

석사학위청구논문

DCGAN을 이용한 영상복원 연구

Image Restoration Study using DCGAN

경상국립대학교 정보통계학과 빅데이터 전공 허지혜

2022.11.11

목차

1. 서론

2. 관련연구

- 2.1 잡음
- 2.2 전통적인 영상복원
- 2.3 딥러닝
 - (1) CNN
 - (2) GAN

3. 딥러닝을 이용한 영상복원

- 3.1 DnCNN
- 3.2 RED-CNN
- 3.3 DCGAN

4. 제안된 RED-DGAN 모형

- 4.1 Deformable Convolution
- 4.2 RED-DGAN 모형 구조

5. 성능실험

- 5.1 실험환경
- 5.2 실험 데이터
- 5.3 성능실험결과
 - (1) 정성적 평가
 - (2) 정량적 평가

6. 결론 및 향후연구

1. 서론

연구배경

- 현대사회는 과학기술의 발전으로 카메라, CCTV, 블랙박스 등 다양한 디지털 영상기기가 보편화됨에 따라 디지털 영상을 자주 접할 수 있음 이러한 발전에도 디지털 영상을 얻는 인코딩 과정에서 잡음(Noise)이 생성되어 영상의 질이 저하됨
- 영상의 정확한 정보를 얻어내기 위해 영상을 재구성하여 고품질 영상을 얻어내는 영상복원은 CCTV 영상 판별, 병명 판독 등 여러 분야에 활용가능성이 높음
- 최근 딥러닝은 컴퓨터 비전, 자연어 처리 등 다양한 분야에서 우수한 성능을 보이고 있음 특히 CNN 모형, GAN 모형을 필두로 영상처리 분야 연구가 활발히 진행되고 있음

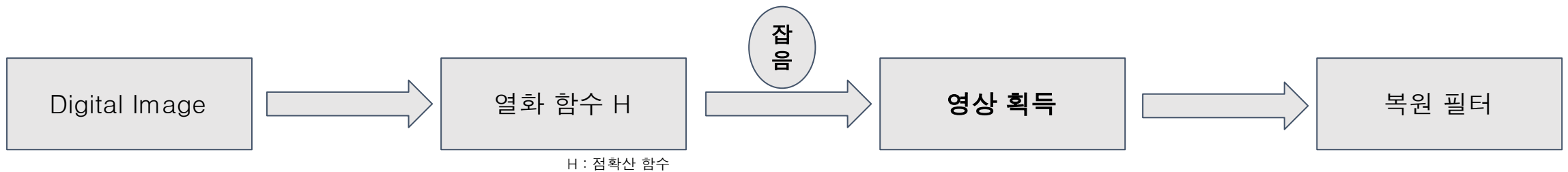
연구내용

- 본 논문에서는 기존 DCGAN 모형의 구조를 변형시킨 **RED-DGAN(Residual Encoder Decoder – Deformable Generative Adversarial Network) 모형**을 통해 영상복원 성능을 향상시키고자 함
- 제안된 RED-DGAN 모형의 성능을 평가하기 위해 다양한 잡음 즉, 가우시안 잡음(Gaussian Noise), 포아송 잡음(Poisson Noise), 스펙클 잡음(Speckle Noise) 으로 실험
- 고전적인 영상복원 방법, 딥러닝 영상복원 방법과 제안된 RED-DGAN 모형을 비교하기 위해 정성적, 정량적 평가 진행

2. 관련연구

영상복원

- 영상의 열화현상에 대한 사전정보를 이용해 열화된 영상을 재구성 또는 복구하는 것



영상복원의 3가지 분야

Image Deblurring

카메라나 손 흔들림으로 발생하는 블러링 제거



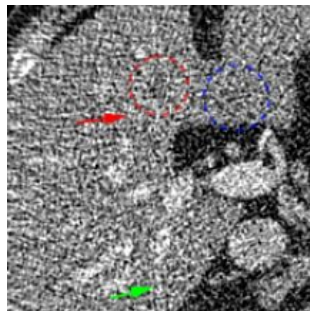
(왼) Blurring Image



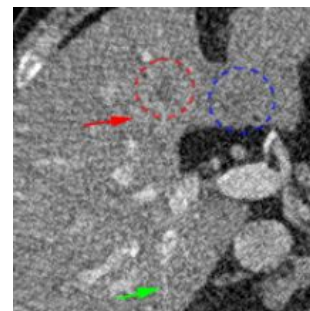
(오) Deblurring Result

Image Denoising

이미지 복원 과정에서 잡음 제거



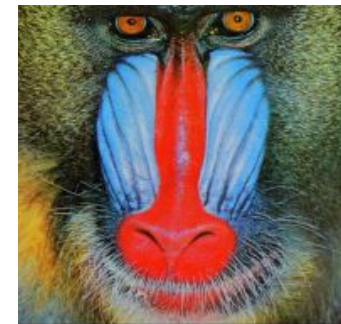
(왼) Noise Image



(오) Denoising Result

Super Resolution

이미지 크기 변화에 따른 해상도 복원



(왼) Original Image



(오) SR (Upsampling*4)

2. 관련연구

2.1 잡음

- 잡음은 크게 3가지로 실험하여 다양한 잡음에 대하여 복원이 가능함을 증명함

Y : 잡음 영상, X : 원 영상

가우시안(Gaussian) 잡음

$$Y = X + a \times Z$$

Z : 가우시안 분포, a : 잡음의 정도

포아송(Poisson) 잡음

$$Y = X + a \times P(\lambda)$$

$P(\lambda)$: 평균 λ 인 포아송 분포, a : 잡음의 정도

스펙클(Speckle) 잡음

$$Y = X + X \times a \times Z$$

Z : 가우시안 분포, a : 잡음의 정도



(a) 원 영상



(b) 가우시안 잡음영상



(c) 포아송 잡음영상



(d) 스펙클 잡음영상

2. 관련연구

2.2 고전적인 영상복원

- 가장 유명한 고전적인 영상복원 방법으로 공간 도메인 필터링이 있음
- 공간 도메인 필터링은 필터를 이용하여 공간을 특정 값으로 바꾸면서 Denoising하는 방법

(1) Mean 필터

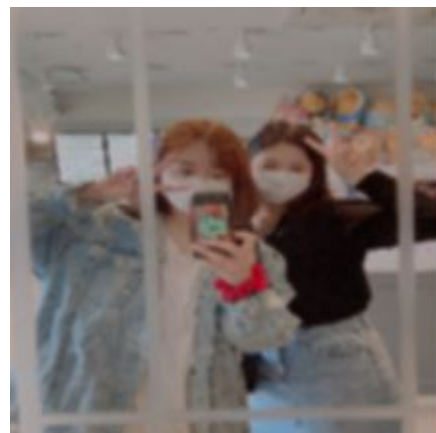
- 평균을 이용한 선형 필터링 기법
- 특정 좌표값을 주위 값들의 산술평균으로 설정하여 날카로운 엣지가 무너지고 잡음제거 효과를 가져옴
- 하지만, 원하는 영상이 아닌 흐릿한 영상을 얻을 수 있음



(A) 원영상



(B) 가우시안 잡음제거 영상



(C) 포아송 잡음제거 영상



(D) 스펙클 잡음제거 영상

2. 관련연구

2.2 고전적인 영상복원

(2) Median 필터

- 중앙값을 이용한 비선형 필터링 기법
- 엣지를 고려하지 않는 Mean 필터의 단점을 극복하기 위해 중앙값으로 설정해줌
- 마스크의 크기에 따라 디테일 잡기 어려움



