

# 흑백 이미지 복원을 위한 Stable Diffusion 및 GAN 기반 하이브리드 프레임워크

박혜인<sup>1</sup>, 이신화<sup>1</sup>, 최승효<sup>1</sup>, 엄다현<sup>1</sup>, 지정원<sup>1</sup>, 김준화<sup>2</sup>

건양대학교 의료인공지능학과<sup>1</sup>, 건양대학교 인공지능학과<sup>2</sup>

22619011@konyang.ac.kr, 22619019@konyang.ac.kr, tmdgy8362@gmail.com,

22619016@konyang.ac.kr, rhtmaehcl739@naver.com, junhwakim@konyang.ac.kr

## Stable Diffusion and GAN-Based Hybrid Framework for Gray-scale Image Recovery

Hyein Park<sup>1</sup>, Sinhwa Lee<sup>1</sup>, Seunghyo Choi<sup>1</sup>, Dahyun Eom<sup>1</sup>, Jeongwon Ji<sup>1</sup>, Junhwa Kim<sup>2</sup>

Department of Medical Artificial Intelligence, Konyang University<sup>1</sup>

Department of Artificial Intelligence, Konyang University<sup>2</sup>

### 요약

본 연구는 손상된 흑백 이미지를 복원하고 색상화하는 방안을 제안한다. 이미지 복원은 Stable Diffusion을 활용하였으며, 손상된 부분을 마스크 기반으로 복원하였다. 색상화는 GAN을 활용하여 진행하였다. 프롬프트 엔지니어링의 중요성과 복원 단계 품질이 색상화 성능에 미치는 영향을 연구하였으며, 복원-색상화 단계의 상호 의존성을 확인하였다. 본 연구는 손상된 흑백 이미지의 복원과 색상화를 통해 디지털 보존 및 인공지능 기반 이미지 처리 기술의 새로운 가능성을 제시한다.

### Abstract

This study proposes a method for restoring and colorizing damaged grayscale images. Restoration uses Stable Diffusion with mask-based inpainting, while colorization employs a GAN. We investigate the importance of prompt engineering and how restoration quality affects colorization performance, confirming interdependencies between the two stages. Through high-quality restoration and colorization, this approach offers new possibilities for digital preservation and AI-based image processing.

### 1. 서론

손상된 이미지 복원은 다양한 분야에서 필수적인 기술로 자리 잡고 있다. 예를 들어 과거의 흑백 사진이나 역사적 가치가 높은 오래된 사진 복원의 경우, 개인적, 문화적, 역사적 가치를 가지고 있다. 하지만 이는 시간이 흐르면서 변색되거나 훼손되는 문제가 발생하기에 복원이 요구된다 [1].

또한, 문화재 보존 분야에서도 유화나 벽화와 같은 예술 작품이 시간이 지나면서 물리적 마모, 환경적 손상(습기, 온도 변화), 혹은 인간에 의한 훼손으로

인해 손상될 수 있다 [2]. 이러한 작품들을 디지털로 변환한 후 손실된 영역을 마스크 처리하여 복원함으로써 기존 가치를 회복할 수 있다.

이미지의 특정 영역이 훼손되거나 손실된 경우 이를 복원하는 기술은 단순히 시각적 정보를 되찾는 것 이상으로, 인류의 문화적 자산을 보존하고, 현대 기술 사회의 데이터 신뢰성을 유지하며, 과거와 현재를 연결하는 다리 역할을 할 수 있다. 이러한 이유로 손상된 이미지 복원은 디지털 기술 발전의 중요한 연구 분야 [3]로 주목받고 있다.

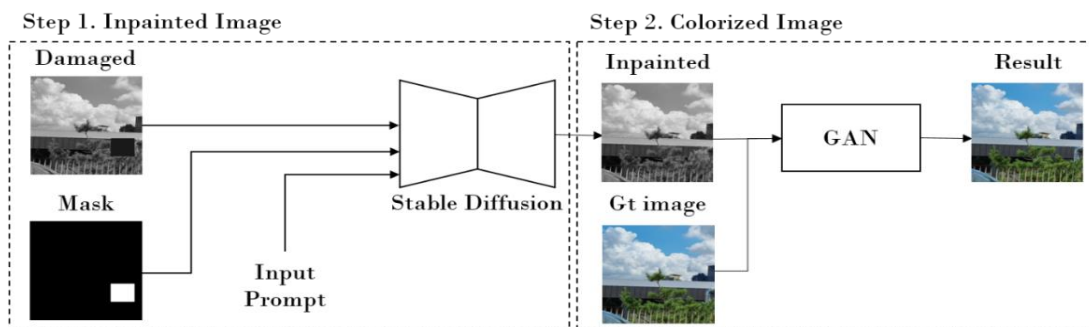


그림 1. 전체 모델 구조도

본 연구에서 제안하는 방법의 전체 구조도는 그림 1과 같다. 손상된 부분을 복원하기 위해, 텍스트 프롬프트를 입력 받는 Stable Diffusion 모델 [4]을 사용하여 마스크 기반의 인페인팅(Inpainting)을 수행하였다. 다음으로, 흑백 이미지의 색상화는 GAN [5] 기반을 활용하여 진행하였다.

