

딥러닝 기반 저해상도 이미지의 텍스트 복원 초해상화 연구

함승재, 성민규, 강재모*

경북대학교

{hamsj99, alsrb0351, *jmkang}@knu.ac.kr

A Study on Text Restoration Super Resolution of Deep Learning-Based Low-Resolution Images

Ham Seung Jae, Sung Min Gyu, Kang Jae-Mo*

Kyungpook National Univ.

요약

저해상도 이미지를 고해상도로 변환하는 초해상화(Super Resolution) 기법은 딥러닝 기술의 발전과 함께 성능이 점차 향상되고 있다. 그러나 텍스트 복원 성능은 여전히 제한적이다. 본 논문에서는 OCR-Diff 모델을 사용하여 저해상도 이미지 내의 텍스트를 추출한 후, 이를 SUPIR 모델의 프롬프트로 활용하여 텍스트 복원 성능을 개선하는 방법을 제안한다. OCR 모델을 통해 추출된 텍스트를 프롬프트로 사용한 결과, 기존 SUPIR 모델보다 우수한 텍스트 복원이 이루어졌음을 확인할 수 있다.

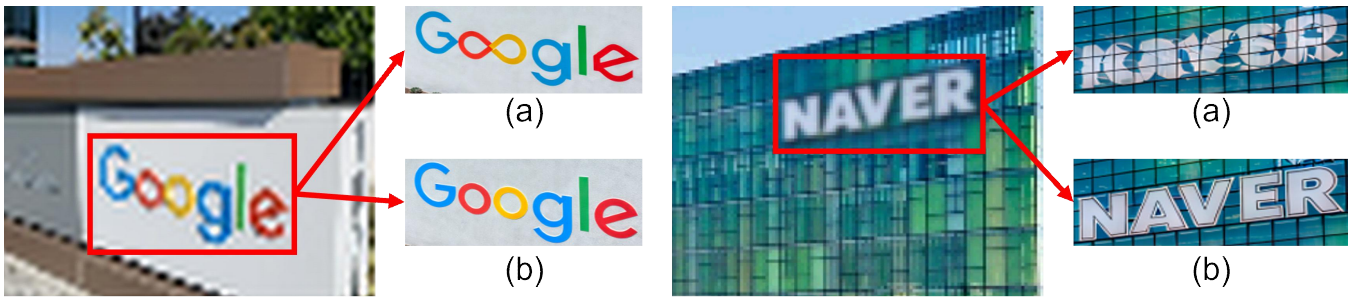


그림 1. OCR-Diff와 SUPIR 모델을 활용한 텍스트 복원 예시

I. 서론

저해상도 이미지를 고해상도로 복원하는 초해상화(Super Resolution, SR) 작업은 딥러닝 및 컴퓨터 비전 분야에서 다루는 중요한 연구 분야이다. 특히, 이미지 내에 있는 텍스트 복원은 OCR(Optical Character Recognition) 기술과 활용될 때, 많은 응용 분야에서 중요한 역할을 한다. 기존의 초해상화 모델들은 이미지의 전반적인 품질을 개선하는 데는 성공하였으나, 텍스트와 같은 세밀한 정보를 복원하는 데에는 한계를 보였다. 예를 들어, 그림 1의 (a)와 같이 Google에서 'oo'가 피비우스의 띠 형태로 있으며, 'G', 'g'와 'e'의 모양이 어색한 점과, NAVER에서는 글꼴이 다르게 복원된다. 이처럼 저해상도 이미지의 텍스트는 SR 모델의 복원 과정에서 종종 왜곡되거나 제대로 복원되지 않는 한계를 보인다. 텍스트 복원은 이미지의 픽셀 수준에서 매우 세밀하게 처리해야 하는 작업으로, 단순히 해상도를 높이는 것만으로는 복잡한 문자 구조를 온전하게 재현하기 어렵다. 따라서, 텍스트가 흐려지거나 왜곡되는 문제는 초해상화 기법의 한계 중 하나로 인식되고 있다.

이러한 문제를 해결하기 위해 본 논문에서는 OCR 기반의 모델을 사용하여 저해상도 이미지에서 텍스트를 정확하게 추출하고, 추출된 텍스트를 활용하여 SR 모델을 개발하는 방법을 제안한다. OCR 기반의 모델은 텍스트 인식과 추출에 있어 뛰어난 성능을 보여주는 OCR-Diff 모델로, 저해상도 이미지에서 효과적으로 텍스트를 식별 및 추출할 수 있다. 추출된 텍스트는 SR 모델인 SUPIR 모델의 프롬프트로 제공되며, 이를 통해 기존

모델의 텍스트 복원 성능을 크게 향상시킬 수 있다.

