Codebook-NeRF: 코드북 기반의 NeRF 해상도 개선

이강현*°, 최성준*, 김정욱
경희대학교 전자정보대학 전자공학과
경희대학교 소프트웨어융합대학 소프트웨어융합학과
경희대학교 소프트웨어융합대학 컴퓨터공학부
khlee01@khu.ac.kr, csj000714@khu.ac.kr, ju.kim@khu.ac.kr

Codebook-NeRF: Improving NeRF resolution based on codebook

Kanghyun Lee*°, Seongjun Choi*, Jung Uk Kim
Department of Electronic Engineering, Kyung Hee University
Software Convergence, Kyung Hee University
Department of Computer Science and Engineering, Kyung Hee University
khlee01@khu.ac.kr, csj000714@khu.ac.kr, ju.kim@khu.ac.kr

요 약

본 논문에서는 참조 이미지 없이도 저해상도 이미지의 고해상도 디테일을 복원할 수 있는 새로운 NeRF[1] 방법을 제안한다. 이를 위해 NeRF-SR[2]의 Super Resolution 과정을 유지하면서, VQ-VAE[3]의 코드북(codebook) 구조를 도입하여 고해상도 이미지의 패턴을 학습하고 Refinement 기법을 개선한다. 코드북의 임베딩 벡터 수를 증가시켜 더 많은 고해상도 정보를 학습하게 하였으며, Imitation Inference를 통해 참조 이미지 없이도 고해상도 잠재 특성을 모방하도록 훈련한다. 실험 결과, 제안한 모델은 NeRF-SR[2]의 PSNR 성능을 유지했고, 선명하고 디테일이 풍부한 이미지를 생성하는 데 성공하였다. 더 많은 정보는 다음의 링크에서 확인할 수 있다: https://drawingprocess.github.io/Codebook-NeRF/

1. 서 론[1]

NeRF[1]는 최근 3차원 생성 인공지능 분야에서 가장 주목받는 기술 중 하나이다. NeRF[1]는 장면을 3차원적으로 표현할 수 있어, 다양한 3D 모델링 및 시각화 분야에서 많은 관심을 끌고 있다. 그러나 NeRF[1]는 저해상도 이미지에서 고해상도의 세부 정보를 복원하는 데 한계가 있으며, 이에 따라 여러 문제점을 보완하거나 새로운 기술을 덧붙인 모델들이 등장하였다.

NeRF-SR[2]은 Super Resolution 기술을 활용해 저해상도의 이미지를 고해상도로 복원하는 대표적인 모델이다. NeRF-SR[2]은 Super Resolution과 Refinement 두가지 단계로 이루어져 있다. 기존의 NeRF[1] 모델은 각 픽셀에 ray marching 과정을 진행하는 반면에, NeRF-SR[2]은 Super Resolution 과정을 통해 각 픽셀을 2x2로 나누어 ray marching을 함으로써 해상도를 높인 이미지를 얻는다.

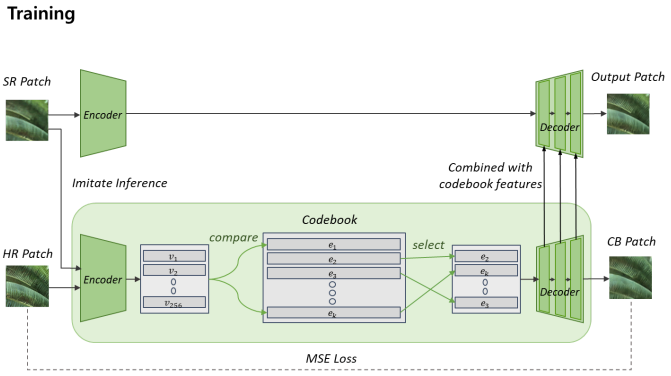

그림 1. 훈련 모델 구조

이후 Refinement 과정을 통해 고해상도 디테일을 복원한다. 이때 NeRF-SR[2] 모델은 학습과 테스트 과정에서 반드시 고해상도의 참조 이미지가 필요하며, 이러한 한계로 인해 참조 이미지가 없는 상황에서는 성능이 제한된다. 이에 본

[1] "본 연구는 과학기술정보통신부 및 정보통신기획평가원의 2024년도 SW중심대학사업의 결과로 수행되었음"(2023-0-00042)
* 해당 연구에 기여한 부분이 동등함

논문에서는 이러한 NeRF-SR[2] 모델의 한계를 해결하기 위해, 참조 이미지 없이도 고해상도 디테일을 복원할 수 있는 새로운 방법을 제안한다. 제안한 방법은 코드북을 사용하여 고해상도