그림 2은 본 연구에서 사용된 데이터셋의 구성을 나타낸다. 학습 단계에서 사용된 Train input은 손상된 흑백 이미지를 나타내며, Train gt는 해당 이미지의 Ground Truth로 원본 컬러 이미지를 제공한다. Train mask는 손상된 영역을 나타내는 마스크 이미지이다. Train mask는 손상된 흑백 이미지와 원본 이미지를 비교하여 그 차이를 마스크로 생성하였다. 테스트 단계에서도 유사한 구조가 적용되며, Test input은 손상된 흑백 이미지를, Test mask는 복원해야 할 영역을 나타낸다. Test Mask는 원본 데이터가 존재하지 않으므로 수작업으로 마스크를 지정해 주었다.

Train input	Train gt	Train mask	Test input	Test mask

그림 2. 데이터셋

## 2. 본론

### 2.1. Stable Diffusion 기반 이미지 복원

이미지 복원을 위해 Stable Diffusion 기술을 활용하였다. Stable Diffusion [4]은 프롬프트(prompt)에 따라 결과물이 달라질 수 있으므로, 본 연구에서는 Mask 영역을 채우기 위해 세 가지 프롬프트를 작성하여 실험을 진행하였다. 각 프롬프트는 표 1과 같다.

표 1. Stable Diffusion 명령어

Name	Prompt
Prompt 1	"Fill vacant space"
Prompt 2	"Please fill the vacant space by considering the pixels around the mask."
Prompt 3	"Please fill the vacant space by considering the pixels around the mask. Please exclude any text or letters in vacant space."

Prompt 1은 빈 공간을 채우라는 기본 명령어를 기반으로 작성하였으며, Prompt 2는 주변 픽셀의 맥락을 고려하라는 조건을 추가하였다. Prompt 3에서는 더욱 구체화하여 빈 공간 내 텍스트 생성을 제외하라는 제약을 추가하였다. 프롬프트 별 결과는

그림 3과 같다.

	Gt	Input	Prompt 1	Prompt 2	Prompt 3
Train					
Val					

그림 3. Stable Diffusion 복원 결과 - 1

프롬프트 별 Stable Diffusion 결과물에 대한 이미지 품질 평가는 원본을 흑백으로 변환한 이미지와 결과물 간의 MSE(Mean Square Error)와 SSIM(Structural Similarity Index)을 각각 계산하고 평균 내어 점수를 통해 평가를 진행하였다. MSE는 두 이미지 간의 픽셀 차이를 제공하여 평균화한 값이고, SSIM은 밝기( $l$ ), 대비( $c$ ), 구조( $s$ )를 종합적으로 고려해 두 이미지 간의 구조적 유사성을 평가하는 지표이다. MSE는 값이 작을수록, SSIM은 값이 1에 가까울수록 유사도가 높다. MSE와 SSIM의 수식은 아래와 같다.

$$MSE = \frac{\sum_{M,N} [I_1(m,n) - I_2(m,n)]^2}{M \times N} \quad (1)$$

$$SSIM = [l(x,y)]^\alpha \cdot [c(x,y)]^\beta \cdot [s(x,y)]^\gamma \quad (2)$$

### 2.2. GAN 기반 이미지 색상화

Mask 기반 이미지 복원 결과 이미지를 사용하여 색상화를 진행한다. 색상화를 위한 모델은 GAN [5]을 활용하였으며 구조는 그림 4와 같다.

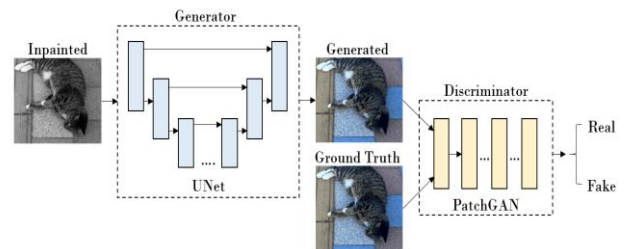


그림 4. 이미지 색상화 모델 구조

Generator는 UNet 구조 [6]를 기반으로 설계되었으며, Encoder-Decoder와 Skip Connection을 통해 복원의 세부 정보를 효과적으로 전달한다. Discriminator는 PatchGAN 구조 [7]를 기반으로 하여, 이미지를 패치 단위로 평가해 진짜/가짜 여부를 판별한다. 그림 5는 프롬프트 별

복원된 이미지를 모델에 적용한 색상화 결과이다.

