

적응형 채널 어텐션 모듈을 활용한 복합 열화 복원 네트워크

이해윤^o 조성현*

대구경북과학기술원 정보통신융합전공

{haeyun, scho}@dgist.ac.kr

Image Restoration Network with Adaptive Channel Attention Modules for Combined Distortions

Haeyun Lee^o Sunghyun Cho*

Department of Information and Communication Engineering, DGIST

요약

자율 주행 자동차나 소방 로봇과 같은 시스템에서 영상을 얻을 때 다양한 요인들로 인해 잡음, 블러와 같은 열화가 발생한다. 이런 열화된 영상에 직접 영상 분류와 같은 기술을 적용하기 어려워 열화 제거가 불가피하나 이러한 시스템들은 영상의 열화를 인식할 수 없어서 열화된 영상을 복원하는데 어려움이 있다. 본 논문에서는 영상에 적용된 열화를 인지하지 못하는 상황에서 여러 방법들로 열화된 영상으로부터 자연스럽게 선명한 영상을 복원하는 방법을 제안한다. 우리가 제안한 방법은 딥러닝 모델에 채널 어텐션 모듈과 스킵 커넥션을 사용하여 영상에 적용된 열화에 따라 복원에 필요한 채널에 높은 가중치를 적용해 복합 열화 영상의 복원을 진행한다. 이 방법은 다른 복합 열화 복원 방법에 비해 학습이 간단하고 기존의 다른 방법들에 비해 높은 복합 열화 복원 성능을 낸다.

Abstract

The image obtained from systems such as autonomous driving cars or fire-fighting robots often suffer from several degradation such as noise, motion blur, and compression artifact due to multiple factor. It is difficult to apply image recognition to these degraded images, then the image restoration is essential. However, these systems cannot recognize what kind of degradation and thus there are difficulty restoring the images. In this paper, we propose the deep neural network, which restore natural images from images degraded in several ways such as noise, blur and JPEG compression in situations where the distortion applied to images is not recognized. We adopt the channel attention modules and skip connections in the proposed method, which makes the network focus on valuable information to image restoration. The proposed method is simpler to train than other methods, and experimental results show that the proposed method outperforms existing state-of-the-art methods.

키워드: 영상 복원, 딥러닝, 채널 어텐션, 합성곱 신경망

Keywords: Image restoration, Deep learning, Channel attention, CNN

1. 서론

열화된 영상을 원래 형태로 복원하는 것을 목표로 하는 영상 복원 문제는 컴퓨터 비전과 영상 처리 분야에서 가장 근본적인 문제 중 하나이다. 열화된 영상 y 는 일반적으로 $y = f(x) + n$ 으로 모델링 되며, 여기서 x 는 원본 영상, f 는 영상 열화 방법, n 은 표준편

차 σ 를 갖는 가산성 백색 가우시안 잡음 (Additive White Gaussian Noise)으로 가정된다. 영상 복원의 목표는 y 로부터 x 를 추정하는 것이다. f 의 정의에 따라 영상 복원은 다른 하위 문제들로 분류될 수 있다. 예를 들어, $f(x) = x$ 인 경우는 잡음 제거, $f(x)$ 가 블러 커널인 경우에는 블러 제거, $f(x)$ 가 Joint Photographic Experts Group (JPEG) 압축인 경우에는 JPEG 영상 복원 문제가 된다.

*corresponding author: Sunghyun Cho/DGIST(scho@dgist.ac.kr)

영상 복원의 중요성 때문에 간단한 필터링 방법부터 정교한 학습 기반의 방법에 이르기까지 수십년 동안 다양한 해결 방법들이 제안되어 왔다. 기존의 이미지 복원 방법은 필터링 기반 접근법 [1], 총 변량 (Total Variation) 기반의 방법 [2, 3], 자기 유사성 기반 방법 [4, 5, 6], 그리고 딥러닝 이전의 학습 기반 방법 [7, 8] 이 제안되었다. 딥러닝의 발전으로 많은 영상 복원 문제 [9, 10] 에 적용되었다. 딥러닝 기반의 방법들은 대규모 학습 데이터셋을 사용하여 열화된 영상으로부터 해당 복원된 영상으로 매핑하는 방법을 학습한다. 딥러닝의 능력 덕분에 고전적인 영상 복원 방법보다 훨씬 뛰어난 결과를 보여준다.

딥러닝을 사용하면서 기존의 방법들에 비해 높은 성능을 내지만, 대부분 딥러닝 기반의 영상 복원 방법들은 각 잡음 정도나 JPEG 압축 정도, 블러 정도에 따라 각각 학습을 해줘야 한다. 즉, 위에서 언급한 딥러닝 기반의 영상 복원 방법들은 한 가지 열화 방법에 특화되어 있다. 그렇기 때문에 열화된 영상을 복원시키려면 열화된 방법을 인지하여 그에 맞는 모델을 이용하여 복원을 해야한다. 하지만 자율주행 자동차나 로봇 시스템과 같은 영상 열화 방법을 인지할 수 없는 상황에는 영상 복원 방법을 적용하는데 무리가 있다. 또한 실제 환경에서는 잡음, 블러 등과 같은 다양한 열화들이 발생하기 쉽다. 즉, 이런 상황에서는 한 가지 열화에 특화된 복원 방법을 적용할 수 없기에, 다양한 열화에 대처가 가능한 영상 복원 방법이 필요하다.

