Journal of KIIT. Vol. 23, No. 6, pp. 1-10, Jun. 30, 2025. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2025.23.6.01

비전-언어 모델을 활용한 블라인드 올인원 영상 복원

이재훈*, 손창환**

Blind All-In-One Image Restoration using Vision-Language Model

Jae-Hun Lee*, Chang-Hwan Son**

요 약

기존의 영상 복원 연구들은 잡음, 빗줄기, 헤이즈 등 특정 결함을 대상으로 개발되었으며, 복원 과정에서 잡음 강도나 다운샘플링 비율과 같은 열화 모델의 사전 정보를 알고 있다는 가정하에 이루어졌다. 그러나 이러한 접근은 실세계에서 발생하는 복잡한 열화 과정을 충분히 반영하지 못해 복원 성능과 실용성에 한계가 있다. 이에 본 연구에서는 비전-언어 모델에서 사전 훈련된 텍스트 프롬프트 벡터를 활용하여, 다양한 열화 종류에 강인한 블라인드 올인원 영상 복원 모델을 제안한다. 특히, 제안한 방법은 텍스트 프롬프트 벡터를 디코더의 가이드 정보로 활용함으로써, 영상 열화에 적응적인 복원을 달성하고자 한다. 실험 결과, 제안한 모델은 기존 PromptIR 대비 PSNR이 평균 0.554dB, SSIM이 평균 0.04 향상되었으며, 결과 영상에서도 결함이 효과적으로 제거되고 텍스처 및 에지 디테일이 더욱 선명하게 복원되었음을 확인하였다.

Abstract

Previous image restoration studies have primarily focused on addressing specific types of degradation, such as noise, rain streaks, and haze, under the assumption that prior information about the degradation model—such as noise level or downsampling ratio—is known. However, such approaches fail to adequately capture the complexity of real-world degradations, leading to limitations in both restoration performance and practical applicability. To overcome these challenges, this study proposes a blind all-in-one image restoration model that leverages text prompt vectors pre-trained in vision-language models to handle various types of degradation robustly. In particular, the proposed method utilizes the text prompt vectors as guidance information for the decoder, enabling adaptive restoration according to the degradation characteristics. Experimental results show that the proposed model achieves an average improvement of 0.554 dB in PSNR and 0.04 in SSIM compared to PromptIR, while also demonstrating more effective artifact removal and clearer restoration of textures and edge details.

Keywords
image restoration, all-in-one model, complex degradation, vision-language

* 국립군산대학교 소프트웨어학과 학사과정
- ORCID: https://orcid.org/0009-0009-9578-6301
** 국립군산대학교 소프트웨어학과 교수(교신저자)
- ORCID: https://orcid.org/0000-0001-7077-3074

· Received: Mar. 17, 2025, Revised: May 22, 2025, Accepted: May 25, 2025
· Corresponding Author: Chang-Hwan Son
Department of Software Science and Engineering, Kunsan National University, Republic of Korea
Tel.: +82-63-469-8915, Email: cson@kunsan.ac.kr

Ⅰ. 서 론

영상 복원은 잡음, 빗줄기, 헤이즈 등의 결함이 포함된 저화질 영상을 고화질 영상으로 변환하는 과정이다[1]. 이러한 기술은 의료 영상 장비, 지능형 감시 시스템, 자율주행 차량 등 다양한 분야에서 광범위하게 활용되고 있다. 예를 들어, 자율주행 차량의 경우, 기상 조건이나 조명 변화에 따라 영상 화질이 크게 저하될 수 있으며, 이는 영상 기반 자율주행 시스템의 성능 저하로 직결될 수 있다. 따라서 다양한 기상 및 촬영 환경에서도 안정적으로 고품질의 영상을 제공하여 시스템 성능을 유지할 수 있는 영상 복원 기술 개발이 필수적이다.

기존의 영상 복원 기법은 열화 모델의 파라미터 정보인 잡음 종류, 다운샘플링 비율, 블러링 유형 등을 알고 있다고 가정한다[2]. 또한 이러한 기법들은 주로 잡음, 헤이즈, 빗줄기와 같은 단일 열화 요소를 제거하는 데 초점을 맞추고 있다. 그러나 실세계 환경에서는 단일 열화가 아닌 복합적인 열화 과정이 발생할 가능성이 높으며 특정 열화 유형에 대한 사전 정보를 확보하기 어려운 경우가 대부분이다. 따라서 기존의 영상 복원 기법은 실세계 환경에서의 적용에 한계가 존재한다.

이러한 문제를 해결하기 위해 최근 블라인드 방식의 올인원 모델이 개발되고 있다. 블라인드(Blind) 방식은 열화 파라미터에 대한 사전 정보 없이 고화질의 영상을 복원하는 기법을 의미한다. 또한, 올인원(All-in-One) 모델은 특정 열화 제거에 국한되지 않고, 다양한 열화를 처리할 수 있도록 설계된 영상 복원 기법을 의미한다. 대표적인 블라인드 올인원 모델에는 AirNet[3]과 PromptIR[4]이 있다. AirNet은 열화 종류에 따른 잠재 특징 추출을 위해 대조 손실(Contrastive loss)을 적용하였고 PromptIR은 열화 종류에 적응적인 프롬프트를 사용하여 올인원 기능을 구현하였다.

