NeRF의 깊이 추정 시 뷰디렉션 설정의 중요성 연구

Exploring the Role of View Direction Settings in Depth Estimation with NeRF: A Comparative Study

요 약 이 논문에서는 Neural Radiance Fields (NeRF)를 사용하여 깊이 추정에서 뷰디렉션의 역할을 탐구한다. 균일한 데이터 세트와 특정 방향의 데이터 세트로 학습된 모델 간의 성능 차이를 비교하여 뷰디렉션의 중요성을 밝힌다. 실험 결과, 특정 방향의 데이터 세트로 학습된 모델이 특히 특정 시점에서의 희소한 관측이 있는 영역에서 우수한 성능을 보였다. 이러한 연구 결과는 NeRF를 활용한 깊이 추정에서 뷰디렉션의 중요성을 강조하며, 향후 더 효율적인 학습 방법과 데이터 세트 구성에 대한 연구를 제안한다.

1. 서 론

딥러닝을 이용한 3D 시각화 기술은 가상 현실(VR), 확장 현실(AR), 로봇학, 영화 및 게임 제작 등 다양한 분야에서 중요한 응용 가능성을 보여주고 있다. 특히 최근에는 Neural Radiance Fields (NeRF)와 같은 모델들이 기하학적 3D 장면을 학습하고 렌더링하는 데에 있어서 많은 관심을 받고 있다. 그러나 NeRF에 사용되는 실제 세계의 데이터 세트는 종종 불균형하고 편향된 특성을 가지고 있다. 특히 특정 방향에서만 풍부한 데이터를 가지고 있거나, 특정 뷰포인트에서만 고해상도 데이터를 가지고 있는 경우가 많다.

이러한 불균형은 특히 데이터 세트의 뷰포인트와 관련하여 주목할만한 문제를 야기한다. NeRF 모델의 성능은 훈련 데이터의 분포에 따라 크게 변동될 수 있다. 특정 시점의 데이터 세트에 대해 학습한 모델과 비교적 균일하게 분산된 데이터 세트에 대해 학습한 모델을 비교할 때 성능의 차이가 발생한다.

이 문제의 근원은 NeRF 모델의 근본적인 기능에 있다. 이러한 모델은 훈련 데이터에 캡처된 여러 뷰포인트의 정보를 합성하여 장면의 기하학적 구조와 모양을 재구성한다. 그러나 데이터 세트에서 뷰포인트 범위의 다양성이 부족한 경우 NeRF 모델은 보이지 않는 부분을 렌더링하는 데 어려움을 겪어 왜곡이 생기거나 세부 정보가 누락될 수 있다.

이러한 문제에 비추어, 이 연구의 주요 목표는 NeRF 모델 훈련에서 뷰디렉션의 중요성을 설명하는 것이다. 데이터 세트 편향이 성능에 미치는 영향을 조사함으로써 우리는 이러한 문제를 완화하고 광범위한 관점에서 NeRF 모델의 견고성과 일반화 성능을 향상시키기 위한 잠재적인 전략을 밝히는 것을 목표로 한다.

2. 관련 연구

NeRF[1]는 기하학적 3D 장면을 처리하기 위한 딥러닝 기반의 새로운 접근 방식이다. 이 모델은 장면의 각 점에 대한 복잡한 광학 특성을 학습하여 렌더링한다.

Mip-NeRF[2]는 NeRF의 성능을 향상시키기 위해 개발된 기술이다. NeRF가 고해상도 이미지를 생성하는데 많은 계산 비용이 소요된다는 단점을 보완하기 위해, Mip-NeRF는 다양한 해상도의 이미지 피라미드를 활용한다. 즉, NeRF는 고해상도 이미지를 생성하는 대신, 더 낮은 해상도의 이미지로부터 시작하여 이미지 피라미드를 구성하고, 각 수준에서 NeRF 모델을 학습하여 성능을 향상시킨다. 이렇게 함으로써 더 빠르고 효율적으로 고해상도 이미지를 생성할 수 있게 된다.

따라서 Mip-NeRF는 NeRF와 비교하여 성능 면에서 우수하고 동시에 계산 비용을 줄일 수 있기에, 본 논문에서는 Mip-NeRF를 활용하고자 한다.

