

폭우 환경에서 촬영된 얼굴 영상의 초고해상화

정다희*, 손창환**

Super Resolution for Face Images Captured in Heavy Rain Conditions

Da-Hee Jeong*, Chang-Hwan Son**

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2020R1A2C1010405)

요약

최근 침입탐지, 객체 카운팅, 용의자 및 이상행동 검출과 같이 객체를 인식하고 장면의 상황까지 분석 가능한 지능형 감시카메라가 활발히 연구되고 있다. 이러한 지능형 감시카메라는 대부분 실외에 설치되므로 다양한 촬영 환경에 놓이게 된다. 특히 폭우 환경에서는 빗줄기나 빗줄기 축적 현상으로 촬영된 영상의 가시성이 현저히 저하되기 때문에 얼굴 인식이 어렵게 된다. 따라서 본 논문에서는 폭우 환경에서 촬영된 영상의 얼굴 인식 개선을 위한 얼굴 초고해상화 딥 러닝 모델을 제안하고자 한다. 특히, 폭우 환경에서 촬영된 저해상도 얼굴 영상에서 빗줄기와 빗줄기 축적 현상을 제거하기 위한 폭우 제거 모듈, 구조화된 얼굴 영상을 분할하기 위한 얼굴 파싱 모듈, 그리고 저해상도 얼굴 영상에서 고해상도로 변환하기 위한 초고해상화 모듈을 제안하고자 한다. 실험 결과를 통해, 제안한 얼굴 초고해상화 기법이 저가시성·저해상도인 얼굴 영상에서도 얼굴 인식이 가능한 선명한 고해상도 영상을 복원할 수 있음을 보이고자 한다.

Abstract

Recently, intelligent surveillance cameras for recognizing objects and analyzing scene situations such as intrusion detection, object counting, suspect detection and abnormal behavior detection are being actively studied. Since most of these cameras are installed outdoors, they are under various shooting environments. In particular, in heavy rain conditions, face recognition becomes difficult because visibility of a captured image is significantly reduced due to rain streaks or rain accumulation. Therefore, in this paper, we propose a face super-resolution deep learning model for improving face recognition in heavy rain conditions. Specifically, we propose a heavy rain removal module for removing rain streaks and rain accumulation from a low-resolution face image taken in a heavy rain condition, a face parsing module for segmenting a structured face image, and a high-resolution module for converting a low-resolution face image to a high-resolution one. Through the experimental results, we want to show that the proposed super-resolution method can restore a clear, high-resolution image that can recognize a face even from a low-visibility, low-resolution face image.

Keywords

image restoration, super resolution, rain removal, generative adversarial network, intelligent cctv

* 군산대학교 소프트웨어융합공학과 학사과정
- ORCID: <https://orcid.org/0000-0002-8019-6693>
** 군산대학교 소프트웨어융합공학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0001-7077-3074>

· Received: Jul. 13, 2021, Revised: Aug. 20, 2021, Accepted: Aug. 23, 2021
· Corresponding Author: Chang-Hwan Son
Department of Software Convergence Engineering, Republic of Korea
Tel.: +82-63-469-8915, Email: cson@kunsan.ac.kr

I. 서 론

단순형 감시카메라는 센서의 해상도를 높여서 고선명의 영상을 획득하는 것을 목적으로 한다. 하지만 최근에는 이런 단순형 감시카메라에 인공지능을 접목한 지능형 감시카메라가 개발되고 있다. 지능형 카메라는 인간의 시각 기능뿐만 아니라 센싱된 정보로부터 객체 인식 및 상황 추론까지 겸비한 카메라를 말한다. 이런 지능형 감시카메라는 입력 영상에서 이상행동 검출, 침입탐지, 방화조기발견, 객체 카운팅, 용의자 얼굴 검출 등과 같이 인간의 고차원적인 사고 및 추론 능력을 모방할 수 있다.

하지만 이러한 지능형 감시카메라는 대부분 실외에 설치되므로 다양한 촬영 환경에 놓이게 된다. 예를 들어, 폭우 환경에서 촬영된 영상은 빗줄기(Rain streaks)[1]나 빗줄기 축적(Rain accumulation)[2] 현상으로 인해 영상의 가시성이 현격히 저하된다. 또한 야간 환경에서는 저조도로 인해 객체의 밝기 및 색상의 구별력이 떨어지게 된다. 이러한 야외 촬영 환경에서의 저가시성 및 저조도 문제는 지능형 감시카메라 개발의 걸림돌로 작용하고 있다. 따라서 본 논문에서는 야외 환경에서의 지능형 감시카메라의 화질 성능을 개선할 수 있는 새로운 연구주제와 기술을 선보이고자 한다. 특히 폭우 환경에서 수반되는 저가시성·저해상도 열화 문제를 해결 가능한 얼굴 초고해상화 기법(Face super resolution)을 처음으로 소개하고자 한다.

제안한 폭우 환경에서의 얼굴 영상 초고해상화 기법은 다음과 같이 네 가지의 기여도를 갖는다.

본 연구를 통해 폭우 환경에서의 얼굴 영상 초고해상화 기법을 처음으로 선보이고자 한다. 저자가 조사한 바에 따르면, 폭우 환경에서 수반되는 저가시성·저해상도 열화 문제를 제기하고 얼굴 해상도를 개선한 연구는 아직 없다. 따라서 본 연구에서 소개된 얼굴 초고해상화 기법은 향후 베이스 모델로 자리 잡을 것으로 기대된다.

제안한 얼굴 초고해상화 기법은 빗줄기 및 빗줄기 축적 제거와 저해상도 열화 문제를 동시에 해결 가능한 딥 러닝 아키텍처를 제안한다. 특히, 폭우 제거(Heavy rain removal) 모듈과 얼굴 초고해상화 모듈을 결합한 조인트 학습(Joint learning) 방식을

소개한다. 또한 얼굴 구성 요소인 눈, 코, 입 영역을 분할 가능한 얼굴 파싱(Face parsing) 모듈을 도입하여 최종 고해상도 변환 시 눈, 코, 입의 선명도를 개선할 수 있음을 보이고자 한다.

폭우 환경에서의 얼굴 영상 생성 모델을 제시한 후, 모델링 기반으로 생성된 저가시성·저해상도 얼굴 영상 데이터 셋을 공개하고자 한다. 아직까지 폭우 환경을 모델링한 얼굴 영상 데이터 셋은 부재인 상황이다. 따라서 본 연구의 결과물인 저가시성·저해상도 얼굴 영상은 향후 연구목적용으로 널리 활용될 것으로 기대된다.