(A) 원영상



(B) 가우시안 잡음제거 영상



(C) 포아송 잡음제거 영상



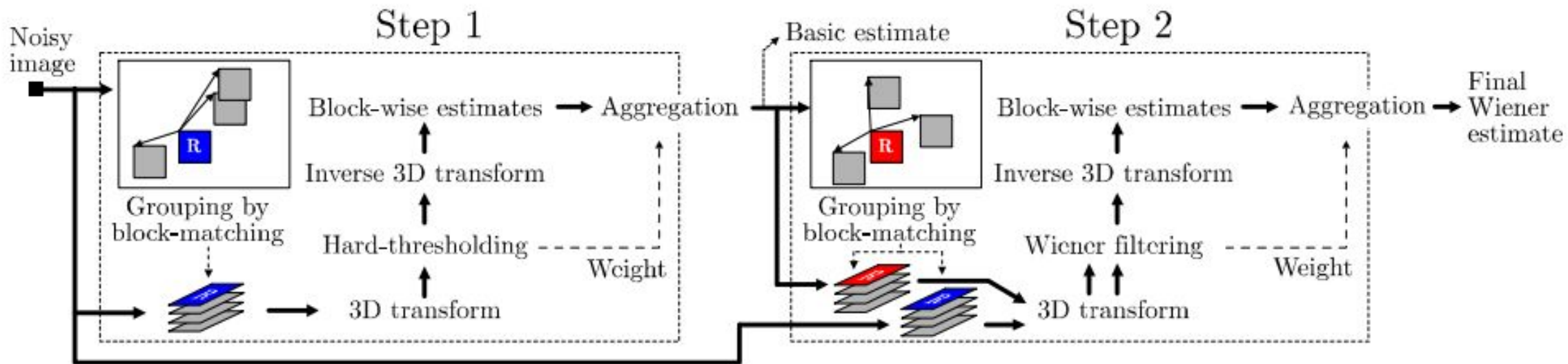
(D) 스펙클 잡음제거 영상

2. 관련연구

2.2 고전적인 영상복원

(3) BM3D 필터

- BM3D(Block-Matching 3D) 필터는 변환필터링과 비지역적 평균(Non-local means)의 확장으로 Denoising에 사용되는 알고리즘
- 협업 필터링을 Step1에서는 Hard-threshold를 기준으로, Step 2에서는 Wiener 필터링을 기준으로 진행
- 비지역적 평균 알고리즘은 비슷한 영역(Block)을 추출하여 영역의 가중평균으로 대체하는 알고리즘임



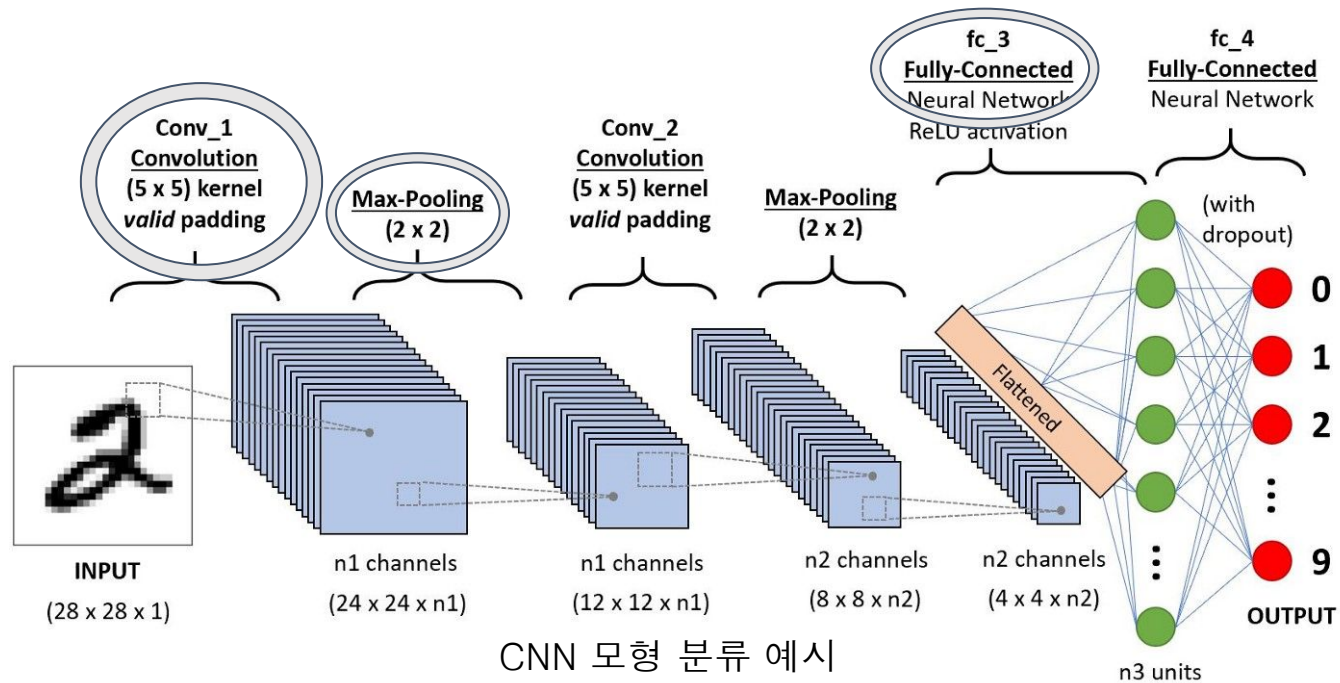
BM3D 알고리즘

2. 관련연구

2.3 딥러닝

(1) CNN

- CNN(Convolutional Neural Network) 모형은 Convolutional Layer, Pooling Layer, Fully Connected Layer으로 이루어진 분류 모형
- Convolutional Layer에서는 이미지의 특징을 추출하는 합성곱 연산을 수행
- Pooling Layer에서는 이미지의 크기를 줄임
- Fully Connected Layer를 거쳐 분류를 수행

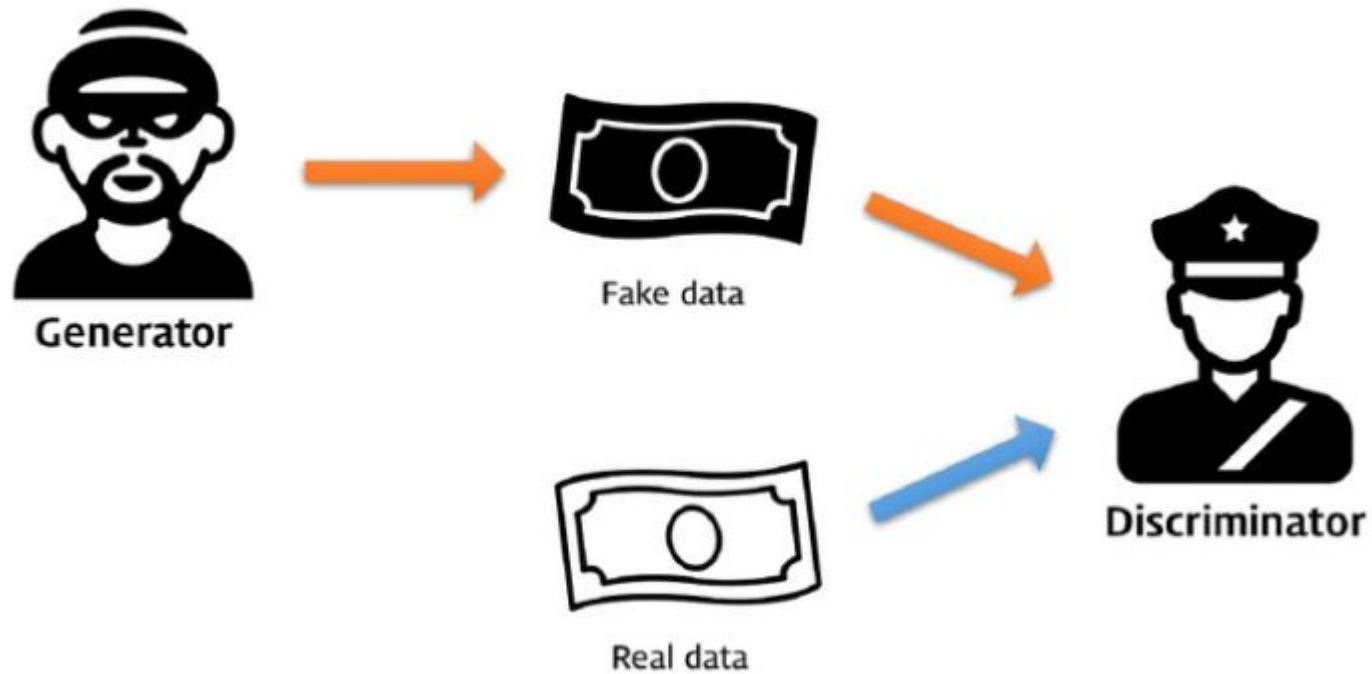


2. 관련연구

2.3 딥러닝

(2) GAN

- GAN(Generative Adversarial Network) 모형은 Generator 모형과 Discriminator 모형이 존재
- Generator 모형에서는 진짜와 비슷한 가짜 영상을 생성
- Discriminator 모형에서는 진짜와 가짜를 분류



