특집논문 (Special Paper)
방송공학회논문지 제28권 제4호, 2023년 7월 (JBE Vol.28, No.4, July 2023)
https://doi.org/10.5909/JBE.2023.28.4.382
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

다중 스케일 확산 모델을 이용한 영상 흐려짐 복원 방법

윤 천 희[a], 김 원 준[b]‡

Image Deblurring Based on Multi-Scale Diffusion Models

Cheonhee Yun[a] and Wonjun Kim[b]‡

요 약

최근 컴퓨터 비전 분야에서 확산 모델(Diffusion Model)은 다양한 작업에 적용되어 뛰어난 성능을 보여주고 있다. 확산 모델은 마르코프 연쇄(Markov Chain)의 성질을 사용한 확률 기반의 학습 방법으로 안정적인 성능을 보이나, 확산 모델의 역방향 과정에서 널리 사용되는 U자형 신경망 구조(U-Net)는 정밀한 잡음 예측을 위해 많은 수의 계층과 채널을 사용한다. 본 논문에서는 이러한 문제점을 극복하기 위해 적은 수의 계층과 채널을 사용하는 다중 스케일 입출력 기반 신경망을 활용한 확산 모델을 제안하고 이를 영상 흐려짐 복원 작업에 적용한다. 제안하는 방법은 다중 스케일의 입력을 통해 잠재 특징 간 복합 관계를 학습하고, 다중 스케일의 출력을 사용하여 학습한 확산 모델을 통해 잡음을 예측한다. 실험 결과를 통해 제안하는 방법이 기존 신경망 기반 방법과 비교하여 적은 연산량에도 성능을 효과적으로 향상시킬 수 있음을 보인다.

Abstract

Recently, in the field of computer vision, diffusion models have been applied to various tasks and has shown promising performance. Diffusion models perform reliably as probability-based learning methods using the properties of Markov Chain, however U-Net, which is widely used in the reverse process of diffusion models, uses a large number of layers and channels for precise noise prediction. To overcome this problem, in this paper, we propose diffusion models using a multi-scale input-output based neural network using a small number of layers and channels and apply it to image deblurring. The proposed method learns complex relationships between latent features through multi-scale inputs and predicts noise through diffusion models learned using multi-scale outputs. Experimental results show that the proposed method can effectively improve performance even with less computation compared to existing neural network based methods.

Keyword : Diffusion Model, Image Deblurring, Deep Neural Network

a) 건국대학교 전자·정보통신공학과(Department of Electronics, Information & Communication Engineering, Konkuk University)
b) 건국대학교 전기전자공학부(Department of Electrical and Electronics Engineering, Konkuk University)
‡ Corresponding Author : 김원준(Wonjun Kim)
E-mail: wonjkim@konkuk.ac.kr
Tel: +82-2-450-3396
ORCID: https://orcid.org/0000-0001-5121-5931
※ 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2023R1A2C1003699).
※ 이 논문은 2023학년도 건국대학교의 연구년교원 지원에 의하여 연구되었음.
※ This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2023R1A2C1003699).
※ This paper was written as part of Konkuk University's research support program for its faculty on sabbatical leave in 2023.
· Manuscript May 31, 2023; Revised July 21, 2023; Accepted July 21, 2023.

Ⅰ. 서 론

최근 컴퓨터 비전 분야에서 생성 모델을 기반으로 한 다양한 연구가 진행되고 있다. 대표적인 생성 모델인 적대적 생성 신경망(Generative Adversarial Network, GAN)[1]은 생성자와 판별자로 구성된 두 개의 신경망이 최소-최대 게임(Min-Max Game)을 통해 학습되는 모델이다. 이와 같은 방법은 실제와 구별하기 어려운 고품질의 데이터를 생성하는 데 효과적이나 경쟁적인 학습으로 인한 불안정성 및 생성자가 판별자를 속이는 방향으로만 학습하는 모드 붕괴(Mode Collapse)가 종종 발생하는 문제가 있다.

확산 모델(Diffusion Model)[2]은 마르코프 연쇄(Markov Chain)의 성질을 사용한 확률 기반의 학습 방법으로 기존 GAN 모델의 문제점을 해결한다. 자세히 살펴보면, 확산 모델은 입력 데이터에 마르코프 연쇄 성질을 적용하여 단계적으로 가우스 잡음(Gaussian Noise)을 주입한다. 이 과정을 통해 데이터에 점점 더 많은 잡음이 주입되어 최종적으로 데이터는 가우스 잡음으로 변한다. 확산 모델은 가우스 잡음에서 원본 데이터로 복원하는 역방향 과정(Reverse Process)을 심층 신경망(Deep Neural Network)을 통해 학습한다. 학습된 모델은 잡음이 주어졌을 때 역방향 과정을 거쳐 원하는 데이터를 생성할 수 있다. 기존 확산 모델은 역방향 과정의 각 시간 단계(Time Step)에서 잡음 제거를 위해 쿨백-라이블러 발산(Kullback-Leibler Divergence)을 사용하였다. 그러나, 쿨백-라이블러 발산 방식으로는 각 단계에서의 데이터 분포의 작은 변화를 정확하게 학습하는 데에 어려움이 있어 고품질 영상을 생성하지 못하였다. 이를 보완하기 위해 잡음 제거 확산 확률 모델(Denoising Diffusion Probabilistic Model, DDPM)[3]이 제안되었다. 기존의 확산 모델과 달리, DDPM은 데이터 분포 변화량에 관계없이 어떤 단계에서든지 잡음을 제거하여 원본 데이터를 예측할 수 있도록 설계되었다.