본 연구를 통해 제안한 얼굴 초고해상화 기법이 저가시성·저해상도 얼굴 영상에서도 얼굴 인식이 가능한 고해상도 영상을 복원할 수 있음을 확인하고자 한다. 또한 정량적·정성적 평가에서도 최신식 초고해상화 기법보다 더 우수한 성능을 달성할 수 있음을 보이고자 한다.

II. 관련 연구

2.1 빗줄기 제거 기법

빗줄기 제거(Rain removal)[3]란 폭우 환경에서 촬영된 영상에서 빗줄기를 제거하여 선명한 영상을 복구하는 과정을 말한다. 이러한 빗줄기 제거는 객체의 특징 추출을 방해하는 빗줄기를 제거하여 객체 검출의 성능을 개선할 수 있다. 즉, 빗줄기 제거는 객체 검출의 전처리 과정으로 활용된다.

이러한 빗줄기 제거 기법은 빗줄기 생성 모델에 따라 크게 두 가지로 분류될 수 있다. 첫째는 가산 혼합 모델[4]이고 둘째는 빗줄기 축적 모델[5]이다. 가산 혼합 모델은 빗줄기 영상을 원본 영상과 빗줄기 레이어(Rain layer)의 합으로 표현한다. 반면 빗줄기 축적 모델은 가산 혼합 모델에 헤이즈 효과처럼 가시성 저하를 유도할 수 있는 생성 과정을 추가한 모델이다. 초기 빗줄기 제거 기법은 가산 혼합 모델에 기반해서 개발되었지만, 최근에는 빗줄기 축적 효과까지 반영한 좀 더 복잡한 기법들이 연구되고 있다. 빗줄기 제거 기법은 얇은 학습인 스파스 코딩[6][7] 및 가우시안 혼합 모델[8]이 주류를 이루다가 최근에는 딥 러닝으로 전환되었다. 딥 러닝 기

반의 주요 아키텍처로는 디테일 네트워크[9], 구조 인식(Structure-aware) 네트워크[1], 밀도 인식(Density-aware) 네트워크[10], 스케일 인식(Scale-aware) 네트워크[11], 점진적(Progressive) 네트워크[12], 폭우 제거 네트워크[5]가 있다.

2.2 초고해상화 기법

초고해상화(Super resolution)란 저해상도 영상에서 고해상도 영상으로 변환하는 기법이다. 초고해상화 기술은 보간법[13]과 같은 적응적 필터링 기반의 고전적인 접근 방식으로 시작해서 얇은 학습인 사전 학습(Dictionary learning)[14]을 사용하다가 최근에는 딥 러닝이 대세를 이루고 있다. 딥 러닝 기반의 초고해상화 기술은 픽셀 및 인지 주도(Perceptual driven)로 분류될 수 있다. 픽셀 주도 초고해상화 기술은 픽셀의 정확한 복원력을 목적으로 삼는다. 반면 인지 주도 초고해상화 기술은 인지화질(Perceptual image quality)을 개선하는 것을 목적으로 한다. 픽셀 주도 초고해상화 기법의 주요 모델로는 VDSR(Very-Deep Super-Resolution)[15], RCAN(Residual Channel Attention Network)[16]이 있다. 그리고 인지 주도 초고해상화 기법의 대표적인 모델로는 SRGAN[17], ESRGAN(Enhanced Super-Resolution Generative Adversarial Networks)[18], SPSR(Structure-Preserving Super Resolution with Gradient Guidance)[19] 등이 있다.

III. 인지 주도의 초고해상화 딥 러닝 모델

이 장에서는 최근 초고해상화 기법에서 각광을 받고 있는 인지 주도 기반의 딥 러닝 모델인 SRGAN에 대해 소개하고자 한다. SRGAN은 기존의 생성적 적대 신경망(Generative adversarial network)을 초고해상화에 접목한 기술이다[17]. 생성적 적대 신경망은 생성자(Generator)와 판별자(Discriminator)라 불리는 두 개의 네트워크로 구성되어 있고 서로 적대적인 관계로 학습을 한다. 컴퓨터 비전 분야의 대표적인 생성적 적대 신경망의 모델로는 Pix2Pix[20], CycleGAN[21] 등이 있다.

SRGAN 모델은 기존의 Pix2Pix 모델을 기반으로 초고해상화 기술에 접목하기 위해 인지 손실(Perceptual loss)을 새롭게 도입하였다. 기존의 널리 사용되던 PSNR(Pixel Signal to Noise Ratio) 손실의 단점인 텍스처 복원 능력을 개선하기 위함이다. 인지 손실은 사전 학습된 VGG(Visual Geometry Group) 네트워크[22]를 통과한 특징 맵의 오차를 최소화하는 것을 목표로 한다. VGG 네트워크는 이미지넷(ImageNet)이라 불리는 대용량 영상 데이터를 분류하기 위해 사전 학습되었기 때문에, 인지 손실은 마치 인간 시각에서 처리되는 특징 추출 과정을 대체해서 화질을 측정하는 것으로 해석될 수 있다.

그림 1은 초고해상화 실현을 위한 SRGAN의 아키텍처를 보여 주고 있다.

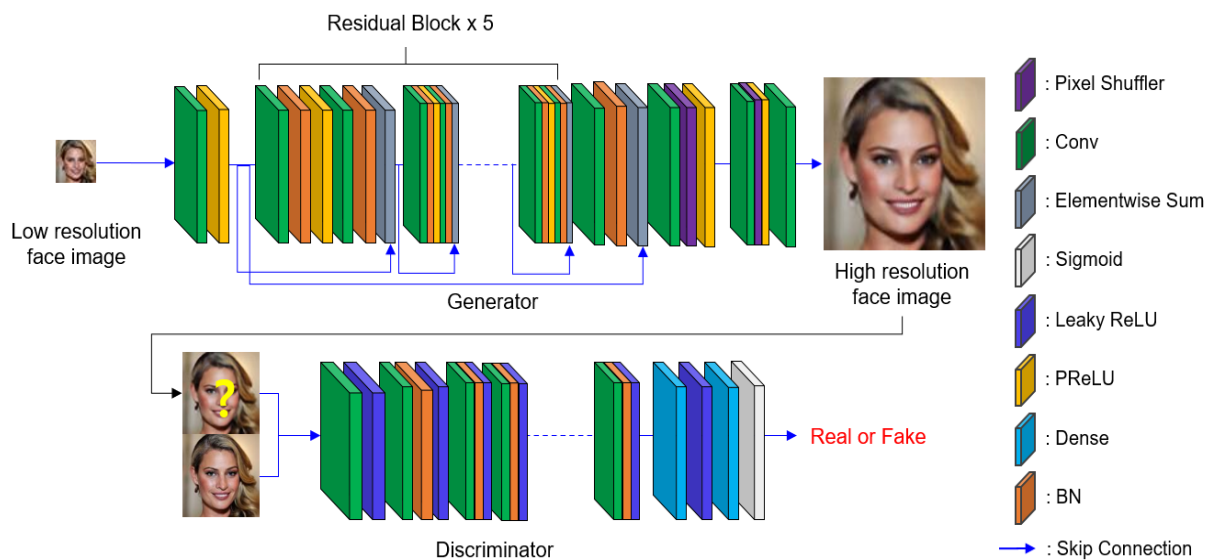


그림 1. 생성적 적대 신경망 기반의 초고해상화 기법[17]

