

Minimal-Step Diffusion 기반 초경량 TinyUNet 의 해상도 복원 성능 분석

황병오^{1*}, 이호성¹, 김다산¹, 강명주²

¹서울대학교 공과대학 협동과정 인공지능전공, ²서울대학교 자연과학대학 수리과학부

[*brian2568@snu.ac.kr](mailto:brian2568@snu.ac.kr)

Performance Analysis of Image Super-Resolution Using Minimal-Step Diffusion-Based Lightweight TinyUNet

Byeongoh Hwang^{1*}, Hosung Lee¹, Dasan Kim¹ and Myungjoo Kang²

¹Interdisciplinary Program in Artificial Intelligence, Seoul National University,

²Department of Mathematical Sciences, Seoul National University

요약

본 논문은 최근 이미지 초해상도 분야에서 높은 성능을 보이고 있는 Diffusion 모델의 연산속도 문제를 해결하기 위해, CIFAR-10 데이터셋에서 초경량 UNet 구조와 극소 스텝(3-5 스텝)을 적용한 Diffusion 모델의 성능을 평가하였다. 제안한 초경량 Diffusion 모델은 기존 Diffusion 기반 모델과의 비교 실험에서 모델 크기와 추론 속도에서 우수한 성능을 보이며, 초소형 모델로도 충분한 품질의 초해상도를 구현할 수 있음을 실험적으로 입증하였다.

I. 서론

이미지 초해상도 기술은 낮은 해상도의 이미지를 고해상도로 복원하는 기술로서, CNN, GAN, Diffusion 등 다양한 모델이 사용되어왔다[1]. 특히 Diffusion 모델은 최근 이미지 생성 분야에서 뛰어난 성능을 보였으나, 높은 연산 비용과 느린 추론 속도가 단점으로 지적된다[2]. 본 연구는 Diffusion 모델의 연산 속도를 해결하고자 극소 스텝과 경량 모델 구조를 결합한 접근을 제안하며, CIFAR-10 데이터셋을 기반으로 성능을 정량적으로 분석하였다. 본 논문의 주요 독창성은 기존의 복잡한 구조나 다수 스텝 기반 Diffusion 초해상도 방식과 달리, 극소 스텝과 초경량 UNet 구조를 결합함으로써 모델의 경량화와 고속 추론을 동시에 달성한 점에 있다.

II. 본론

본 논문에서는 극소 스텝(3~5 스텝) 기반의 Diffusion 모델과 초경량 UNet 구조를 결합하여 CIFAR-10 데이터셋의 이미지 초해상도 문제에 적용하였다. 제안한 TinyUNet 은 전체 구조가 입력 이미지에 대해 두 개의 합성곱 층을 차례로 적용하는 인코더 부분과, 두 개의 전치 합성곱 층을 활용하는 디코더 부분으로 이루어진다. 인코더 블록은 저해상도 입력 이미지로부터 점진적으로 추상적이고 압축된 표현을 추출하는 역할을 하며, 디코더 블록은 이를 다시 고해상도 이미지로 복원하는 기능을 수행한다. 각 합성곱 층과 전치 합성곱 층 뒤에는 활성화 함수를 적용하여 비선형성을 도입하였다. 활성화 함수의 경우 마지막 단계에서만 sigmoid 를 활용하였고 이전

단계에서는 모두 ReLU 를 활용하였다. 이러한 간소화된 구조는 전체 파라미터 수를 크게 줄이면서도, CIFAR-10 과 같은 저해상도 데이터셋에 대해 빠른 학습 및 추론이 가능하도록 설계되었다.

실험에서는 CIFAR-10 의 원본 이미지를 32×32 에서 16×16 으로 downsampling 한 후, 이를 다시 32×32 로 복원하는 초해상도 태스크를 수행하였다. 모델의 성능 평가는 PSNR, SSIM 과 같은 화질 지표와, 이미지당 추론 속도(ms), 그리고 전체 파라미터 수(MB 단위)로 진행하였다. 실험 결과, 제안한 TinyUNet 기반 극소 스텝 Diffusion 모델은 기존의 대표적인 경량화 초해상도 모델과 비교해 유사한 수준의 화질(PSNR, SSIM)을 유지하면서도, 모델 크기와 추론 속도 면에서 뚜렷한 이점을 보였다. 본 논문의 실험 결과는, 복잡한 구조나 다수의 타임스텝 없이도 효율적인 초해상도 복원이 가능함을 시사한다.