복합 열화 복원을 위한 연구는 거의 진행되지 않았다. Yu *et al.* [11]는 강화학습을 사용하여, 네트워크 내에서 영상 열화를 인지하여 그에 맞는 열화 복원 네트워크를 사용하여 복합 열화를 복원하는 방법인 RL-Restore를 제안했다. 하지만 이 방법을 사용하기 위해서는 각 열화 정도 별로 복원 네트워크를 학습 시켜줘야하고, 학습된 열화별 복원 네트워크를 가지고 영상에 적용된 열화를 인지하기 위한 LSTM도 따로 학습을 시켜줘야한다. 이는 네트워크를 학습시키기 위해 많은 시간이 필요하고, 복합 영상 복원 성능 또한 좋지는 못하다. Saganuma *et al.* [12]는 복합 열화 복원을 위해 operation attention과 group attention을 적용하여 Operation-wise Attention Network (OWAN)을 제안했다. 한 레이어에서 총 8가지의 operation을 사용하여 특징 맵을 얻은 뒤, 열화된 영상으로부터 각 operation간의 어텐션 정보를 얻어 곱해 주게 된다. 어텐션 정보가 곱해진 8개의 특징맵을 concatenation을 시켜 1개의 특징맵으로 합친 뒤, 1×1 합성곱을 사용하여 원래 채널수와 맞춰준다. 다양한 operation을 사용하여 얻은 많은 정보중에서 필요한 정보만 가지고 영상 복원을 진행하여 효율적으로 보인다. 하지만 이 네트워크는 한 층에서 여러가지 operation을 사용하기 때문에 많은 파라미터로 구성되어 있고, 논문의 분석 내용을 보면 거의 사용되지 않는 operation도 존재한다. 이는 비효율적이고, 학습 및 테스트를 진행할때에도 많은 시간을 소모하게 된다.

본 논문에서 우리는 영상 열화 방법을 인지할 수 없는 상황에서 적용할 수 있는 복합 열화 복원 네트워크를 제안한다. 우리는 네트워크에 채널간 어텐션 모듈과 스킵 커넥션을 적용하여, 영

상에 적용된 열화에 따라 복원에 필요한 채널에 높은 가중치를 적용하여 효율적인 영상 복원이 가능하도록 하였다. 본 논문에서 제안하는 네트워크는 기존에 제안되었던 최신 (state-of-the-art) 방법들 보다 높은 성능을 내고, 단 한번의 학습으로 복합 열화 복원이 가능하도록 하였다.

이 연구의 기여는 다음과 같다.

1. 우리는 채널 어텐션 모듈을 결합한 단일 end-to-end 네트워크를 통해 기존 방법들에 비해 좋은 복합 영상 열화 복원 성능을 낸다.
2. 우리는 추가 실험을 통해 채널 어텐션 모듈의 적용에 대해 분석하였다.

2. 관련 연구

이번 절에서는 다양한 열화에 대한 영상 복원 및 채널 어텐션의 기존 연구에 대해 설명한다. 영상 복원은 영상 처리 분야에서 가장 근본적인 문제 중 하나이기 때문에 방대한 양의 연구 결과가 있다. 여기서 우리는 각 열화별로 대표적인 작품들 중 일부만을 다루고 있다.

2.1 영상 복원

주변 픽셀이 복원할 픽셀과 비슷한 픽셀을 가지고 있다고 가정하여 가우시안 필터링이나 바이레터럴 필터링과 같은 필터링 기반의 방법들 [1]이 제안되었다. Freeman *et al.* [7]은 초해상도 복원을 위해 외부 데이터베이스를 활용하여 선명한 영상과 열화된 영상사이의 관계를 학습하는 방법을 제안했다. Elad와 Aharon [8]은 잡음 제거를 위해 사전을 학습하는 방법을 제안하였다. Buades *et al.* [4]는 영상내의 비국부적 자기유사성을 이용한 영상 복원 방법을 처음으로 제안했다. Dabov *et al.* [5]는 3D DCT 변환과 블락 매칭을 결합하여 BM3D라는 영상 잡음 제거 방법을 제안했다. BM3D는 매우 효과적인 잡음 제거 방법으로, 여전히 영상 잡음제거에서는 비고 방법 중 하나로 사용된다. Glasner *et al.* [13]는 단일 이미지 내에서 블락 레벨 통계를 활용함으로써 단일 영상을 외부 데이터베이스나 사전의 필요 없이 초해상도 복원을 할 수 있음을 보여준다. Cho *et al.* [14]는 위해 기댓값 최대화 알고리즘을 사용하여 outlier들을 조절하는 방법을 제안하여 영상의 블러제거에서 뛰어난 결과를 보여준다.

최근 몇 년간 영상 복원 문제에 대해 딥러닝이 활발히 적용되었고, 고전적인 방법에 비해 우수한 결과를 보여주고 있다. Dong *et al.* [15]는 처음으로 딥러닝 기반의 초해상도 복원 방법인 SRCNN을 제안했다. SRCNN은 단순히 3개의 합성곱 층으로 구성되어 있지만, 고전적인 복원 방법들보다 우수한 결과를 보여준다. Kim *et al.* [16]은 단일 이미지 초해상도 복원을 위한 매우 깊은 네트워크와 잔여 학습(residual learning)을 제안했다. Xu *et*

al. [17]는 딥러닝 기반의 deconvolution CNN을 제안하여, 블러 제거 문제에서 딥러닝이 적용되지 않은 방법들에 비해 좋은 결과를 보여준다. Zhang *et al.* [9]은 잡음제거를 위해 잔여학습과 배치 정규화 (batch normalization)를 사용하여 DnCNN을 제안했다. DnCNN은 잡음제거 뿐만 아니라, JPEG 영상 복원 문제 및 초해상도 복원 문제에서도 좋은 성능을 보였다. Tai *et al.* [18]는 많은 재귀 유닛과 게이트 유닛으로 구성된 80개의 합성곱 층을 가진 메모리 네트워크를 제안하여 잡음 제거, 초해상도 복원 및 JPEG 영상 복원에 적용했다. Liu *et al.* [19]는 영상 복원을 위해 반복되는 비국부적 모듈로 구성된 Non-local recurrent network를 제안했다.

위에서 언급한 대부분 방법들은 열화 종류를 인지해야만 적용 가능하다. 영상의 열화 종류를 알 수 없는 경우에, 영상내 열화를 제거하기 위한 소수의 방법들도 제안되었다. Yu *et al.* [11]는 강화학습을 사용하여 복합 열화에 대처 가능한 RL-Restore를 제안했다. Saganuma *et al.* [12]는 다양한 operation간의 어텐션을 이용하여 영상내의 복합 열화를 제거하는 모델인 OWAN을 제안했다.