Fig. 1. Super resolution based on generative adversarial networks(SRGAN)[17]

그림에서 보듯이, SRGAN은 생성자와 판별자로 구성되어 있다. 생성자는 크게 잔차 블록(Residual block) 5개, 연결 레이어(Skip connection), 그리고 업샘플링 블록으로 구성된다. 여기서 블록이란 레이어의 묶음을 의미한다.

잔차 블록은 합성곱 레이어(Conv, Convolution layer), 배치 정규화(BN, Batch Normalization) 레이어, 매개변수 정류선형유닛((PReLU, Parametric Rectified Linear Unit) 레이어, 합성곱 레이어, 배치 정규화 레이어로 구성된다. 그리고 블록의 입력 값과 출력 값을 연결 레이어를 통해 더한 형태의 구조를 가진다. 즉, ResNet(Residual neural Network)[23]에서 사용된 잔차 구조를 블록 단위로 확장했다고 보면 된다. 그리고 업샘플링 블록은 특징 맵의 해상도를 증가시키는 역할을 한다. 업샘플링 블록은 합성곱 레이어, 픽셀 셔플러(Pixel shuffler), 그리고 매개변수 정류선형유닛 레이어로 구성된다. 이러한 업샘플링을 통해 영상을 2배로 증가할 수 있다. 총 업샘플링 블록은 2개이다. 이는 입력 영상의 크기를 4배로 증가시킨다는 것을 말한다.

본 연구에서는 32×32 입력 저해상도 영상을 128×128 크기의 고해상도 영상으로 변환한다. 판별자는 합성곱 레이어, 배치 정규화 레이어, 기울기 정류선형유닛(Leaky ReLU) 레이어를 쌓은 후, 밀집(Dense) 레이어를 거쳐 최종 시그모이드(Sigmoid) 레이어를 통과한다. 여기서 밀집 레이어란 완전 연결(Fully connected) 레이어를 말한다.

밀집 레이어는 생성자의 가짜 영상과 원본 영상을 분류하기 위한 하나의 노드 개수로 맞춰 주고, 시그모이드 레이어는 이진 분류를 위해 0에서 1사이의 확률 값으로 매핑한다.

IV. 제안한 폭우 환경에서 촬영된 얼굴 영상의 초고해상화

4.1 제안한 연구 동기

제안한 폭우 환경에서의 얼굴 영상 초고해상화 기법은 폭우 환경에서 수반된 저가시성과 저해상도 문제를 동시에 해결하여 고해상도의 얼굴 영상을 복원하는 과정을 말한다. 관련연구에서 소개된 기존

의 빗줄기 제거와 초고해상화를 결합한 과정으로 이해하면 된다. 기존의 빗줄기 제거 및 폭우 제거는 빗줄기나 빗줄기 축적에 따른 저가시성 문제만을 해결하고 초고해상화 기법은 저해상도 문제만을 다룬다. 하지만, 실제 폭우 환경에서 촬영된 영상은 저가시성 문제와 저해상도 문제가 동시에 발생할 수 있다. 얼굴 영상의 경우 가시성과 해상도가 동시에 나빠질 수 있는 대표적인 사례이다. 뿐만 아니라 얼굴 영상은 눈, 코, 입과 같은 구성 요소로 이루어진 구조화된 특성을 가지기 때문에 얼굴 파싱을 통해 얼굴 특징의 구별력을 제고할 수 있다. 따라서 본 연구에서는 폭우 환경에서 촬영된 영상의 저가시성·저해상도 문제를 동시에 해결하고 얼굴의 구조화 특성을 활용할 수 있는 얼굴 초고해상화 기법을 제안하고자 한다. 특히 폭우 환경에서의 저해상도 얼굴 영상 생성 모델을 소개한 후, 폭우 제거, 얼굴 파싱, 초고해상화 모듈로 구성된 제안한 아키텍처를 제시하고자 한다.

4.2 저해상도 폭우 영상 생성 모델

각 모듈을 설명하기 전에 먼저 폭우 환경에서의 저해상도 얼굴 영상 생성 과정에 대해 설명하고자 한다. 본 연구에서는 지도 학습 기법을 사용하기 때문에 원본 영상을 수집해야 된다. 하지만 실제 폭우 환경에서 촬영된 영상은 원본 영상이 존재하지 않는다. 따라서 폭우 환경에서의 저해상도 영상 생성 과정을 모델링할 필요가 있다. 폭우 환경에서의 저해상도 영상 생성 과정은 크게 저해상도 변환 과정과 폭우 영상 생성 과정으로 이루어진다. 첫 번째 단계는 고해상도 원본 영상을 저해상도 영상으로 변환하는 과정이다[13].

$$J = (H \otimes K) \downarrow \quad (1)$$

여기서 H 와 J 는 각각 원본 고해상도 영상과 저해상도 영상을 의미하고 K 는 가우시안 필터이다. 그리고 \otimes 와 \downarrow 는 각각 합성곱 연산(Convolution)과 다운 샘플링을 의미한다.

두 번째 단계는 저해상도 영상 J 에 폭우 환경에서 동반될 수 있는 시각적 효과를 추가하는 과정이

다[2][5]. 폭우 환경에서 촬영된 영상에서는 빗줄기와 빗줄기 축적 효과가 나타난다. 수식적으로 다음과 같이 표현된다.

$$I = T \odot (J + \sum_i^m S_i) + (1 - T) \odot A \quad (2)$$

여기서 I 는 최종 생성된 저해상도 폭우 영상이고 J 는 식 (1)의 저해상도 영상에 대응한다. 그리고 S_i 는 i 번째의 빗줄기만을 포함한 레이어를 말한다. 그리고 T 는 투과율 맵(Transmission map)이고 A 는 대기 산란광(Airlight)을 의미한다. \odot 는 원소 별 곱에 해당한다. 식 (2)는 저해상도 영상 J 에 빗줄기 레이어를 더하고 대기 산란광과의 가중치 합을 통해 저해상도 폭우 영상 I 를 만들 수 있다는 것을 의미한다.

