

NeRF 에서 커리큘럼 학습의 효과

이재석, 이혜진, 이재구*
국민대학교

*jaekoo@kookmin.ac.kr

Effects on Curriculum Learning in Neural Radiance Fields

Jaeseok Lee, Hyejin Lee, Jaekoo Lee*

College of Computer Science, Kookmin University

요약

NeRF(Neural Radiance Fields)는 다층 퍼셉트론을 사용하여 새로운 시점에서 장면 생성이 가능하다. 그러나 학습 수렴 속도가 느리고 복잡한 기하를 가진 데이터일수록 일반화가 어렵다는 단점이 있다. 따라서 본 논문에서는 쉬운 데이터에서 어려운 데이터로 점차 학습을 진행하여 학습 수렴 속도를 높이고 일반화 성능을 높이는 커리큘럼 학습(Curriculum Learning)을 NeRF 에 적용하였다. 커리큘럼 학습 효과를 확인하기 위해 무작위로 학습하는 기존 방식과 이전 학습 사진과 유사한 방향에서 바라본 사진부터 학습하는 커리큘럼 방식, 극단적으로 정반대 방향의 사진부터 학습하는 반 커리큘럼 방식을 비교하였다. 결과적으로 커리큘럼 학습을 적용한 방식은 기존 방식과 유사한 성능이었고, 반 커리큘럼 방식은 최대 82% 성능이 낮아졌다. 이는 100 장 이하의 학습 데이터 개수를 가지는 NeRF 는 커리큘럼 학습이 적용되기 어려우며, 오히려 무작위 순서로 학습을 진행하는 것이 과대적합(overfitting)을 방지할 수 있기 때문이다. 또한 NeRF 에서의 커리큘럼 학습의 과업 난이도를 정의하는 것은 단순한 방향보다는 잡음의 유무나 복잡한 기하 정보를 추가해야 함을 시사한다.

I. 서론

Neural Radiance Fields(NeRF)[1]는 소수의 사진과 카메라 센서 자세를 이용하여 3 차원 표현이 가능하기 때문에, 증강현실이나 가상현실, 자율주행 등의 분야에 응용될 수 있다. NeRF 는 합성곱 신경망을 사용하거나[2,3] 순환 신경망을 사용하는[4] 이전의 다른 모델들과 비교하여 다층 퍼셉트론만을 사용하지만, 더 높은 생성 성능을 보여준다. 하지만, 학습 수렴 속도가 느리고 복잡한 기하를 가진 데이터일수록 일반화가 어렵다. 따라서 학습 수렴 속도를 높이고 일반화 성능이 좋아지는 커리큘럼 학습[5,6]을 NeRF 에 적용하여 학습의 효율성을 제고할 방안을 모색하고자 하였다.

II. 본론

NeRF 는 사진과 카메라 자세에서 추출한 3 차원 점의 좌표와 시야각을 입력으로 하고 RGB 색상과 체적밀도(Volume Density) σ 를 출력으로 한다. 각 픽셀로부터 나온 출력값을 이용하여 새로운 시야의 사진, 동영상, 3 차원 형상을 렌더링할 수 있는 신경망 모델이다. [그림 1]은 NeRF 의 모델 구조와 전체적인 렌더링 방식을 보여주고 있다.

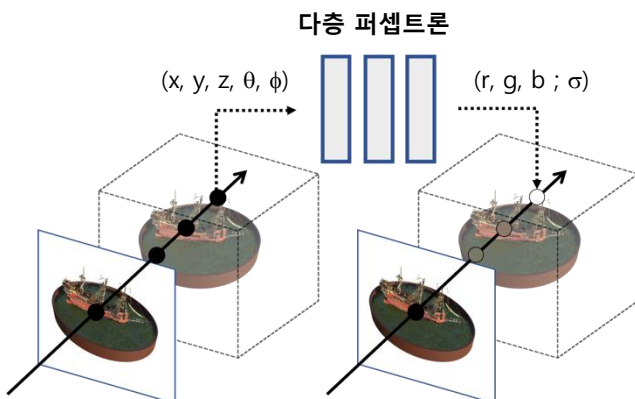


그림 1. NeRF 의 구조와 렌더링 방식

커리큘럼 학습(Curriculum Learning)은 기계학습 모델을 쉬운 데이터로 시작하여 점차 어려운 데이터를 학습하는 방법이다. 학습 방식은 [그림 2]와 같다. Q_T 는 전체 데이터 집합 분포를 가리킨다. 본 논문에서는 무작위(Random)로 학습하는 기존 방식과 커리큘럼 학습 방식, 어려운 데이터들부터 학습하는 반 커리큘럼(Anti-Curriculum) 방식을 NeRF 학습에 각각 적용하여 성능을 비교하였다.