2.2 채널 어텐션 모듈

Hu *et al.* [20]가 처음으로 채널 어텐션을 제안함으로써 몇몇 연구에 채널 어텐션을 도입하였다. Hu *et al.*이 제안한 채널 어텐션 모듈을 적용한 네트워크는 2017 ILSVRC 영상 분류 (image classification) 분야에서 우승했다. Park *et al.* [21]은 bottleneck 구간에서 정보량이 줄기 전에 채널 어텐션 모듈을 추가하여, 적은 파라미터 수의 증가로 영상 분류에서 높은 성능 향상을 이끌어 냈다. Woo *et al.* [22]은 채널 어텐션과 공간 어텐션 (spatial attention)을 결합한 CBAM (convolutional block attention module)을 제안하여 영상 분류 및 검출 (image detection)과 같은 여러 영상인식 문제에서 높은 성능을 이끌어냈다. 위에서 언급한 것처럼 고레벨 비전문제뿐 만 아니라, 영상 복원에서도 채널 어텐션 모듈이 적용되었다. Zhang *et al.* [23] 과 Cheng *et al.* [24]은 초해상도 복원을 위해 채널 어텐션을 사용하여 다른 딥러닝 기반의 초해상도 복원 방법들에 비해 우수한 결과를 보여준다.

3. 방법

이번 절에서는 우리가 제안한 복합 열화 복원을 위한 네트워크를 소개한다. 3.1절에서는 여기서 사용한 채널 어텐션 모듈을 소개하고, 3.2절에서는 본 논문에서 제안하는 네트워크 구조, 3.3절에서는 네트워크의 학습방법을 소개한다.

3.1 채널 어텐션 모듈

채널 어텐션 모듈은 Hu *et al.* [20]가 처음으로 squeeze-and-excitation block이란 이름으로 제안하였다. 일반적인 합성곱 연

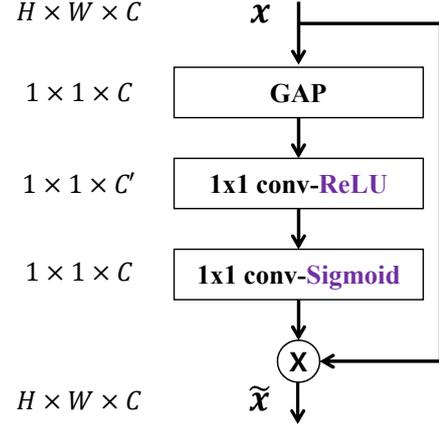


Figure 1: The architecture of channel attention module [20]

산은 국부적 정보만 이용하지만, 채널 어텐션 모듈은 global average pooling (GAP)를 적용하여 비국부적인 정보까지 이용할 수 있게 한다. Fig. 1는 채널 어텐션 모듈을 보여준다.

Fig. 1에서 보는거와 같이, 먼저 GAP를 이용하여 채널별 입력 특징맵의 각 채널인 x_c 에 대해 압축된 전체 공간 정보 (global spatial information) z_c 를 얻을 수 있다.

$$z_c = GAP(x_c) = 1/(H \times W) \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (1)$$

위의 식으로부터 얻은 정보를 두 개의 1×1 합성곱 층과 두 개의 비선형 활성화 함수인 ReLU와 Sigmoid를 통해 채널 사이의 의존성을 계산하는 데 사용한다.

$$s_c = \sigma(W_2 \delta(W_1 z_c)) \quad (2)$$

여기서 σ 및 δ 는 각 Sigmoid와 ReLU 비선형 활성화 함수를 나타내며, $W_1 \in R^{C \times C/r}$ 및 $W_2 \in R^{C/r \times C}$ 는 1×1 합성곱 층의 가중치들을 의미한다. W_1 의 필터 개수를 C/r 만큼 줄이고 W_2 의 필터 개수를 다시 C 만큼 증가시킨다. Fig. 1의 C' 은 C/r 을 의미한다. 감소비 r 를 사용하여, C 개의 필터를 갖는 1×1 합성곱 층 하나만 사용하는 것보다 파라미터를 적게 사용할 수 있다. 그뿐만 아니라 [20]에서는 r 이 16일 때 r 이 1일 때와 성능 차이가 거의 없으며, 파라미터의 개수를 줄일 수 있음을 입증하였다. 우리도 마찬가지로 제안한 방법에서 r 을 16으로 사용한다. 아래의 방정식으로 구한 스케일링 팩터 s 를 각 채널의 입력 특징맵에 곱해주게 된다.

$$\hat{x}_c = s_c * x_c \quad (3)$$

여기서 \hat{x}_c 는 재조정된 특징맵의 c 번째 채널이다.

우리는 입력 영상의 비국부적 정보를 이용하기 위해 모델에서 채널 어텐션 모듈을 사용하였다. 이 뿐만 아니라, 채널 어텐션 모듈을 통해 재조정된 스케일링 팩터 s 를 이용하여 열화 제거에 중요한 채널에는 높은 값들이, 덜 중요한 채널에는 낮은 값들이

곱해지게 된다.

3.2 네트워크 구조

Fig. 2는 우리가 제안한 네트워크의 전체적인 구조이다. 제안한 모델은 열화된 칼라 영상을 입력으로 받고 네트워크의 결과와 입력 영상을 더하여 최종적인 복원 결과를 도출한다. 이와 같이 우리는 제안한 네트워크에 잔여학습 방법을 적용하였고, 이미 잔여학습에 대한 효과는 여러 논문들 [9, 23]에서 입증되었다.