DDPM의 성능에 힘입어, 많은 연구자들은 다양한 분야에 이를 적용하고 있으며, 그 성과를 통해 확산 모델의 유용성이 증명되고 있다. Baranchuk[4] 등은 의미론적 분할(Semantic Segmentation) 분야에서 확산 모델이 픽셀 레벨의 의미론적 특징 추출에 효과적임을 보였다. 이러한 방법은 기존에 사용되었던 GAN 또는 변이형 자동 압축기(Variational Autoencoder, VAE) 기반 모델보다 더 뛰어난 성능을 보였다. Wyatt[5] 등은 이상 탐지(Anomaly Detection) 분야에서 입력 영상을 일부 훼손하고, 그 손상을 복구하는 과정에서 확산 모델을 활용하는 새로운 이상 탐지 방법을 제시하였다. Saharia[6] 등은 조건부 확산 모델(Conditional Diffusion Model)을 기반으로 영상 변환을 위한 통합 프레임워크를 개발하였다. 이 프레임워크는 영상 채색(Image Colorization), 자른 영상 복원(Image Uncropping), 영상 인페인팅(Image Inpainting), 그리고 JPEG 영상 복원(JPEG Image Restoration) 등에 활용되어 성능을 향상시켰다. Li[7] 등은 영상 초해상도(Image Super-Resolution) 분야에서 조건부 확산 모델을 기반으로 저해상도 영상 압축기를 사용하여, 효과적으로 성능을 향상시켰다. 하지만 DDPM의 역방향 과정을 학습하기 위한 심층 신경망으로 U자형 신경망 구조(U-Net)[8]가 일반적으로 사용되는데, 잡음 예측 과정을 정밀하게 학습하기 위해 많은 수의 계층과 채널이 요구된다. 또한, 특징 간 전역적 관계를 고려하기 위해 자기주의(Self-Attention) 계층[9]을 사용하기도 하는데 이로 인해 매

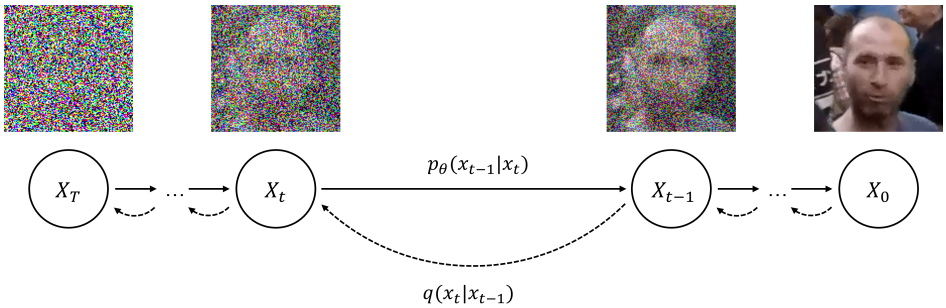

그림 1. 확산 모델의 확산 과정(실선 화살표)과 역방향 과정(점선 화살표)의 개요
Fig. 1. Overview of the forward process (solid arrow) and reverse process (dotted arrow) of the diffusion models

개변수의 수가 크게 증가한다.

이러한 문제를 해결하기 위해 본 논문에서는 적은 수의 계층과 채널을 사용하여 다중 스케일에서 추출된 잠재 특징 간 복합 관계를 활용한 다중 스케일 확산 모델을 제안한다. 제안하는 방법은 확산 모델의 역방향 과정을 학습할 때, MIMO-UNet[10] 구조의 다중 스케일의 입력을 통해 잠재 특징 간 복합 관계를 학습하고, 다중 스케일의 출력을 사용한 손실 함수를 통해 잡음을 예측한다.

제안하는 구조를 영상 흐려짐 복원(Image Deblurring)에 적용하여 기존 신경망 구조 대비 제안하는 방법이 적은 연산량으로 효과적으로 성능 개선할 수 있음을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 제안하는 다중 스케일 확산 모델에 대해 자세히 설명하며, 3장에서는 실험을 통해 제안하는 방법의 성능을 검증한다. 마지막으로 4장에서는 본 논문의 결론을 서술한다.

Ⅱ. 제안하는 방법

제안하는 방법은 기존 확산 모델에서 사용되는 U자형 신경망 구조를 개량하여 영상 흐려짐 복원의 계산 비용을 효과적으로 줄이고 성능을 개선하고자 한다. 본 장에서는 먼저 잡음 제거 확산 확률 모델의 학습 과정에 대해 설명한다. 이어서 제안하는 신경망 구조를 자세히 설명한 후 마지막으로 영상 흐려짐 복원을 위한 확산 모델의 학습 과정과 추론 과정에 관해 설명한다.