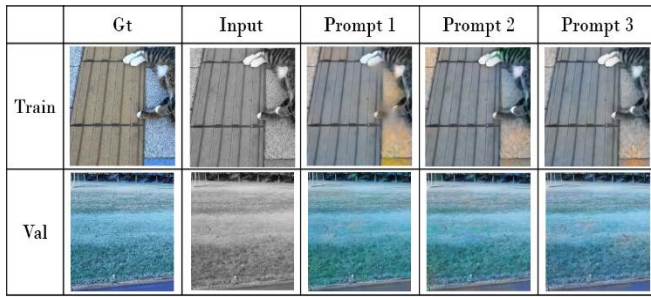


그림 5. GAN 색상화 결과 - I

### 3. 실험 결과

#### 3.1 데이터 세트 소개

데이콘(Daicon)에서 진행된 대회인 “이미지 색상화 및 손실 부분 복원 AI 경진대회”에서 제공한 데이터셋 [8]을 활용하였다. 데이터는 손상된 흑백 이미지, 원본 이미지, 테스트를 위한 손상된 흑백 이미지로 구성되어 있으며, 학습 셋은 29604장이고, 테스트 셋은 100장으로 되어있다. 이를, 학습 데이터셋과 테스트 셋에 대해 Mask를 추출하고, 학습 셋을 8:2 비율로 학습 데이터셋과 검증 데이터셋으로 나누어 실험을 진행하였다.

#### 3.2 프롬프트에 따른 이미지 복원 비교

본 연구에서는 Prompt 1, Prompt 2, Prompt 3의 세 가지 프롬프트를 사용하여 동일한 입력 이미지에 대한 복원 결과를 비교하였다. Stable Diffusion [4]에서 복원된 결과물은 MSE와 SSIM에서 사용된 프롬프트에 따른 차이를 보임을 확인했다.

표 2는 학습 데이터셋과 검증 데이터셋의 복원 결과를 나타낸다.

표 2. 학습 및 검증 데이터셋 복원 결과

Train Set		
Name	MSE	SSIM
Prompt 1	61.0180	<b>0.6440</b>
Prompt 2	61.0889	0.6427
Prompt 3	<b>60.6899</b>	0.6429
Validation Set		
Name	MSE	SSIM
Prompt 1	61.5680	<b>0.6475</b>
Prompt 2	61.5744	0.6466
Prompt 3	<b>61.3046</b>	0.6466

실험 결과, 학습 데이터셋에서 Prompt 3의 MSE는 60.6899로 가장 낮고, SSIM은 0.6429로 세 프롬프트 간의 차이는 소수점 셋째 자리 이하에서 나타나

큰 차이가 없음을 확인할 수 있었다. 검증 데이터셋에서도 Prompt 3은 MSE에서 61.3046으로 가장 낮은 값을 보이고, SSIM은 0.6466으로 다른 프롬프트와의 유사한 성능을 보였다. 이를 통해 MSE 기준으로는 Prompt 3이 가장 우수하며, SSIM 기준으로는 모든 프롬프트 간 성능 차이가 크지 않음을 확인할 수 있었다.

Prompt 1은 빈 공간을 채우라는 지시만을 포함하고 있어, 그림 6과 같이 복원된 이미지에 불필요한 글자, 아이콘, 또한 원본 이미지에 존재하지 않았던 부적합한 맥락(Contextual) 영역이 생성되는 문제가 발생하였다. 단순한 지침보다 구체적이고 세부적인 조건을 명시한 프롬프트를 사용할수록, 원본 이미지와 유사한 복원 결과를 얻을 수 있음을 확인할 수 있다.

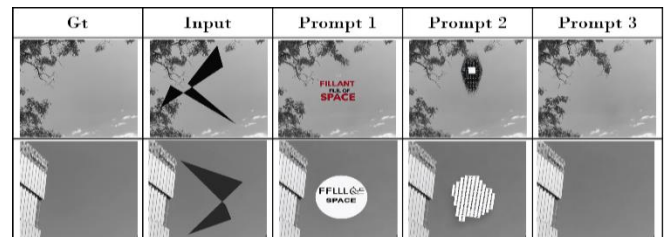


그림 6. Stable Diffusion 복원 결과 - II

#### 3.3 프롬프트에 따른 색상화 성능 비교

색상화는 GAN [5]을 기반으로 Stable Diffusion에서 복원된 프롬프트별 결과물 데이터를 사용하여 학습을 진행했다. GAN 학습을 위한 최적화 함수로는 Adam Optimizer를 사용하였으며, Generator loss는 생성된 이미지와 원본 이미지 간의 픽셀 차이(MSE Loss)를 기반으로 하였으며, Discriminator loss는 Generator에서 생성된 이미지와 원본 이미지에 대한 적대적 손실(BCELoss)의 평균으로 계산했다.