본 논문에서 제안한 모델은 특징 추출기(feature extractor), 채널간 어텐션 블록, 특징 재구성기(feature reconstruction)로 구성되어 있다. 특징 추출기는 128개의 필터를 갖는 $3 \times 3 \times 3$ 크기의 합성곱 층과 ReLU 활성화 함수로 구성되어 있다. 여러 가지 열화를 다루기 위해서 128개의 필터를 사용하였다. 채널간 어텐션 블록은 총 4개의 128개의 필터를 갖는 $3 \times 3 \times 128$ 크기의 합성곱 층을 사용하였고 마지막에는 3.1절에서 소개한 채널 어텐션 모듈을 적용하였다. 앞 3개의 합성곱 층에서는 배치 정규화와 ReLU 활성화 함수를 같이 사용하였고, 마지막 4번째 합성곱 층은 합성곱 연산만 적용되도록 구성하였다. 우리는 총 5개의 채널 어텐션 블록을 사용하였다. 마지막으로 3개의 필터를 갖는 $3 \times 3 \times 128$ 크기의 합성곱 층으로만 구성된 특징 재구성기를 사용하여 입력 영상의 채널수와 맞추어 주었고 특징 재구성기에서 나온 특징맵과 입력 영상을 더해줘 최종 결과를 얻는다.

3.3 학습 방법

손실 함수. 우리는 제안한 모델을 학습하기 위해 L1 손실 함수를 사용하였다. i 번째 열화 영상과 깨끗한 영상을 각각 $I^{(i)}$ 와 $J^{(i)}$ 로, 학습 데이터 셋을 $D = \{\dots, (I^{(i)}, J^{(i)}), \dots\}$ 라 할때, 아래 식을 최소화하여 학습을 진행하였다.

$$L(\Theta; D) = \sum_i \left\| \left(I^{(i)} - f(I^{(i)}; \Theta) \right) - J^{(i)} \right\|_1 \quad (4)$$

위의 식에서 Θ 는 네트워크의 파라미터, $f(I^{(i)}; \Theta)$ 는 네트워크 출력값을 나타낸다.

학습 및 테스트 데이터셋. 제안한 네트워크를 학습하기 위해 [11]에서 설명한 실험 절차를 따랐다. 800개의 높은 질의 영상이 담긴 DIV2K 데이터셋을 이용하였다. 이 800개 데이터셋을 학습을 위한 750개의 영상과 테스트를 위한 50개 영상으로 나누어서 사용하였다. 먼저 학습을 위한 750장의 영상을 1/2, 1/3, 1/4배 다운 샘플링한다. 그런 다음 학습을 위한 영상과 테스트를 위한 영상 모두 56 픽셀 간격으로 63×63 크기로 나눈다. 우리가 얻는 총 학습 데이터셋은 230,080장, 테스트 데이터셋은 3,594장이다. 우리는 학습 및 테스트 데이터셋을 얻기 위해 Yu *et al.* [11]가 제공한 코드를 이용하였다. Yu *et al.* [11]가 제공한 코드를 사용했지만 저자가 사용한 249,344장보다 적은 학습 데이터셋을 얻었다. 위와 같이 영상을 분할한 다음 각 영상에 여러 가지 열화를 적용했다.

가우시안 블러, 가우시안 잡음, JPEG 압축을 순서대로 열화를 적용했으며 가우시안 블러와 가우시안 잡음은 각 표준 편차가 [0, 5]와 [0, 50] 사이 임의의 값으로, JPEG 압축의 질은 [10, 100] 사이 임의의 값을 사용하였다. Yu *et al.* [11]은 복합 열화를 extremely mild부터 extremely severe까지 총 5가지로 구분하였다. Table 1은 열화 정도에 따라 각 열화마다 level를 나타낸 표로, 각 열화마다 정도에 따라 1부터 10까지로 level을 분류하였다. 이 표를 기준으로 3가지 열화의 level 합이 [3, 10]이면 extremely mild, [11, 13]이면 mild, [14, 19]는 moderate, [20, 22]는 severe, [23, 30]은 extremely severe라고 명명하였다. 우리는 RL-Restore와 OWAN 방법들과 마찬가지로 학습 데이터셋에는 열화 정도를 moderate만 적용하였고, 테스트 데이터셋에는 일반화 능력을 검증하기 위해 moderate 뿐만 아니라 mild, severe 총 3개 종류의 열화 정도를 적용하였다.

우리는 최적화 방법으로는 $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$ 의 변수를 가지는 Adam 최적화 방법 [25]을 배치 사이즈는 64 사용하여, 위에서 언급한 학습 데이터셋을 가지고 학습을 진행하였다. 총 60 에폭(epoch) 동안 학습을 진행하였으며, 초기 학습률(learning rate)를 $1e^{-3}$ 으로 설정하고 15 에폭마다 5배씩 감소하도록 하였다. 우리는 모델을 구현하기 위해 PyTorch [26]를 사용했다. 학습에는 약 2일 정도가 걸렸으며, 학습 및 테스트시 CPU는 Intel Zeon E5-2620 @ 2.0 GHz을, GPU는 NVIDIA TITAN RTX (24GB)을 사용했다.

4. 실험 결과 및 분석

4.1 네트워크 구조 분석

이번 절에서 우리는 제안한 네트워크의 구조를 분석한다. 네트워크 구조는 성능을 결정하는 중요한 요소 중 하나이다. 이에 따라, 우리는 여러 네트워크 구조에 대해 비교를 진행하였다. 여러 네트워크 구조의 성능을 비교하기 위해 테스트 데이터셋중 moderate를 사용하여 영상 평가 척도인 Peak Signal-to-Noise Ratio (PSNR)을 측정하였다.

먼저, Table 2은 합성곱 층의 필터 개수를 64개로 고정하고 다른 4가지 네트워크 구조에 따른 성능을 보여준다. 첫 번째와 두 번째 네트워크 구조는 채널 어텐션 블록안에 합성곱 층이 2개인 구조로 4개인 구조와 총 합성곱 층의 개수를 같게 하기 위해 총 10개의 블록을 사용하였다. 두 번째와 네 번째 네트워크 구조는 채널 어텐션 블록안의 마지막 합성곱 층에 배치 정규화와 ReLU 활성화 함수를 사용하지 않았고, 반대로 첫 번째와 세 번째 네트워크 구조는 마지막 합성곱 층에 배치 정규화와 ReLU 활성화 함수를 사용했다. 실험을 통해 우리는 네 번째 네트워크 구조의 성능이 가장 좋음을 입증하였고, 이 구조를 기반으로 다음 실험을 진행하였다.