1. 잡음 제거 확산 확률 모델(Denoising Diffusion Probabilistic Model)의 학습 과정

확산 모델은 변형 추론(Variational Inference)을 사용하여 마르코프 연쇄를 통해 단순한 분포의 잠재 변수(Latent Variable) x_T로부터 복잡한 분포의 데이터 x_0를 단계적으로 생성하는 모델이다. 여기서 T는 확산 단계의 총 개수이며, 각 확산 시간 단계 $t\in\{1,2,...,\mathrm{T}\}$의 결과로 $x_t\in\mathrm{R}^{\mathrm{d}}$를 설정하고, x_0는 x_t와 동일한 차원 d를 갖는다. 그림 1에서 볼 수 있듯이 확산 모델은 확산 과정과 역방향 과정의 두 가지 과정으로 구성된다.

확산 과정에서 사후 확률(Posterior) $q(x_1,...,x_T \mid x_0)$는 분산 스케줄 $\beta_1,...,\beta_T$에 따라 데이터에 단계적으로 가우스 잡음(Gaussian Noise) ϵ을 추가하는 마르코프 연쇄에 아래와 같이 계산된다.

$$\begin{aligned} q(x_1,...,x_T|x_0) &:= \prod_{t=1}^{T} q(x_t|x_{t-1}), \\ q(x_t|x_{t-1}) &:= N(x_t;\sqrt{1-\beta_t}\,x_{t-1},\beta_t \mathrm{I}), \end{aligned} \tag{1}$$

여기서 I는 단위행렬이고, N은 가우스 분포(Gaussian Distribution)를 나타내며, β_t는 초매개변수(Hyperparameter)인 작은 양의 상수이다. $\alpha_t := 1-\beta_t, \overline{\alpha}_t := \prod_{s=1}^{t}\alpha_s$로 설정하면 다음 수식과 같이 확산 과정을 통해 임의의 시간 간격(time step) t에서 x_t를 샘플링(Sampling)할 수 있다.

$$q(x_t|x_0) = N(x_t;\sqrt{\overline{\alpha}_t}\,x_0,(1-\overline{\alpha}_t)\mathrm{I}). \tag{2}$$

이 수식은 다음과 같이 매개변수를 재조정할 수 있다.

$$x_t(x_0,\epsilon) = \sqrt{\overline{\alpha}_t}\,x_{0+}\sqrt{1-\overline{\alpha}_t}\,\epsilon, \epsilon \sim N(0,\mathrm{I}), \tag{3}$$

여기서 0은 영행렬이다. 역방향 과정은 잠재 변수 분포 θ로 매개변수화된 $p_\theta(x_T)$를 데이터 분포 $p_\theta(x_0)$로 변환한다. 이는 학습된 가우스 전이(Gaussian Transition)를 포함하는 마르코프 연쇄로 정의되며, $p(x_T) = N(x_T;0,\mathrm{I})$로부터 시작된다. 과정은 다음 수식과 같다.

$$\begin{aligned} p_\theta(x_0,...,x_{T-1}|x_T) &:= \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t), \\ p_\theta(x_{t-1}|x_t) &:= N(x_{t-1};\mu_\theta(x_t,t),\sigma_\theta^2(x_t,t)\mathrm{I}), \end{aligned} \tag{4}$$

여기서 $\mu_\theta(x_t,t)$는 t 역방향 단계의 가우스 분포의 평균이고, $\sigma_\theta^2(x_t,t)$는 t 역방향 단계의 가우스 분포의 분산이다.

학습 단계에서는 로그 우도(log likelihood) $\log p_\theta(x_0)$에 대한 변동 하한(Variational Lower Bound)을 최대화하고 쿨백-라이블러 발산 및 분산 감소를 사용하며 그 과정은 다음과 같다.

$$
\begin{aligned}
\mathrm{E}[\log p_\theta(x_0) &\geq \mathrm{E}_q[\log \frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}] \\
&= \mathrm{E}_q[\log p(x_T) + \sum_{t \geq 1} \log \frac{p_\theta(x_{t-1} \mid x_t)}{q(x_t \mid x_{t-1})}] \\
&= \mathrm{E}_q[\log p \frac{(x_T)}{q(x_T \mid x_0)} + \sum_{t>1} \log \frac{p_\theta(x_{t-1} \mid x_t)}{q(x_{t-1} \mid x_t, x_0)} + \log p_\theta(x_0 \mid x_1)] \\
&= \mathrm{E}_q[\underbrace{D_{KL}(q(x_T \mid x_0) \parallel p(x_T))}_{L_T} \\
&+ \sum_{t>1} \underbrace{D_{KL}(q(x_{t-1} \mid x_t, x_0) \parallel p_\theta(x_{t-1} \mid x_t))}_{L_{t-1}} \underbrace{-\log p_\theta(x_0 \mid x_1)}_{L_0}].
\end{aligned} \tag{5}
$$

수식 (5)를 계산하기 위해서 쿨백-라이블러 발산을 사용하여 $p_\theta(x_{t-1}|x_t)$와 해당 확산 과정의 사후 확률 분포 사이의 차이를 직접적으로 측정한다. 여기서 D_{KL}은 쿨백-라이블러 발산이고 L_T는 정규화 과정이고 L_{t-1}는 잡음 제거 과정이고 L_0은 복원 과정이다. 다음 수식과 같이 설정하면,

$$
\begin{aligned}
\tilde{\mu}_t(x_t, x_0) &:= \frac{\sqrt{\overline{\alpha}_{t-1}}\beta_t}{1-\overline{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t} x_t, \\
\tilde{\beta}_t &:= \frac{1-\overline{\alpha}_{t-1}}{1-\overline{\alpha}_t} \beta_t,
\end{aligned} \tag{6}
$$

