2023년도 한국통신학회 동계종합학술발표회

경량화된 초해상화 방법을 통한 NeRF 렌더링 효율 개선

이재석, 서정현, 이재구*
국민대학교

*jaekoo@kookmin.ac.kr

NeRF Efficiency Improvement with Lightweight Super-Resolution Method

Jaeseok Lee, Junghyeon Seo, Jaekoo Lee*
College of Computer Science, Kookmin University

요 약

신경 방사 필드(Neural Radiance Fields; NeRF)는 다층 퍼셉트론(Multi-Layer Perceptron; MLP)을 사용하여 물체의 임의 시야각 형상을 학습 추론으로 렌더링할 수 있다. 그러나 NeRF 의 볼륨 렌더링(Volume Rendering) 방식은 이미지 픽셀별로 카메라 광선을 투사함으로써 진행되기 때문에 렌더링 속도가 느리다. 따라서 본 논문에서는 렌더링 시간에서의 효율을 높이기 위하여 NeRF 렌더링 단계에서는 저해상도 이미지를 렌더링하고, 경량화된 초해상화(Super-Resolution) 모듈을 추가하여 원본 해상도로 업스케일링(Upscaling) 하는 기법을 제안한다. 실험 결과 렌더링 성능은 미비한 수준의 하락이 있었지만, 렌더링 시간은 최대 4.13 배 빨라져 NeRF 렌더링 효율을 개선할 수 있었다.

Ⅰ. 서 론

3 차원 재구성(3D Reconstruction)은 여러 장의 2 차원 이미지와 이미지가 가진 카메라 매개변수를 통해 3 차원 모델을 만들어내는 과업이다. 3 차원 재구성을 위해 제안된 심층 신경망을 이용한 방법 중, 신경 방사 필드(Neural Radiance Fields; NeRF)[1]는 다층 퍼셉트론만을 사용한다. 신경 방사 필드는 단순한 구조에도 불구하고 학습 시 사용한 이미지뿐만 아니라 새로운 시점에서의 장면 생성에서 높은 재구성 화질을 보였다[1].

그러나 신경 방사 필드에서 사용하는 볼륨 렌더링 방식은 이미지 픽셀 당 수백 번의 모델 추론이 필요하다.[2] 즉, 생성할 이미지의 해상도가 커질수록 렌더링 속도가 느려진다.

따라서 본 논문에서는 신경 방사 필드의 이미지를 저해상도로 렌더링하고, 경량화 된 초해상화 모델을 추가한 제안 방법을 통해 고해상도 이미지 직접 렌더링한 결과와 화질은 최대한 유지하면서 렌더링 시간은 감소하여 렌더링 효율을 개선하고자 하였다.

Ⅱ. 본 론

신경 방사 필드는 100 장 정도의 이미지와 이미지가 가진 카메라 자세(Camera Pose) 행렬을 훈련 데이터 집합으로 학습하여 가상의 카메라 광선을 따라 생성한 샘플 점들이 갖고 있는 체적밀도(Volume Density) σ 와 색상 c 를 추론하고, 이를 적분함으로 해당 픽셀에서의 색상 $\hat{C}(\boldsymbol{r})$을 계산하는 볼륨 렌더링 기법을 사용한다. $\boldsymbol{r}$은 카메라 광선, δ는 인접한 샘플 점 사이의 간격을, N은 샘플 점의 개수를, T_i는 i번째 샘플 점까지 축적된 불투명도를 의미한다.

$$\hat{C}(\boldsymbol{r}) = \sum_{i=1}^{N} T_i\left(1 - exp(-\sigma_i\delta_i)\right)c_i \qquad \text{식 (1)}$$

$$T_i = exp\left(-\sum_{j=1}^{i-1}\sigma_i\delta_i\right) \qquad \text{식 (2)}$$

일반적으로 초해상화는 저해상도의 이미지로부터 고해상도의 이미지를 복원하는 과업이다. 성능이 뛰어난 합성곱 신경망을 이용한 초해상화 기법의 등장 이후로 심층 신경망을 사용한 초해상화 기법들이 주로 사용하고 있다. 또한 경량화 된 초해상화 모델들은 이미지의 해상도를 수십 밀리초(ms) 내에 높일 수 있어 실시간 초해상화를 가능하게 하였다.