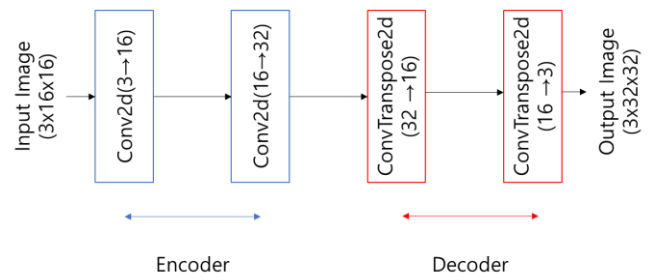


그림 1. TinyUNet 구조의 전체 흐름도

III. 결론

본 논문에서는 CIFAR-10 데이터셋을 대상으로, PSNR 과 SSIM 등 화질 지표, 추론 속도, 모델 파라미터 수를 기준으로 초경량 TinyUNet 기반 극소 스텝 Diffusion 모델의 성능을 평가하였다. 표 1 에서 확인할 수 있듯이, 제안한 모델은 SinSR[3]에 비해 PSNR(28.5 vs 27.2)와 SSIM(0.841 vs 0.832)에서 소폭 높은 화질을 기록하면서도, 추론 속도(6.8ms)는 SinSR(12.4ms)의 절반 수준으로 매우 빠른 것이 특징이다. 또한 ResShift[4], ResDiff[5]와 같이 더 많은 타임스텝과 큰 모델을 사용하는 기존 방법들과 비교했을 때, 제안한 모델은 훨씬 적은 파라미터(0.15M)와 타임스텝(5) 만으로도 경쟁력 있는 화질과 가장 빠른 추론 속도를 달성하였다. 이처럼 TinyUNet 기반 극소 스텝 Diffusion 모델은 단순한 구조와 적은 연산량만으로도 기존 경량화 모델에 비해 실용적 수준의 초해상도 복원이 가능함을 실험적으로 확인하였다. 이처럼 TinyUNet 기반 극소 스텝 Diffusion 모델은 단순한 구조와 적은 연산량 만으로도 기존 경량화 모델에 비해 실용적 수준의 초해상도 복원이 가능함을 실험적으로 확인하였다.

모델	Step	Param(M)	추론(ms)	PSNR	SSIM
SinSR	1	2.1	12.4	27.2	0.83
ResShift	15	1.7	20.2	28.9	0.85
ResDiff	50	3.4	45.5	29.8	0.86
TinyUNet	5	0.15	6.8	28.5	0.84

표 1. 실험 결과

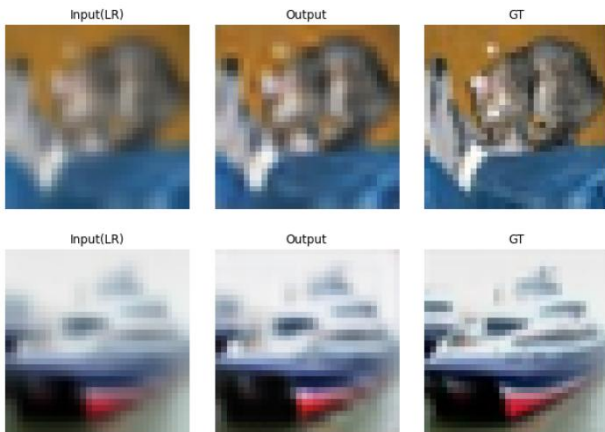


그림 2. 입력(좌), 제안한 모델의 출력(중), 원본(우) 예시

위 그림은 CIFAR-10 테스트 이미지에 대해, 저해상도 입력(16×16, 좌측), TinyUNet 기반 모델이 복원한 초해상도 결과(32×32, 가운데), 원본 고해상도(우측)를 비교한 것이다. 제안 모델의 출력은 입력에 비해 객체의 형태와 경계가 뚜렷하게 복원되었으며, 원본과 유사한 수준의 시각적 품질을 확인할 수 있다.

제안 모델이 SinSR[4], ResShift[5], ResDiff[6] 대비 적은 스텝과 파라미터에도 유사 화질을 유지할 수 있었던 주요 이유는 다음 세 가지로 요약할 수 있다.

첫째, CIFAR-10 과 같은 저해상도 데이터의 분포가 단순하여 극소 스텝에서도 기본적인 노이즈 제거와 형태 복원이 가능하다[1]. 둘째, TinyUNet 의 적절한 용량 배분과 업샘플링 기반 residual 복원 전략을 통해 전역 구조 부담을 줄이고 주로 고주파 보정에 집중함으로써 0.15M 파라미터만으로도 핵심 정보를 충분히 복원할 수 있었다[5][6]. 셋째, DDIM 기반 극소 스텝 샘플링 및 선형·코사인 스케줄 비교를 통한 노이즈 스케줄 최적화와, 동일 환경에서의 세밀한 하이퍼파라미터 조정으로 적은 스텝에서도 효율적 복원을 이끌어냈다[3]. 이러한 분석은 “극소 스텝 Diffusion + TinyUNet” 조합이 저연산 환경에서도 실용적 SR 솔루션이 될 수 있음을 뒷받침하며, 더 큰 도메인으로 확장할 때의 가이드라인으로 활용될 수 있다.

또한 일관된 벤치마크 환경에서 SinSR, ResShift, ResDiff 등 여러 경량 SR 기법을 동일 조건으로 비교·분석함으로써, 후속 연구자들이 “극소 스텝 Diffusion 성능-속도 trade-off”를 재현하고 확장할 수 있는 기준점을 제공했다. 이와 함께 구현이 간단한 TinyUNet 기반 파이프라인을 제시하여, 연구자나 학생들이 Diffusion SR 개념을 빠르게 프로토타입하고 실험을 진행할 수 있도록 돕는다. 이러한 통합적 비교와 교육적 프로토타입 제공은 연구생태계의 효율성과 확장 가능성을 높이는 데 기여한다.

ACKNOWLEDGMENT

This work was supported by MSIT/IITP[2021-001343-004, Artificial Intelligence Graduate School Program (Seoul National University)].

참 고 문 헌

- [1] Saharia, C., et al. (2023). Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 475-488.
- [2] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 6840-6851.
- [3] Kong, L., et al. (2023). SinSR: Diffusion-Based Image Super-Resolution in a Single Step. arXiv preprint arXiv:2304.10556.
- [4] Li, Y., Yue, X., Luo, Y., Yu, L., & Li, P. (2023). Residual Shifting for Efficient Diffusion-Based Image Super-Resolution. *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv preprint arXiv:2307.12348.
- [5] Shang, W., Li, Y., Zhang, H., Zhu, J., & Wu, Y. (2023). ResDiff: Combining CNN and Diffusion Model for Image Super-Resolution. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(3), 3121-3129. arXiv preprint arXiv:2303.08714.