사후 확률은 다음 수식과 같다.

$$
q(x_{t-1}|x_t, x_0) = N(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t \mathrm{I}) \quad , \tag{7}
$$

여기서 $\tilde{\mu}_t(x_t, x_0)$는 t 확산 단계에서 가우스 분포의 평균이고, $\tilde{\beta}_t$는 t 확산 단계에서 가우스 분포의 분산이다. 수식 (2), (4), (7)은 수식 (5)의 쿨백-라이블러 발산이 가우스 분포 간의 비교임을 보여준다. $t>1$, $\tilde{\beta}_1 = \beta_1$에 대한 $\sigma_t^2 = \tilde{\beta}_t = \frac{1-\overline{\alpha}_{t-1}}{1-\overline{\alpha}_t}\beta_t$와 상수 C를 사용하면 다음 수식과 같다.

$$
L_{t-1} = \mathrm{E}_q[\frac{1}{2\sigma_t^2} \parallel \tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t) \parallel^2] + C. \tag{8}
$$

손실 함수(Loss Function)를 간단히 하기 위해, 학습 과정에서 x_0와 t를 입력으로 사용하여 변동 하한의 변형을 최소화하는 다음 수식을 사용한다.

$$
\min_\theta L_{t-1}(\theta) = \mathrm{E}_{\mathrm{x}_0, \epsilon, \mathrm{t}}[\|\epsilon - \epsilon_\theta(\sqrt{\overline{\alpha}_\mathrm{t}}\,\mathrm{x}_0 + \sqrt{1-\overline{\alpha}_\mathrm{t}}\,\epsilon, \mathrm{t})\|^2], \tag{9}
$$

여기서 ϵ_θ는 잡음 예측 모델이다. 추론 과정에서 먼저 $x_T \sim N(x_T; 0, \mathrm{I})$를 샘플링한 다음, 수식 (4)에 따라 $x_{t-1} \sim p_\theta(x_{t-1}|x_t)$를 샘플링한다. x_{t-1}은 다음 수식과 같이 매개변수화 될 수 있다.

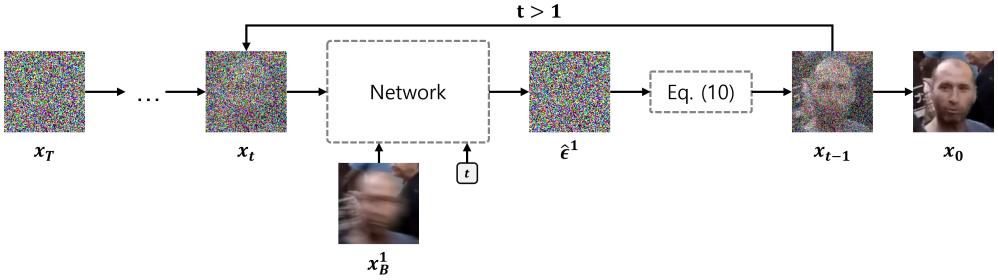

그림 2. 영상 흐려짐 복원을 위한 확산 모델의 추론 과정
Fig. 2. The inference procedure for diffusion models for image deblurring

$$
\begin{aligned}
&x_{t-1}(x_t,t)\\
&=\mu_\theta(x_t,t)+\sigma_\theta(x_t,t)^2 z\\
&=\frac{1}{\sqrt{\alpha_t}}(x_t-\frac{\beta_t}{\sqrt{1-\overline{\alpha_t}}}\epsilon_\theta(x_t,t))\widetilde{\beta}_t z, z\sim N(0,\mathrm{I})
\end{aligned}
\tag{10}
$$

2. 확산 모델을 통한 영상 흐려짐 복원 과정 및 다중 스케일 입출력 기반 U자형 신경망 구조

확산 모델을 이용한 영상 흐려짐 복원 방법은 확산 과정과 역방향 과정의 두 가지 과정을 포함하는 T 단계의 확산 모델을 기반으로 한다. 확산 과정은 수식 (3)에서 알 수 있듯이 가우스 잡음 ϵ을 점진적으로 추가하여 x_0를 가우스 분포 기반 잠재 변수 x_t로 변환한다. 역방향 과정은 조건부 잡음 예측 모델 ϵ_θ를 사용하여 T 단계의 반복적인 잡음 제거를 통해 흐려짐을 복원한 영상(x_0)을 생성한다(그림 2 참조).