본 논문에서는 [그림 1]처럼 경량화 된 초해상화 모델인 XLSR[3] 모델 기반의 초해상화 모듈을 신경 방사 필드 모델에 덧붙이는 방법을 제안한다. 우선 신경 방사 필드는 저해상도 이미지를 렌더링한다. 저해상도 이미지는 8 채널을 가진 4 개의 3 x 3 합성곱 층과 16 채널을 가진 1 개의 3 x 3 합성곱 층을 통과한다. 8 채널의 합성곱 연산을 수행한 이미지를 연결(Concat) 후 32 채널의 1 x 1 합성곱 연산을 한 뒤 블록 구간을 통과한다. 블록 구간에서는 입력 채널을 4 개로 분할한 뒤 분할한 채널마다 8 채널의 3 x 3 합성곱 연산을 한 뒤 연결 후 32 채널의 1 x 1 합성곱 연산을 수행한다. 3 개의 블록을 거친 특성 맵은 앞서 16 x 16 합성곱 연산을 한 특성 맵과 연결시켜 32 채널의 1 x 1 합성곱 연산을 진행한 뒤, 업스케일을 위한 Depth2Space[4] 연산을 수행하고 27 채널의 3 x 3 합성곱 층을 통과하여 고해상도의 이미지를 얻는다. 활성화 함수는 마지막 27 채널의 3 x 3 합성곱 연산은 하한값을 1 로 설정한

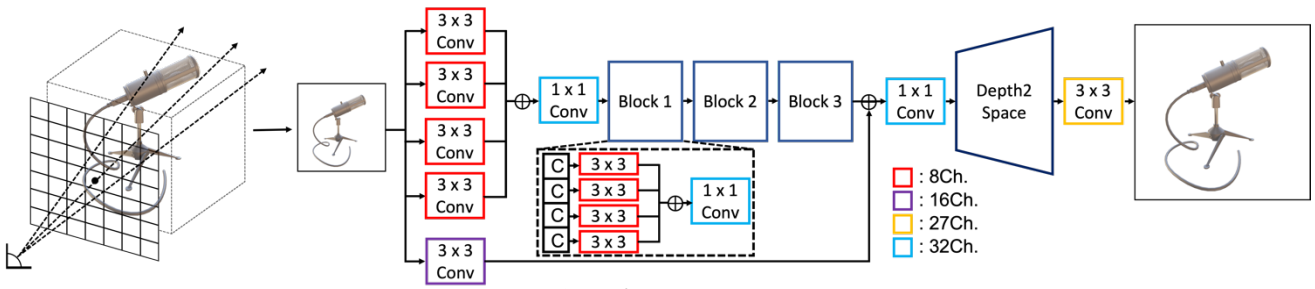

그림 1. 제안 모델 구조

표 3. 데이터 집합에 따른 렌더링 성능, 속도 비교

		기본 (고해상도)	초해상화	보간
NERF synthetic dataset [1]	PSNR(↑)	**37.682**	37.232	37.139
	SSIM(↑)	**0.948**	0.939	0.935
	LPIPS(↓)	**0.021**	0.029	0.031
	Time(↓)	12.864	**3.306**	-
LLFF Dataset [7]	PSNR(↑)	**31.921**	31.594	31.556
	SSIM(↑)	**0.784**	0.765	0.759
	LPIPS(↓)	**0.087**	0.114	0.115
	Time(↓)	12.516	**3.03**	-

Clipped ReLU 함수를 사용하고, 나머지는 모두 ReLU 함수를 사용하였다.

Ⅲ. 실 험 및 논 의

렌더링 성능은 PSNR(Peak Signal-to-noise Ratio)[5], SSIM(Structural Similarity Index Measure)[5], LPIPS(Learned Perceptual Image Patch)[6]로 측정하였다. PSNR과 SSIM은 높을수록, LPIPS는 낮을수록 성능이 좋다. 데이터 집합은 NeRF synthetic 데이터 집합[1]과 LLFF 데이터 집합[7]을 사용하였다.