그림 2는 폭우 환경에서의 저해상도 얼굴 영상 생성 과정을 보여 주고 있다. 맨 좌측의 원본 고해상도 영상은 해상도 변환(Resolution conversion)과 폭우 생성(Heavy rain generation) 과정을 통해 최종 저해상도 폭우 영상으로 변환된다. 그림 2에서 생성된 저해상도의 얼굴 폭우 영상은 빗줄기 축적 효과로 인해 영상 전체가 뿌옇게 된 것을 볼 수 있다. 또한 빗줄기도 포함된 것을 볼 수 있다. 생성된 얼굴 폭우 영상은 저해상도와 저가시성 특성을 동시에 지닌다.

식 (1)과 (2)가 주어졌을 때, 본 연구의 목적은 저해상도 얼굴 폭우 영상 I 에서 고해상도 영상 H 를 복원하는 것이다.

$$H = f_{\theta}(I) \quad (3)$$

역문제(Inverse problem)를 풀기 위해 본 연구에서는 학습 가능한 함수 f_{θ} 를 추정하고자 한다. f_{θ} 는 신경망으로 구현되고 다시 세 개의 모듈인 폭우 제거, 얼굴 파싱, 초고해상화 모듈로 이루어진다.

$$f_{\theta} = \{f_{\theta_n}, f_{\theta_f}, f_{\theta_s}\} \quad (4)$$

$f_{\theta_n}, f_{\theta_f}, f_{\theta_s}$ 는 각각 폭우제거, 얼굴 파싱, 초고해상화 모듈을 의미한다.

4.3 폭우 제거 모듈

폭우 제거는 저해상도 얼굴 폭우 영상에서 폭우를 제거하는 과정을 말한다. 식 (2)의 역과정이다.

$$\hat{J} = f_{\theta_n}(I) \quad (5)$$

f_{θ_n} 는 입력 저해상도의 얼굴 폭우 영상 I 를 폭우가 제거된 \hat{J} 영상으로 변환한다. 여기서 \hat{J} 은 추정된 저해상도의 얼굴 영상이다. f_{θ_n} 를 학습하기 위해 손실은 아래와 같이 정의된다.

$$L(\hat{I}, I, J, \hat{J}) = \|\hat{J} - J\|_2^2 + \gamma_1 \|\hat{I} - I\|_2^2 \quad (6)$$

L 은 손실함수이며 두 개의 항목으로 구성된다. 첫 번째 항목은 추정된 저해상도 얼굴 영상 \hat{J} 과 원본 저해상도 영상 J 와의 오차를 측정한다. 두 번째 항목은 입력 저해상도 얼굴 영상 I 와 복원된 저해상도 얼굴 영상 \hat{I} 과의 오차를 측정한다.

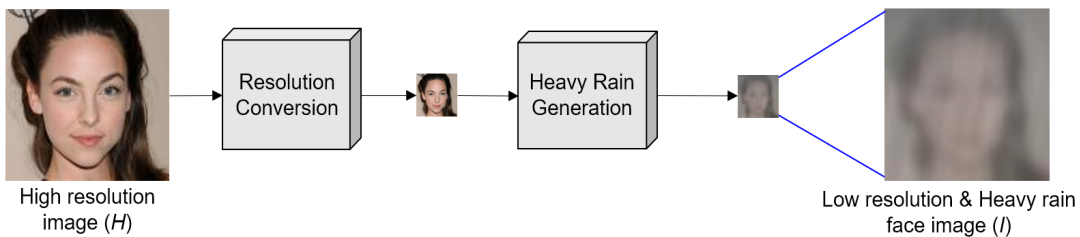


그림 2. 저해상도 얼굴 폭우 영상 생성
Fig. 2. Low resolution and heavy rain face image generation

γ_1 는 두 항목에 대한 가중치이고 $\|\cdot\|_2^2$ 는 유클리디언 놈(Norm)의 제곱이다. \hat{I} 은 다음과 같이 추정된 파라미터를 사용해서 복원될 수 있다.

$$\hat{I} = \hat{T} \odot (J + \sum_i^m \hat{S}_i) + (1 - \hat{T}) \odot \hat{A} \quad (7)$$

여기서 $\hat{S}_i, \hat{T}, \hat{A}$ 은 추정된 빗줄기 레이어, 투과율 맵, 대기 산란광이다. 원본 저해상도 영상 J 는 주어 젤 있기 때문에 추정된 $\hat{S}_i, \hat{T}, \hat{A}$ 을 사용해서 \hat{I} 을 복원할 수 있다.

식 (6)과 (7)에 따르면, $\hat{S}_i, \hat{T}, \hat{A}, \hat{J}$ 총 4개의 파라미터를 추정하기 위한 네트워크가 필요하다. 본 연구에서는 4개의 파라미터를 추정하기 위해, 밀집 네트워크(Dense network)[24]를 사용했다.

$$\hat{S} = f_{\theta_{DS}}(I), \quad \hat{T} = f_{\theta_{DT}}(I), \quad \hat{A} = f_{\theta_{DA}}(I) \quad (8)$$

여기서 $f_{\theta_{DS}}, f_{\theta_{DT}}, f_{\theta_{DA}}$ 는 $\hat{S}_i, \hat{T}, \hat{A}$ 을 추정하기 위한 밀집 네트워크이다.

그림 3은 제안한 폭우 제거 모델을 보여 준다. 그림에서 보듯이, 특징을 추출하기 위한 인코더와 4개의 파라미터를 추정하기 디코더로 구성되어 있다.

$$\hat{S} = f_{\theta_{DS}}(f_{\theta_E}(I)) \quad (9)$$

$$\hat{T} = f_{\theta_{DT}}(f_{\theta_E}(I)) \quad (10)$$

$$\hat{A} = f_{\theta_{DA}}(f_{\theta_E}(I)) \quad (11)$$

$$\hat{J} = f_{\theta_{DJ}}(f_{\theta_E}(I)) \quad (12)$$

여기서 f_{θ_E} 는 인코더이고 $f_{\theta_{DJ}}$ 는 \hat{J} 을 추정하기 위한 디코더이다. 이는 f_{θ_h} 가 결국 총 5개의 파라미터로 $\theta_h = \{\theta_E, \theta_{DS}, \theta_{DT}, \theta_{DA}, \theta_{DJ}\}$ 로 구성된다는 것을 말한다. 그리고 이 파라미터는 식 (6)에 정의된 손실을 최소화함으로써 구해진다.