조건부 잡음 예측 모델 ϵ_θ의 목표는 수식 (9)에 따라 흐려진 영상의 정보를 통해 각 확산 과정의 시간 간격에 추가되는 잡음 ϵ을 예측하는 것이다. 그림 3은 MIMO-UNet[10]을 기반으로 다중 스케일의 입출력을 활용하여 조건부 잡음을 예측하는 신경망의 구조이다. 세 가지 스케일의 x_t, 확산 시간 간격 $t\in\{1,2,...,\mathrm{T}\}$, 조건부 영상 압축기(Conditional Image Encoder, CIE)의 출력을 입력으로 하여 구성된다. 그림 3의 제안하는 구조를 자세히 살펴보면, 다중 스케일의 각 x_t가 하나의 2차원 합성곱 계층을 통과한 후, 다중 스케일의 흐려진 입력 영상이 CIE를 통과한 출력과 채널 방향으로 결합(Concatenation)된다. 다중 스케일 입력 활용 모듈인 SCM, FAM, AFF는 기존 MIMO-UNet의 모듈을 그대로 사용했으며, CIE는 ResBlock[11] 2개를 이용하여 간단하게 구축하였다. 모델에서의 시간 간격은 트랜스포머(Transformer)[12]에서 제안한 위치 부호화(Positional Encoding)를 사용하여 시간 간격 t를 시간 임베딩 값으로 변환하여 모든 ResBlock에 입력으로 더해지는 방식으로 계산된다.

3. 학습 과정 및 추론 과정

학습 과정은 다음과 같은 순서대로 진행된다. 학습 데이터셋의 배치 쌍(P)에서 흐려진 영상의 배치(x_B)와 선명한 영상의 배치(x_0)를 샘플링한다. 정수 집합 $\{1,...,\mathrm{T}\}$ 에서

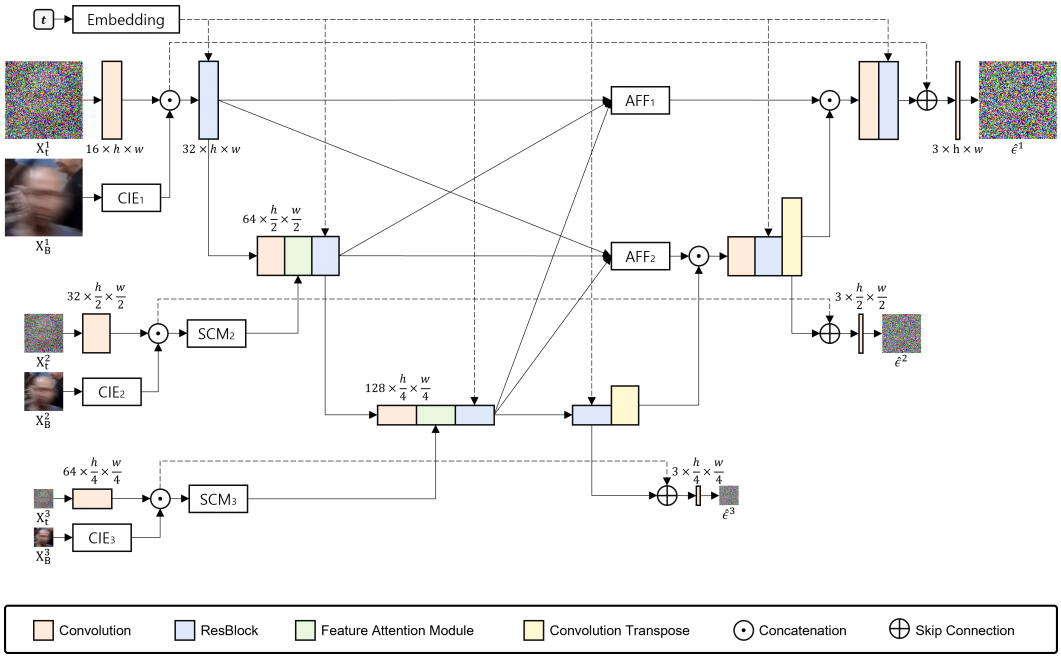

그림 3. MIMO-UNet을 활용한 다중 스케일 확산 모델의 자세한 신경망 구조
Fig. 3. Detailed neural network architecture of a multi-scale diffusion model utilizing MIMO-UNet

t를 구하고, 표준 가우스 분포에서 ϵ^1을 샘플링한다. 그런 다음 ϵ^1을 $\frac{1}{2}$ 다운샘플링하여 ϵ^2을 만들고, ϵ^2을 $\frac{1}{2}$ 다운샘플링하여 ϵ^3을 만든다. 수식 (3)에 의해 구한 x_t와 t, x_B를 다중 스케일의 잡음을 예측하는 모델 ϵ_θ에 입력으로 주어 다중 스케일의 잡음 $\hat{\epsilon}^1$, $\hat{\epsilon}^2$, $\hat{\epsilon}^3$을 예측한다. L_{cont}를 계산하기 위해 평균 절대 오차(Mean Absolute Error, MAE)를 사용하였으며 다음과 같이 계산할 수 있다.

$$L_{cont} = \sum_{k=1}^{3} \frac{1}{t_k} \parallel \hat{\epsilon}^k - \epsilon^k \parallel , \tag{11}$$

여기서 t_k는 전체 픽셀 수를 의미한다. 또한, 주파수 공간의 차이를 줄이기 위해 다중 스케일 주파수 재구성(Multi-scale Frequency Reconstruction) 손실 함수를 사용하며 다음과 같이 계산할 수 있다.

$$L_{MSFR} = \sum_{k=1}^{3} \frac{1}{t_k} \| F(\hat{\epsilon}^k) - F(\epsilon^k) \|, \tag{12}$$

여기서 F는 고속 푸리에 변환(Fast Fourier Transform, FFT)을 나타낸다. 신경망 학습을 위한 최종 손실 함수는 다음과 같이 정의된다.

$$L_{total} = L_{cont} + \lambda L_{MSFR}, \tag{13}$$

여기서 λ는 상수를 나타내며, MIMO-UNet에서 사용된 0.1로 설정하였다.