그 다음 실험으로는 채널 어텐션 모듈의 효과를 검증하기 위해 채널 어텐션 모듈이 없는 네트워크 구조와 있는 네트워크 구조를

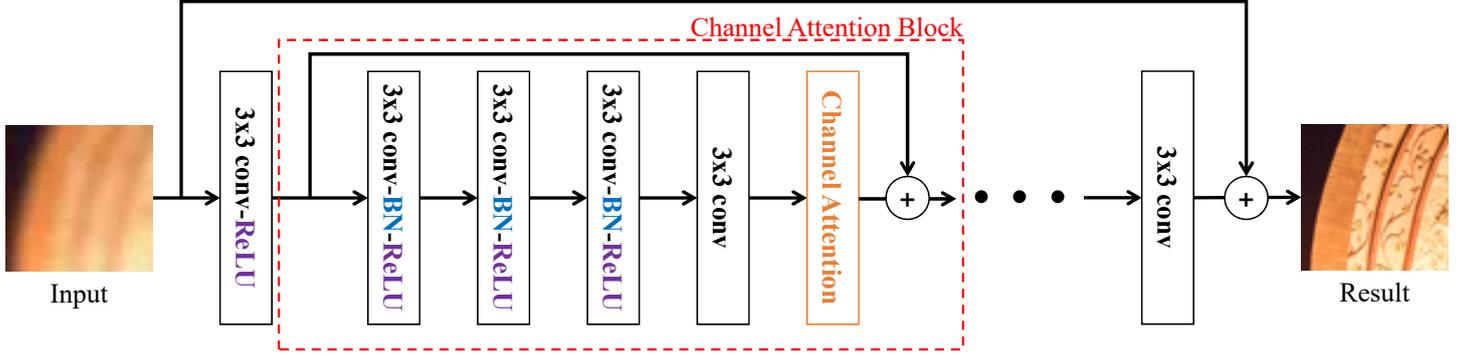


Figure 2: The architecture of the proposed method

Table 1: Degradation levels of individual distortion [11]

Degradation Level	1	2	3	4	5	6	7	8	9	10
Gaussian blur	[0, 0.5]	[0.5, 1]	[1, 1.5]	[1.5, 2]	[2, 2.5]	[2.5, 3]	[3, 3.5]	[3.5, 4]	[4, 4.5]	[4.5, 5]
Gaussian noise	[0, 5]	[5, 10]	[10, 15]	[15, 20]	[20, 25]	[25, 30]	[30, 35]	[35, 40]	[40, 45]	[45, 50]
JPEG compression	[80, 100]	[60, 80]	[50, 60]	[40, 50]	[35, 40]	[30, 35]	[25, 30]	[20, 25]	[15, 20]	[10, 15]

Table 2: A comparison of different architecture with 64 filters of convolution layers on the test set of moderate. The best performance is in **bold**.

Architecture of network	PSNR
Conv-BN-ReLU $\times 2$	26.90
Conv-BN-ReLU + Conv	26.52
Conv-BN-ReLU $\times 4$	26.79
Conv-BN-ReLU $\times 3$ + Conv	26.97

Table 3: A comparison of different architecture on the test set of moderate class. The best performance is in **bold**.

Architecture of network	PSNR
Without channel attention module & with 64 channels	26.82
Without channel attention module & with 128 channels	26.91
With channel attention module & with 64 channels	26.97
With channel attention module & with 128 channels	27.12

비교하였다. 추가적으로 합성곱 층의 필터 개수에 대한 효과도 검증하기 위해 각 64개와 128개의 필터 개수를 사용하여 비교하였다. 먼저 첫 번째와 두 번째 네트워크 구조는 채널 어텐션 모듈을 사용하지 않았고, 세 번째와 네 번째 네트워크 구조에서는 사용하였다. 첫 번째와 세 번째 네트워크 구조에서는 합성곱 층의 필터 개수를 64개를 사용하였고, 두 번째와 네 번째 네트워크 구조에서는 128개를 사용하였다. Table 3는 네트워크 구성에 따른 성능을 나타낸 것이다. 세 번째와 네 번째 네트워크 구조의 성능이 첫 번째, 두 번째 네트워크 구조 성능에 비해 각 0.15dB, 0.21dB가 상승되었다. 또한 합성곱 층의 필터 개수가 2배로 증가하면서, 각 0.1dB 이상씩 성능이 증가하였다. 우리는 이 실험을 통해 합성곱 층의 필터 개수가 128개이고, 채널 어텐션 모듈을 적

용한 네 번째 네트워크 구조의 성능이 가장 좋음을 입증하였고, 그래서 우리의 최종 네트워크 구조로 채택하였다.

우리는 3.1절에 소개한 채널 어텐션 모듈뿐만 아니라, 다른 채널 어텐션 모듈을 추가적으로 적용해보았다. [22]에서 제안한 average pooling과 max pooling을 같이 적용한 채널 어텐션 모듈을 적용하여 실험을 진행하였다. 이 모듈을 사용한 결과 오히려 average pooling만 있는 모듈을 적용한 네트워크 비해 성능이 0.04dB 하락하였다. 이를 통해, 복합 열화 복원에서는 average pooling으로부터 얻은 정보가 max pooling의 정보보다 큰 역할을 한다고 볼 수 있고, 최종적으로 우리는 average pooling만 사용한 채널 어텐션 모듈을 사용했다.

4.2 결과 비교

이번 절에서는 우리가 제안한 네트워크의 타당성을 입증하기 위해 다른 방법과 성능을 비교한다. 성능을 측정하기 위해 3.3절에서 언급한 세 종류의 테스트 데이터셋을 이용하였고, 비교 대상으로는 최신 복합 열화 제거 방법인 RL-Restore [11]와 OWAN [12]과 RL-Restore 논문에서 비교 방법으로 사용한 DnCNN [9]를 채택했다. 성능 비교를 위해 질적 평가와 양적 평가를 진행하였다. 양적 평가와 질적 평가를 위해 사용된 코드 및 저장된 파라미터는 저자가 직접 올려 놓은 공식 코드를 사용하였다.