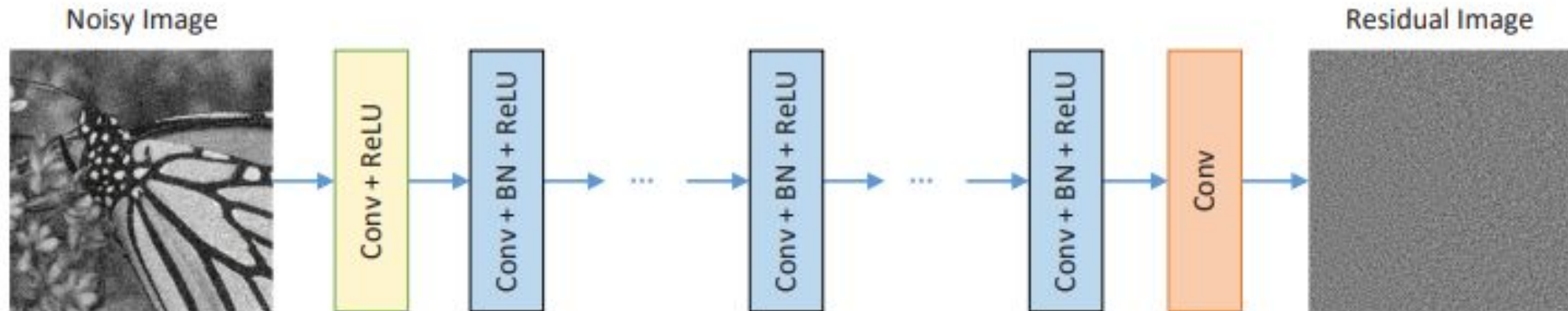
GAN 모형 분류 예시

3. 딥러닝을 이용한 영상복원

3.1 DnCNN

- DnCNN(Denoising Convolutional Neural Network) 모형은 CNN 모형의 대표적인 영상복원 모형
- 다른 딥러닝 영상복원 모형과 다르게 잡음영상에서 원 영상과 잡음을 구분하여 잡음을 출력
- 우리는 원 영상과 잡음을 구분하여 원 영상을 출력하는 형태로 학습 진행

$$\ell(\theta) = \frac{1}{2N} \sum_{i=1}^N \|F(y_i; \theta) - (y_i - x_i)\|_F^2$$

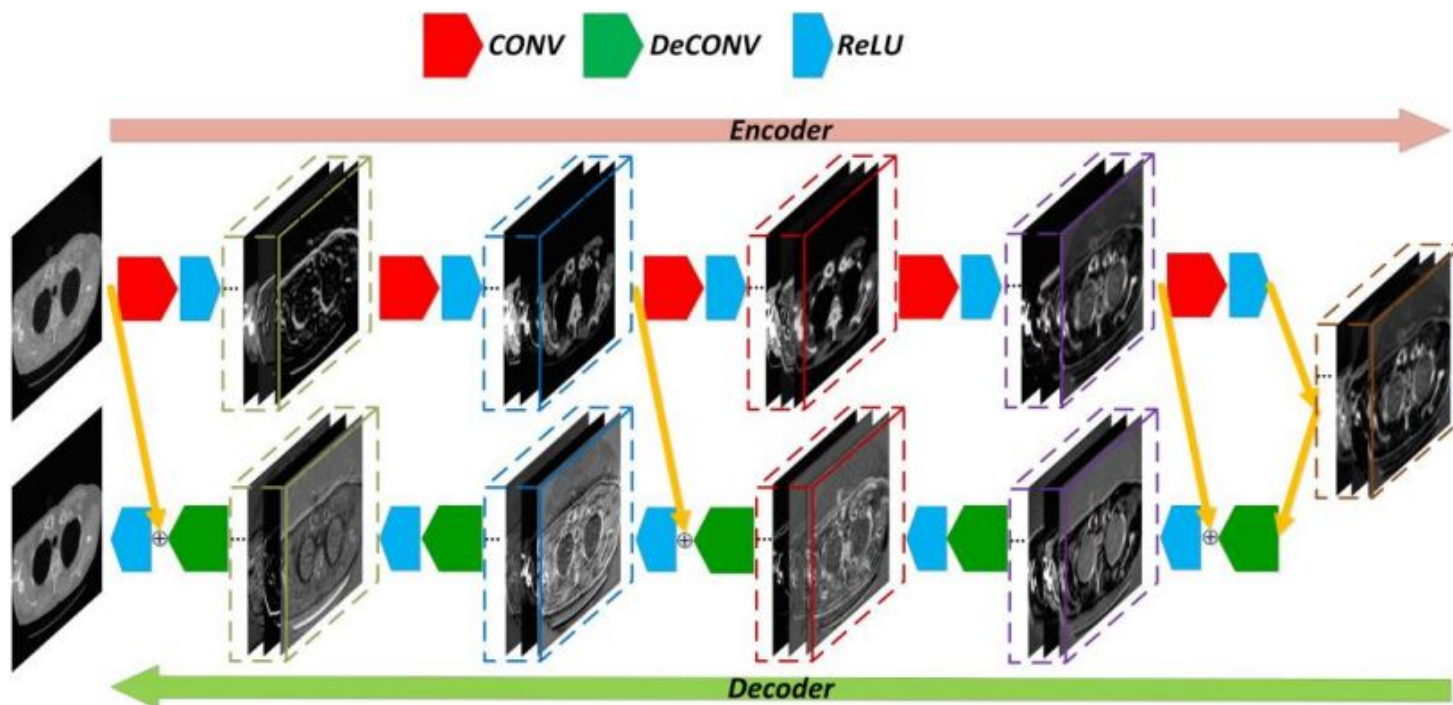


DnCNN 모형 구조

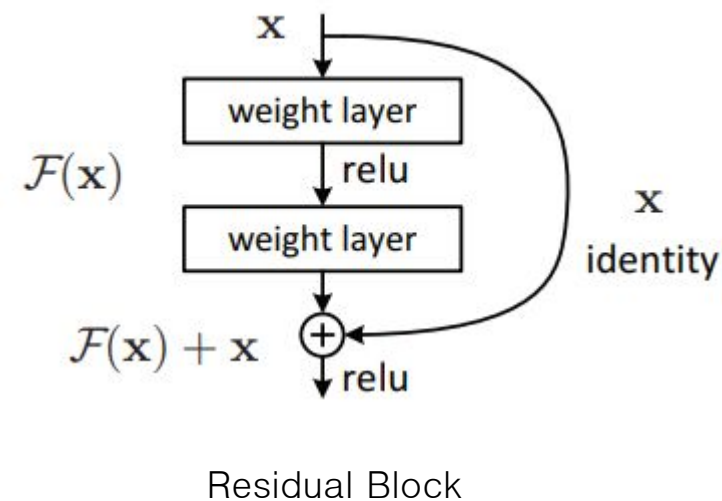
3. 딥러닝을 이용한 영상복원

3.2 RED-CNN

- RED-CNN(Residual Encoder Decoder – Convolutional Neural Network) 모형은 Encoder 모형과 Decoder 모형에 Residual Block을 추가하여 영상복원을 실시함
- Encoder 모형에서는 Convolutional Layer를 이용해 특징을 추출하고 Decoder 모형에는 Deconvolutional Layer로 영상을 복원함



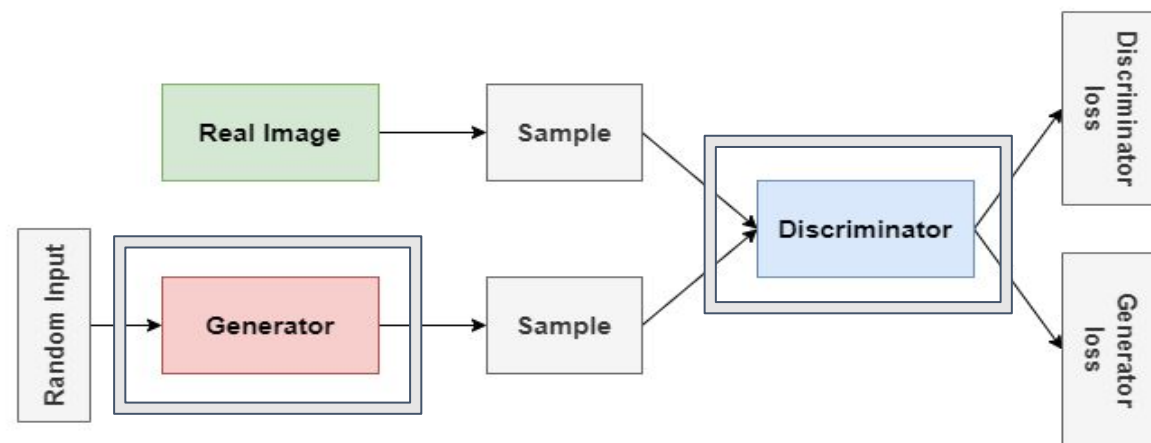
RED-CNN 모형
구조



3. 딥러닝을 이용한 영상복원

3.3 DCGAN

- DCGAN(Deep Convolutional Generative Adversarial Network) 모형은 GAN 모형에 Convolutional Layer를 추가하여 영상의 특징을 잘 활용하고자 함
- Generator 모형에서는 잡음영상을 입력으로 받아 복원영상을 출력함
- Discriminator 모형에서는 복원영상과 원 영상을 비교하여 분류함