추론 과정은 다음과 같은 순서대로 진행된다. 흐려진 영상 x_B를 입력으로 받는다. 표준 가우스 분포에서 잠재 변수 x_T를 샘플링한다. 반복은 $t=T$ 단계에서 시작되어 $t=1$단계에서 마친다. 각 반복에서 잡음 $\hat{\epsilon}^1$을 예측하고 t가 감소함에 따라 매 단계에서 다음과 같은 수식으로 모델(ϵ_θ)에 x_t, t, x_B를 입력으로 주어 x_{t-1}을 추론할 수 있다.

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t, x_B)) + \sigma_t z, \tag{14}$$

$t=1$일 때, $z=0$으로 설정하고 최종적으로 흐려짐을 제거한 영상(x_0)을 출력한다.

Ⅲ. 실험 결과 및 분석

본 논문에서는 제안하는 방법의 성능 평가를 위해 두 개의 벤치마크 데이터셋을 사용하였다. 먼저, 영상 흐려짐 복원(Image Deblurring) 분야에서 가장 대표적으로 사용되는 GoPro 데이터셋[13]을 사용하여 성능을 평가하였다. GoPro 데이터셋은 GOPRO4 Hero Black 카메라를 사용하여 240fps의 동영상을 촬영한 다음 연속 프레임의 평균화를 통해 흐려진 영상을 생성하였다. GoPro 데이터셋은 총 2,103개의 흐려진 영상과 선명한 영상 쌍의 학습 데이터셋과 총 1,111개의 영상 쌍의 테스트 데이터셋으로 구성되어 있다. 다음으로 성능 평가에 사용한 HIDE 데이터셋[14]은 GoPro 데이터셋과 같은 방법으로 촬영되었으며, 영상 속 사람의 근접 촬영 여부와 군집 여부에 대한 정보가 포함되어 있다. HIDE 데이터셋[14]은 총 6,397개의 영상 쌍의 학습 데이터셋과 총 2,025개의 영상 쌍의 테스트 데이터셋으로 구성되어 있다. 본 논문에서는 제안하는 다중 스케일 입출력 기반 U자형 신경망 구조의 영상 흐려짐 복원을 위한 확산 모델에서의 효율성과 성능 향상을 비교하기 위해 ViT[9]의 자기주의 계층이 존재하는 U자형 신경망 구조[8]를 사용한 확산 모델과 성능 비교를 수행하였다. 제안하는 방법의 성능 개선 효과를 검증하기 위해 GoPro 데이터셋과 HIDE 데이터셋의 테스트 데이터셋 중, 총 11개의 장면에서 1개씩을 선정하여 11개의 영상 쌍의 검증(Validation) 데이터셋을 사용하였다.

자기주의 계층을 포함한 U자형 신경망 구조를 사용한 확산 모델과 제안하는 다중 스케일 입출력 기반 신경망을 사용한 확산 모델의 정성적인 영상 흐려짐 복원 결과를 그림 4와 5에 나타내었다. 그림에서 볼 수 있듯이, 제안하는 방법을 적용했을 때 흐려진 부분이 더 선명하게 복원된다. 특히 그림 4에서 흐려짐이 강하게 발생한 사람의 얼굴이 더 효과적으로 복원되는 것을 확인할 수 있다. 다음으로는 영상 흐려짐 복원 성능의 정량적 평가에 널리 사용되는 최대 신호대 잡음비(Peak Signal-to-Noise Ration, PSNR)와 구조적유사도(Structural Similarity Index Measure, SSIM)를 활용하여 성능 비교를 수행하였으며, 해당 결과를 표 1과 2에 나타내었다. 표 1과 2의 결과에서 볼 수 있듯이 제안하는 방법을 사용하였을 때 성능 개선 효과를 확인할 수 있다. 또한, 매개변수 수의 비교를 통해 제안하는 모델이 4배가

그림 4. GoPro 데이터셋에 대한 영상 흐려짐 복원 결과. (a): 입력 영상, (b): 정답 영상, (c): U-Net+Self-Attention 기반 모델의 복원 결과, (d): 제안하는 모델의 복원 결과
Fig. 4. Deblurring result of GoPro dataset. (a): input image, (b): ground truth, (c): restoration result of U-Net+Self-Attention based model, (d): restoration result of proposed method