본 연구에서는 PromptIR[4] 기반으로 입력 저화질 영상의 열화 종류에 적응적으로 대처할 수 있는 학습 가능한 텍스트 프롬프트를 생성하는 학습 방안을 제시한다. 특히, 비전-언어 모델인 CLIP[5]을 사용해서 저화질의 열화 종류와 텍스트 프롬프트 간의 유사도를 최대화하여 학습 가능한 프롬프트 벡터를 생성하는 과정을 제안한다. 또한 생성된 텍스트 프롬프트 벡터를 기존의 PromptIR[4] 모델에 적용하여 그 효과를 검증하였다. 실험 결과를 통해, 제안한 모델이 3종류의 열화 상황에 대해 PromptIR 보다 PSNR이 평균 0.554dB, SSIM이 평균 0.04만큼 향상된 정량적 화질 개선을 달성하였다. 그리고 정성적 평가에서도 텍스처의 디테일과 선명도를 제고할 수 있었고 결과 영상에서 영상 결함이 보다 깨끗하게 제거됨을 확인하였다.

Ⅱ. 관련 연구

2.1 논블라인드 영상 복원

논블라인드(Nonblind) 영상이란 열화 모델의 파라미터 값을 안다고 가정하고 영상 복원을 개발하는 방식을 말한다[6]. 예를 들면, 잡음 제거의 경우, 잡음의 강도나 잡음 종류에 대한 사전 지식을 안다고 가정한다. 초해상화의 경우, 다운샘플링 비율을 안다는 전제하에 저해상도 영상을 고해상도 영상으로 변환한다. 하지만, 이런 논블라인드 영상 복원은 실세계 적용에는 적합하지 않다. 왜냐하면, 저화질 영상에서 잡음의 특성과 샘플링 정보에 대한 사전 지식을 확보하는 것이 어렵기 때문이다.

2.2 블라인드 올인원 영상 복원

최근 논블라인드 영상 복원의 한계를 극복하기 위해 블라인드 올인원 복원 방식이 활발히 연구되고 있다. 블라인드 올인원 방식은 저화질 영상에서 사전 정보 없이 다양한 영상 결함을 처리할 수 있도록 하나의 통합된 모델로 개발된다. 이를 통해 기존 논블라인드 방식의 적용 한계를 극복하고, 보다 범용적인 영상 복원 성능을 확보하는 것이 가능하다. 대표적인 모델로는 AirNet[3]과 PromptIR[4] 모델이 있다. AirNet은 인코더-디코더 구조로 설계된 모델로, 인코더는 저화질 영상에 포함된 다양한 영상 결함 정보를 효과적으로 분석하며, 디코더는 이를 활용하여 복원 프로세스를 수행한다.

특히, 인코더에서는 잠재 손실을 도입하여 동일 유형의 영상결함 간 패치 유사도를 최대화하고, 이종 유형의 영상 결함 간 패치 유사도를 최소화한다. 이를 통해, 열화 종류에 대한 잠재 벡터 표현을 효과적으로 학습할 수 있으며, 최종 디코더에서 해당 잠재 벡터를 활용하여 영상 복원 성능을 향상시킨다.

PromptIR[4]는 다양한 영상 결함 종류에 대한 차별화된 정보를 효과적으로 인코딩하기 위해, 학습 가능한 파라미터의 집합, 즉 프롬프트를 활용한다. 복원 프로세스에서는 프롬프트를 통해 학습된 각 결함에 대한 정보를 기반으로, 입력 저화질 영상에 적응적으로 대응하여 복원 성능을 최적화한다. 그러나 PromptIR에서 사용되는 프롬프트는 랜덤 함수를 통해 초기화된다. 이는 최적 성능이 아닌, 추가적인 개선 가능성이 있다는 것을 말한다. 따라서 본 연구에서는 기존 PromptIR의 프롬프트 성능 개선을 위해, CLIP 모델을 통해 학습된 텍스트 프롬프트 벡터를 적용하는 방안을 새롭게 제시하고자 한다.

AirNet과 PromptIR 모델 외에도, 최근에는 잡음, 빗줄기, 헤이즈와 같은 다양한 영상 결함의 주파수 특성을 활용한 올인원 모델[7], 잡음·빗줄기·헤이즈 제거를 위한 멀티태스크 학습 기법을 적용한 올인원 모델[8], 그리고 영상 결함에 적응적인 프롬프트 튜닝 기반의 올인원 모델[9] 등이 개발되고 있다.