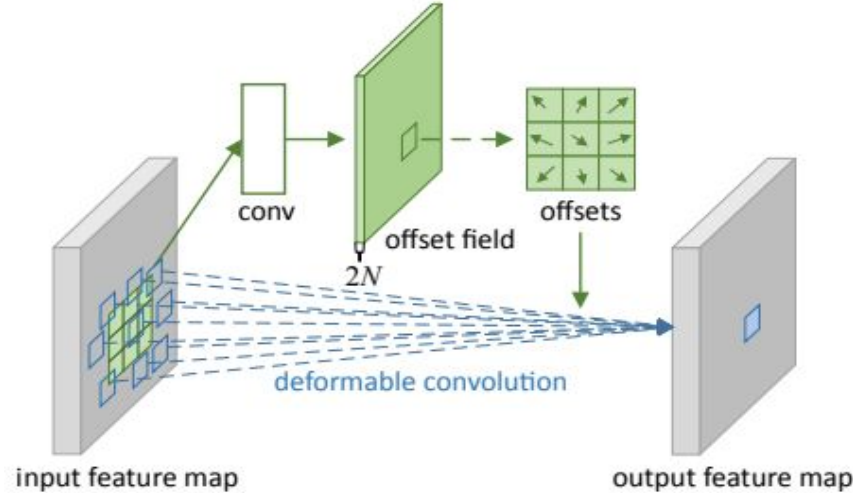


DCGAN 모형 진행

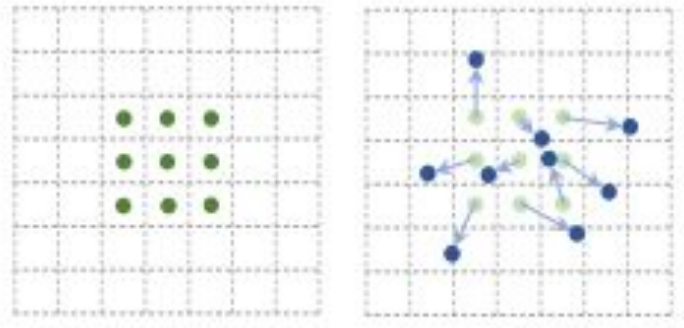
4. 제안된 RED-DGAN 모형

4.1 Deformable Convolution

- 기존의 Convolution 연산은 일정한 패턴을 가지고 있어 복잡한 업무에 유연하게 대처하기 어려움
- Deformable Convolution 연산을 통해 조금 더 넓은 범위로 학습 가능
- Deformable Convolution 연산 여부에 따른 성능차이가 있음



Deformable Convolution Layer



(왼) 기존의 Convolution (오) Deformable Convolution

$$y(p_0) = \sum_{p_n \in \mathbb{R}} w(p_n) \cdot x(p_0 + p_n)$$

Convolution 연산 식

$$y(p_0) = \sum_{p_n \in \mathbb{R}} w(p_n) \cdot x\left(p_0 + p_n + \underbrace{\Delta p_n}_{\text{실수}}\right) * \Delta m_k$$

Deformable Convolution 연산 식

p_0 Feature Map
 p_n Grid
 $w(p_n)$ 가중치

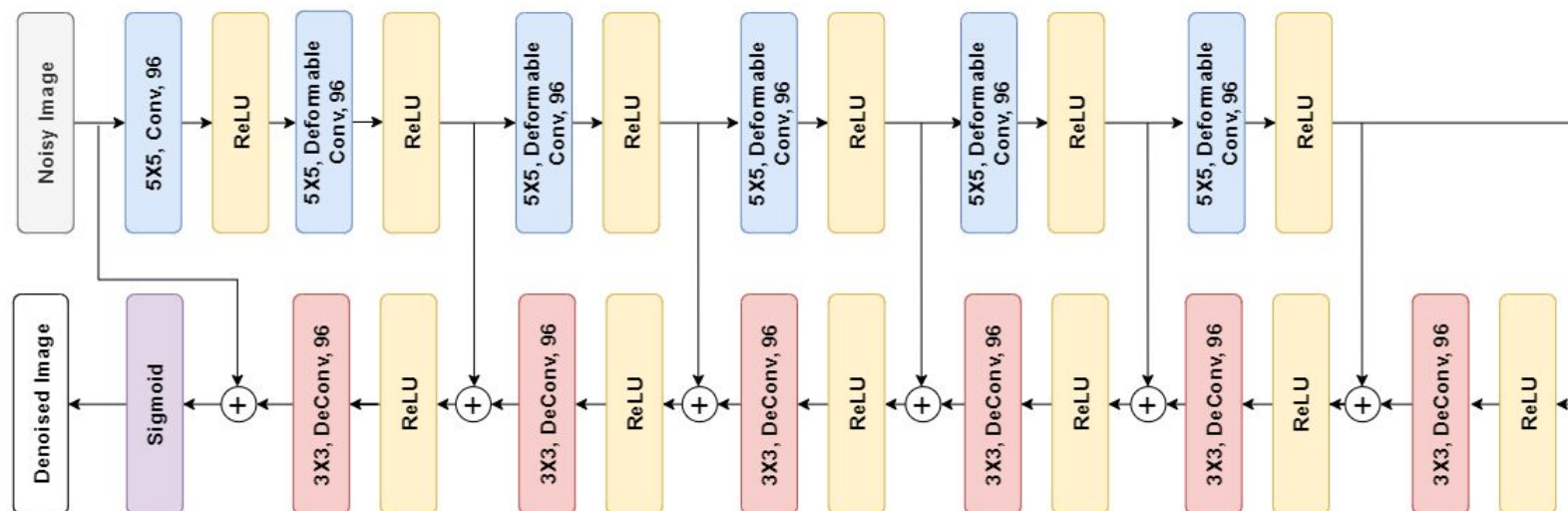
4. 제안된 RED-DGAN 모형

4.2 RED-DGAN 모형 구조

- RED-DGAN 모형은 GAN 모형 기반으로 Generator 모형과 Discriminator 모형으로 나뉨

Generator 모형

- 6개의 Encoder 모형과 6개의 Decoder 모형으로 이루어져 있음
- Encoder 모형의 각 출력값은 Decoder 모형의 각 출력값에 더해지는 Residual 구조
- Encoder 모형** : 1개의 Convolutional Layer와 5개의 Deformable Convolutional Layer를 이용하여 영상의 특징을 압축
- Decoder 모형** : 6개의 Deconvolutional Layer를 이용하여 복원영상 생성
- 입력** : 잡음영상 / **출력** : 복원영상



RED-DGAN 모형 Generator 구조

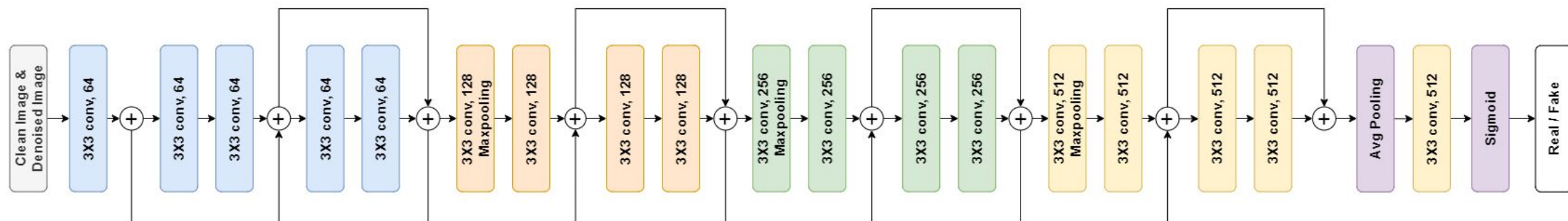
4. 제안된 RED-DGAN 모형

4.2 RED-DGAN 모형 구조

- RED-DGAN 모형은 GAN 모형이 베이스이므로 Generator 모형과 Discriminator 모형으로 나뉨

Discriminator 모형

- 18개 Convolutional Layer로 이루어짐
- Resnet18 모형**의 Fully Connected Layer를 완전히 제거하고 Convolutional Layer와 Sigmoid 함수 추가
- 입력 : 복원영상 / 출력 : 1
- 입력 : 원 영상 / 출력 : 0



