 뷰 불일치 렌더링으로 다중 뷰에서 일관성이 없는 영상을 합성할 수 있다. 둘째, 표현력이 충분하지 않은 표현 네트워크 아키텍처에 의존하는 경우가 많으며 결과적으로 영상 품질이 떨어진다. pi-GAN[21]은 고품질 3D 인식 영상 합성을 위해 π -GAN 또는 pi-GAN(Periodic Implicit Generative Adversarial Networks)이라는 새로운 생성 모델을 제안한다. π -GAN은 주기적인 활성화 기능과 체적 렌더링을 통한 신경 표현을 활용하여 장면을 미세한 세부 사항으로 일관된 3D로 표현하여 고품질의 다중 뷰 일관된 영상을 <그림 25>와 같이 생성한다. <그림 26>은 역 렌

더링 및 독창적인 합성을 통해 암시적 표현의 매개변수를 고정한 후, 렌더링될 때 대상 영상과 가장 잘 일치하는 광도 필드(radiance field)를 생성하는 컨디셔닝 매개변수에 대해 최적화한 실험 결과이다.



<그림 25> ρ -GAN은 128 x 128에서 CelebA, Cats 및 CARLA에서 3D 인식 이미지 합성에 대한 결과를 달성



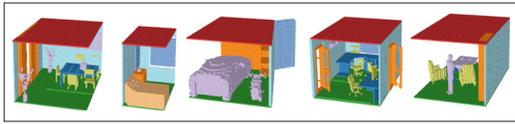
<그림 26> 역 렌더링 및 뷰 합성

5. 3D-FCR-alphaGAN

GAN 또는 VAE(Variational Auto-Encoder)를 활용한 생성 모델은 활발한 연구가 이루어지는 방법 중 하나이다. 이는 고품질 생성을 가능하게 할 뿐만 아니라 확률적 공간 및 다양체를 사용하여 감독 없이 표현 학습, 특징 추출 및 일부 인식 작업에 적용된다.

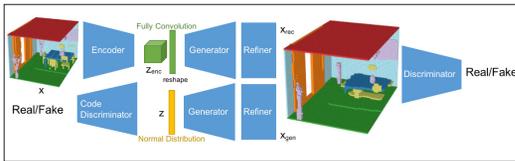
3D-FCR-alphaGAN[22]은 특히 다양한 3D 다중 객체를 합성하고 모양, 객체, 레이아웃을 포함한 객체를 인식할 수 있는데 3D 다중 객체 생성 모델은 AR/VR 및 그래픽 분야에서 매우 중요한 작업이다. 데이터셋은 12개의 클래스(빈, 천장, 바닥, 벽, 창, 의자, 침대, 소파,

테이블, TV, 가구, 오브제)가 있는 185,000개 이상의 장면으로 구성된다.



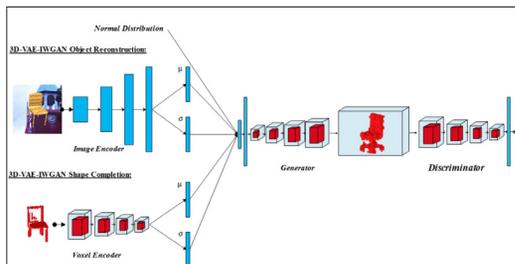
<그림 27> 수집한 데이터셋

네트워크 아키텍처는 자동 인코딩 생성적 적대 네트워크이다. 이것은 3DGAN, alphaGAN 및 SimGAN에서 영감을 받았다. 그리고 완전 컨볼루션 레이어와 다중 객체 분류는 생성 모델의 새로운 아키텍처로 생성적 적대 네트워크와 변형 자동 인코더를 결합했다.