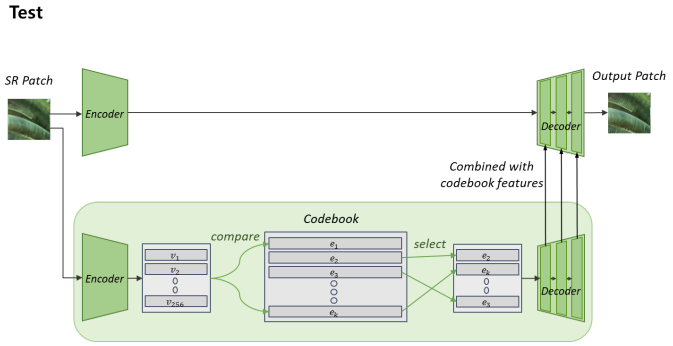

그림 2. 테스트 모델 구조

이미지의 패턴 학습하고, 이를 통해 고해상도 디테일을 복원하는 새로운 Refinement 기법을 도입한다.

2. 본론

2-1. 제안한 모델 구조

그림 1은 훈련 모델의 구조를 보여준다. 제안한 모델은 기존 NeRF-SR[2]의 Super Resolution 구조를 유지하면서 Refinement 모델을 개선한다. 우선 NeRF-SR[2]와 동일하게 이미지를 64x64 크기의 패치 단위로 분할하여 학습한다. 고해상도 디테일 복원을 위해 VQ-VAE[3]의 코드북 구조를 도입하며, 코드북은 고해상도 이미지의 잠재 특성을 학습하고 저장하는 임베딩 테이블로 구성되어 있다. 테이블의 각 임베딩 벡터 $e_i(i = 1, 2, \dots, k)$는 고해상도 이미지를 표현하는 데 필요한 정보를 담고 있으며, 기존 VQ-VAE에서 사용된 임베딩 벡터의 개수 k를 512에서 1024로 증가시켜 더 많은 정보를 담을 수 있도록 개선한다. 이러한 개선을 통해 고해상도 디테일을 복원하는 성능이 향상된다. 코드북을 학습할 때, Imitation Inference를 도입하여 HR patch(고해상도 참조 패치)와 SR patch(저해상도 Super Resolution 패치)를 모두 사용한다. 디코딩 과정에서는 코드북에서 얻은 고해상도 디테일을 입력으로 사용하여 SR patch를 고해상도 패치로 복원할 때, UNet 구조의 아이디어를 적용하였다. 각 deconvolution 레이어에서 코드북의 출력과 결합하여 고해상도 이미지를 복원함으로써, 더 나은 복원 성능을 얻는다.

그림 2는 테스트 모델의 구조를 보여준다. 그림 1의 Imitation Inference 과정을 통해 SR patch가 고해상도의 디테일을 모방하도록 학습하므로, 테스트 시에는 코드북에 학습된 고해상도 잠재 특성을 통해 SR patch만으로 고해상도 이미지를 복원한다.

2-2. 학습 과정

훈련 과정에서는 HR patch와 SR patch를 코드북의 입력으로 사용하여 고해상도 특성을 학습한다. 코드북을 통과한 HR patch의 출력은 고해상도 잠재 특성으로 저장되며, SR patch 역시 코드북 학습에 사용되어 Imitation Inference 과정을 통해 HR patch의 특성을 모방하도록 훈련한다. 이와 동시에, HR patch의 잠재 특성을 디코더로 복원할 때, 각 deconvolution 레이어 입력에 코드북의 출력과 결합하여 고해상도 이미지를 복원한다. UNet

구조를 적용하여 각 단계에서 얻어진 고해상도 디테일을 추가 입력으로 결합함으로써 복원력을 더욱 높인다.

2-3. 테스트 과정

테스트 과정에서는 훈련 과정과 유사하게 진행되지만, 그림 2와 같이 고해상도 참조 패치(HR patch)를 사용하지 않고 SR patch만을 사용한다. SR patch를 코드북에 입력하여 고해상도

특성을 가진 잠재 표현을 얻고, 이를 디코더를 통해 최종 고해상도 이미지로 복원한다. 이 과정에서 코드북을 통해 학습된 고해상도 디테일이 SR patch에 적용되며, 이를 통해 참조 이미지 없이도 고해상도의 디테일을 복원할 수 있는 능력을 확보한다.

훈련 과정과 테스트 과정 모두 코드북을 중심으로 SR patch가 HR patch처럼 고해상도 특성을 가지도록 구성되어 있어, 참조 이미지 없이도 고해상도 이미지 복원 성능을 얻는다.