RED-DGAN 모형 Discriminator 구조

5. 성능실험

5.1 실험환경

- <표1>의 환경에서 딥러닝 구현
- <표2>처럼 동일한 조건 내에서 실험

GPU	NVIDIA GeForce RTX 3090
파이썬 버전	Python 3.8.5
프레임워크	Pytorch 1.12.0+cu113

<표1>

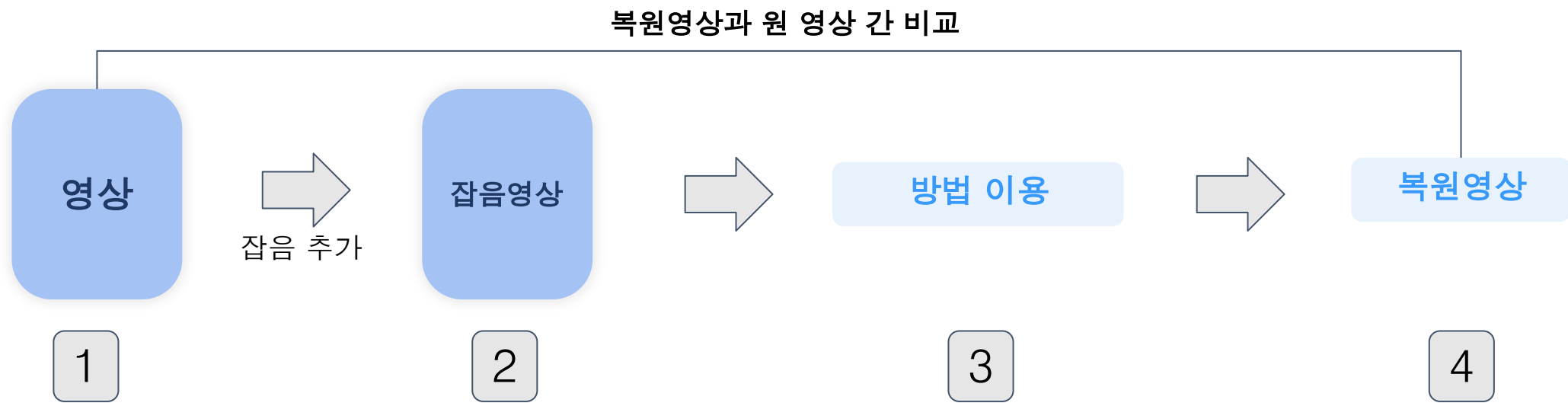
LOSS	MSE(Mean Square Error)
Generator LOSS	MSE(Mean Square Error)
DISCRIMINATOR LOSS	BCE(Binary Cross Entropy Error)
OPTIMIZER	ADAM
LEARNING RATE	0.0001
EPOCH	100

<표2>

5. 성능실험

5.1 실험환경

- 영상복원 모형의 실험은 다음과 같은 순서로 진행



5. 성능실험

5.2 실험 데이터

- <https://github.com/hromi/SMILEsmileD>
- 웃는 얼굴 데이터를 이용하여 학습 진행
- 웃는 얼굴 데이터는 성별, 나이를 가리지 않은 확대 얼굴 영상임

TOTAL	TRAIN	TEST
3,690	2,952	738

IMAGE SIZE	CHANNEL	BATCH SIZE
128	3	4



5. 성능실험

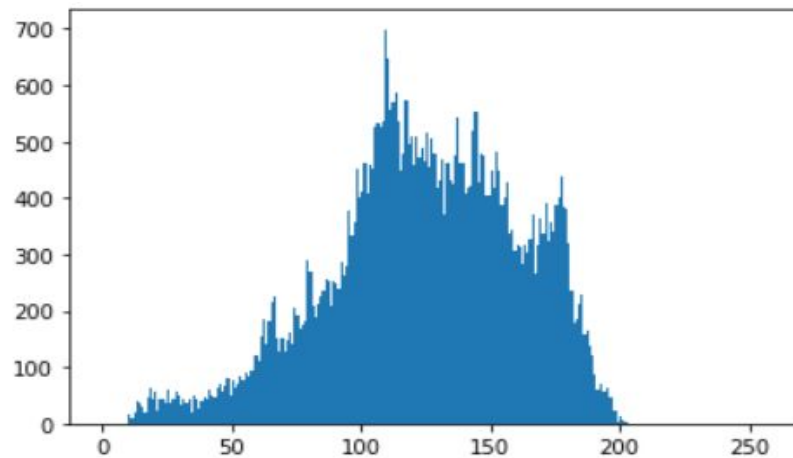
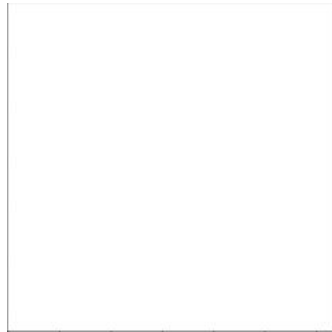
5.3 성능실험결과