우리는 우선 영상 평가에 널리 사용되는 척도인 PSNR 및 structural similarity (SSIM)를 이용해 우리의 모델을 최신 방법들과 정량적으로 비교한다. 각 방법들의 PSNR 및 SSIM 수치는 [12]에서 가져온 수치이다. Table 4는 각 방법들과의 정량적으로 비교를 한 결과이다. 이 표를 통해 모든 테스트 데이터셋에 대해 본 논문에서 제안한 방법의 성능이 가장 우수함을 알 수 있다. 특히, mild 테스트 데이터셋에서는 제안한 방법이 다른 방법들을 큰 폭으로

Table 4: A quantitative comparison of different methods. The best performance are in **bold**.

Dataset	DnCNN [9]	RL-Restore [11]	OWAN [12]	Ours
Mild (unseen)	27.51/0.7315	28.04/0.7313	28.33/0.7455	28.54/0.7493
Moderate	26.50/0.6650	26.45/0.6557	27.07/0.6787	27.12/0.6793
Severe (unseen)	25.26/0.5974	25.20/0.5915	25.88/0.6167	25.89/0.6170

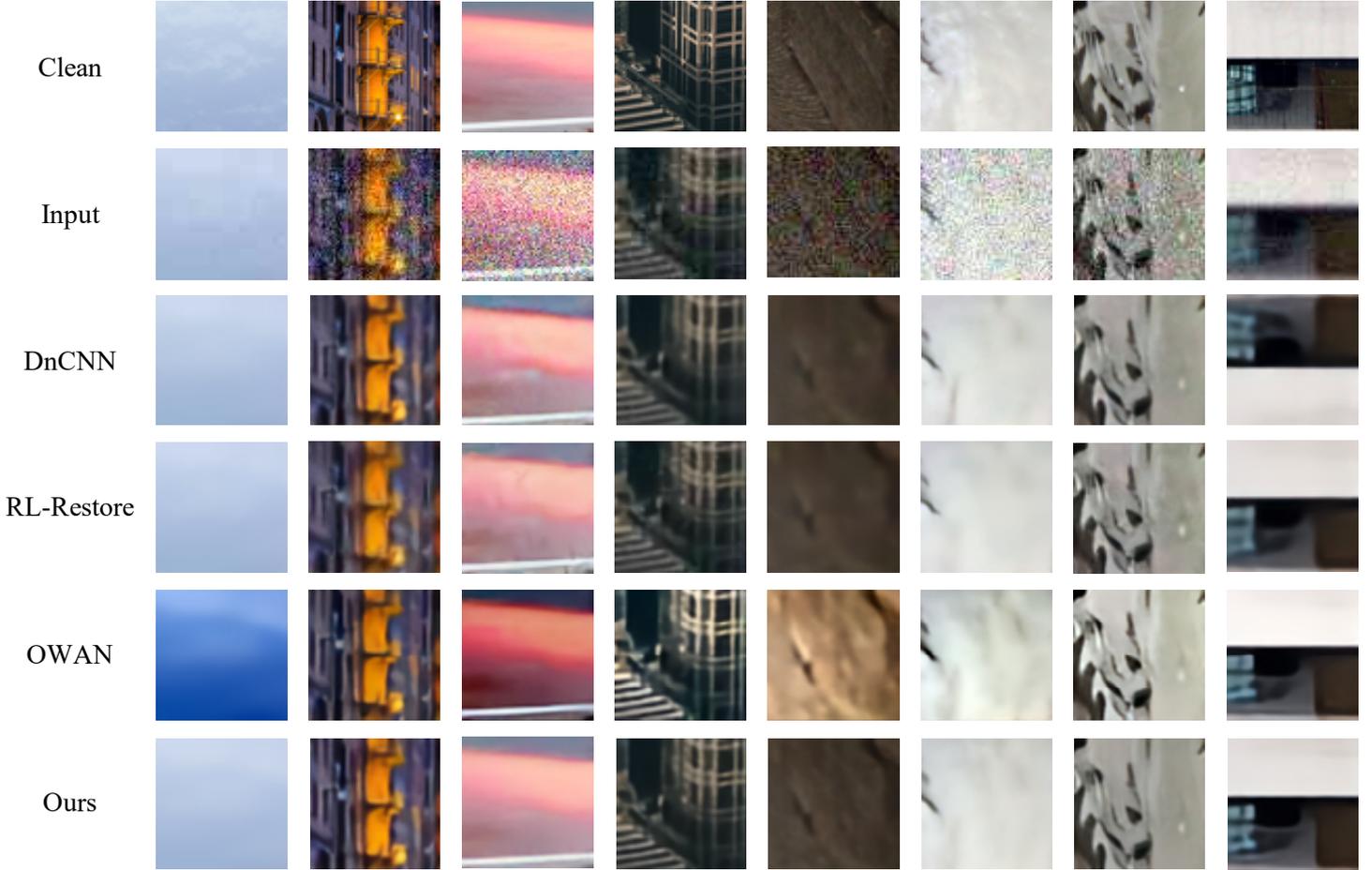


Figure 3: A qualitative comparison of different methods on a mild dataset.

능가한다는 것을 보여준다.

Fig. 3는 mild 테스트 데이터셋에 대한 여러 방법들과의 질적 비교 결과이다. 첫 번째 행은 원본 영상이고, 두 번째 행은 열화된 영상, 세 번째와 네 번째, 다섯 번째 행은 각 DnCNN 방법과 RL-Restore 방법, OWAN 방법의 결과이다. 마지막으로 여섯 번째 행은 본 논문에서 제안한 네트워크의 결과이다. DnCNN 방법은 주로 블러한 영상을 만들어낸다. RL-Restore 방법의 결과들은 대체적으로 아티팩트들이 많다. 또한 다른 방법들에 비해 블러한 영상들이 많다. OWAN 방법은 색 부분에 대한 복원이 잘 이루어지지 않는다. 첫 번째와 세 번째 열을 보면, 원본 영상보다 진한 색으로 복원이 이루어지는 것을 볼 수 있다. 우리 방법의 결과는 두 방법에서 문제된 부분이 해결된 것을 볼 수 있다.