III. 실험

비교를 위해 NeRF 모델을 각각 학습시킬 때 정확도와 손실값의 변화를 측정하였다. 커리큘럼 학습 방식은 전체 데이터 집합 100 개 중 초기 사진과 유사한 방향의 사진 20 개로 시작한다. 학습을 4 만 번 반복할 때마다 점차 다른 각도의 사진을 20 개씩 추가하여 마지막 4 만 번은 데이터 집합 전체를 훈련하도록 한다. 반 커리큘럼 방식은 초기 사진과 유사한 방향의 사진 20 개로 시작하는 것은 같다. 그러나 학습을 4 만 번 반복할 때마다 극단적으로 다른 각도의 사진을 20 개씩 추가하여 마지막 4 만 번은 데이터 집합 전체를 훈련하도록 한다.

정량적 지표로는 PSNR(Peak Signal-to-noise Ratio)[7], SSIM(Structural Similarity Index

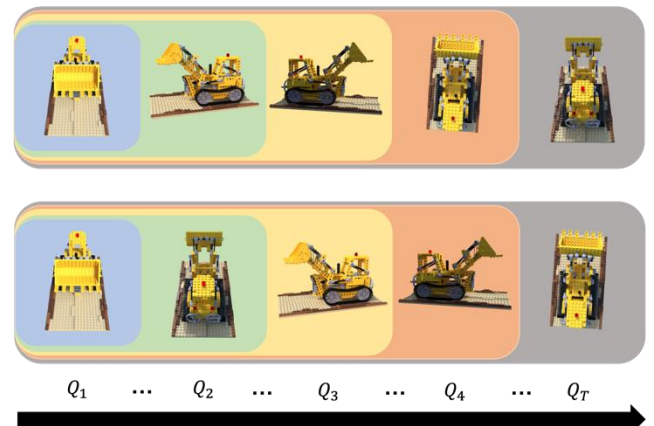


그림 2. 커리큘럼(위), 반 커리큘럼(아래) 학습 예시

표 1. 데이터 집합에 따른 학습 방법의 성능 비교

		무작위	커리큘럼	반커리큘럼
선박	PSNR(↑)	27.709	26.344	26.070
	SSIM(↑)	0.841	0.826	0.821
	LPIPS(↓)	0.165	0.179	0.185
레고	PSNR(↑)	31.241	30.544	26.898
	SSIM(↑)	0.951	0.946	0.908
	LPIPS(↓)	0.044	0.049	0.082

Measure)[7], LPIPS(Learned Perceptual Image Patch)[8]를 사용하였다. PSNR 과 SSIM 은 높을수록, LPIPS 는 낮을수록 성능이 좋다. 데이터 집합은 NeRF 에서 제공하는 다양한 방향에서 찍은 합성 데이터 중에서 대표적인 레고, 선박 데이터 집합[1]을 사용하였다.

선박 데이터 집합의 학습 결과는 [표 1]의 1 행부터 3 행까지를 가리킨다. 커리큘럼 방식과 반 커리큘럼 방식 모두 SSIM 은 무작위 방식과 유사한 성능을 보였으나, PSNR 은 반 커리큘럼 방식이 기존 방식보다 6% 하락하고, LPIPS 는 12% 상승하였다. 레고 데이터 집합의 학습 결과는 [표 1]의 4 행부터 6 행까지를 가리킨다. 커리큘럼 방식은 기존의 방식과 유사한 성능을 보였으나 반 커리큘럼 방식은 PSNR 이 12%, SSIM 이 4% 하락하고, LPIPS 는 82% 상승하였다.

[그림 3]은 선박 모델과 레고 모델의 학습 과정에서의 손실 추이를 가리킨다. 손실은 1000 번 반복하는 동안의 평균을 계산하여 총 200 번을 나타냈다. 기존 방식은 모두 안정적으로 손실이 줄어든 반면, 커리큘럼 방식과 반 커리큘럼 방식은 학습 데이터 집합이 추가될 때마다 손실이 증가하였다. 특히 레고 데이터 집합은 반 커리큘럼 방식으로 학습할 경우 전체 데이터 집합으로 학습할 때 손실이 매우 많이 증가했음을 확인할 수 있다.