4.4 얼굴 파싱 모듈

얼굴 영상은 자연 영상(Natural images)과는 달리, 눈, 코, 입과 같은 구성 요소로 이루어진 구조화된 특성을 가진다. 따라서 얼굴 파싱 모듈을 통해 구성 요소를 구분할 수 있다면 추후 초고해상화 모듈에서 얼굴 특징의 구별력을 높여서 최종 얼굴 영상을 복원력을 개선하는 데 도움을 줄 수 있다.

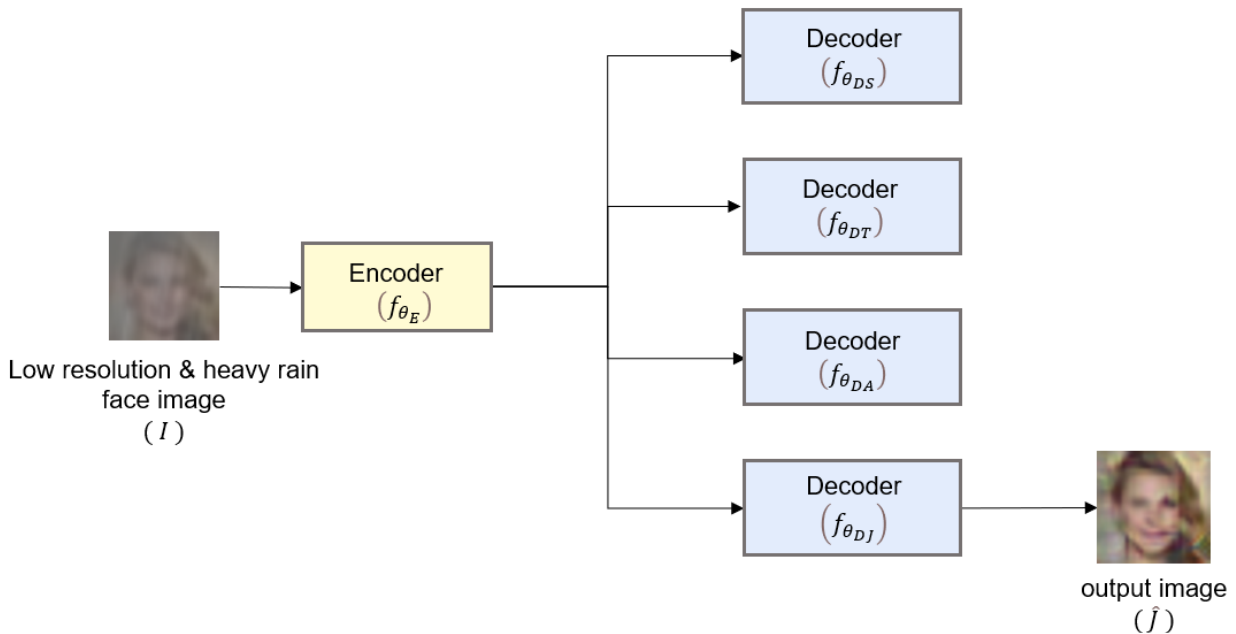
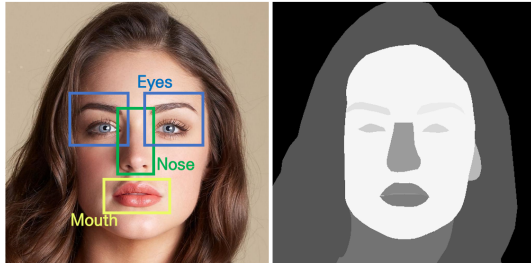


그림 3. 폭우 제거 모듈
Fig. 3. Heavy rain removal module

얼굴 파싱 과정은 다음과 같이 표현된다.

$$\hat{F} = f_{\theta_f}(I, \hat{J}) \quad (13)$$

\hat{F} 은 추정된 얼굴 파싱 영상이며 눈, 코, 입, 머리, 눈썹 등을 포함한 총 18개의 영역으로 분할된다. 그리고 f_{θ_f} 는 얼굴 파싱 모듈이고 U-Net[25]으로 구현되었다. f_{θ_f} 의 입력 폭우 얼굴 영상(I)과 추정된 저해상도 영상(\hat{J})이다. f_{θ_f} 학습을 위해 평균제곱오차를 손실함수로 사용하였다. 참고로 그림 4는 원본 얼굴 파싱 영상(F)의 예시이다.



(a) 원본 영상 (b) 파싱된 영상
그림 4. 얼굴 파싱

Fig. 4. Face parsing, (a) Original image, (b) Parsed image

4.5 초고해상화 모듈

초고해상화 모듈은 I , \hat{J} , \hat{F} 을 사용해서 최종 고해상도 영상인 \hat{H} 을 복원하는 과정이다.

$$\hat{H} = f_{\theta_s}(I, \hat{J}, \hat{F}) \quad (14)$$

여기서 f_{θ_s} 는 초고해상화 모듈이며 SRGAN 아키텍처를 기본으로 한다. 기존의 SRGAN의 차이점은 입력 영상이 I , \hat{J} 으로 구성되고 업샘플링 레이어를 통과하기 전의 특징 맵을 얼굴 파싱 모듈의 결과인 \hat{F} 와 퓨전한다는 점이다.

f_{θ_s} 를 학습하기 위한 사용된 손실 함수는 다음과 같다.

$$L = L_{MSE} + \gamma_2 L_A + \gamma_3 L_P \quad (15)$$

손실 함수는 총 3개의 항목으로 구성되고 각각의 항목은 다음과 같이 계산된다.

$$L_{MSE} = \frac{1}{2} \sum_{i=1}^n (H_i - \hat{H}_i)^2 \quad (16)$$

L_{MSE} 항목은 평균제곱오차를 나타내고 원본 고해상도 영상 H 와 추정된 고해상도 영상 \hat{H} 과의 오차를 측정한다. i 는 배치의 샘플 영상을 의미한다.

$$L_A = \sum_{i=1}^n -\log f_{\theta_{s,D}}(f_{\theta_{s,G}}(I, \hat{J}, \hat{H})) \quad (17)$$

L_A 는 적대적 손실(Adversarial loss)이다. 즉, 생성자 $f_{\theta_{s,G}}$ 가 판별자 $f_{\theta_{s,D}}$ 를 속이도록 유도한다.

$$L_P = \frac{1}{2} \sum_{i=1}^n (VGG(H) - VGG(\hat{H}))^2 \quad (18)$$

L_P 는 인지 손실을 의미하며, 사전학습된 VGG 네트워크를 통과한 특징 맵의 오차를 측정한다. 이 논문에서는 VGG-16 네트워크를 사용했고 5번째 풀링 레이어에서 특징 맵을 추출했다.