색상화된 이미지에 대한 평가는 테스트 셋을 활용하여 데이콘에 제출한 점수이며, 평가 지표는 SSIM이다. 표 3은 테스트 셋에 대한 색상화 결과이다. 세 가지 프롬프트의 SSIM 측정 결과, Prompt 3에서 0.5185로 가장 우수한 결과를 보였다. 복원 과정에서 프롬프트 별 성능은 근소한 차이를 보였지만, 이를 색상화에 적용했을 경우 성능 차이가 크게 나타나는 것을 확인할 수 있다.

표 3. Test 색상화 결과

Name	SSIM
Prompt 1	0.4844
Prompt 2	0.4952
Prompt 3	<b>0.5185</b>

이러한 결과를 얻은 이유로, 그림 7의 Prompt 1,2에

서 마스크 영역에 부적합한 맥락 영역이 더 자주 발생하여 이러한 부분이 색상화에 부정적으로 작용했을 가능성을 예상할 수 있다. 반면, Prompt 3의 결과물은 상대적으로 마스크 영역에 부적합한 맥락 영역이 적게 발생하여 색상화 성능이 상대적으로 높은 것으로 예상된다.




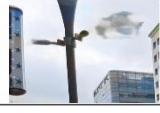


	Prompt 1	Prompt 2	Prompt 3
Input			
Colorized			

그림 7. GAN 색상화 결과 -II

Prompt 3을 사용한 테스트 데이터의 최종 복원 및 색상화 결과는 그림 8과 같다.

Input				
Result				

그림 8. 최종 복원 및 색상화 결과

#### 4. 결론 및 향후계획

본 연구에서는 손상된 흑백 이미지 복원을 위해 Stable Diffusion을 활용하여 Mask 기반으로 손상된 부분을 복원시킨 후 GAN을 활용하여 색상화를 진행하였다.

Stable Diffusion을 활용한 마스크 기반 복원에서는 프롬프트의 구체성과 제약조건이 복원 품질에 미치는 영향을 확인할 수 있었으며, 이는 프롬프트 엔지니어링이 복원 품질에 영향을 미치는 중요한 요소임을 보여준다.

복원된 이미지의 색상화를 위해 적용한 GAN 기반 모델은 UNet 구조의 Generator와 PatchGAN 구조의 Discriminator를 통해 자연스러운 색상화를 달성하였다. 특히 초기 복원 품질이 높은 이미지(Prompt 3로 처리된 이미지)가 색상화 결과에서도 더 나은 성능을 보였으며, 복원 단계의 품질이 후속 색상화 과정의 성능에도 미치는 영향을 검증하였다. 이는 복원-색상화의 두 단계가 서로 독립적이지 않으며, 초기 단계의 성능이 전체 파이프라인의 품질에 결정적인 기여를 한다는 점을 시사한다.

향후 연구에서는 Stable Diffusion과 GAN의 조합에 더해, 최신 이미지 복원 및 색상화 기술을 비교 분석하여 복원 품질을 향상시키는 방안을 모색할 계획이

다. 또한 최적의 프롬프트 설정을 위한 추가적인 연구가 필요함을 확인하였다.

#### ACKNOWLEDGEMENT

본 과제(결과물)는 2024년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다. (2021RIS-004)

본 연구는 2025년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음(2024-0-00047)

#### 참고문헌

- [1] 박찬우, 김희권, 이재호, & 지형근. 오픈소스 기반의 딥러닝을 활용한 흑백 이미지 문화유산 복원방법 연구. 대한전자공학회 학술대회, 2139-2142. (2023)
- [2] 김지수. 기후변화와 지류·직물 문화유산의 보존을 위한 영향과 과제에 관한 연구. 저널 국가유산, 8(2), 113-126. (2023).
- [3] 강상현, 이명희, and 김종기. "문화 원형의 디지털 복원과 활용 방안에 관한 연구: 상황 정보처리에 따른 Virtual Interactive Diorama 구성을 중심으로: 상황 정보처리에 따른 Virtual Interactive Diorama 구성을 중심으로." 한국디자인학회 국제학술대회 논문집 : 38-39. (2006)
- [4] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10684-10695). (2022).
- [5] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems 27 (2014).
- [6] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." Medical image computing and computer-assisted intervention-MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015.
- [7] Li, Chuan, and Michael Wand. "Precomputed real-time texture synthesis with markovian generative adversarial networks." Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14. Springer International Publishing, 2016.
- [8] "이미지 색상화 및 손실 부분 복원 AI 경진대회" (<https://dacon.io/competitions/official/236420/data>)