III. 제안한 학습 가능한 텍스트 프롬프트 기반 블라인드 올인원 복원 모델

CLIP[5]은 대규모 텍스트-영상 쌍 데이터셋을 활용하여 텍스트와 영상의 특징을 정렬하는 모델이다. 본 연구에서는 저화질 영상에 포함된 결함 유형을 파악하기 위해 학습 가능한 텍스트 프롬프트 벡터를 사용하고자 한다. 또한 이 텍스트 프롬프트를 블라인드 올인원 모델에 적용하여 다양한 영상 결함에 적응적인 영상 복원을 실현하고자 한다.

3.1 텍스트 프롬프트 학습

그림 1은 학습 가능한 텍스트 프롬프트 벡터를 생성하는 과정을 보여준다[10]. 학습 가능한 텍스트 프롬프트란 기존 텍스트 문장을 벡터화하여 표현한 형태를 말한다. 일반적으로 텍스트 문장은 입력 표현에 따라 최종 결과에 상당한 영향을 미치는 경향이 있다. 이런 문제를 완화하기 위해, 텍스트 문장을 벡터로 변형한 기법이 최근 연구되고 있다 [11].

그림 1에서 텍스트 프롬프트 벡터를 학습하기 위해, CLIP 모델에서 사전 학습된 텍스트 인코더와 영상 인코더가 필요하다. 텍스트 인코더는 텍스트 입력에 대한 특징을 추출하고 영상 인코더는 저화질 영상으로부터 고수준의 특징을 추출한다.

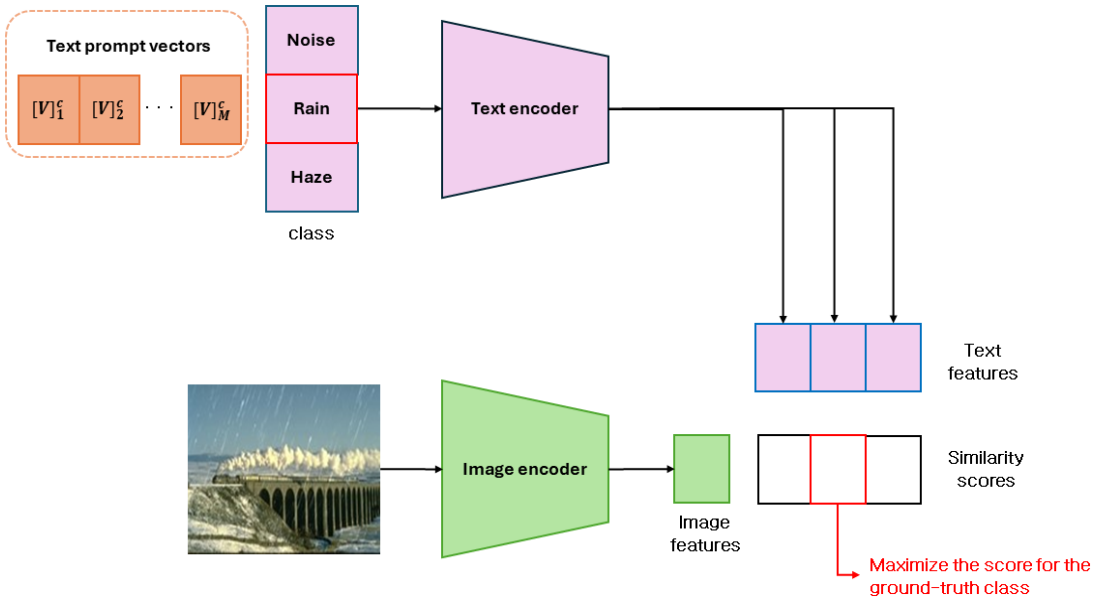


그림 1. 학습 가능한 텍스트 프롬프트 벡터 생성 과정
Fig. 1. Process of generating learnable text prompt vectors

본 연구에는 텍스트 문장 대신 $V_1, V_2, ..., V_m$ 벡터와 CLASS 정보를 결합한 텍스트 프롬프트 벡터를 사용한다. CLASS 정보는 잡음, 헤이즈, 빗줄기를 포함한 총 3가지로 구성된다.

$$P = [V]_1^c[V]_2^c...[V]_m^c[CLASS] \quad (1)$$

P는 벡터 방식으로 표현된 텍스트 프롬프트이고, $[V]_1^c[V]_2^c...[V]_m^c$은 클래스별로 학습될 M개의 벡터로서 난수로 초기화된다. 그리고 $[CLASS]$는 해당 클래스에 대한 라벨 정보를 나타낸다. 본 연구에서 M은 16으로 설정되었다.