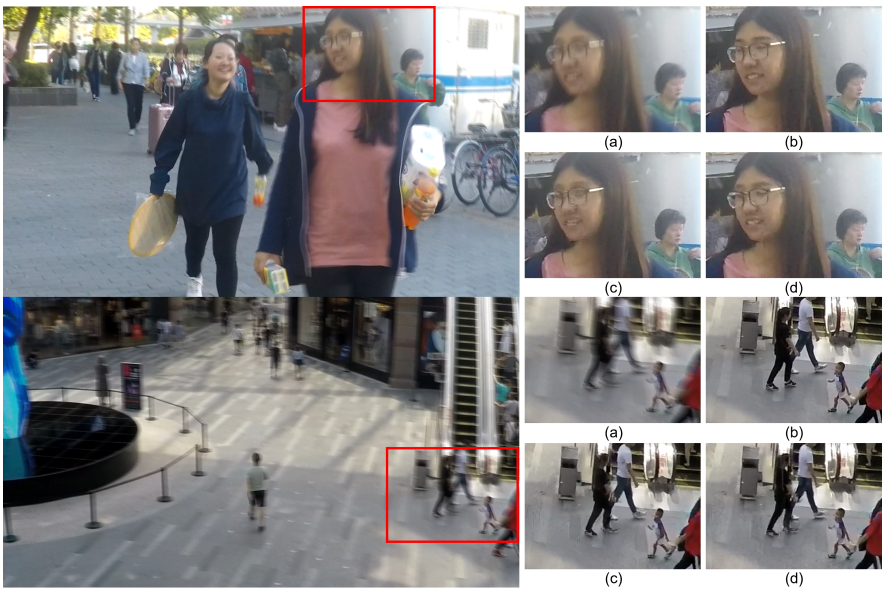

그림 5. HIDE 데이터셋에 대한 영상 흐려짐 복원 결과. (a): 입력 영상, (b): 정답 영상, (c): U-Net+Self-Attention 기반 모델의 복원 결과, (d): 제안하는 모델의 복원 결과
Fig. 5. Deblurring result of HIDE dataset. (a): input image, (b): ground truth, (c): restoration result of U-Net+Self-Attention based model, (d): restoration result of proposed method

량 적음에도 불구하고 더 흐려진 영상을 잘 복원하는 것을 확인할 수 있다.

표 1. GoPro 데이터셋(검증 셋)에서의 정량적 성능 비교
Table 1. Performance comparison on the GoPro dataset (Validation set)

Structure	Params	PSNR	SSIM
U-Net[8]+Self-Attention[9]	33.02M	28.06	0.914
Proposed Method	7.12M	29.05	0.932

표 2. HIDE 데이터셋(검증 셋)에서의 정량적 성능 비교
Table 2. Performance comparison on the HIDE dataset (Validation set)

Structure	Params	PSNR	SSIM
U-Net[8]+Self-Attention[9]	33.02M	26.14	0.889
Proposed Method	7.12M	26.35	0.893

추가로 다른 생성 모델을 사용한 영상 흐려짐 복원 방법과의 비교를 위해 GoPro 데이터셋의 테스트 데이터셋을 사용하여 표 3에 나타냈다. 표 3에 있는 세 개의 방법은 기존의 대표적인 생성 모델인 GAN 기반의 영상 흐려짐 복원 방법을 사용하였다. 그러나, 제안하는 방법은 DeblurGAN-v2[15]와 DBGAN[16]에 비해 낮은 성능을 보여준다. 이는 사전 학습을 수행하지 않은 결과에 따른 차이로 보인다. 제안하는 방법은 사전 학습을 수행하지 않은 Ghost-DeblurGAN[17]과 비교하여 더 높은 성능을 보여준다.

표 3. GoPro 데이터셋(테스트 셋)에서의 정량적 성능 비교
Table 3. Performance comparison on the GoPro dataset (Test set)

Structure	PSNR	SSIM
DeblurGAN-v2[15]	29.55	0.934
DBGAN[16]	30.10	0.942
Ghost-DeblurGAN[17]	28.75	0.919
Proposed Method	29.43	0.930

제안하는 방법은 PyTorch[18] 프레임워크를 기반으로 구현되었다. 본 논문에서는 신경망 가중치를 최적화하기 위한 알고리즘으로 Adam[19]을 사용하였고, 파워(Power)와 가속도(Momentum) 값은 각각 0.9와 0.999로 설정하였다. 학습 스케줄러(Scheduler)를 사용하여 선형 예열(Linear Warm-up) 수행 후, 학습 속도(Learning Rate)를 2×10^{-4}에 수렴하도록 설정하였으며, 총 3,500 에포크 동안 학습을 진행하였다. 학습 영상은 원본 영상에서 무작위로 256×256 픽셀 크기로 잘라 생성하였고, 과적합(Overfitting) 문제를 해결하기 위해 영상을 수평으로 반전하는 데이터 증강 방법을 적용하였다. 확산 모델의 분산 스케줄 β는 1×10^{-4}에서 2×10^{-2}까지 단계적으로 늘어난다. 또한 학습 모델에 지수 가중 평균(Exponentially Weighted Moving Averages, EMA)을 사용하였고, β 값은 0.9999로 설정하였다. 학습과 성능 평가에는 Intel(R) Core(TM) i7-6850K @3.60GHz CPU와 NVIDIA RTX 2080Ti GPU 2대가 이용되었다.

Ⅳ. 결 론

본 논문에서는 영상 흐려짐 복원을 위해 다중 스케일 확산 모델을 제안하였다. 제안하는 방법은 다중 스케일의 입출력을 사용하는 신경망을 통해 흐려진 영상을 복원하기 위한 잡음을 학습한다. 이를 통해 흐려진 영상을 입력하여 흐려짐을 제거한 영상을 효율적으로 출력하도록 설계하였다. 실험을 통해 제안하는 방법이 연산량을 줄임에도 영상 흐려짐 복원 성능을 효과적으로 향상시킬 수 있음을 확인하였다.

