

일반논문 (Regular Paper)

방송공학회논문지 제28권 제1호, 2023년 1월 (JBE Vol.28, No.1, January 2023)

<https://doi.org/10.5909/JBE.2023.28.1.90>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## 코렉터 어텐션 네트워크를 이용한 로우 센서 영상 초해상화 기법

신 바 울<sup>a)</sup>, 김 태 하<sup>b)</sup>, 이 의 진<sup>c)†</sup>

### Raw Sensor Single Image Super Resolution Using Color Corrector-Attention Network

Paul Shin<sup>a)</sup>, Teaha Kim<sup>b)</sup>, and Yeejin Lee<sup>c)†</sup>

#### 요 약

본 연구에서는 전통적인 RGB 영상보다 데이터양이 적은 로우 센서 영상을 이용한 초해상화 네트워크를 제안하고 이에 대한 실험 결과를 정리하였다. 로우 센서 영상의 초해상화는 일반적인 RGB 초해상화와 달리 카메라에서 일어나는 후처리 가공이 없는 무손실 영상을 이용하기 때문에 결과물의 성능이 일반 RGB 초해상화 연구와 달리 색상 보정에 따라 많이 좌우된다. 따라서, 본 연구에서는 색상 보정을 위한 모듈을 개발하여 기존 RGB 기반 네트워크에 삽입해 이를 이용해 성능 결과를 비교하였다. 연구 결과 색상 보정 모듈을 적용함으로써 성능 지표의 향상이 있음을 확인하였다. 다만, 출력 영상의 의도하지 않은 아티팩트가 발생하는 현상을 확인하였고, 성능 지표 중 PSNR의 향상이 분명하나 SSIM의 성능이 일부 떨어지는 것으로 확인하였다.

#### Abstract

In this paper, we propose a super resolution network for raw sensor image which data size is lower comparatively to RGB image. But the actual capabilities of raw image super resolution depends on color correction because its absent of camera post processing that leads to unintended result having different white balance, saturation, etc. Thus, we introduce novel color corrector attention network by adopting the idea of precedent raw super resolution research, and tune to the our faced problem from data specification. The result is not superior to former researches but shows decent output on certain performance matrix. In the same time, we encounter new challenging problem of unexpected shadowing artifact around image objects that cause performance declination despite its good result overall. This problem remains a task to be solved in the future research.

Keyword : Super resolution, raw sensor image, image restoration, attention network, color correction

a) 서울과학기술대학교 스마트에너지시스템학과(Dept. of Smart Energy System Engineering, Seoul National University of Science and Technology)

b) 서울과학기술대학교 전기정보공학과(Dept. of Electrical and Information Engineering, Seoul National University of Science and Technology)

c) 서울과학기술대학교 전기정보공학과, 스마트에너지시스템학과(Dept. of Electrical and Information Engineering, Dept. of Smart Energy System Engineering, Seoul National University of Science and Technology)

† Corresponding Author : 이의진(Yeejin Lee)

E-mail: yeejinlee@seoultech.ac.kr

Tel: +82-2-970-6443

ORCID: <https://orcid.org/0000-0002-3439-5042>

\* 이 논문은 2022년도 정부(교육부)의 재원으로 한국연구재단 개인연구사업의 지원을 받아 수행된 연구임. (No. 2022R1F1A1062950)

· Manuscript February 7, 2022; Revised November 21, 2022; Accepted December 13, 2022.

## 1. 서론

영상 초해상화(super resolution, SR)는 저해상도의 영상을 고해상도의 영상으로 변환하는 기술이다. 영상 초해상화는 컴퓨터 비전(computer vision) 분야의 하위 주제인 영상 복원(image restoration) 문제의 일종으로 다양한 산업에서 응용기술로 사용되고 있어 그 연구 가치가 크다. 대표적인 응용 분야로는 감시 카메라의 영상 해상도를 높여 보안(surveillance) 분야에 활용하거나 위성 영상의 해상도를 높여 원격 탐사(remote sensing)에 활용 가능하며, 디스플레이 장치에서도 저해상도로 송신된 영상을 장치에 적합한 해상도로 변경하거나 콘텐츠 스트리밍 서비스에서도 과거의 저해상도 콘텐츠를 송신하기 위하여 관련 기술이 필요하다. 이러한 이유로 그 중요성을 인정받아 CVPR(IEEE Conference on Computer Vision and Pattern Recognition)과 함께 열리는 NTRIE Challenge(New Trends in Image Restoration and Enhancement)<sup>[1]</sup>에서 매년 초해상화 대회를 열고 있을 정도로 초해상화 연구는 많은 연구자들과 산업계의 관심을 받는 분야이다.