<그림 28> 3D-FCR-alphaGAN 네트워크 아키텍처

이외 생성적 기법의 3D 모델링으로 3D-IWGAN[23], GIRAFFE[24] 등을 들 수 있는데 3D-IWGAN은 GAN을 적용한 3D 연구로 건물 형태의 합성을 대상으로 한다. 특히 NYC 시내의 700개 고층 건물 세트를 사용하여 3D GAN을 훈련한 다음 GAN을 사용하여 새로운 디자



<그림 29> 리포지토리에 사용된 3개의 GAN을 설명하는 다이어그램

인을 생성하고 특정 유형의 객체에 대한 3D 포인트 클라우드를 생성하도록 설계되었다. 그리고 3개의 GAN 프로젝트를 진행하였는데 3D 객체의 생성, 영상 입력에 따라 3D 객체를 생성, 깊이 스캔에서 3D 객체를 재구성하는 기능을 한다.

심층 생성 모델을 사용하면 고해상도에서 사실적인 영상 합성이 가능하다. 그러나 많은 응용 프로그램의 경우 이것은 쉽지 않은데, 이유는 콘텐츠 생성도 제어할 수 있어야 하기 때문이다. 이에 최근 몇 가지 연구 중에서 GIRAFFE는 합성 3D 장면 표현을 생성 모델에 통합하면 더 제어 가능한 영상 합성으로 이어지도록 하였다. 합성 특수 영상으로 추가 감독 없이 구조화되지 않은 영상 컬렉션에서 학습하면서 배경과 개별 객체의 모양 및 모양 내에서 하나 이상의 객체를 분리할 수 있다. 이 장면 표현을 신경 렌더링 파이프라인과 결합하면 빠르고 사실적인 영상 합성 모델이 생성된다.

<그림 30>은 2D 기반의 GAN과 비교한 것으로 (우)측에 있는 물체는 (좌)측에 있는 물체에 비해 일관된 움직임과 크기를 보여준다.



<그림 30> (좌) 2D 기반 GAN을 위한 단일 객체 번역, (우) GIRAFFE방법에 대한 단일 객체 번역의 비교 실험

VII. 3D 스타일트랜스퍼(3D style transfer)

3D 스타일트랜스퍼[25]는 3D 특징 임베딩을 스타일 전송에 의해 새로운 뷰 합성을 한다. 여기에는 DeepVoxels, SRN(Scene Representation Networks) 및 Neural Meshes 등이 있다. 이러한 신경망은 2D 영상 세트에서 학습한 객체의 새로운 3D 영상을 생성할 수 있다. 하지만 이러한 3D 구조에 2D 스타일을 전달하는 연구는 거의 수행되지 않았는데 3D 스타일트랜스퍼는 이러한 3D 모델에 2D 영상의 스타일을 전달한다.



<그림 31> 2D 영상의 스타일을 3D에 전달

3D 응용 프로그램에 대한 개발 시간을 단축하기 위해 2D 영상을 3D로 변환할 수 있는 가속화 도구로는 NVIDIA Kaolin 라이브러리[26]와 GAN기법으로 3D 게임 캐릭터를 실감나게 표현하는 키메라 페인터(Chimera Painter)[27]가 있다. NVIDIA Kaolin 라이브러리는 다양한 3D 표현 작업을 위한 PyTorch API를 제공하며 모듈식 미분 렌더링, 표현 간의 빠른 변환, 데이터 로딩, 3D 체크포인트 등과 같은 GPU 최적화 작업을

포함하고 메쉬, 포인트 클라우드 및 복셀 그리드에 대한 3D 체크포인트를 USD 형식으로 내보내 모델 교육을 대화형으로 시각화할 수 있다. 키메라 페인터는 몸 부위마다 다른 색과 선으로 몬스터를 그리면 현실감 있는 질감의 가상 몬스터를 생성할 수 있게 해주는 도구이다. 키메라 페인터의 장점은 게임 아티스트(사용자)를 위해 실제 붓 질감으로 작동하는 것이다.