텍스트 프롬프트 벡터를 학습하기 위해, 텍스트와 영상 인코더는 사전 학습된 파라미터를 고정한 상태에서 텍스트와 영상 특징을 추출한다.

$$\{PC_i\}_{i=1}^K = E_T(P) \quad (2)$$

$$f = E_I(X) \quad (3)$$

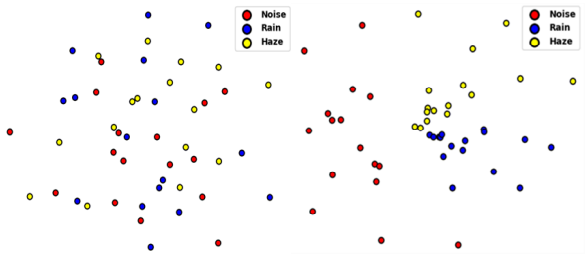
E_T와 E_I는 각각 사전 학습된 텍스트 인코더와 영상 인코더이다. 그리고 P는 최종 학습될 텍스트 프롬프트 벡터이고 X는 저화질 입력 영상이다. 텍스트 인코더의 출력은 K개의 텍스트 특징 벡터 PC_i이고, 영상 인코더의 출력은 영상 특징 벡터 f이다. 텍스트 프롬프트 벡터 P를 학습하기 위해, 코사인 유사도(Cosine similarity)를 계산한다.

$$p(Y=i \mid X) = \frac{\exp(\cos(PC_i, f))}{\sum_{j=1}^{K}\exp(\cos(PC_j, f))} \quad (4)$$

여기서 cos는 PC_i와 f의 유사도를 계산하는 코사인 유사도 함수이다. p는 이미지 벡터 X가 주어졌을 때, 출력 Y가 i번째 클래스가 될 확률이다. 수식 (4)는 코사인 유사도를 입력으로 갖는 소프트맥스 함수의 예측치이다. 이미지 특징 f에 대응하는 텍스트 특징 벡터 PC_i와 유사도가 최대가 되도록, 예측치 p를 교차 엔트로피에 적용하여 최종 텍스트 프롬프트를 학습한다.

3.2 텍스트 프롬프트 시각화

그림 2는 t-SNE[12] 기법을 사용해서 학습 전후의 텍스트 프롬프트 벡터를 시각화 결과이다. t-SNE는 고차원 벡터의 분포를 저차원 공간에서도 보존할 수 있는 시각화 기술이다. 그림에서 표현된 텍스트 프롬프트 벡터는 식 (1)에서 $CLASS$ 정보를 뺀 순수한 16개의 벡터에 해당한다. 그리고 각 벡터의 차원은 512이다. 그림 2에서 확인할 수 있듯이, 초기 프롬프트 벡터는 난수로 초기화되었기 때문에 잡음, 빗줄기, 헤이즈에 대응하는 텍스트 프롬프트 벡터들이 서로 뒤섞여 있다. 하지만, 학습이 진행된 후에는 클래스별로 텍스트 프롬프트가 군집을 형성한 것을 볼 수 있다. 이는 텍스트 프롬프트 벡터가 입력 저화질 영상의 결함 유형을 잘 구분할 수 있는 벡터로 변환되었음을 의미한다. 또한, 이러한 텍스트 프롬프트 벡터는 입력 저화질 영상의 결함 유형에 적응적으로 대응하기 위한 사전 정보로 유용하게 활용될 수 있음을 말해준다.



(a) (b)

그림 2. 텍스트 프롬프트 벡터 시각화
(a) 학습 적용 전, (b) 학습 적용 후
Fig. 2. Visualization of text prompt vectors
(a) before training, (b) after training

3.3 텍스트 프롬프트 기반 올인원 모델

그림 3은 제안된 텍스트 프롬프트 벡터를 적용한 블라인드 올인원 영상 복원 모델의 아키텍처를 나타낸다. 제안한 모델은 기존 PromptIR[4] 구조와 동일하게 인코더-디코더 기반의 U-Net 아키텍처를 따르며, 전역적 특징을 추출하기 위해 트랜스포머 블록(Transformer block)[13]을 포함하고 있다. 그림 3에서 제안한 올인원 모델의 핵심 요소는 프롬프트 블록(Prompt block)이다.

