

상대 좌표 입력을 통한 Instant-NGP 의 구조 인식 능력 향상

김민¹, 김광수²

¹성균관대학교 소프트웨어학과 학부생

²성균관대학교 소프트웨어융합대학 교수

kimmin8758@g.skku.edu, kim.kwangsu@skku.edu

Enhancing Structural Awareness of Instant-NGP via Relative Coordinate Input

Min Kim¹, Kwangsu Kim²

¹Dept. of Computer Science and Engineering, Sungkyunkwan University

²College of Computing and Informatics, Sungkyunkwan University

요 약

NeRF 는 고품질 view synthesis 를 가능하게 했지만, 모든 좌표별 속성을 MLP 가 직접 학습해야 하므로 학습 속도가 느리다는 단점이 있다. 이를 보완한 Instant-NGP 는 해시 인코딩을 도입해 학습 속도를 개선했으나, 입력으로 절대 좌표만을 사용해 구조적 관계를 학습하기 어렵다는 한계가 있다. 본 연구는 이를 개선하기 위해 Instant-NGP 의 입력에 상대 좌표를 추가하는 방식을 제안한다. 상대 좌표는 구조적 연관성을 반영할 수 있도록 평균 위치를 기준으로 계산되며, 기존 인코딩과의 스케일 차이를 보정해 두 입력이 균형 있게 학습되도록 한다. 실험 결과, 제안한 방법은 LPIPS 지표에서 우수한 성능을 보이며, 구조 복원과 노이즈 제거 측면에서 유의미한 개선을 나타냈다.

1. 서론

View synthesis 는 여러 장의 2D 이미지로부터 3D 장면의 구조와 색을 학습하여 새로운 시점의 이미지를 생성하는 기술이다. 이 분야의 대표적인 기법인 NeRF(Neural Radiance Fields)[1]는 고품질의 view synthesis 를 가능하게 했지만, 모든 위치 정보를 하나의 MLP 가 직접 학습하기 때문에, 학습 속도가 느리다는 한계가 있다.

Instant-NGP[2]는 해시 기반 인코딩을 도입하여 NeRF 보다 빠른 학습 및 실시간 렌더링을 가능하게 했다. 하지만, 입력 좌표로 절대 좌표만을 사용하기 때문에 물체의 구조나 좌표 간의 상대적인 관계 등을 명시적으로 학습하기 어렵다.

한편, 3D semantic segmentation 분야에서 물체의 구조를 효과적으로 학습하기 위해 비슷한 점들을 하나의 슈퍼 포인트로 만드는 방식이 제안되었다[3]. 해당 방식은 슈퍼 포인트를 부모 슈퍼 포인트의 중심점을 기준으로 상대 좌표로 표현함으로써 공간적 문맥과 구조적 관계를 학습할 수 있도록 한다.

본 연구에서는 이러한 아이디어에서 영감을 받아 Instant-NGP 의 기존 입력 벡터에 상대 좌표를 추가

하는 방식을 제안한다. 이를 통해 물체의 정확한 위치를 학습함과 동시에 물체의 구조적 의미도 학습할 수 있도록 한다. 실험은 Synthetic NeRF 데이터셋을 사용하여 기존 입력 벡터만 사용했을 때와 상대 좌표 값을 추가하였을 때의 성능을 비교 분석했다.

2. 방법론

본 연구는 물체의 구조적 의미를 효과적으로 학습할 수 있도록 Instant-NGP 의 입력 벡터에 상대 좌표를 추가하는 방식을 제안한다. 기존 방식은 절대 좌표를 입력으로 사용하기 때문에 각 점이 독립적으로 표현되어 점들 간의 구조적 관계나 연관성을 학습하기 어렵다. 반면, 상대 좌표는 기준점을 중심으로 각 점의 좌표를 표현하므로, 모델이 구조 내에서 샘플링된 점들의 위치를 정확하게 인식할 수 있도록 한다. 이를 통해 물체의 구조를 정밀하게 학습할 수 있으며, 구조 외부에 불필요한 점들이 나타나는 현상을 줄일 수 있다.

상대 좌표(x_i^{rel})는 샘플링된 점들의 평균 위치를 기준으로 설정한 후, 각 점의 절대 좌표에서 기준점 값을 빼는 방식으로 계산한다. 이는 식 (1)과 같다.

$$x_i^{rel} = x_i - \frac{1}{N} \sum_{j=1}^N x_j \quad (1)$$

또한, 기존 인코딩 벡터값과 스케일을 맞추기 위해, 인코딩 벡터의 평균 표준편차(Std(z))와 유사한 스케일을 갖도록 상대 좌표 값을 조정한다. 이는 식 (2)와 같다.