VIII. 결론

2D를 3D로 변환하는 작업이 고가의 장비인 3D 스캐너나 3D 프로그램에 의존했지만, 인공지능의 발전으로 사용자가 온라인을 통해 손쉽게 2D 영상을 3D 영상으로 변환할 수 있게 되었다. 이러한 3D 딥러닝 연구를 가속화하기 위한 도구로 NVIDIA Kaolin 라이브러리, GAN기법으로 3D 게임 캐릭터를 실감나게 만들 수 있는 키메라 페인터를 소개하였다. 2D로부터 3D로 변환하는 방법으로서 강화학습을 이용한 3D 모델 자동 생성 기법, 라이트 필드 네트워크, SRN 장면 표현 네트워크와 ShaRF에서 사용된 모양 네트워크(shape network) 등이 소개되었다. 또한 GAN을 이용한 생성적 기법 중 GANverse3D, DIB-R, 3D-GAN, pi-GAN, 3D-FCR-alphaGAN, 3D-IWGAN 등이 소개되었다. 향후 2D로부터 3D 변환하는 연구의 방향을 예상해 보면, 현실과 같은 정교하고 높은 해상도를 보이는 3D 모델들이 개발될 것으로 기대되고 또한 적은 데이터 또는 하나의 영상에서 나머지 모습을 추론하여 3D를 생성하는 연구가 더욱 활발할 것으로 예상된다. 다양한 3D 가속화 도구의 발전을 통해 일반인도 쉽게 한 장의 2D영상으로부터 다양하고 높은 품질의 3D 모델을 생성할 수 있을 것으로 기대한다.