2-4. Loss

최종 Loss는 다음과 같다.

$$L = \lambda_{Refine} \cdot L_{Refine} + \lambda_{Recon} \cdot L_{Recon} + \lambda_{VQ} \cdot L_{VQ} + \lambda_{Imit} \cdot L_{Imit}$$

L_{Refine}는 실측 자료(G.T)와 최종 출력 사이의 L1 Loss이다. L_{VQ}는 HR patch를 입력으로 넣었을 때의 codebook loss와 commitment loss를 더한 loss이다. L_{Recon}는 reconstruction loss로, HR patch와 CB patch 사이의 MSE loss이다. 이 loss를 통해 고해상도 디테일을 잘 복원하도록 학습한다. VQ Loss는 codebook loss와 commitment loss를 더한

표 1. PSNR 성능 평가. 고해상도의 참조 이미지 없이도 NeRF-SR[2]만큼 잘 동작함을 보여준다.

	fern	flower	fortress	horns	
NeRF-SR[2]	19.7	24.59	26.95	24.93	
Ours	**20.45**	**24.88**	**27.38**	**24.97**	
	leaves	orchids	room	trex	**Avg.**
NeRF-SR[2]	**16.99**	**20.21**	28.78	**24.81**	**23.37**
Ours	16.1	18.37	**30.26**	24.52	**23.37**

값으로, VQ-VAE[3]에서 코드북을 학습하기 위해 사용한 loss를 그대로 사용한다. codebook loss와 commitment loss는 다음과 같다.

$$L_{VQ} = ||sg[z_e(x)] - e||_2^2 + \beta||z_e(x) - sg[e]||_2^2$$

마지막으로, L_{Imit}는 Imitation Loss로, 인퍼런스를 모방하기 위해 추가했으며, SR patch를 입력으로 넣어서 얻은 codebook loss와 commitment loss를 더한 값이다. 여기서 λ_{Refine}, λ_{Recon}, λ_{VQ}, λ_{Imit}, β는 각각 1, 10, 1, 1, 0.25로, NeRF-SR[2]와 VQ-VAE[3]에서 사용한 값을 동일하게 사용한다.

3. 실험 결과

3-1. 데이터셋

실험에는 Local Light Field Fusion(LLFF) 데이터셋을 사용했으며, LLFF는 일정 거리에서 찍은 전방 이미지 데이터셋이다. 데이터셋 내부에는 fern, flower, leaves 등 8개의 장면으로 구성되며, 모든 장면에서 기존 모델인

(a) G.T (b) NeRF-SR[2] (c) Ours

그림 3. 정성 평가: 테스트를 통해 얻은 patch image

NeRF-SR[2]과 제안한 모델과의 성능 평가를 진행했다. 이때 저해상도 입력 이미지를 만들고자 크기가 1008×756인 원본 이미지를 2배 축소시켜 504×378 크기의 이미지를 만들어 모델의 입력으로 사용했다.

3-2. 성능

정량 평가 지표로 사용된 Peak Signal-to-Noise Ratio(PSNR)은 화질 손실 정도를 평가하는 데 널리 사용되며, PSNR 값이 높을수록 화질이 좋음을 의미한다. 에포크 단위로 가장 값이 큰 PSNR을 성능으로 채택했으며, 표 1에서 보이듯 많은 장면에서 제안한 모델이 NeRF-SR[2]보다 성능이 소폭 개선되었으며, 모든 장면의 PSNR을 평균낸 결과 기존 모델과 같음을 확인했다. 그림 3에서는 제안한 모델이 생성한 이미지를 육안으로 확인했을 때 더 선명하고 디테일이 풍부한 것을 확인할 수 있다. 이는 PSNR 수치의 개선뿐만 아니라, 고해상도 이미지의 세부 정보를 보다 효과적으로 보존하고 재구성하는 데에 중점을 둔 접근 방식의 차이에 기인한다. 제안한 방법은 다중 이미지의 고해상도 정보를 통합적으로 활용함으로써, 참조이미지 없이도 NeRF-SR[2]만큼의 성능을 이끌어냈다.

4. 결론

본 논문에서는 참조 이미지를 사용하지 않고도 고해상도 디테일을 복원할 수 있는 새로운 NeRF[1]를 제안한다. 제안한 모델은 코드북을 이용하여 고해상도 이미지를 학습하고 이를 통해 더욱 선명한 이미지를 생성하지만, 코드북은 많은 파라미터 수를 갖고 있어 학습이 오래

걸린다는 단점이 존재한다. 향후에 코드북 최적화를 통해 학습 시간을 줄이는 연구를 진행하고자 한다.

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] Chen Wang, Xian Wu, Yuan-Chen Guo, Song-Hai Zhang, Yu-Wing Tai, Shi-Min Hu. NeRF-SR: High-Quality Neural Radiance Fields using Supersampling. MM, 2022

[3] Aaron van den Oord, Oriol Vinyals, Koray Kavukcuoglu. Neural Discrete Representation Learning. NIPS, 2017