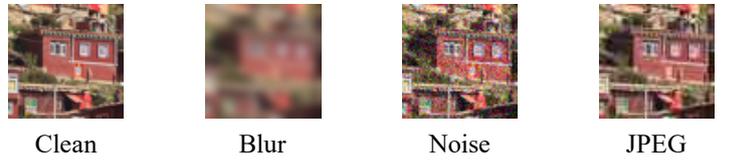


Figure 4: Examples for the analysis of channel attention module

4.3 채널 어텐션 모듈 효과

이번 절에서는 채널 어텐션 모듈에 대해 분석을 진행한다. 우리는 Table 3에서의 실험을 통해 복합 열화 복원에 채널 어텐션 모듈이 효과적임을 입증하였다.

채널 어텐션 모듈이 어떻게 동작하는지 확인하기 위해, 우리는 네트워크에 각 하나의 열화만 적용된 영상을 입력으로 받아 채널 어텐션 모듈의 가중치 값을 비교한다. Fig. 4은 채널 어텐션을 분

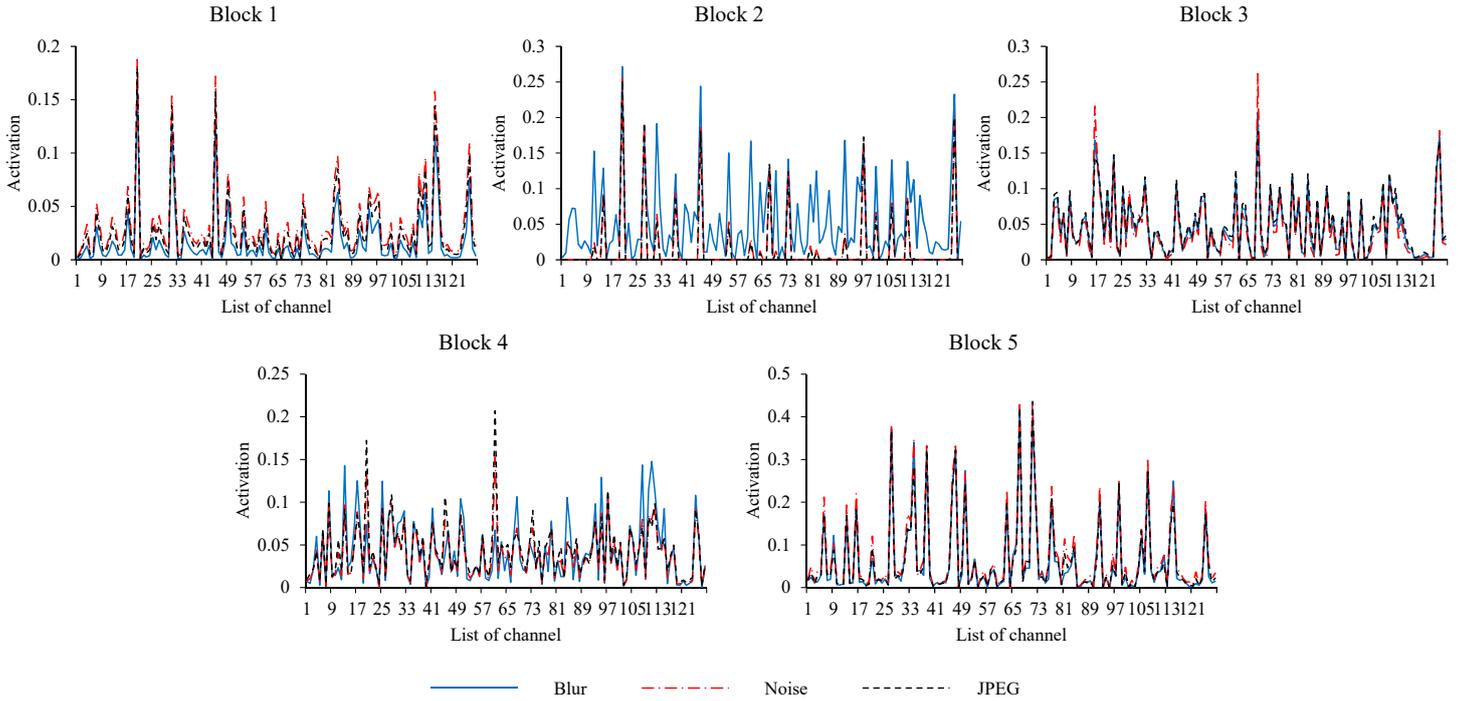


Figure 5: The analysis of channel attention module

석하기 위해 사용한 영상들이다. 각 열화정도는 Table 1에서의 강도 5를 적용하였다.

Fig. 5은 열화별로 각 채널 어텐션 블록별 채널 어텐션 모듈에서 얻은 가중치들을 나타낸다. 먼저 첫 번째와 세 번째, 다섯 번째 블록을 보면 각 열화별로 같은 형태의 가중치 분포를 가지고 가중치들의 값에서 약간씩만 차이가 있다. 특히, 다섯 번째 블록에서는 값에도 거의 차이가 없어, 높은 가중치들을 갖는 채널들이 열화를 복원하기 위해 특징 재구성에 주로 사용된 채널이라고 유추할 수 있다. 두 번째 블록에서는 블러를 제외한 잡음과 JPEG 압축의 경우 대부분의 가중치들이 0의 값에 수렴해있다. 이를 통해, 두 번째 블록에서의 채널들은 대부분 블러를 제거하기 위해 작동한다고 볼 수 있다. 네 번째 블록에서도 열화 별로 약간씩 다른 가중치 분포를 갖는다. 4. 1절에서의 실험을 통해 복합 열화 복원에 채널 어텐션 모듈을 적용한 네트워크가 높은 성능을 내는 것을 보였고 이번 절의 분석을 통해 우리는 채널 어텐션 모듈이 열화에 따라 다르게 작용하는 것을 입증하였다.