참고 문헌

- [1] 인공지능과 최적설계, <https://www.koreascience.or.kr/article/JAKO202017054987861.pdf>(Accessed September 26, 2021)
- [2] 딥 러닝 생성 설계의 외계인 스타일, <https://medium.com/intuitionmachine/the-alien-look-of-deep-learning-generative-design-5c5f871f7d10>(Accessed September 26, 2021)
- [3] Shaoshuai Shi Xiaogang Wang Hongsheng Li, PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud, <https://arxiv.org/pdf/1812.04244.pdf>, arXiv:1812.04244v2[cs.CV] 16, May, 2019
- [4] Mingyang Jiang, Yiran Wu, Tianqi Zhao, Zelin Zhao, PointSIFT: A SIFT-like Network Module for 3D Point Cloud Semantic Segmentation, <https://arxiv.org/pdf/1807.00652.pdf>, arXiv:1807.00652v2 [cs.CV] 24, Nov, 2018
- [5] 임보트넷(ImVoteNet), <https://medium.com/codex/imvotenet-paper-review-and-code-analysis-bf103117b32e>, (Accessed September 26, 2021)
- [6] Loic Landrieu, Martin Simonovsky, Universite Paris-Est, LASTIG MATIS IGN, ENSG, Universite Paris-Est, Ecole des Ponts ParisTech, Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs, <https://arxiv.org/pdf/1711.09869.pdf>, arXiv:1711.09869v2[cs.CV] 28, Mar, 2018
- [7] Martin Simon, Stefan Milz, Karl Amende, Horst-Michael Gross, Complex-YOLO: An Euler-Region-Proposal for Real-time 3D Object Detection on Point Clouds, <https://arxiv.org/pdf/1803.06199.pdf>, arXiv:1803.06199v2 [cs.CV] 24, Sep, 2018
- [8] Timo Hackel, Nikolay Savinov, Lubor Ladicky, Jan D. Wegner, Konrad Schindler, Marc Pollefeys, SEMANTIC3D.NET: A NEW LARGE-SCALE POINT CLOUD CLASSIFICATION BENCHMARK, <https://arxiv.org/pdf/1704.03847.pdf>, arXiv:1704.03847v1 [cs.CV] 12, Apr, 2017
- [9] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, Andrew Markham, University of Oxford, Sun Yat-sen University, National University of Defense Technology, RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds, https://openaccess.thecvf.com/content_CVPR_2020/papers/Hu_RandLA-Net_Efficient_Semantic_Segmentation_of_Large-Scale_Point_Clouds_CVPR_2020_paper.pdf, ICCV 2021: October 11th - 17th, Virtual
- [10] Quang-Hieu Pham, Duc Thanh Nguyen, Binh-Son Hua, Gemma Roig, Sai-Kit Yeung, Singapore University of Technology and Design, Deakin University, The University of Tokyo, Hong Kong University of Science and Technology, JSIS3D: Joint Semantic-Instance Segmentation of 3D Point Clouds with Multi-Task Pointwise Networks and Multi-Value Conditional Random Fields, <https://arxiv.org/pdf/1904.00699.pdf>, arXiv:1904.00699v2 [cs.CV] 5, Apr, 2019
- [11] Guandao Yang, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, Bharath Hariharan, Cornell University, Cornell Tech, NVIDIA, PointFlow: 3D Point Cloud Generation with Continuous Normalizing Flows, <https://arxiv.org/pdf/1906.12320.pdf>, arXiv:1906.12320v3 [cs.CV] 2, Sep, 2019
- [12] Tai Wang, Xinge Zhu, Jiangmiao Pang, Dahua Lin, FCOS3D: Fully Convolutional One-Stage Monocular 3D Object Detection, <https://arxiv.org/pdf/2104.10956v3.pdf>, arXiv:2104.10956v3 [cs.CV] 24, Sep, 2021
- [13] Despoina Paschalidou, Angelos Katharopoulos, Andreas Geiger, Sanja Fidler, Max Planck Institute for Intelligent Systems Tübingen, University of Tübingen, Idiap Research Institute, Switzerland, Ecole Polytechnique Fédérale de Lausanne (EPFL), Max Planck ETH Center for Learning Systems, NVIDIA, University of Toronto, Vector Institute, Neural Parts: Learning Expressive 3D Shape Abstractions with Invertible Neural Networks, <https://arxiv.org/pdf/2103.10429.pdf>, arXiv:2103.10429v1 [cs.CV] 18, Mar, 2021
- [14] Alex Yu Vickie Ye Matthew Tancik Angjoo Kanazawa, pixelNeRF: Neural Radiance Fields from One or Few Images, <https://arxiv.org/pdf/2012.02190.pdf>, arXiv:2012.02190v3 [cs.CV] 30, May, 2021
- [15] Vincent Sitzmann, Semon Rezhikov, William T. Freeman, Joshua B. Tenenbaum, Frédo Durand, MIT CSAIL, Columbia University, NSF IAFI, MIT BCS, NSF CBMM, Light Field Networks: Neural Scene Representations with Single-Evaluation Rendering, <https://arxiv.org/pdf/2106.02634.pdf>, arXiv:2106.02634v1 [cs.CV] 4, Jun, 2021
- [16] Vincent Sitzmann, Michael Zollhofer, Gordon Wetzstein, Scene Representation Networks: Continuous Neural Scene 3D-Structure-Aware Representations, <https://arxiv.org/pdf/1906.01618.pdf>, arXiv:1906.01618v2 [cs.CV] 28, Jan, 2020