(1) 정성적 평가 - 가우시안 잡음

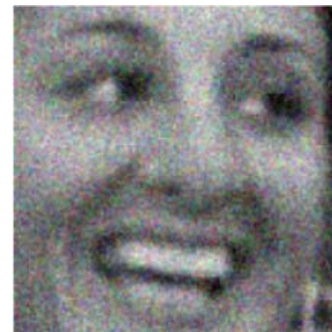
원 영상



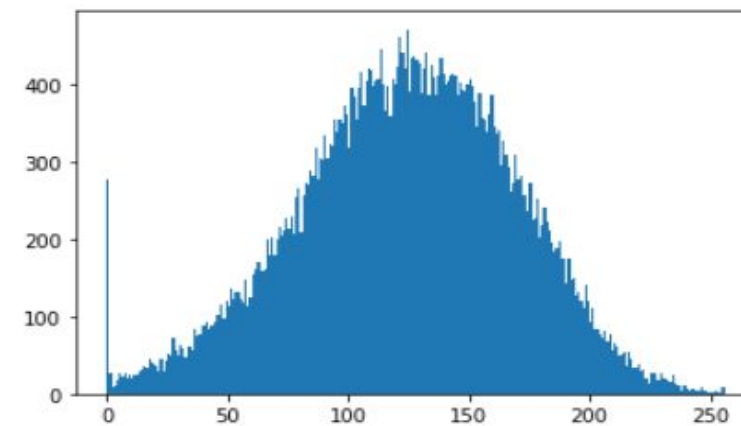
DIFF



잡음영상



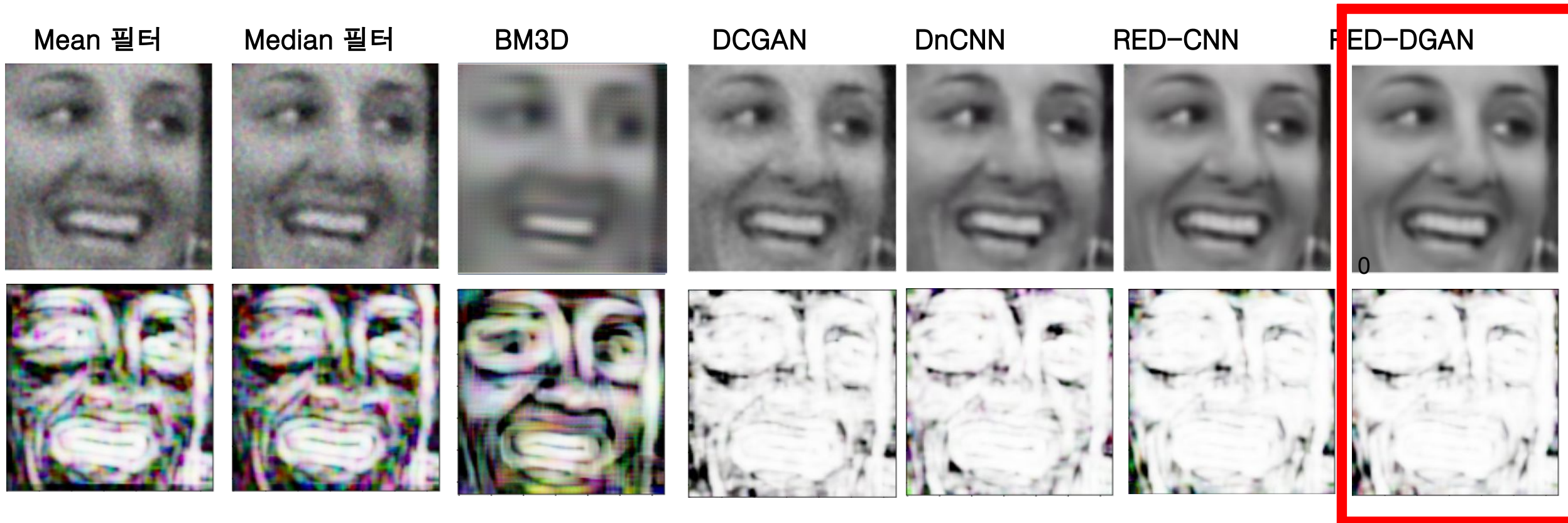
DIFF



5. 성능실험

5.3 성능실험결과

(1) 정성적 평가 - 가우시안 잡음

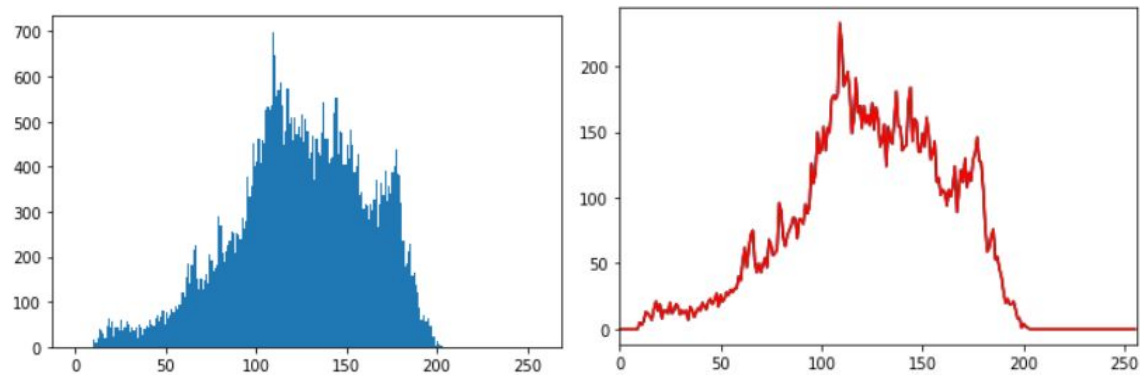


5. 성능실험

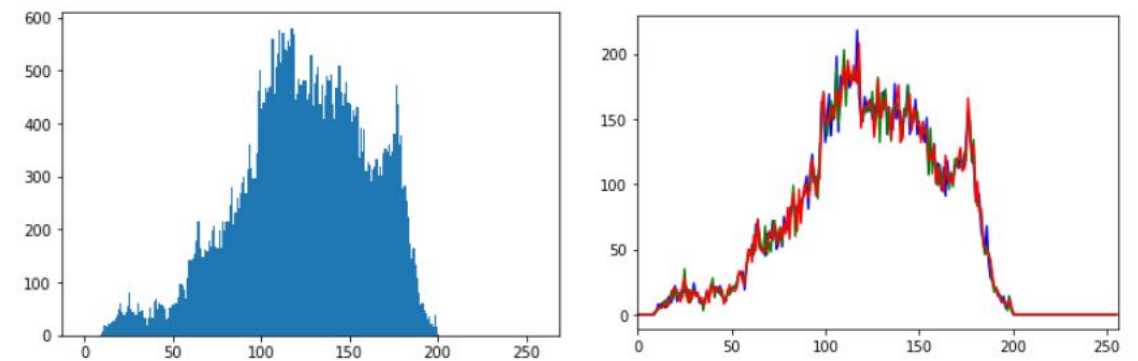
5.3 성능실험결과

(1) 정성적 평가 - 가우시안 잡음

원본



제안한 RED-DGAN 모형



5. 성능실험

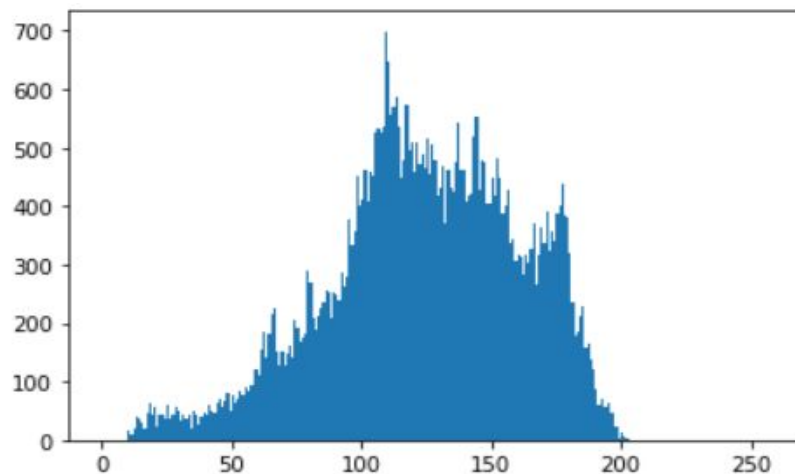
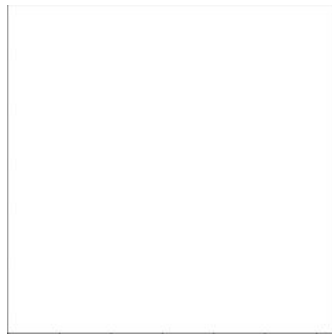
5.3 성능실험결과

(1) 정성적 평가 - 포아송 잡음

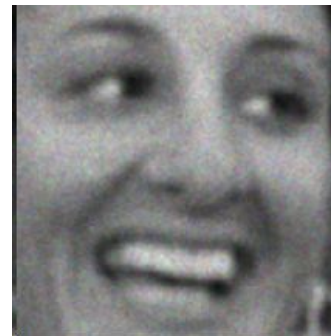
원 영상



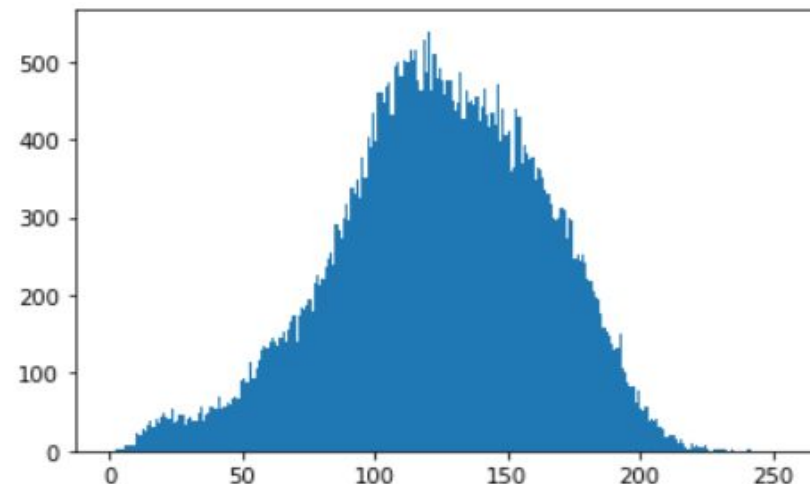
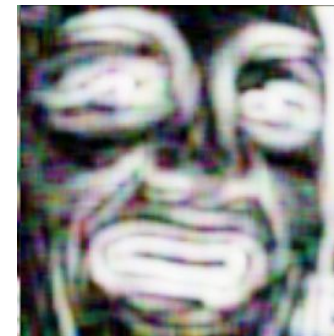
DIFF