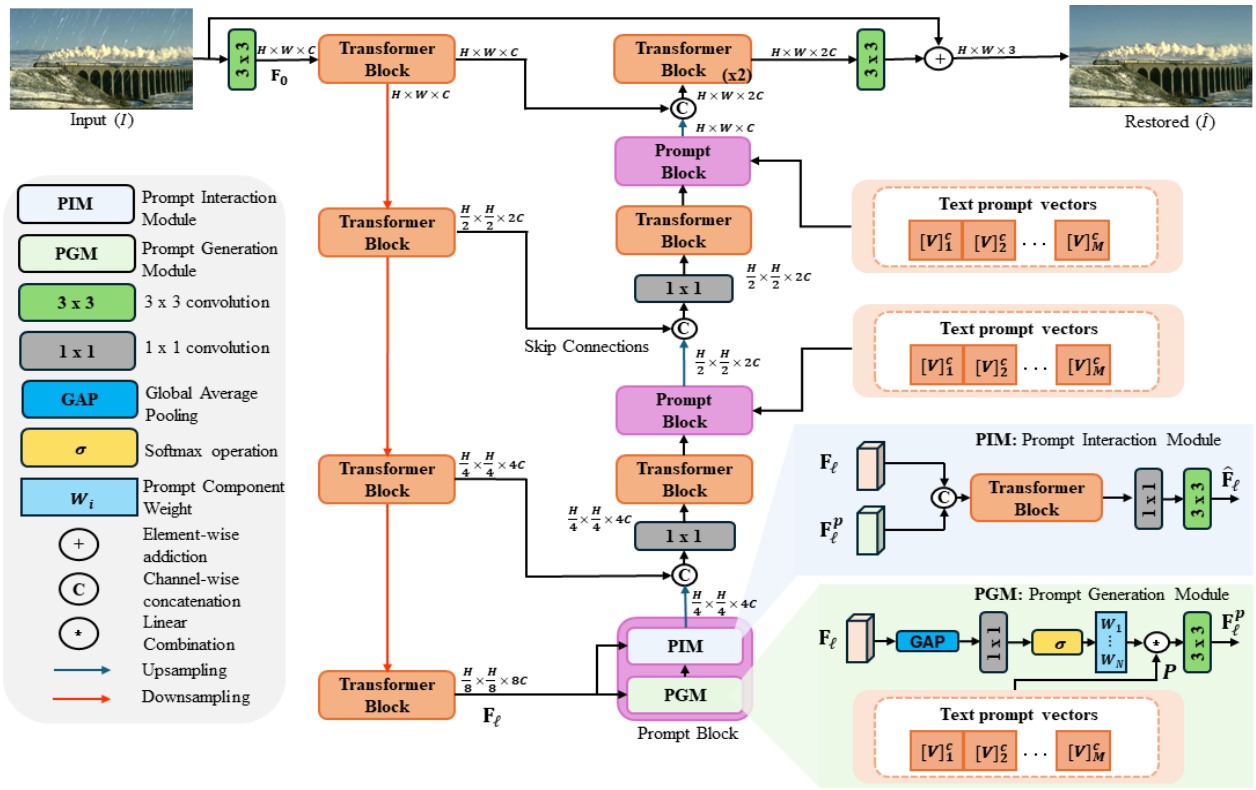


그림 3. 제안한 텍스트 프롬프트 벡터를 활용한 블라인드 올인원 복원 모델
Fig. 3. Proposed blind all-in-one restoration model using text prompt vectors

기존의 PromptIR은 프롬프트 블록에 난수로 초기화된 프롬프트 벡터를 활용했으나, 제안된 프롬프트 블록에서는 사전 학습된 텍스트 프롬프트 벡터로 대체하였다. 그림 2에서 보듯이, 사전 학습된 텍스트 프롬프트 벡터는 저화질 영상의 결함 유형을 효과적으로 구분하는 사전 정보로 활용될 수 있다. 따라서 제안한 텍스트 프롬프트 벡터를 적용함으로써, 기존 난수 방식보다 결함 유형에 효과적으로 대응할 수 있을 것으로 기대된다.

제안된 프롬프트 블록은 PGM(Prompt Generation Module)과 PIM(Prompt Interaction Module)으로 구성된다. 먼저, PGM 모듈은 U-Net 인코더에서 전달된 특징에 대해 전역 평균 풀링(GAP, Global Average Pooling), 컨볼루션 및 소프트맥스 계층을 순차적으로 적용하여 가중치 벡터를 계산한다.

$$w = Softmax(Conv_{1\times 1}(GAP(F_\ell))) \tag{5}$$

$$F_\ell^p = Conv_{3\times 3}(wP) \tag{6}$$

여기서 F_ℓ는 인코더의 ℓ번째 트랜스포머 블록에서 전달된 특징 맵이고 GAP, $Conv_{1\times 1}$, $Softmax$는 각각 전역 평균 풀링, 1×1 컨볼루션, 소프트맥스 계층을 의미한다. 가중치 벡터 w는 이후 사전 학습된 텍스트 프롬프트 벡터와 선형 결합(Linear combination)하여 입력 영상 결함 유형에 최적화된 텍스트 프롬프트 벡터를 새롭게 생성한다. 이 텍스트 프롬프트 벡터는 3×3 컨볼루션 연산을 거쳐 입력 영상 결함 유형을 대표하는 프롬프트 특징 벡터로 변환된다.

PIM 모듈은 입력 특징 F_ℓ과 새롭게 생성된 텍스트 프롬프트 벡터 F_ℓ^p를 퓨전하여 입력 저화질의 결함 유형에 적응적인 디코더 프로세스를 수행한다. PIM 모듈은 연결계층(Concatenation), 트랜스포머 블록 및 컨볼루션 계층으로 구성된다. 먼저, 연결 계층은 두 종류의 벡터인 F_ℓ^p와 F_ℓ을 채널 축으로 쌓고 트랜스포머 블록은 이 입력 특징에 대해 어텐션 메커니즘을 적용하여 중요한 영역을 강조한다. 그리고, 컨볼루션 계층을 거쳐 최종 특징 맵 $\hat{F}_\ell$을 생

성한다. 즉, 텍스트 프롬프트 벡터를 디코더의 가이드 정보로 활용하여 최종 저화질 영상에 적응적인 올인원 영상 복원을 달성하게 한다.