본 연구는 이러한 방식이 텍스트 복원의 성능을 개선하는 데 있어 높은 잠재력을 가지고 있음을 입증하고자 한다. 그림 1의 (b)는 본 논문에서 제안된 방법의 결과물을 보여준다.

II. 관련 연구

초해상화 및 이미지 복원: 최근 수년간 초해상화 및 이미지 복원 기술은 급격히 발전해 왔으며, SUPIR 모델은 이러한 발전의 중요한 사례 중 하나이다[1]. SUPIR 모델은 텍스트 기반 프롬프트를 사용하여 이미지 복원 성능을 제어하는 메커니즘을 도입하였고, 이를 통해 더 구체적이고 세밀한 이미지 복원을 가능하게 하였다. 그러나 이러한 모델이 텍스트와 같은 세밀한 구성 요소를 복원하는 데에는 한계가 존재한다. 여러 연구에서 SUPIR 모델과 유사한 초해상화 모델들이 텍스트 복원에서 겪는 어려움을 해결하기 위해 다양한 방법을 제안하고 있으나, 기존 기법들은 여전히 충분한 성능을 보여주지 못하고 있다.

OCR 기술: OCR 기술은 텍스트를 인식하고 추출하는 데 매우 효과적인 도구로, 최근에는 저해상도 이미지에서 텍스트를 인식하는 문제를 해결하기 위해 많은 연구가 이루어지고 있다. 그 중, 특히 OCR-Diff 모델은 두 단계 학습 구조를 통해 저해상도 이미지에서도 높은 텍스트 인식 성능을 제공하며, 이를 통해 텍스트를 더욱 정확하게 추출할 수 있다[2]. OCR 기술이 텍스트 복원 성능에 미치는 영향을 보았을 때, OCR-Diff 모델은 그

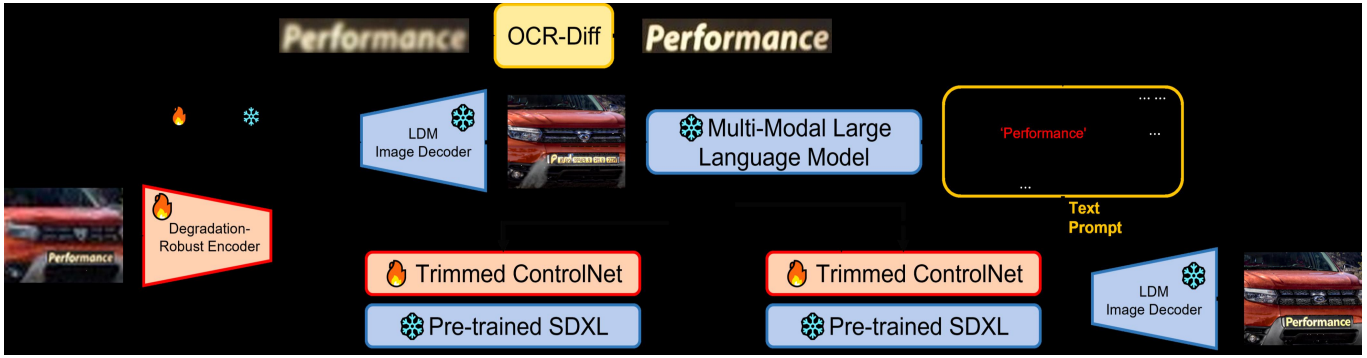


그림 2. 제안된 방법의 모델 구조도

안에서 가장 주목할 만한 성과를 보인다.