잡음영상



DIFF



5. 성능실험

5.3 성능실험결과

(1) 정성적 평가 - 포아송 잡음

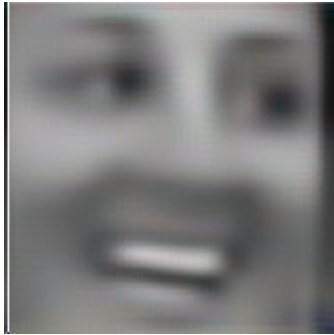
Mean 필터



Median 필터



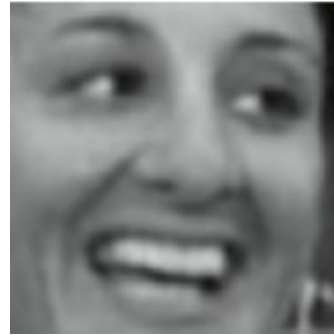
BM3D



DCGAN



DnCNN



RED-CNN



RED-DGAN

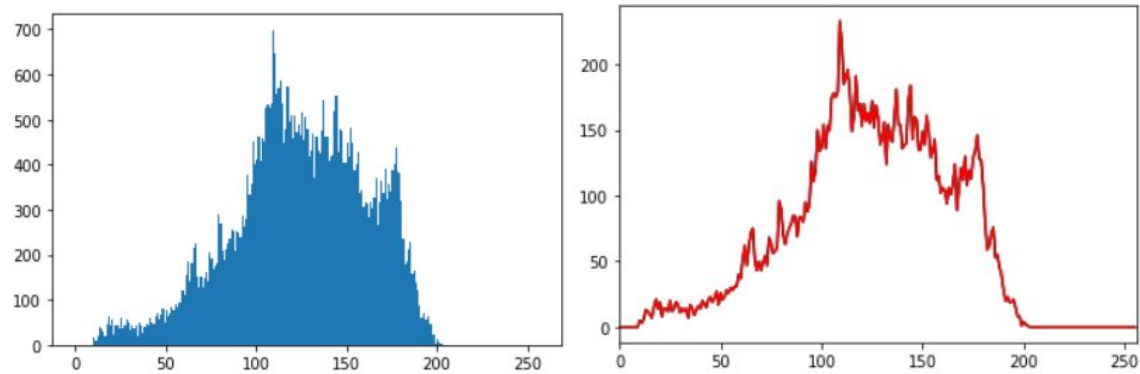


5. 성능실험

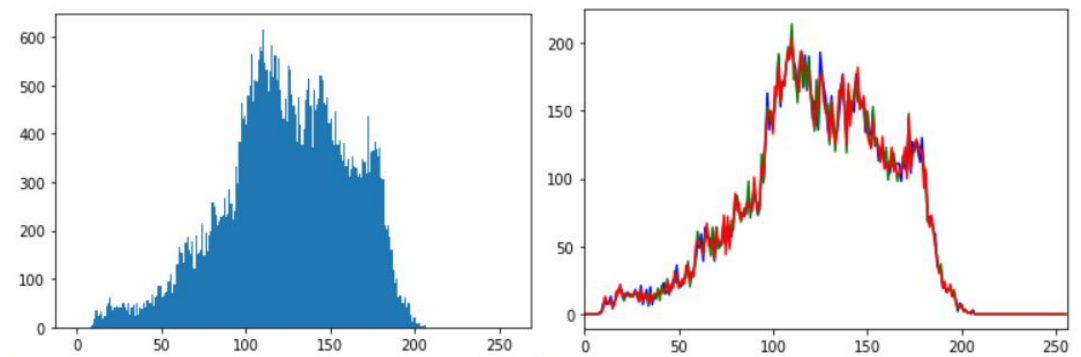
5.3 성능실험결과

(1) 정성적 평가 - 포아송 잡음

원본



제안한 RED-DGAN 모형



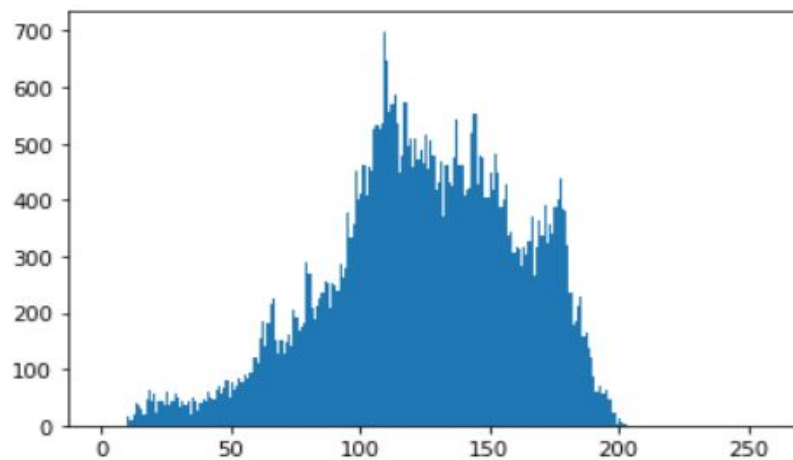
5. 성능실험

5.3 성능실험결과

(1) 정성적 평가 - 스펙클 잡음

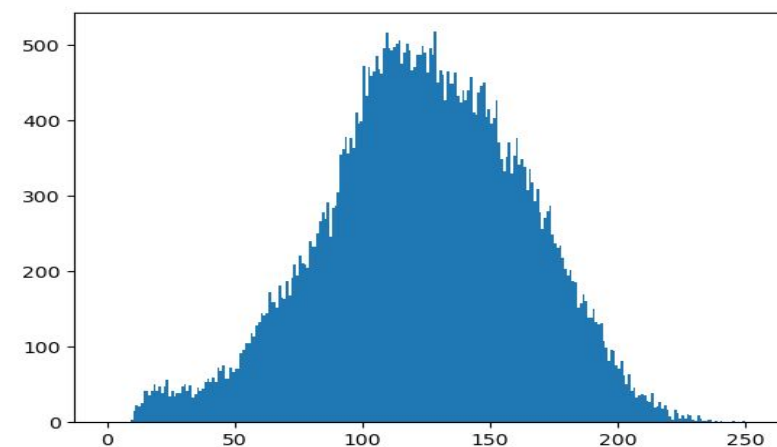
원 영상

DIFF



잡음영상

DIFF



5. 성능실험

5.3 성능실험결과

(1) 정성적 평가 - 스펙클 잡음

Mean 필터



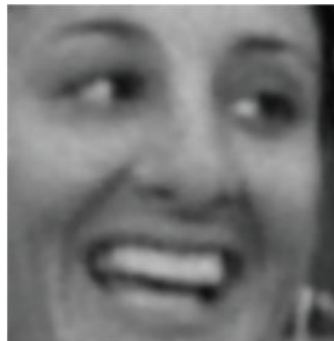
Median 필터



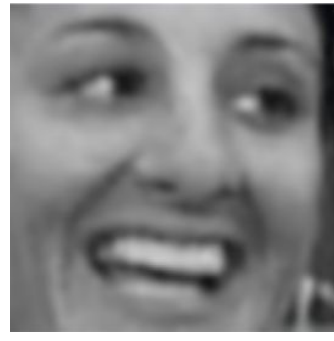
BM3D



DCGAN



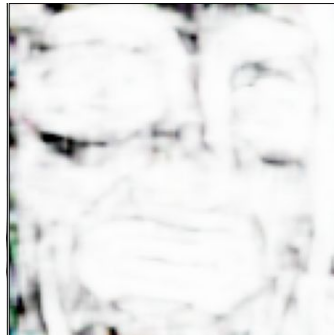
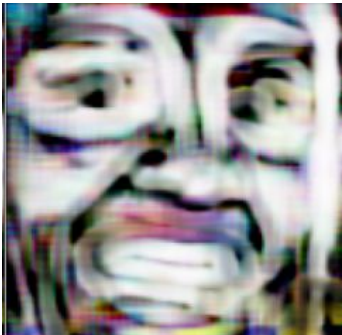
DnCNN