3. 본 론

NeRF 모델링 영역에서 훈련 데이터의 다양성은 모델의 성능에 중요한 영향을 미친다. 그러나 실제 데이터 세트는 특정 방향에서만 풍부한 데이터를 가지고 있어 데이터 불균형 문제가 있다. 이 문제를 해결하려면 데이터 세트 분류 및 관점 선택에 대한 전략적 접근 방식이 필요하다.

본 연구에서는 Mip-NeRF 모델 훈련을 위해 Synthetic NeRF 데이터 세트를 사용하였고, 그 중 엣지가 많이 포함되어 있는 Lego 클래스를 모델 훈련에 사용한다.

이때 앞, 뒤, 오른쪽, 왼쪽, 위쪽(front, back, right, left, top)의 5가지 개별 시점을 기반으로 Lego 데이터 세트를 체계적으로 분류하여 관점 편향 문제를 해결한다. 이 분류 체계를 통해 NeRF 모델 성능에 대한 각 관점의 영향을 분리하고 분석할 수 있다.

이 분류의 근거는 특정 시점이 다른 시점보다 장면 재구성에 더 중요한 정보를 포함할 수 있다는 관찰에 있다. 데이터 세트를 별개의 관점으로 나누어서 중요한 패턴을 식별하고 어떤 관점이 모델의 학습 프로세스에 가장 두드러지게 기여하는지 식별하는 것을 목표로 한다.

이 분석을 용이하게 하기 위해 훈련 목적으로 각 시점 범주에서 20개의 이미지 하위 집합을 선택하여 입력으로 사용한다. 여기서 그림 1은 전면에 해당하는 이미지 20장 중 일부이며, 말 그대로 데이터 세트 중 전면에서 촬영한 이미지의 집합이다.

우리의 접근 방식을 통해 각 시점과 관련된 고유한 특성 집중적으로 조사할 수 있다. 다양한 관점에서 모델 성능을 체계적으로 평가함으로써 NeRF 모델링에 대한 관점 편향의 영향에 대한 통찰력을 얻고 모델 일반화 기능을 향상시키고자 한다.

그림 1. front 뷰포인트 이미지 20장 중 일부

4. 실험

4.1. 실험 환경, 모델, 성능 평가 지표

실험 환경은 Intel(R) Xeon(R) E-2334, GeForce RTX 3090 환경에서 진행하였으며, 실험 결과는 iter 40,000을 기준으로 비교 분석한다.

모델은 특정 관점에서만 훈련된 모델의 성능을 종합적으로 평가하기 위해 두 종류로 구성했다. 먼저, 전체 데이터 세트의 20%를 구성하는 20개의 데이터로 학습한 lego_20 모델을 학습하였다. 이는 전체 데이터 세트를 샘플링하여 특정 뷰디렉션에서만 훈련된 모델을 평가하는 기준선 역할을 한다. 또한 우리는 lego_<view> 모델을 구성하였다. 각 모델은 동일한 20개의 데이터를 사용하여 훈련되었지만 특정 방향(front, back, top, right, left)의 뷰디렉션 정보만으로 학습하였다.

모델 평가 지표는 Depth Estimation 논문의 성능 지표로 사용되는 Absolute Relative Error (AbsRel), Square Relative Error (SqRel), Root Mean Square Error (RMSE), Accuracy under a threshold (δ)를 사용하였다.

단, 데이터 세트 내부 Depth의 Ground Truth가 없기에 lego 데이터 세트를 모두 학습한 모델이 예측한 Depth 정보를 Ground Truth로 하여 성능평가를 진행하였다.

4.2. 전면 이미지 학습 결과

표 1은 lego_20, lego_front 모델을 균일한 방향의 데이터 세트에 대해 테스트한 결과이며, 표2는 전면 방향의 데이터 세트에 대해 테스트한 결과입니다.

	AbsRel↓	SqRel↓	RMSE↓	δ↑
lego_20	**0.18097**	**0.33305**	**0.86261**	**0.78688**
lego_front	0.19793	0.35750	0.91179	0.74853

표 1. 전면 이미지 학습 (테스트 세트: 균일한 방향)

표 1.의 결과로 보아 균일한 데이터 세트에 대해서는 균일한 데이터를 학습한 lego_20 모델이 좋은 성능을 보였다.