Ⅳ. 실험 및 결과

4.1 실험 환경

결함 대상과 데이터셋은 기존 올인원 연구와 동일하게 설정되었다. 즉, 결함 대상은 잡음, 빗줄기, 헤이즈 총 3종류로 국한되었고 잡음 제거를 위해서는 BSD400/68[14], Urban100[15], WED[16] 데이터셋이, 빗줄기 제거를 위해서는 Rain100L[17] 데이터셋이, 헤이즈 제거를 위해서는 RESIDE[18]의 OTS와 SOTS 데이터셋이 사용되었다. 잡음 제거를 위해, 가우시안 잡음이 적용되었고 잡음 레벨은 15, 25, 50 표준편차 값으로 설정되었다. 잡음 제거를 위해서는 BSD400과 WED 데이터셋의 5,144장을 훈련 데이터로, BSD68과 Urban100 데이터셋의 168장을 테스트 데이터로 사용하였다. 빗줄기 제거를 위해서는 Rain100L 데이터셋의 200장을 훈련 데이터로, 100장을 테스트 데이터로 사용했다. 헤이즈 제거를 위해서는 OTS 데이터셋의 72,135장을 훈련 데이터로, SOTS 데이터셋 500장을 테스트 데이터로 사용하였다. 학습에 사용된 배치 크기는 32이고 에폭은 120이다. 학습률은 0.0002로 설정되었고 가중치 갱신을 위해 아담 옵티마이저를 사용하였다.

4.2 정성적 평가

그림 4는 제안한 학습 가능한 텍스트 프롬프트 벡터를 적용한 블라인드 올인원 영상 복원의 결과이다. 정성적 평가를 위해, 최신 올인원 모델인 AirNet[3]과 PromptIR[4]를 비교하였다. 그림에서 첫 번째 행은 원본 영상이고 두 번째 행은 저화질 입력 영상이다. 그리고 나머지 행은 차례대로 각각 AirNet, PromptIR 및 제안한 방법으로 복원된 결과이다. 각 열은 서로 다른 열화 유형을 나타내며, 각각 잡음, 빗줄기, 헤이즈에 대한 실험 결과를 보여준다. 결과 영상의 세부 비교를 위해, 빨간색 박스 영역을 확대해서 우측 하단에 추가하였다. 먼저, 잡음 영상에 대해서는 제안한 기법이 암석 표면의 텍스처를 더 디테일하게 복원한 것을 볼 수 있다. 그러나 AirNet의 경우 텍스처가 잡음과 함께 제거된 것을 볼 수 있다. 빗줄기의 경우, 제안한 모델이 배경 영역의 빗줄기를 더 말끔하게 제거한 것을 볼 수 있다. 반면, AirNet과 PromptIR의 경우, 빗줄기가 여전히 남아 있는 것을 볼 수 있다. 헤이즈도 제안한 모델이 헤이즈로 뿌옇게 된 머리 부분을 더욱 선명하게 복원한 것을 확인할 수 있다.

4.3 정량적 평가

화질 평가를 위해, 최대 신호 대 잡음 비(PSNR, Peak Signal-to-Noise Ratio)와 구조적 유사성(SSIM, Structural Similarity)[19]을 사용하였다. 최대 신호 대 잡음 비는 원본 영상과 복원 영상의 화소 값이 얼마나 유사한지 평가하는 척도이고 구조적 유사성은 영상 구조나 텍스처의 복원 능력을 판별하는 척도이다. 제안한 모델의 성능 평가를 위해, 기존의 단일 영상 복원 모델인 BRDNet[20], LPNet[21], FDGAN[22], MPRNet[23]과 올인원 모델인 DL[24], AirNet[3], PromptIR[4]을 비교하였다. 표 1은 올인원 영상 복원에 대한 정량적인 평가 결과를 보여준다. 표에서 보듯이, 제안한 모델이 PSNR과 SSIM 평가 지표에서 결함 유형에 상관없이 가장 높은 수치를 달성하였다. 특히, 제안한 모델이 PromptIR보다 PSNR 수치가 평균 0.554dB 개선되었고, SSIM 수치가 평균 0.04 향상된 것을 볼 수 있다. 이는 학습 가능한 텍스트 프롬프트 벡터가 기존의 난수 벡터보다 영상 결함 유형에 적응적인 올인원 영상 복원에 보다 효과적임을 말해준다.

Ⅴ. 결 론

본 논문에서는 잡음, 빗줄기, 헤이즈와 같은 다양한 영상 결함을 하나의 모델로 복원할 수 있는 블라인드 올인원 영상 복원 모델을 제안하였다.