5. 결론 및 향후 계획

본 논문에서는 채널 어텐션 모듈을 적용한 합성곱 신경망을 이용하여 영상내 복합 열화를 복원하기 위한 방법을 제안하였다. 기존에 제안된 영상 복원 방법들은 영상의 열화 종류를 인지하는 상황에만 적용이 가능하기 때문에 자율주행 자동차나 로봇 시스템과 같은 영상 열화 방법을 인지할 수 없는 상황에는 기존의 영상 복원 방법을 적용하는데 무리가 있다. 열화를 인지할 수 없는 상황을 해결하기 위해 여러 복합 열화 복원 방법들이 제안되었지

만 낮은 영상 품질로 인해 실생활에 적용하기에는 무리가 있다. 우리는 추가 실험을 통해 채널 어텐션 모듈이 복합 열화 복원에 매우 효과적임을 입증하였고 채널 어텐션 모듈을 적용한 방법의 성능이 기존의 방법들보다 뛰어난임을 입증하였다.

하지만 본 논문에서 제안한 방법의 한계점도 몇 가지 존재한다. 자율주행 자동차나 소방 로봇과 같은 시스템에 적용하기 위해서는 실시간으로 동작해야지만 본 논문에서 제안한 방법은 많은 파라미터가 필요하기 때문에 실시간에 사용하기엔 아직 무리가 있다. 또한 실제로 사용하기에는 성능이 부족하기도 하다. 영상 열화 종류도 잡음, 블러, JPEG 압축등과 같이 세 가지만 다루어 해상도 저하와 같은 다른 열화에 대한 문제는 여전히 남아있다.

향후에는 복합 열화에 효율적인 새로운 채널 어텐션 모듈을 적용함으로써 성능을 향상시키고 네트워크의 경량화를 통해 실시간으로 적용될 수 있음을 기대한다.

References

- [1] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, Jan 1998, pp. 839–846.
- [2] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.
- [3] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image

- restoration,” *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 460–489, 2005.
- [4] A. Buades, B. Coll, and J.-M. Morel, “A non-local algorithm for image denoising,” in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 60–65.
- [5] 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, 2007.
- [6] T. Michaeli and M. Irani, “Nonparametric blind super-resolution,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 945–952.
- [7] W. T. Freeman, T. R. Jones, and E. C. Pasztor, “Example-based super-resolution,” *IEEE Computer graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.
- [8] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *IEEE Transactions on Image processing*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [9] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian denoiser: Residual learning of deep cnn for image denoising,” *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.
- [10] T. Remez, O. Litany, R. Giryes, and A. M. Bronstein, “Deep class-aware image denoising,” in *Sampling Theory and Applications (SampTA), 2017 International Conference on*. IEEE, 2017, pp. 138–142.
- [11] K. Yu, C. Dong, L. Lin, and C. Change Loy, “Crafting a toolchain for image restoration by deep reinforcement learning,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2443–2452.
- [12] M. Suganuma, X. Liu, and T. Okatani, “Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [13] D. G. S. B. M. Irani, “Super-resolution from a single image,” in *Proceedings of the IEEE International Conference on Computer Vision, Kyoto, Japan*, 2009, pp. 349–356.
- [14] S. Cho, J. Wang, and S. Lee, “Handling outliers in non-blind image deconvolution,” in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 495–502.
- [15] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2016.
- [16] J. Kim, J. Kwon Lee, and K. Mu Lee, “Accurate image super-resolution using very deep convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.
- [17] L. Xu, J. S. Ren, C. Liu, and J. Jia, “Deep convolutional neural network for image deconvolution,” in *Advances in Neural Information Processing Systems*, 2014, pp. 1790–1798.
- [18] Y. Tai, J. Yang, X. Liu, and C. Xu, “Memnet: A persistent memory network for image restoration,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4539–4547.
- [19] D. Liu, B. Wen, Y. Fan, C. C. Loy, and T. S. Huang, “Non-local recurrent network for image restoration,” *arXiv preprint arXiv:1806.02919*, 2018.
- [20] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [21] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, “Bam: Bottleneck attention module,” *arXiv preprint arXiv:1807.06514*, 2018.
- [22] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, “Cbam: Convolutional block attention module,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.
- [23] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham: Springer International Publishing, 2018, pp. 294–310.
- [24] X. Cheng, X. Li, J. Yang, and Y. Tai, “Sesr: Single image super resolution with recursive squeeze and excitation networks,” in *2018 24th International Conference on Pattern Recognition (ICPR)*, Aug 2018, pp. 147–152.
- [25] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

- [26] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.

〈 저자 소개 〉



이 해 운

- 2016년 2월 전북대학교 수학과 학사
- 2018년 2월 대구경북과학기술원 정보통신융합전공 석사
- 2018년 2월 ~ 현재 대구경북과학기술원 정보통신융합전공 박사
- 관심분야 : 영상 처리, 영상 복원, 계산 사진학
- <https://orcid.org/0000-0002-7572-1705>



조 성 현

- 2005년 8월 포항공과대학교 컴퓨터공학과 학사
- 2012년 2월 포항공과대학교 컴퓨터공학과 박사
- 2006년 8월 ~ 2007년 2월 Microsoft Research Asia 인턴
- 2010년 7월 ~ 2010년 11월 Adobe Research 인턴
- 2012년 3월 ~ 2014년 3월 Adobe Research 연구원
- 2014년 4월 ~ 2017년 4월 삼성전자 책임연구원
- 2017년 4월 ~ 현재 대구경북과학기술원 조교수
- 관심분야: 영상 및 비디오 처리, 계산 사진학, 영상 복원
- <https://orcid.org/0000-0001-7627-3513>