RED-CNN



RED-DGAN

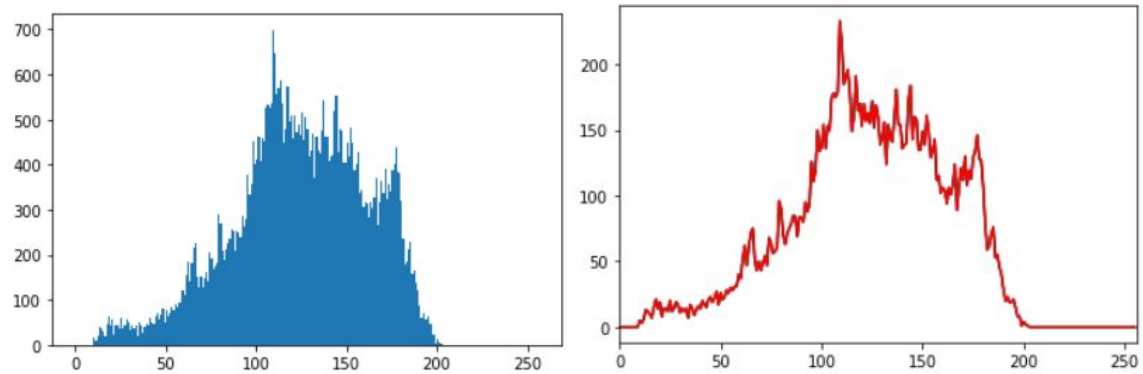


5. 성능실험

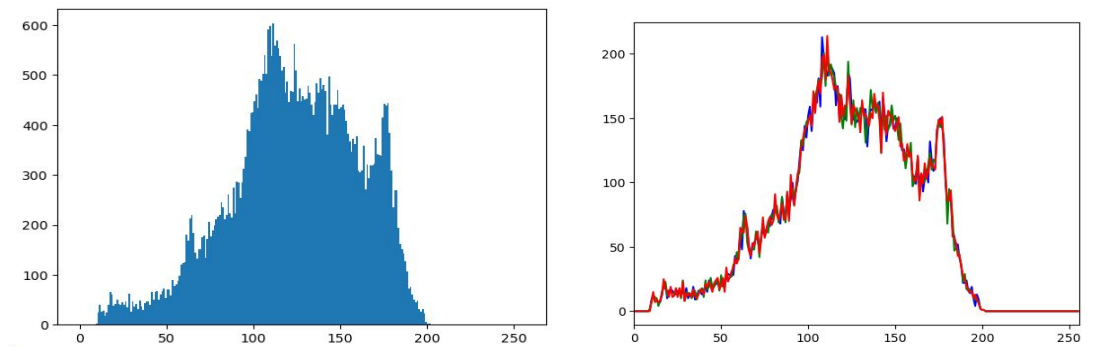
5.3 성능실험결과

(1) 정성적 평가 - 스펙클 잡음

원본



제안한 RED-DGAN 모형



5. 성능실험

5.3 성능실험결과

(2) 정량적 평가

PSNR(Peak Signal to Noise Ratio, 최대 신호 대 잡음비)	SSIM(Structural Similarity Index Measure, 구조적 유사도 지수)
<div><ul style="list-style-type: none">영상의 화질 또는 압축에 대한 손실 정보를 평가$PSNR = 20 \times \log_{10} \left(\frac{MAX}{\sqrt{MSE}} \right)$<p><i>MAX</i> : 픽셀의 최댓값</p></div>	<div><ul style="list-style-type: none">시각적 화질 차이를 평가하기 위한 평가 척도Luminance, Contrast, Structural 측면에서 품질 평가$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$</div>
MAE(Mean Absolute Error, 평균 절대값 오차)	MSE(Mean Square Error, 평균 제곱 오차)
<div><ul style="list-style-type: none">원 영상과 복원 영상의 절대값 차의 평균$MAE = \frac{1}{mn} \sum_{i \in m \times n} X_i - \hat{Y}_i$</div>	<div><ul style="list-style-type: none">원 영상과 복원 영상 차의 제곱의 평균$MSE = \frac{1}{mn} \sum_{i \in m \times n} \{X_i - \hat{Y}_i\}^2$</div>

5. 성능실험

5.3 성능실험결과

(2) 정량적 평가 - 가우시안 잡음

	PSNR	MAE	MSE	SSIM
Mean Filter	27.0944	0.0329	0.00195	0.6387
Median Filter	27.1673	0.0342	0.00192	0.6171
BM3D	33.8216	0.0158	0.00047	0.9285
DCGAN	33.7585	0.0167	0.00042	0.9385
DnCNN	37.2811	0.0106	0.00019	0.9585
RED-CNN	37.3768	0.0102	0.00018	0.9604
RED-DGAN	38.1832	0.0096	0.00015	0.9669

5. 성능실험

5.3 성능실험결과

(2) 정량적 평가 - 포아송 잡음

	PSNR	MAE	MSE	SSIM
Mean Filter	30.7689	0.01694	0.000846	0.8466
Median Filter	33.3220	0.01600	0.000467	0.8645
BM3D	36.5780	0.01123	0.000221	0.9529
DCGAN	33.2260	0.01958	0.000477	0.9734
DnCNN	42.3996	0.00588	0.000058	0.9838
RED-CNN	42.6457	0.00564	0.000055	0.9840
RED-DGAN	42.7473	0.00568	0.000053	0.9847

5. 성능실험

5.3 성능실험결과

(2) 정량적 평가 - 스펙클 잡음

	PSNR	MAE	MSE	SSIM
Mean Filter	30.1667	0.01901	0.000918	0.8162
Median Filter	32.0338	0.01842	0.000627	0.8267
BM3D	36.1474	0.01186	0.000245	0.9501
DCGAN	29.5120	0.03150	0.001121	0.9676
DnCNN	41.7399	0.00637	0.000067	0.9817
RED-CNN	41.7123	0.00617	0.000068	0.9813
RED-DGAN	41.9381	0.00616	0.000065	0.9827

6. 결론 및 향후연구

결론

- DCGAN 모델을 변형한 RED-DGAN 모델을 이용하여 영상복원 문제에 접근함 영상복원을 통해 잡음이 제거된다면 정확한 정보를 이용하여 영상을 볼 수 있기 때문에 영상복원은 꼭 필요함
- 지금까지 진행된 영상복원 연구를 고전적인 영상복원과 딥러닝 영상복원 방법으로 나누어 설명함
- 성능 실험 결과, 정성적 평가와 정량적 평가 모두 RED-DGAN 모델이 대부분 압도하는 것을 볼 수 있음 포아송 잡음에서 MAE 수치가 RED-CNN 모델이 더 좋지만 아주 작은 수치 차이를 보이며 Robust한 모델임을 알 수 있음

향후연구

- 성능에 초점을 두어 Deformable Convolution 연산을 이용하였으나 계산량이 많다는 단점이 있음 추후에 시간을 단축시키면서 성능을 올릴 수 있는 영상복원 모델 연구를 희망함

참고문헌

- [1] M. Biemond and A. Katsaggelos, "Digital image restoration", IEEE Signal Processing Magazine, vol. 14, no. 2, pp.24–41. March 1997.
- [2] R. C. Gonzalez, R. E. Woods and S. L. Eddins, Digital Image Processing using MATLAB, Prentice–Hall, 2003
- [3] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, W. T. Freeman, "Removing camera shake from a single photograph," ACM Transactions on Graphics (TOG), Vol. 25, no. 3, pp. 787–794, July 2006.
- [4] Low Dose CT Image Denoising Using a Generative Adversarial Network with Wasserstein Distance and Perceptual Loss
- [5] Ledig, Christian et al. "Photo–Realistic Single Image Super–Resolution Using a Generative Adversarial Network." *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*(2017): 105–114.
- [6] Burger, H. C., Schuler, C. J., Harmeling, S., 2012. Image denoising: Can plain neural networks compete with bm3d? In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 2392–2399.
- [7] denoising a
- [8] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," to appear in IEEE transactions on Pattern Analysis and Machine Intelligence, 2016.
- [9] Zhang, K. et al. "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising." *IEEE Transactions on Image Processing* 26 (2017): 3142–3155.
- [10] H. Chen et al., "Low–Dose CT With a Residual Encoder–Decoder Convolutional Neural Network," in *IEEE Transactions on Medical Imaging* vol. 36, no. 12, pp. 2524–2535, Dec. 2017, doi: 10.1109/TMI.2017.2715284.
- [11] Zhang, K. et al. "FFDNet: Toward a Fast and Flexible Solution for CNN–Based Image Denoising." *IEEE Transactions on Image Processing* 27 (2018): 4608–4622.
- [12] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non–local sparse models for image restoration," in IEEE International Conference on Computer Vision, 2009, pp. 2272–2279.
- [13] K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, "Image Denoising by Sparse 3–D Transform–Domain Collaborative Filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007, doi: 10.1109/TIP.2007.901238.
- [14] Radford, Alec et al. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks." *CoRR* abs/1511.06434 (2016): n. pag.
- [15] He, Kaiming et al. "Deep Residual Learning for Image Recognition." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016): 770–778.
- [16] Dai, Jifeng et al. "Deformable Convolutional Networks." *2017 IEEE International Conference on Computer Vision (ICCV)*(2017): 764–773.

감사합니다