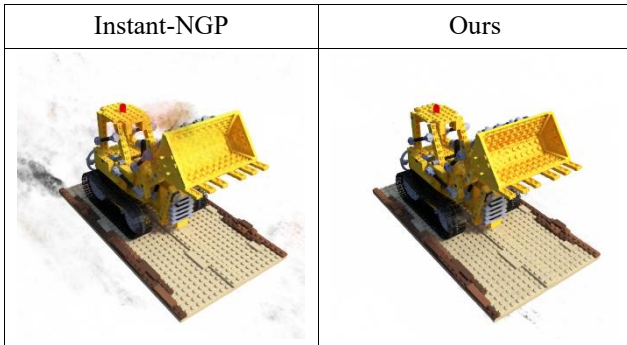
$$x_i^{scaled} = \left(\frac{Std(z)}{Std(x^{rel}) + \epsilon} \right) \cdot x_i^{rel} \quad (2)$$

이를 통해 모델이 기존 인코딩 벡터와 추가된 상대 좌표 값을 균형 있게 학습할 수 있도록 한다.

3. 실험

본 연구에서는 NeRF에서 제공한 Synthetic NeRF 데이터셋 중 Lego, Materials, Ficus를 사용하여 실험을 수행했다. 각 데이터셋은 총 400 장으로, 이 중 100 장은 train에, 100 장은 validation에, 나머지 200 장은 test에 사용했다.

성능 평가는 PSNR과 LPIPS를 지표로 사용했다. PSNR은 두 이미지 간의 픽셀 단위 차이를 바탕으로, 복원된 이미지가 원본과 얼마나 비슷한지를 측정하는 지표다. LPIPS는 두 이미지를 사전 학습된 네트워크에 입력하여 feature 공간에서 차이를 계산하는 지표다. 해당 실험을 통해 입력 좌표로 절대 좌표만을 사용했을 때와 상대 좌표를 추가로 사용했을 때의 성능 차이를 확인하고자 했다.



(그림 1) Lego 데이터에 대한 렌더링 결과.

그림 1은 Lego test 이미지에 대한 렌더링 결과를 시각적으로 비교한 예시다. 기존 방식은 입력 좌표로 절대 좌표만 사용했기 때문에, 물체의 구조가 흐릿하게 나타나고 노이즈가 많이 관찰된다. 반면, 제안된 방식은 상대 좌표를 도입하여 샘플링된 점들 간의 연관성을 학습함으로써 물체의 구조를 더욱 세밀하게 복원할 수 있었다. 구조를 보다 정확하게 학습한 결과, 구조 외부에 불필요한 점들이 나타나는 현상이 줄어들어 전체적인 노이즈도 감소한 것을 확인할 수 있다.

<표 1> PSNR 및 LPIPS 결과

Dataset	PSNR ↑		LPIPS ↓	
	[2]	Ours	[2]	Ours
Lego	28.16	27.00	0.097	0.073
Materials	25.29	23.03	0.229	0.206
Ficus	26.86	24.88	0.143	0.132

표 1은 PSNR과 LPIPS 지표를 통해 성능 분석한 결과다. PSNR 지표는 기존 모델이 제안 모델보다 다소 높은 수치를 보였으나, LPIPS 지표에서는 제안한 상대 좌표 기반 모델이 더 우수한 성능을 기록했다. 이는 PSNR이 밝기 변화나 해상도 차이에 민감한 반면, LPIPS는 사람이 인지하는 시각적 품질을 더 잘 반영하기 때문이다. 따라서 PSNR 수치만으로는 평가가 어려운 구조적 왜곡이나 노이즈 제거 측면에서, 제안한 방식이 시각적으로 더 자연스러운 결과를 생성했음을 LPIPS와 그림 1을 통해 확인할 수 있다.

4. 결론 및 향후 연구

본 연구는 Instant-NGP의 입력으로 상대 좌표를 추가하여 물체의 구조 학습을 더욱 가능하게 했다. 실험 결과 상대 좌표를 추가하는 것이 렌더링 품질 향상에 기여함을 확인했다. 그러나, PSNR 지표에서는 다소 낮은 성능을 보였다. 이는 본 연구에서 사용한 MSE 손실 함수가 픽셀 단위 오차에 더 민감하게 반응하기 때문으로 해석된다. 따라서 향후에서는 LPIPS 기반 손실 함수를 적용하는 한편, 상대 좌표 입력 방식을 개선하여 추가 성능 향상을 기대할 수 있다.

ACKNOWLEDGMENTS

이 논문은 2025년도 정부(개인정보보호위원회)의 재원으로 한국인터넷진흥원의 지원을 받아 수행된 연구임 (No.RS-2023-00231200, 자율주행 환경에서 AI 학습 가능한 개인영상정보 프라이버시 보존 기술개발)

참고문헌

- [1] Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." Communications of the ACM, Vol. 65, No. 1, pp. 99–106, 2021.
- [2] Müller, Thomas, et al. "Instant neural graphics primitives with a multiresolution hash encoding." ACM Transactions on Graphics, Vol. 41, No. 4, pp. 1–15, 2022.
- [3] Robert, Damien, Hugo Raguét, and Loïc Landrieu. "Efficient 3d semantic segmentation with superpoint transformer." Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2023, pp. 17195–17204.