- [17] Konstantinos Rematas, Ricardo Martin-Brualla, Vittorio Ferrari, ShaRF: Shape-conditioned Radiance Fields from a Single View, <https://arxiv.org/pdf/2102.08860.pdf>, arXiv:2102.08860v2 [cs.CV] 23, Jun, 2021
- [18] Yuxuan Zhang, Wenzheng Chen, Huan Ling, Jun Gao, Yinan Zhang, Antonio Torralba, Sanja Fidler, NVIDIA, University of Toronto, Vector Institute, University of Waterloo, IMAGE GANS MEET DIFFERENTIABLE RENDERING FOR INVERSE GRAPHICS AND INTERPRETABLE 3D NEURAL RENDERING, <https://arxiv.org/pdf/2010.09125.pdf>, arXiv:2010.09125v2 [cs.CV] 20, Apr, 2021
- [19] Wenzheng Chen, Jun Gao, Huan Ling, Edward J. Smith, Jaakko Lehtinen, Alec Jacobson, Sanja Fidler, NVIDIA, University of Toronto, Vector Institute, McGill University, Aalto University, Learning to Predict 3D Objects with an Interpolation-based Differentiable Renderer, https://nv-tlabs.github.io/DIB-R/files/diff_shader.pdf, arXiv:1908.01210v2 [cs.CV] 21, Nov, 2019
- [20] Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T. Freeman, Joshua B. Tenenbaum, Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling, http://3dgan.csail.mit.edu/papers/3dgan_nips.pdf, arXiv:1610.07584v2 [cs.CV] 4, Jan, 2017
- [21] Eric R. Chan, Marco Monteiro, Petr Kellnhofer, Jiajun Wu, Gordon Wetzstein, Stanford University, pi-GAN: Periodic Implicit Generative Adversarial Networks for 3D-Aware Image Synthesis, <https://arxiv.org/pdf/2012.00926.pdf>, arXiv:2012.00926v2 [cs.CV] 5, Apr, 2021
- [22] 3D-FCR-alphaGAN, <https://github.com/yunishi3/3D-FCR-alphaGAN> (Accessed September 26, 2021)
- [23] Edward J. Smith, David Meger, Improved Adversarial Systems for 3D Object Generation and Reconstruction, <https://arxiv.org/pdf/1707.09557.pdf>, arXiv:1707.09557v3 [cs.CV] 30, Oct, 2017
- [24] Michael Niemeyer, Andreas Geiger, Max Planck Institute for Intelligent Systems, Tubingen, University of Tubingen, GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields, <http://www.cvlibs.net/publications/Niemeyer2021CVPR.pdf>, CVPR 2021 (oral, best paper award)
- [25] Mattia Segu, Margarita Grinvald, Roland Siegwart, Federico Tombari, 3DSNet: Unsupervised Shape-to-Shape 3D Style Transfer, <https://arxiv.org/pdf/2011.13388.pdf>, arXiv:2011.13388v4 [cs.CV] 18, May, 2021
- [26] NVIDIA Kaolin, <https://developer.nvidia.com/nvidia-kaolin> (Accessed September 26, 2021)
- [27] GAN을 사용하여 환상적인 생물 만들기, <https://ai.googleblog.com/2020/11/using-gans-to-create-fantastical.html>, (Accessed September 26, 2021)

필자소개



조형래

- 2015년 : 세종대학교 회화과(서양화 전공) 학사
- 2021년 : 서울과학기술대학교 일반대학원 미디어IT공학과 석사
- 2021년 ~ 현재 : 서울과학기술대학교 일반대학원 미디어IT공학과 박사과정
- ORCID : <https://orcid.org/0000-0002-2777-2073>
- 주관심분야 : 데이터시각화, 인공지능을 활용한 비주얼라이제이션, GAN, 실감형콘텐츠



박구만

- 1984년 : 한국항공대학교 전자공학과 공학사
- 1986년 : 연세대학교 대학원 전자공학과 석사
- 1991년 : 연세대학교 대학원 전자공학과 박사
- 1991년 ~ 1996년 : 삼성전자 신호처리연구소 선임연구원
- 1999년 ~ 현재 : 서울과학기술대학교 전자IT미디어공학과 교수
- 2006년 ~ 2007년 : Georgia Institute of Technology, Dept. of ECE. Visiting Scholar
- 2016년 ~ 2017년 : 서울과학기술대학교 나노IT디자인융합대학원 원장
- ORCID : <https://orcid.org/0000-0002-7055-5568>
- 주관심분야 : 컴퓨터비전, 실감미디어