

# 확산 모델을 이용한 효과적인 영상 인페인팅 기법

이은솔, 이윤서, 심유현, 김종호  
순천대학교 인공지능공학부  
e-mail: jhkim@scnu.ac.kr

## Effective Image Inpainting using the Diffusion Model

Eunsol Lee, Yunseo Lee, Yuhyun Sim, Jongho Kim  
Dept. of Artificial Intelligence Engineering, Sunchon National University

### 요약

본 논문에서는 영상의 일부가 훼손되거나 삭제된 경우 시각적으로 자연스럽게 복원하는 기술인 영상 인페인팅 기법의 성능을 향상시키기 위하여 딥러닝 기반의 확산 모델을 기반으로 한 방법을 제안한다. 주로 보간 기술을 이용한 방법인 기존의 예제 기반 방법과 확산 기반 방법은 복원해야 할 영역이 클 때 복원이 부자연스럽고 에지 등 영상의 구조 정보를 효과적으로 복원하지 못하는 한계가 있었다. 이러한 문제를 개선하기 위해 오토인코더 구조와 확산 모델을 결합하여 훼손된 영상을 시각적으로 자연스럽게 복원하는 인페인팅 방법을 제안하고, 영역 기반 연산을 통해 확산 모델의 계산 복잡도를 낮추어 활용성을 향상시켰다. 다양한 영상에 적용한 결과 제안한 모델은 기존 방법에 비해 시각적으로 자연스럽게 복원된 영상을 생성하였고, 학습 및 영상 복원에 필요한 계산 복잡도 또한 실제 활용이 가능한 수준인 것으로 나타났다.

### 1. 서론

영상 인페인팅(inpainting)은 다양한 원인으로 인해 훼손된 영상 또는 배경 제거 등의 목적으로 영상의 일부분이 삭제되었을 때 시각적으로 자연스럽게 복원하는 기법이다. 영상 인페인팅 기법은 크게 두 가지 방법으로 구분할 수 있는데, 예제(exemplar) 기반 방법과 확산(diffusion) 기반 방법을 들 수 있다[1].

예제 기반 방법은 주로 텍스처 합성을 이용하는데, 주변 영역의 텍스처를 샘플링하여 훼손된 영역을 채워나가는 기법이다. 이는 영상의 텍스처 영역에서 자연스러운 결과를 보이지만, 영상의 구조 정보를 보존하는 데 한계가 있어 에지가 끊어지거나 시각적으로 부자연스러운 텍스처를 합성하는 경우가 있다. 확산 기반 방법은 주변 픽셀값에서 복원할 영역으로 영상의 구조 정보를 중심으로 확산하면서 영역을 채워나가는 방법으로 많은 종류의 영상에서 우수한 성능을 나타내지만, 복원할 영역이 크거나 복잡한 텍스처가 존재하는 경우 이를 잘 반영하기 어렵고, 확산 방법의 특징으로 인해 흐림 현상(blurring effect)이 발생하여 시각적으로 부자연스러운 영상을 만드는 단점이 있다[1].

영상에서 정보를 추출하여 훼손된 영역을 복원하는 방식

은 성능의 측면에서 주목할 만한 향상을 이룬 것이 사실이지만, 방법에 따라 뚜렷한 단점이 존재한다. 이러한 단점을 극복하기 위하여 최근 딥러닝의 생성 모델을 활용하려는 움직임이 두드러지고 있고[2], 본 논문에서는 학습의 복잡도 측면에서 뚜렷한 장점이 있는 오토인코더(autoencoder) 모델과 확산 기법의 결합을 통해 저 복잡도를 가지면서 효과적으로 훼손된 영역을 복원하는 영상 인페인팅 기법을 제안한다.

### 2. 제안하는 확산 모델을 이용한 영상 인페인팅

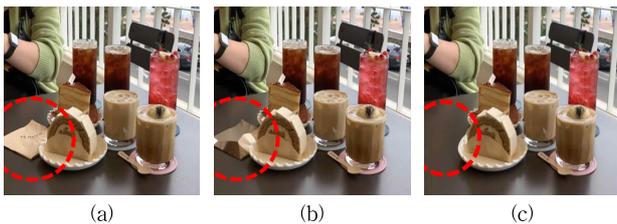
최근 영상을 활용하는 다양한 분야에서 뛰어난 성능으로 인해 CNN(convolutional neural network) 기반 및 생성 기반 딥러닝 기법이 활발하게 활용되고 있고, 영상 인페인팅 분야에서도 뛰어난 성능을 보이고 있다. 이는 다양한 영상을 이용한 학습의 결과로 획득한 특징과 패턴을 활용하여 복원하기 때문에 기존의 인페인팅 기법의 단점을 극복하고 자연스러운 영상을 복원한다. 본 논문에서는 픽셀 단위의 세밀한 복원이 가능한 확산 기반 영상 인페인팅의 성능을 향상시켜 비교적 넓은 영역 및 복잡한 형태의 텍스처의 자연스러운 복원을 위한 딥러닝 기반 모델을 제안한다.

고해상도 영상 합성에 대한 확산 모델의 학습에 필요한

계산량을 낮추기 위하여 손실 조건을 샘플링하여 지각적으로 관련 없는 세부 사항을 무시할 수 있지만, 여전히 픽셀 영역에서의 함수 평가는 비용이 많이 필요하므로 본 논문에서는 생성 학습 단계에서 압축을 명시적으로 분리하여 이러한 단점을 피하는 구조를 제안한다. 이를 달성하기 위해 영상 공간과 지각적으로 동일한 공간을 학습하지만 계산 복잡도를 크게 줄이는 오토인코더 모델을 활용[3]한다. 이러한 구조는 다음과 같은 장점이 있다. (i) 샘플링이 저차원 공간에서 수행되기 때문에 계산 복잡도가 훨씬 줄어든 확산 모델을 얻을 수 있다. (ii) 공간 구조 정보가 없는 데이터에 특히 효과적인 UNet 구조를 상속한 확산 모델을 활용하기 때문에 화질을 크게 낮출 수 있는 강한 압축에 대한 필요성을 줄일 수 있다. (iii) 잠재 공간이 다수의 생성 모델을 학습시키고 다운스트림 응용에도 활용할 수 있는 범용 압축 모델을 얻을 수 있다.

### 3. 실험 결과

기존의 영상 인페인팅 기법은 Telea 방법[4]을 구현하여 비교하였고, 제안하는 기법에 의한 방법은 실제 촬영한 영상에 적용하여 결과를 비교하여 [그림 1]에 나타내었다.



[그림 1] 영상 인페인팅 결과 (a) 원본 영상, (b) Telea 방법, (c) 제안하는 방법

#### 감사의 글

본 논문은 2022년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력 기반 지역혁신 사업의 결과임 (2021IRIS-002)

#### 참고문헌

- [1] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever, “Zero-shot text-to-image generation,” *CoRR*, abs/2102.12092, 2021
- [2] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Ward-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio, “Generative adversarial

networks,” *CoRR*, 2014

- [3] Arash Vahdat, Karsten Kreis, and Jan Kautz, “Score-based generative modeling in latent space,” *CoRR*, abs/2106.05931, 2021
- [4] A. Telea, “An image inpainting technique based on the fast matching method,” *J. Graphics Tools*, vol. 9, Iss. 1, pp. 23-34, 2004