그림 5는 제안한 폭우 환경에서의 얼굴 영상 초고해상화 기법의 전체 네트워크를 보여 준다. 먼저, 폭우 제거 모듈은 입력 저해상도-저가시성 폭우 영상 I 를 사용해서 빗줄기 레이어 \hat{S}_i , 투과율 맵 \hat{T} , 대기 산란광 \hat{A} , 저해상도 영상 \hat{J} 을 각각 추정한다. 그리고 입력 폭우 영상 I 와 저해상도 영상 \hat{J} 이 얼굴 파싱 모듈을 통과한 후, 얼굴 파싱된 결과 영상 \hat{F} 을 출력한다. 마지막으로 폭우 영상, 추정된 얼굴 파싱 영상, 추정된 저해상도 영상이 초고해상화 모듈을 거쳐 최종 고해상도 영상 \hat{H} 을 복원한다.

그림 5에서의 밀집 블록(Dense block)의 아키텍처는 밀집 네트워크의 변형이며 자세한 사항은 참고 문헌[24]을 보기 바란다. 그리고 그림 5의 제안한 전체 네트워크를 효율적으로 학습하기 위해, 폭우 제거 모듈과 얼굴 파싱 모듈을 사전 학습 후에 전체 네트워크를 학습하였다.

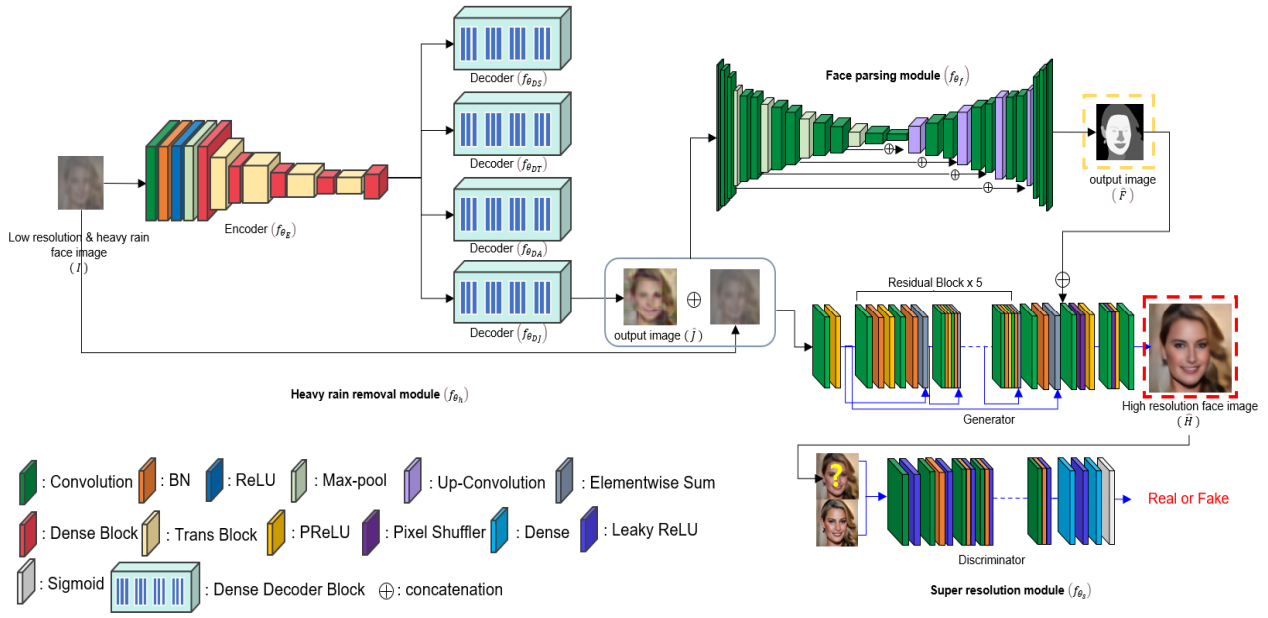


그림 5. 제안한 전체 네트워크 구조
 Fig. 5. Architecture of the proposed entire network

V. 실험 및 결과

5.1 실험환경

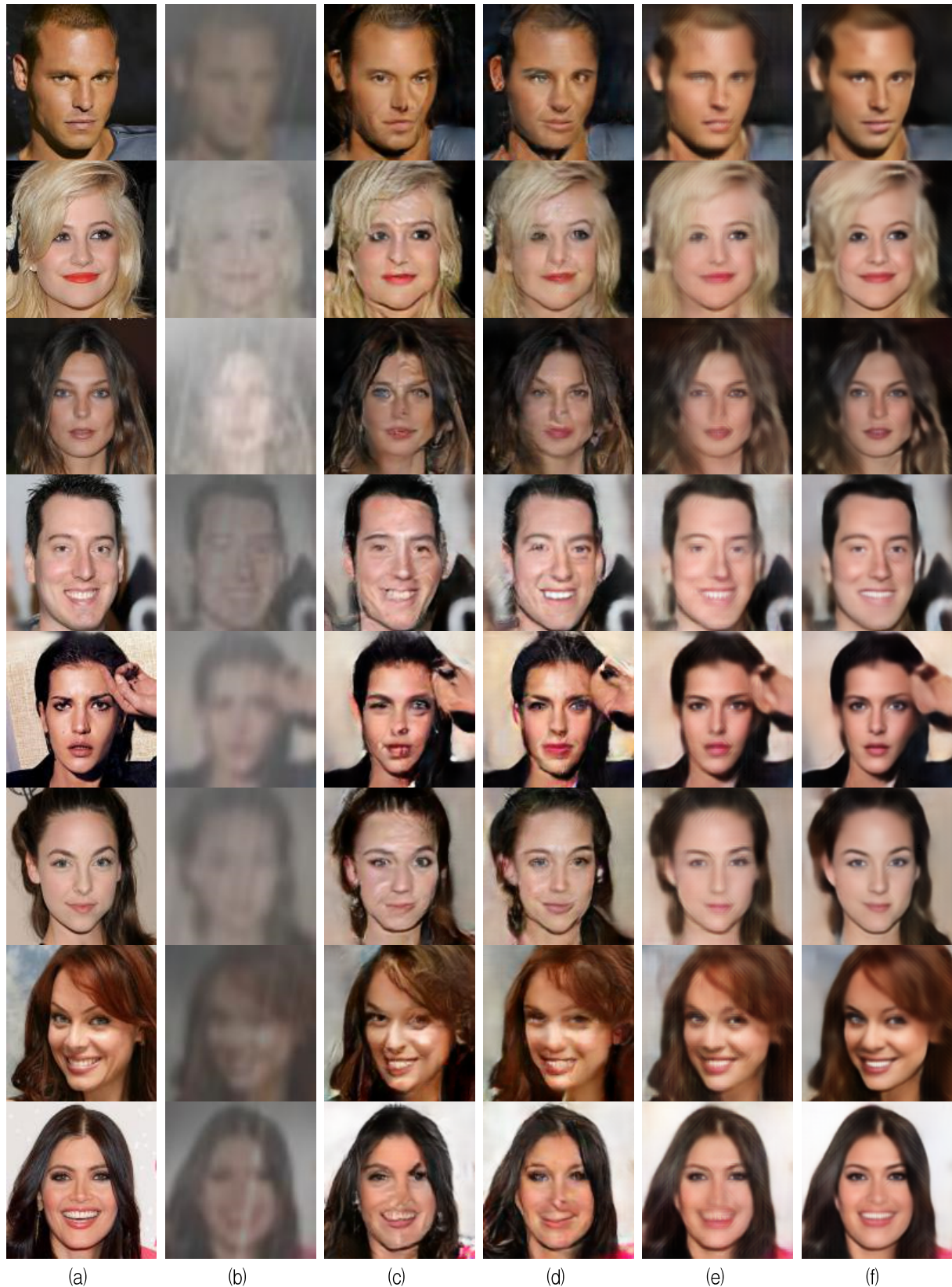
전체 네트워크를 학습하기 위해, 유명인의 얼굴 영상으로 구성된 CelebA-HQ[26]를 사용했다. 총 30,000장으로 구성된 데이터 중에서 18,000장을 훈련 집합으로 그리고 1,800장과 100장을 검증 및 테스트 집합으로 사용했다. 식 (1)의 다운샘플링 배율은 4로 설정했다. 그리고 손실함수의 가중치 값인 $\gamma_1, \gamma_2, \gamma_3$ 는 각 $10^{-2}, 10^{-4}, 10^{-4}$ 로 설정했다. 전체 네트워크 학습을 위해 배치 크기는 32, 에폭 (Epochs)은 100회로 설정했다. 빗줄기 레이어의 개수 m 은 하나로 고정했다. 사용한 딥 러닝 툴은 PyTorch이고 윈도우10 운영체제에서 실행했다.