그림 4. 실험 결과 ; 원본 영상 (첫 번째 행), 저화질 입력 영상 (두 번째 행), AirNet[3] 결과 영상 (세 번째 행), PromptIR[4] 결과 영상 (네 번째 행), 제안한 모델 결과 영상 (마지막 행)

Fig. 4. Experimental results ; Original images (first row), low-quality input images (second row), resulting images with AirNet[3] (third row), resulting images with PromptIR[4] (fourth row), and resulting images with proposed model (last row)

표 1. 올인원 영상 복원을 위한 정량적 평가

Table 1. Quantitative evaluation for all-in-one image restoration

Type	Method	Denoising σ = 15	Denoising σ = 25	Denoising σ = 50	Deraining	Dehazing
All-In-One	BRDNet[20]	32.26/0.897	29.76/0.835	26.34/0.693	27.42/0.895	23.23/0.895
	LPNet[21]	26.47/0.778	24.77/0.747	21.26/0.552	24.88/0.783	20.84/0.827
	FDGAN[22]	30.25/0.910	28.81/0.868	26.43/0.775	29.89/0.932	24.71/0.929
	MPRNet[23]	33.54/0.927	30.89/0.879	27.56/0.779	33.57/0.954	25.28/0.954
	DL[24]	33.05/0.914	30.41/0.860	26.90/0.740	32.62/0.931	26.92/0.931
	AirNet[3]	33.92/0.932	31.26/0.888	28.00/0.797	34.90/0.967	27.94/0.961
	PromptIR[4]	33.98/0.933	31.31/0.888	28.06/0.799	36.37/0.972	30.58/0.974
	Proposed model	**34.12/0.963**	**31.47/0.939**	**28.20/0.892**	**38.12/0.992**	**31.16/0.980**

특히, 기존의 난수 기반의 프롬프트 대신 학습 가능한 텍스트 프롬프트 벡터를 학습하는 방법과 이를 부가정보로 활용하여 입력 저화질 영상에 적응적인 영상 복원을 위한 디코더 프로세스를 제안하였다. 실험 결과, 제안한 모델이 PromptIR보다 PSNR 수치가 평균 0.554dB 개선되었고, SSIM 수치가 평균 0.04 향상된 것을 볼 수 있다. 이를 통해 잡음, 빗줄기, 헤이즈와 같은 결함 유형을 효과적으로 복원할 수 있음을 입증하였다. 또한, 학습 가능한 텍스트 프롬프트가 기존의 난수 기반 프롬프트보다 올인원 영상 복원에 더 효과적임을 확인하였다.

본 연구에서는 블라인드 올인원 모델에서 제시한 평가 방법을 그대로 적용하였다. 구체적으로, 합성된 저화질 영상에 올인원 모델을 적용한 후, 복원된 영상과 원본 고화질 영상 간의 PSNR 및 SSIM을 측정하여 정량적 성능 평가를 수행하였다. 향후 연구에서는 실제 환경에서 촬영한 영상을 수집하고, 제안한 모델을 적용 및 실험할 계획이다.

Acknowledgement

'2024년도 한국정보기술학회 추계종합학술대회에서 발표한 논문(학습 가능한 컨텍스트 벡터를 활용한 블라인드 올인원 영상 복원)[25]을 확장한 것임'

References

[1] L. Zhai, Y. Wang, S. Cui, and Y. Zhou, "A comprehensive review of deep learning-based real-world image restoration", IEEE Access, Vol. 11, pp. 21049-21067, Mar. 2023. https://doi.org/10.1109/ACCESS.2023.3250616.

[2] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations", Proc. IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 3262-3271, Jun. 2018. https://doi.org/10.1109/CVPR.2018.00344.

[3] B. Li, X. Liu, P. Hu, Z. Wu, J. Lv, and X. Peng, "All-in-one image restoration for unknown corruption", Proc. IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, pp. 17452-17462, Jun. 2022. https://doi.org/10.1109/CVPR52688.2022.01693.

[4] V. Potlapalli, S. W. Zamir, S. Khan, and F. S. Khan, "PromptIR: Prompting for all-in-one blind image restoration", arXiv:2306.13090, Jun. 2023. https://doi.org/10.48550/arXiv.2306.13090.

[5] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision", Proc. International Conference on Machine Learning, Virtual, pp. 8748-8763, Jul. 2021. https://doi.org/10.48550/arXiv.2103.00020.

[6] K. Uchida, M. Tanaka, and M. Okutomi, "Non-blind image restoration based on convolutional neural network", Proc. IEEE Global Conference on Consumer Electronics, Taipei, Taiwan, pp. 40-44, Oct. 2018. https://doi.org/10.1109/GCCE.2018.8574671.

[7] Y. Cui, S. W. Zamir, S. Khan, A. Knoll, M. Shah, and F. S. Khan, "AdaIR: Adaptive all-in-one image restoration via frequency mining and modulation", arXiv:2403.14614, Mar. 2024. https://doi.org/10.48550/arXiv.2403.14614.

[8] K. Yu, X. Wang, C. Dong, X. Tang, and C. C. Loy, "Path-Restore: Learning Network Path Selection for Image Restoration", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 44, No. 10, pp. 7078-7092, Oct. 2022. https://doi.org/10.1109/TPAMI.2021.3096255.