	AbsRel↓	SqRel↓	RMSE↓	δ↑
lego_20	0.20228	0.37289	0.91566	0.75220
lego_front	**0.17645**	**0.31777**	**0.87269**	**0.77637**

표 2. 전면 이미지 학습 (테스트 세트: 전면 방향)

표 2의 결과로 보아 전면 방향의 데이터 세트에 대해서는 전면 방향의 데이터만 학습한 lego_front 모델이 좋은 성능을 보였다. 또한 lego_front 모델의 경우, 테스트 데이터가 균일한 분포를 가질때 보다 전면 방향으로 구성되어 있는 경우 더욱 좋은 성능을 보였다.
추가로 시각적인 결과를 확인하고자 전면 방향 이미지의 Depth를 추출하였으며, 그림 2 우측과 같은 결과 이미지를 확인하였다.

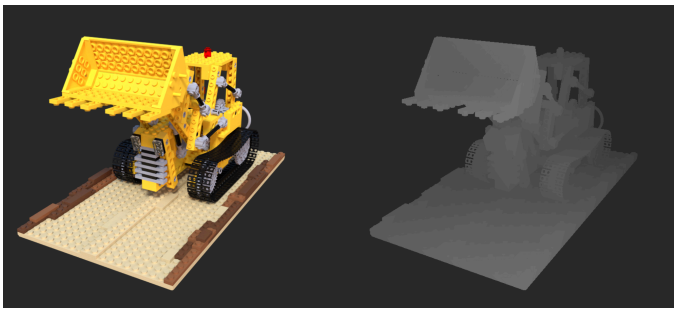

그림 2. front 뷰포인트 이미지와 Depth 이미지

4.3. 특정 방향의 이미지 학습 결과

표 3은 각 뷰포인트로 학습한 모델인 lego_<view>을 균일한 방향의 데이터 세트에 대해 테스트한 결과이며, 표 4는 해당 뷰포인트의 데이터 세트에 대해 테스트한 결과입니다.

	AbsRel↓	SqRel↓	RMSE↓	δ↑
lego_front	0.19793	0.35750	0.91179	0.74853
lego_back	0.17433	0.30545	0.85076	0.76831
lego_right	0.21099	0.37974	0.93041	0.74316
lego_left	0.16688	0.29090	0.84072	0.77612
lego_top	0.24183	0.45494	1.0006	0.71460

표 3. 특정 방향의 이미지 학습 (테스트셋: 균일한 방향)

	AbsRel↓	SqRel↓	RMSE↓	δ↑
lego_front	0.17645	0.31777	0.87269	0.77637
lego_back	0.17390	0.30508	0.85270	0.77075
lego_right	0.19959	0.37196	0.92359	0.78830
lego_left	0.16523	0.29026	0.83972	0.78489
lego_top	0.21771	0.40943	0.96441	0.71926

표 4. 특정 방향의 이미지 학습 (테스트셋: 특정 방향)

특정 방향에 대해서 테스트를 진행한 결과인 표 4에서의 결과가 표 3에서의 결과보다 전반적으로 좋은 성능을 나타내었다.

5. 결론 및 분석

본 연구에서는 NeRF 모델을 사용하여 균일한 데이터 세트와 특정 방향의 데이터 세트로 학습한 결과를 비교하고 불균형한 데이터 세트에서 뷰디렉션의 중요성을 고찰하였다. 실험 결과에서 우리는 특정 방향의데이터 세트에서 학습된 NeRF 모델이 해당 방향에서 더 뛰어난 성능을 보였음을 확인할 수 있었다.
이러한 결과는 불균형한 데이터 세트에서 뷰디렉션의 중요성을 명확히 보여주며, 3D 시각화 및 렌더링 분야에서의 실제 응용에 대한 지식을 확장하는데 기여할 것으로 기대된다.

참고 문헌

[1] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik,Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF:Representing scenes as neural radiance fields for view synthesis. ECCV, 2020

[2] Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, PeterHedman, Ricardo Martin-Brualla, andPratul P. Srinivasan. Mip-NeRF: A MultiscaleRepresentation for Anti-AliasingNeural RadianceFields. ICCV, 2021.