5.2 정성적 평가

제안한 폭우 환경에서의 얼굴 영상 초고해상화 기법의 성능을 평가하기 위해, 폭우 제거 기법[5], SRGAN 초고해상화 기법[17], Pix2Pix[20] 기법과 비교하였다. 그림 6은 기존과 제안한 기법을 사용해서 획득된 결과 영상을 보여 준다. 좌측 두 번째 영상은 입력 저해상도·저가시성 얼굴 영상이다. 폭우 환

경에서 생성 가능한 얼굴 영상을 모델링한 결과로써, 얼굴 인식이 거의 불가능하다. 하지만, 이러한 얼굴 인식이 어려운 입력 얼굴 영상일지라도 제안한 기법을 적용했을 때, 얼굴 인식이 가능한 얼굴 영상을 복원할 수 있다.

예를 들어, 첫 번째 영상에서 제안한 기법은 눈, 코, 입과 같은 얼굴 구성 요소를 선명하게 복원한 것을 볼 수 있다. 또한 표정, 입꼬리, 얼굴형과 같은 세밀한 부분도 잘 표현하였고 원본 얼굴과 유사하도록 복원하였다. 즉, 저가시성이면서 저해상도인 얼굴 영상에서도 얼굴 인식이 가능한 얼굴 복원 능력이 있음을 입증해 준다. 반면 기존의 기법들은 이목구비가 뚜렷하지 않거나 실령 얼굴이 복원되었다 할지라도 얼굴 왜곡이 나타난다. 또한 얼굴 인식 성능이 여전히 낮은 것을 알 수 있다. 이는 기존의 기법은 폭우 제거 또는 고해상도 개선에 적합하도록 설계되었기 때문이다. 반면 제안한 기법은 상대적으로 높은 얼굴 인식 성능을 보인다. 제안한 기법은 폭우 제거와 고해상도를 동시에 개선할 수 있는 조인트 학습 접근 기법을 채택했기 때문이다. 또한 얼굴 파싱 모듈을 도입해서 눈, 코, 입과 같은 구성 요소를 분리하여 특징의 구별력을 높여 최종 얼굴 복원 능력을 개선했기 때문이다. 나머지 결과 영상들도 시각적으로 유사한 효과가 나타나는 것을 알 수 있다.



(a) (b) (c) (d) (e) (f)
그림 6. 결과 영상, (a) 원본 영상, (b) 입력 폭우 영상, (c) 폭우 제거 기법[5], (d) Pix2Pix[20], (e) SRGAN[17],
(f) 제안한 기법

Fig. 6. Resulting images, (a) Original image, (b) Input heavy rain image, (c) Heavy rain removal[5], (d) Pix2Pix[20],
(e) SRGAN[17], (f) Proposed method

5.3 정량적 평가

기존 기법과 제안한 기법의 성능을 정량적으로 평가하기 위해, 최대 신호 대 잡음 비(PSNR, Peak Signal-to-Noise Ratio)와 구조적 유사성(SSIM, Structural Similarity)[27]을 화질 평가 척도로 사용하였다. 최대 신호 대 잡음 비는 복원 영상과 원본 영상의 화소 값이 얼마나 유사한지를 판별하는 척도이고 구조적 유사성은 영상 구조, 즉 에지나 텍스처의 복원 능력을 평가하는 척도이다.

표 1은 테스트 집합에 대해 PSNR과 SSIM을 정량적으로 평가한 결과이다. 평가 결과를 보면, 제안한 기법을 적용한 복원 영상이 SRGAN에 비해 PSNR 수치가 약 0.25dB 정도가 개선되었고 SSIM 수치는 0.0129가 향상된 것을 볼 수 있다. 또한 전체적으로 기존 기법보다 더 개선된 것을 볼 수 있다. 이 결과는 제안한 기법이 원본 영상과 더 유사하게 복원했고 눈, 코, 입의 구조적 선명도를 더 향상했음을 의미한다.

표 1. 정량적 평가

Table 1. Quantitative evaluation

	PSNR	SSIM
Heavy rain removal [5]	20.8678	0.5640
Pix2Pix [20]	21.8714	0.6052
SRGAN [17]	22.4878	0.6852
Proposed method	22.7340	0.6981

VI. 결 론

본 연구에서는 폭우 환경에서 수반된 저가시성·저해상도 문제를 동시에 해결 가능한 얼굴 초고해상화 기법을 제안하였다. 제안한 기법은 폭우 제거, 얼굴 파싱, 초고해상화 모듈로 구성된 새로운 아키텍처를 제시하였고 폭우와 저해상도 문제를 동시에 해결 가능한 조인트 학습을 소개하였다. 실험 결과에서 제안한 기법은 얼굴 인식이 거의 어려운 입력 얼굴 영상에서도 눈, 코, 입 및 얼굴 윤곽까지도 정확하게 복원할 수 있음을 보여 주었다.