초해상화 모듈은 신경 방사 필드 학습이 다 진행된 뒤 원본 해상도와 동일하게 생성한 이미지와 원본 해상도의 절반으로 이미지를 생성한 이미지를 학습 데이터로 하여 학습을 진행하였다. 이미지를 생성할 때 시야각은 데이터 집합이 허용하는 각도 내에서 무작위로 600 장을 생성하였다.

렌더링 성능은 원본 이미지와 동일한 해상도로 바로 렌더링한 이미지(기본), 절반 해상도로 렌더링한 뒤 초해상화를 통해 생성한 이미지(초해상화), 절반 해상도로 렌더링한 뒤 3 차 회선 보간법을 통해 업스케일링 한 이미지(보간)를 비교하였다. 렌더링 속도 측정은 초(sec) 단위로 비교하였다.

실험 결과는 [표 1]과 같다. 초해상화 이미지의 경우 원본에 비해 두 가지 데이터 집합에서 모두 PSNR, SSIM, LPIPS 가 미비한 수준으로 하락하였으나, 렌더링 속도가 각각 3.89 배, 4.13 배씩 빨라졌다. 또한 초해상화의 렌더링 성능은 보간과 비교해서 더 뛰어난 것을 확인할 수 있었다.

[그림 2]는 원본과 초해상화, 보간을 각각 시각화했을 때의 비교 그림이다. 보간의 경우 초해상화보다 더 많은 지글거림이 존재하였다.

Ⅳ. 결 론

본 논문에서는 저해상도로 렌더링한 신경 방사 필드 이미지를 경량화한 초해상화 모듈을 이용하여 원본 해상도와 유사한 렌더링 성능을 유지하면서 렌더링 효율을 개선하고자 하였다.

실험 결과, 초해상화를 통해 렌더링 성능은 소폭 하락하였지만 렌더링 속도가 최대 4.13 배 유의미한 향상을 이루었다. 이를 통해, 초해상화 모듈이 신경 방사 필드의 렌더링 효율을 개선할 수 있음을 확인하였다.

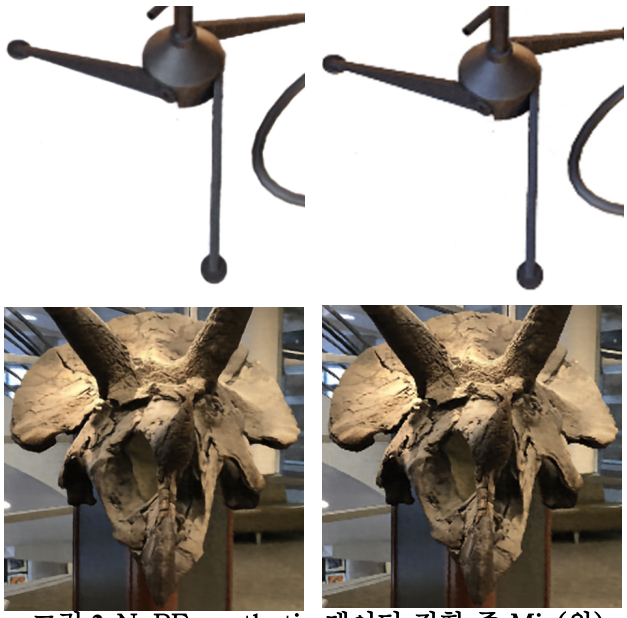

그림 2. NeRF synthetic 데이터 집합 중 Mic(위), LLFF 데이터 집합 중 Horns(아래)의 렌더링 방식에 따른 결과 비교. 왼쪽(초해상화), 오른쪽(보간)

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." Communications of the ACM 65.1(2021): 99-106.

[2] Fridovich-Keil, Sara, et al. "Plenoxels: Radiance Fields Without Neural Networks." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2002.

[3] Ayazoglu, Mustafa. "Extremely lightweight quantization robust real-time single-image super resolution for mobile devices." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[4] Shi, Wenzhe, et al. "Real-time single image and video super-resolution using and efficient sub-pixel convolutional neural network." Proceedings of the IEEE Conference on computer vision and pattern recognition. 2016.

[5] Wang, Zhou, et al. “Image quality assessment: from error visibility to structural similarity.” IEEE transactions on image processing 13.4 (2004): 600-612.

[6] 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.

[7] 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.