III. 본 론

SUPIR 모델은 초해상화 및 이미지 복원에서 매우 효과적인 기법으로, 특히 대규모 생성 프라이어어를 사용하여 이미지의 질감을 복원하는 데 강점을 지니고 있다. SUPIR는 Stable Diffusion XL(SDXL)이라는 대규모 모델을 기반으로 하며, 텍스트 기반 프롬프트를 활용하여 이미지의 특정 속성을 조정할 수 있다. 이때 LLaVA 모델이 사용되어 텍스트를 제공한다 [3]. 이러한 프롬프트 기반 제어는 이미지 복원의 정확도를 높이기 위해 유용하지만, 텍스트와 같은 세밀한 부분의 복원에는 여전히 어려움이 있다. OCR-Diff 모델은 저해상도 이미지 내에서 텍스트를 추출하는 데 매우 효과적인 도구로, 두 단계 학습 과정을 거쳐 더욱 정확한 텍스트 인식을 가능하게 한다. 첫 번째 단계에서는 기본적인 텍스트 구조를 파악하고, 두 번째 단계에서는 더 세밀한 텍스트 특징을 학습하여 복잡한 문자 패턴을 더 잘 인식할 수 있게 한다. 본 연구에서는 OCR-Diff 모델을 사용하여 저해상도 이미지에서 텍스트를 추출한 후, 이를 SUPIR 모델의 프롬프트로 활용하여 텍스트 복원 성능을 향상시키는 방법을 제안한다.

제안하는 방법은 그림 2와 같이 OCR-Diff를 사용해 저해상도 이미지에서 텍스트를 추출하고, 이 정보를 SUPIR 모델에 제공하여 텍스트 복원 성능을 개선하는 새로운 프로세스를 구축하는 것이다. 추출된 텍스트는 SUPIR 모델의 프롬프트로 사용되며, 이를 통해 텍스트 복원 과정에서 더 나은 결과를 얻을 수 있다.

그림 3은 OCR-Diff 모델을 통해 텍스트 정보를 추출하여 SUPIR 모델의 프롬프트에 제공했을 때(w/)와 안 했을 때(w/o)의 결과를 나타낸다. 실험 결과, SUPIR 모델 단독 사용 시에는 텍스트 복원에서 흐림 현상, 왜곡이나 손실이 자주 발생하는 반면, OCR-Diff 모델이 텍스트 정보를 제공한 경우, 텍스트 복원 성능이 크게 향상되었다. 이 방식은 특히 저해상도 이미지 내의 텍스트 복원이 어려운 상황에서 유용하다는 것을 보여준다.

그러나 OCR-Diff 모델과 SUPIR 모델을 함께 활용하여도, 저해상도 이미지에서 복잡한 텍스트 구조를 복원하는 데는 여전히 한계가 존재한다. 예를 들어, OCR-Diff 모델에서 텍스트 정보를 제공하더라도 기존의 SUPIR 모델에서 이미지 복원에 활용되는 LLaVA 기반 프롬프트의 의존도가 높기에 해당 프롬프트 결과의 질에 따라 텍스트 복원 성능이 떨어질 수 있으며, 이러한 한계를 해결하기 위해 추가적인 연구가 필요하다.

V. 결 론

본 논문에서는 OCR-Diff 모델과 SUPIR 모델을 활용하여 텍스트 복원 성능을 향상시키는 방법을 제안하였다. OCR-Diff를 통해 텍스트를 추출하고, 이를 SUPIR 모델의 프롬프트로 사용함으로써 기존 SUPIR 모델이 단독으로 텍스트 복원했을 때보다 더욱 정확하게 이루어졌음을 실험을 통해 입증하였다. 이후 LLaVA 기반 프롬프트 결과의 질에 대한 성능을 향상시키기 위한 연구를 진행할 계획이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터 지원사업의 연구결과로 수행되었음 (IITP-2024-2020-0-01808)

참 고 문 헌

- [1] Yu. F., Gu. J., Li. Z., Hu. J., Kong. X., Wang. X., He. J., Qiao. Y., and Dong. C., "Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 25669-25680, 2024.
- [2] Park. C. W., Palakonda. V., Yun. S., Kim. I. M., and Kang. J. M., "OCR-Diff: A Two-Stage Deep Learning Framework for Optical Character Recognition Using Diffusion Model in Industrial Internet-of-Things," IEEE Internet of Things Journal, Vol. 11, No. 15, pp. 25997-26000, 2024.
- [3] Liu. H., Li. C., Wu. Q., and Lee. Y. J., "Visual Instruction Tuning," Advances in neural information processing systems, 36, 2023.

	LR	SUPIR w/o OCR-Prompt	SUPIR w/ OCR-Prompt
Performance			
SAMSUNG			
Apple Park Visitor Center			

그림 3. 저해상도 이미지의 텍스트 복원 결과

IV. 실험 및 평가