향후 연구로는 실제 폭우 영상에서 촬영된 얼굴 영상 데이터 셋을 수집해서 제안한 기법의 성능을

평가하고자 한다. 아직 제안한 기법은 실제 폭우 얼굴 영상에 대해서는 몇 가지 한계점을 지니고 있다. 첫째는 폭우 영상 모델링이 실제 폭우 생성 모델과 차이가 있다는 점이고 두 번째는 실제 환경에서는 얼굴 촬영 각도가 다양하기 때문에 기하학적인 왜곡 문제가 발생한다는 점이다. 향후에는 이러한 문제까지 개선할 수 있도록 제안한 기법을 고도화하고자 한다.

References

- [1] J. C. Kim and C. H. Son, "Structure-aware residual network for rain streak removal", The Journal of Korean Institute of Information Technology, Vol. 18, No. 10, pp. 87-100, Oct. 2020, <https://doi.org/10.14801/jkiit.2020.18.10.87>.
- [2] P. H. Ye and C. H. Son, "Image captioning via semantic visual feature matching in heavy rain condition", The Journal of Korean Institute of Information Technology, Vol. 19, No. 5, pp. 19-29, May 2021, <https://doi.org/10.14801/jkiit.2021.19.5.19>.
- [3] C. H. Son and X. P. Zhang, "Rain detection and removal via shrinkage-based sparse coding and learned rain dictionary", Journal of Imaging Science and Technology, Vol. 64, No. 3, pp. 30501-1-30501-17(17), May 2020. <https://doi.org/10.2352/J.ImagingSci.Technol.2020.64.3.030501>.
- [4] L. W. Kang, C. W. Lin, and Y. H. Fu, "Automatic single-image-based rain streaks removal via image decomposition", IEEE Transactions on Image Processing, Vol. 21, No. 4, pp. 1742-1755, Apr. 2012. <http://dx.doi.org/10.1109/TIP.2011.2179057>.
- [5] R. Li, L. F. Cheong, and R. T. Tan, "Heavy rain image restoration: Integrating physics model and conditional adversarial learning", In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, pp. 1633-1642, Jun. 2019. <https://doi.org/10.1109/CVPR.2019.00173>.
- [6] Y. Luo, Y. Xu, and H. Ji, "Removing rain from a

- single image via discriminative sparse coding", in Proc. IEEE International Conference on Computer Vision, Santiago, Chile, pp. 3397-3405, Dec. 2015. <https://doi.org/10.1109/ICCV.2015.388>.
- [7] C. H. Son and X. P. Zhang, "Rain removal via shrinkage of sparse codes and learned rain dictionary", in Proc. IEEE International Conference on Multimedia and Expo Workshop, Seattle, WA, USA, pp. 1-6, Jul. 2016. <https://doi.org/10.1109/ICMEW.2016.7574752>.
- [8] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Single image rain streak decomposition using layer priors", IEEE Transactions on Image Processing, Vol. 26, No. 8, pp. 3874-3885, Aug. 2017. <http://dx.doi.org/10.1109/TIP.2017.2708841>.
- [9] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network", in Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 1-9, Jul. 2016. <https://doi.org/10.1109/CVPR.2017.186>.
- [10] H. Zhang and V. M. Patel, "Density-aware single image deraining using a multi-stream dense network", in Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, pp. 695-704, Jun. 2018. <https://doi.org/10.1109/CVPR.2018.00079>.
- [11] W. Yang, J. Liu, S. Yang, and Z. Guo, "Scale-free single image deraining via visibility-enhanced recurrent wavelet learning", IEEE Transactions on Image Processing, Vol. 28, No. 6, pp. 2948-2961, Jun. 2019. <https://doi.org/10.1109/tip.2019.2892685>.
- [12] 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>.
- [13] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion", IEEE Transactions on Image Processing, Vol. 15, No. 8, pp. 2226-2238, Aug. 2006. <https://doi.org/10.1109/TIP.2006.877407>.
- [14] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation", IEEE Transactions on Image Processing, Vol. 19, No. 11, pp. 2861-2873, May 2010. <https://doi.org/10.1109/tip.2010.2050625>.
- [15] J. Kim, J. L. Kwon, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks", In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, pp. 1646-1654, Jun. 2016. <https://doi.org/10.1109/CVPR.2016.182>.
- [16] Y. Zhang, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks", In Proc. European Conference on Computer Vision, Munich, Germany, pp. 286-301, Sep. 2018.
- [17] C. Leidig et al., "Photo-realistic single image super-resolution using a generative adversarial network", In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, pp. 4681-4690, Jul. 2017. <https://doi.org/10.1109/CVPR.2017.19>.
- [18] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "Esrgan: Enhanced super-resolution generative adversarial networks", In Proc. European Conference on Computer Vision, Munich, Germany, pp. 63-79, Sep. 2018.
- [19] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, "Structure-preserving super-resolution with gradient guidance", In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, pp. 7769-7778, Jun. 2020. <https://doi.org/10.1109/CVPR42600.2020.00779>.
- [20] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks", In Proc. IEEE Conference

- on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, pp. 1125-1134, Jul. 2017. <https://doi.org/10.1109/CVPR.2017.632>.
- [21] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks", In Proc. IEEE International Conference on Computer Vision, Venice, Italy, pp. 2223-2232, Oct. 2017. <https://doi.org/10.1109/ICCV.2017.244>.
- [22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", Sep. 2014. <https://arxiv.org/abs/1409.1556>.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, pp. 770-778, Jun. 2016. <http://dx.doi.org/10.1109/CVPR.2016.90>.
- [24] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, pp. 4700-4708, Jun. 2016.
- [25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation", In Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, pp. 234-241, Oct. 2015.
- [26] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation", In Proc. International Conference on Learning Representations, Vancouver, Canada, Feb. 2018.
- [27] 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>.

저자소개

정 다 희 (Da-Hee Jeong)



2018년 3월 ~ 현재 : 군산대학교
소프트웨어융합공학과 학사과정
관심분야 : 컴퓨터 비전, 영상처리,
딥 러닝

손 창 환 (Chang-Hwan Son)



2017년 4월 ~ 현재 : 군산대학교
소프트웨어융합공학과 부교수
관심분야 : 컴퓨터 비전, 영상처리,
딥 러닝, 기계학습, 색 재현