참 고 문 헌 (References)

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Proc. Neural Inf. Process. Syst.*, pp. 2672-2680, Dec. 2014. doi: https://doi.org/10.48550/arXiv.1406.2661

[2] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguil, “Deep unsupervised learning using nonequilibrium thermodynamics,” in *Proc. Int. Conf. Mach. Learn.*, pp. 2256-2265, Jul. 2015. doi: https://dl.acm.org/doi/10.5555/3045118.3045358

[3] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in *Proc. Neural Inf. Process. Syst.*, pp. 6840-6851, Dec. 2020. doi: https://dl.acm.org/doi/abs/10.5555/3495724.3496298

[4] D. Baranchuk, I. Rubachev, A. Voynov, V. Khrulkov, and A. Babenko, “Label-efficient semantic segmentation with diffusion models,” in *Proc. Int. Conf. Learn. Represent.*, Apr. 2022. doi: https://doi.org/10.48550/arXiv.2112.03126

[5] J. Wyatt, A. Leach, S. M. Schmon, and C. G. Willcocks, “AnoDDPM: Anomaly detection with denoising diffusion probabilistic models using simplex noise,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, pp. 650-656, Jun. 2022. doi: https://doi.org/10.1109/CVPRW56347.2022.00080

[6] C. Saharia, W. Chan, H. Chang, C. A. Lee, J. Ho, T. Salimans, D. Fleet, and M. Norouzi, “Palette: Image-to-image diffusion models,” in *Proc.*

ACM SIGGRAPH, pp. 1-10, Aug. 2022.
doi: https://doi.org/10.1145/3528233.3530757

[7] H. Li, Y. Yang, M. Chang, S. Chen, H. Feng, Z. Xu, Q. Li, and Y. Chen, "SRDiff: Single image super-resolution with diffusion probabilistic models," *Neurocomputing*, vol. 479, pp. 47-59, Mar. 2022.
doi: https://doi.org/10.1016/j.neucom.2022.01.029

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention*, pp. 234-241, Oct. 2015.
doi: https://doi.org/10.1007/978-3-319-24574-4_28

[9] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent.*, May 2021.
doi: https://doi.org/10.48550/arXiv.2010.11929

[10] S. J. Cho, S. W. Ji, J. P. Hong, S. W. Jung, and S. J. Ko, "Rethinking coarse-to-fine approach in single image deblurring," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pp. 4641-4650, Oct. 2021.
doi: https://doi.org/10.1109/ICCV48922.2021.00460

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, pp. 770-778, Jun. 2016.
doi: https://doi.org/10.1109/CVPR.2016.90

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Neural Inf. Process. Syst.*, pp. 5998-6008, Dec. 2017.
doi: https://dl.acm.org/doi/10.5555/3295222.3295349

[13] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, pp. 3883-3891, Jul. 2017.
doi: https://doi.org/10.1109/CVPR.2017.35

[14] Z. Shen, W. Wang, X. Lu, J. Shen, H. Ling, T. Xu, and L. Shao, "Human-aware motion deblurring," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pp. 5572-5581, Oct. 2019.
doi: https://doi.org/10.1109/ICCV.2019.00567

[15] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pp. 8878-8887, Oct. 2019.
doi: https://doi.org/10.1109/ICCV.2019.00897

[16] K. Zhang, W. Luo, Y. Zhong, L. Ma, B. Stenger, W. Liu, and H. Li, "Deblurring by realistic blurring," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Pattern Recognit.*, pp. 2737-2746, Jun. 2020.
doi: https://doi.org/10.1109/CVPR42600.2020.00281

[17] Y. Liu, A. Haridevan, H. Schofield, and J. Shan, "Application of ghost-deblurGAN to fiducial marker detection," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 6827-6832, Oct. 2022.
doi: https://doi.org/10.1109/IROS47612.2022.9981701

[18] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S.Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, highperformance deep learning library," in *Proc. Conf. Neural Inf. Process. Syst.*, pp. 8024 – 8035, Dec. 2019.
doi: https://dl.acm.org/doi/10.5555/3454287.3455008

[19] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, pp. 1-15, May 2015.
doi: https://doi.org/10.48550/arXiv.1412.6980

저 자 소 개

윤 천 희

- 2023년 2월 : 건국대학교 학사
- 2023년 3월 ~ 현재 : 건국대학교 전자·정보통신공학과 석사과정
- ORCID : https://orcid.org/0009-0004-1507-7016
- 주관심분야 : 컴퓨터 비전, 기계학습, 패턴 인식

김 원 준

- 2012년 8월 : 한국과학기술원(KAIST) 박사
- 2012년 9월 ~ 2016년 2월 : 삼성종합기술원 전문연구원
- 2016년 3월 ~ 2020년 2월 : 건국대학교 전기전자공학부 조교수
- 2020년 3월 ~ 현재 : 건국대학교 전기전자공학부 부교수
- ORCID : https://orcid.org/0000-0001-5121-5931
- 주관심분야 : 영상이해, 컴퓨터 비전, 기계학습, 패턴 인식