즉, 커리큘럼 방식은 기존 학습 방식에서 개선되지 않았고, 반 커리큘럼 방식은 오히려 학습을 방해하여 테스트 시 결과가 다른 방식에 비해 크게 떨어짐을 알 수 있다. 또한, 선박 데이터 집합은 레고 데이터 집합보다 반 커리큘럼 방식으로 학습 시 테스트 성능이 더 낮음을 알 수 있다.

IV. 결론

본 논문에서는 커리큘럼 학습이 NeRF 학습에 미치는 영향을 파악하고자 무작위로 학습 데이터를 선택하는 기존 방식, 커리큘럼 학습 방식, 반 커리큘럼 학습 방식을 비교하였다. 실험한 결과, 커리큘럼 방식은 기존 방식과 성능이 비슷하고, 반 커리큘럼 방식은 최대 82% 성능이 낮아졌다. 이는 학습 데이터 개수가 100 장 이하로 적은 NeRF 의 특성상 커리큘럼 학습이 적용되기 어려우며, 오히려 무작위 순서로 학습을 진행하는 것이 과대적합을 방지할 수 있기 때문이다. 또한, 레고 데이터 집합은 간단한 기하 정보를 가지고 있어 초기에 학습한 데이터 집합 때문에 과대적합이 발생하여 전체 데이터 집합을 학습할 때 손실이 많이 증가하였고, 선박 데이터 집합은 복잡한 기하 정보를 가지고 있기 때문에 반 커리큘럼 학습이 레고 데이터 집합보다 학습 결과에 더 악영향을 끼쳤다. 이를 통해 NeRF 에서의 학습 과업 난이도를 정의하는 것은 단순한 방향이 아닌 잡음의 유무나 복잡한 기하 정보를 고려한 방향을 고려해야 함을 시사한다.

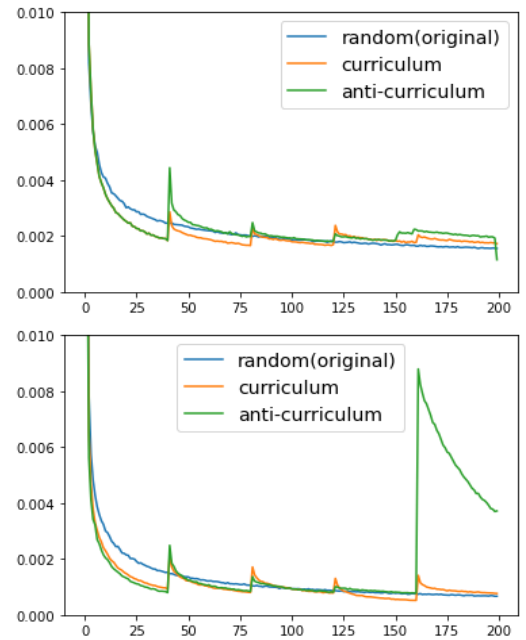


그림 3. 선박(위), 레고(아래) 데이터 집합의 학습 전략에 따른 손실 추세 비교

ACKNOWLEDGMENT

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.RS-2022-00167194, 미션 크리티컬 시스템을 위한 신뢰 가능한 인공지능)

참 고 문 헌

- [1] Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." European conference on computer vision. Springer, Cham, 2020.
- [2] Mildenhall, Ben, et al. "Local light field fusion: Practical view synthesis with prescriptive sampling guidelines." ACM Transactions on Graphics(TOG) 38.4 (2019): 1-14.
- [3] Lombardi, Stephen, et al. "Neural volumes: Learning dynamic renderable volumes from images." arXiv preprint arXiv:1906.07751(2019).
- [4] Sitzmann, Vincent, Michael Zollhofer, and Gordon Wetzstein. "Scene representation networks: Continuous 3d-structure-aware neural scene representations." Advances in Neural Information Precessing Systems 32 (2019).
- [5] Bengio, Yoshua, et al. "Curriculum learning." Proceedings of the 26th annual international conference on machine learning. 2009.
- [6] Wang, Xin, Yudong Chen, and Wenwu Zhu. "A survey on curriculum learning." IEEE Transactions on Pattern Analysis and Machine Intelligence (2021).
- [7] Whang, Zhou, et al. "Image quality assessment: from error visibility to structural similarity." IEEE transactions on image precessing 13.4 (2004): 600-612.
- [8] Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.