[9] J. Ma, T. Cheng, G. Wang, Q. Zhang, X. Wang, and L. Zhang, "ProRes: Exploring degradation-aware visual prompt for universal image restoration", arXiv:2306.13653, Jun. 2023. https://doi.org/10.48550/arXiv.2306.13653.

[10] K. Zhou, J. Yang, C. C. Loy, and Z. Liu, "Learning to prompt for vision-language models",

International Journal of Computer Vision, Vol. 130, No. 7, pp. 2337-2348, Jul. 2022. https://doi.org/10.1007/s11263-022-01653-1.

[11] M. Li, T. Lv, J. Chen, L. Cui, Y. Lu, D. Florencio, C. Zhang, Z. Li, and F. Wei, "TrOCR: Transformer-based optical character recognition with pre-trained models", Proc. Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence, Washington, DC, USA, pp. 13094-13102, Jun. 2023. https://doi.org/10.1609/aaai.v37i11.26538.

[12] L. van der Maaten and G. Hinton, "Visualizing data using t-sne", Journal of Machine Learning Research, Vol. 9, No. 11, pp. 2579-2605, Nov. 2008.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need", Proc. The Conference on Neural Information Processing Systems, Long Beach, CA, USA, pp. 6000-6010, Aug. 2017. https://doi.org/10.48550/arXiv.1706.03762.

[14] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human-segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics", Proc. International Conference on Computer Vision, Vancouver, BC, Canada, pp. 416-423, Jul. 2001. https://doi.org/10.1109/ICCV.2001.937655.

[15] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars", Proc. IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 5197-5206, Jun. 2015. https://doi.org/10.1109/CVPR.2015.7299156.

[16] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang, "Waterloo exploration database: New challenges for image quality assessment models", IEEE Transactions on Image Processing, Vol. 26, No. 2, pp. 1004-1016, Feb. 2017. https://doi.org/10.1109/TIP.2016.2631888.

[17] W. Yang, R. T. Tan, J. Feng, Z. Guo, S. Yan, and J. Liu, "Joint rain detection and removal from a single image with contextualized deep networks", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 42, No. 6, pp. 1377-1393, Jun. 2020. https://doi.org/10.1109/TPAMI.2019.2895793.

[18] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single image dehazing and beyond", IEEE Transactions on Image Processing, Vol. 28, No. 1, pp. 492-505, Jan. 2019. https://doi.org/10.1109/TIP.2018.2867951.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", IEEE Transactions on Image Processing, Vol. 13, No. 4, pp. 600-612, Apr. 2004. https://doi.org/10.1109/TIP.2003.819861.

[20] C. Tian, Y. Xu, and W. Zuo, "Image denoising using deep CNN with batch renormalization", Neural Networks, Vol. 125, pp. 461-473, Mar. 2020. https://doi.org/10.1016/j.neunet.2019.08.022.

[21] X. Fu, B. Liang, Y. Huang, X. Ding, and J. Paisley, "Lightweight pyramid networks for image deraining", IEEE Transactions on Neural Networks and Learning Systems, Vol. 31. No. 6, pp. 1794-1807, Jun. 2020. https://doi.org/10.1109/TNNLS.2019.2926481.

[22] Y. Dong, Y. Liu, H. Zhang, S. Chen, and Y. Qiao, "FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing", Proc. AAAI Conference on Artificial Intelligence, New York, NY, USA, pp. 10729-10736, Jul. 2020. https://doi.org/10.1609/aaai.v34i07.6701.

[23] S. W. Zamir, A. Radwan, M. A. Khaleel, M. R. Amer, and M. Shah, "Multi-stage progressive image restoration", Proc. IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, pp. 14821-14831, Jun. 2021. https://doi.org/10.1109/CVPR46437.2021.01458.

[24] Q. Fan, J. Yang, Y. Xu, and Z. Zhang, "A

general decoupled learning framework for parameterized image operators", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 41, No. 1, pp. 33-47, Jun. 2019. https://doi.org/10.1109/TPAMI.2019.2925793.

[25] J. H. Lee, C. H. Kim, and C. H. Son, "Blind all-in-one image restoration using learnable context vectors", Proc. Korea Information Technology Conference, Jeju, Korea, pp. 1480-1484, Nov. 2024.

저자소개

이 재 훈 (Jae-Hun Lee)

2020년 3월 ~ 현재 :
국립군산대학교 소프트웨어학과
학사과정
관심분야 : 컴퓨터 비전, 영상처리, 기계학습, 딥러닝

손 창 환 (Chang-Hwan Son)

2002년 2월 : 경북대학교
전자전기공학부(공학사)
2004년 2월 : 경북대학교
전자공학과(공학석사)
2008년 8월 : 경북대학교
전자공학과(공학박사)
2017년 4월 ~ 현재 :
국립군산대학교 소프트웨어학과 교수
관심분야 : 컴퓨터 비전, 영상처리, 기계학습, 딥러닝