초해상화 문제는 수학적으로 유일한 해인 고해상도 영상이 존재하지 않는 문제(ill-posed problem)로 해 공간을 줄이기 위해 사전정보(prior information)를 이용한다. 과거에는 샘플을 통한 학습 전략(sample-based strategy)을 사용하여 사전정보를 제한하는 방법으로 연구가 행해져 왔으며, 이와 같은 방법을 활용하는 Total Variation<sup>[2]</sup>, Sparse Coding<sup>[3][4]</sup> 등이 초해상화 연구에서 좋은 성능을 보여주었다. 최근 심층신경망(deep neural network) 발전과 함께 심층학습(deep learning) 기법이 영상 복원에 도입<sup>[5]</sup>되어 동 분야에 비약적인 성능 향상이 있었으며, 이후로 영상 초해상도 분야에서도 대부분의 연구가 심층신경망을 활용하고 있다.

로우 센서 데이터란 디지털 카메라로 촬영된 영상을 후처리(post-processing)하는 과정 없이 센서 출력 그대로 저장된 영상으로 센서 배열에 따라 화소마다 3채널의 컬러 정보 중 R 혹은 G 혹은 B 요소 하나만 가지고 있는 데이터이다. 따라서, 디지털 영상 센서가 획득한 로우 데이터는 카메라 출력 형태인 RGB 영상으로 변환되기 위하여 디모

자이킹(demosaicking), 화이트 밸런스(white balance), 채도(saturation) 보정, 명암비(contrast) 보정, 선명도(sharpness) 보정과 같은 후처리 과정을 통과한 후 JPEG 파일 형식 등으로 압축(compression)되어 저장된다. 그러나 로우 데이터를 후처리하는 과정에서는 12비트 혹은 14비트로 획득된 원래 영상 센서 정보가 유지되지 않으며, 압축 후의 RGB 영상은 일반적으로 8비트의 부호없는 정수형으로 저장되어 정보의 손실이 발생한다.

후처리 과정 중 하나인 디모자이킹은 화소마다 하나의 컬러 정보만 가지고 있는 로우 센서 데이터를 화소마다 3채널 컬러 정보를 모두 가지도록 존재하지 않는 컬러 값을 주변 정보를 이용하여 생성하는 보간(interpolation) 방법이다. 이와 같은 디모자이킹 과정은 영상 초고해상화 과정에서 화소수를 증가하기 위하여 저해상도 영상의 주변 영역 정보를 이용하여 새로운 화소를 만들어내는 기법과 매우 유사하다. 하지만 현재까지의 영상 초해상화 연구는 대부분 디지털 카메라 내에서 후처리 과정을 통과한 압축된 형식의 영상을 학습과 추론을 위해 사용하고 있으며, 로우 센서 데이터를 이용하여 초고해상화를 수행하는 연구는 많지 않다<sup>[28][29]</sup>. 이는 과거에는 디지털 카메라의 최종 출력인 RGB 영상을 제외한 로우 데이터 접근이 쉽지 않았으며, 이에 따라 로우 센서 데이터 포함된 실험 가능한 벤치마크 데이터가 존재하지 않기 때문이다. 그러나 앞서 언급한 초해상화 문제의 특성상 입력 영상에서 더 많은 사전 정보를 추출할 수 있다면 모델 성능에서도 좋은 결과를 얻을 수 있다고 추론할 수 있다. 또한, 초해상화 문제에서 입력 영상과 출력 영상은 낮은 주파수(low frequency) 정보는 공유하는 한편, 높은 주파수(high frequency) 정보를 복원하는 작업이기 때문에 정보가 손실되지 않은 로우 센서 영상을 활용하여 초해상화를 진행하는 것이 성능의 측면에서 유리할 것으로 추정하는 것은 합리적인 접근이라고 할 수 있다. 본 논문에서는 이러한 이유로 기존 초해상화 연구들이 제시한 모델들을 이용하여 로우 센서 영상에서 네트워크의 학습 및 성능을 비교 평가하여 로우 센서 영상에서 기존 RGB 형식의 영상에서 나타내던 수준의 성능을 나타내는지 비교하는 연구를 진행하고자 한다.

## II. 관련 연구

### 1. 딥러닝(deep learning) 기반 초해상화 연구

초해상화 문제에 딥러닝을 활용한 최초의 연구는 C. Dong et al.의 초고해상화 합성곱 신경망(Super-Resolution Convolutional Neural Network, SRCNN)<sup>[5][6]</sup>에 관한 연구이다. 제안된 초고해상화 합성곱 신경망은 기존의 초해상화 연구에서 사용하던 특징 추출(feature extraction), 비선형 맵핑(non-linear mapping), 복원(reconstruction)의 세 단계 모두에 합성곱 층(convolution layer)을 적용한 심층학습 구조를 제시하였으며, 기존 연구들을 뛰어넘는 성과로 이후 딥러닝을 통한 초해상화 연구들이 주류를 이루는데 기여하였다. 이 연구를 통하여, 심층 학습 기반의 신경망이 영상 초해상화 분야에 도입되고, 고전적인 방식에 비해 비약적인 성능 향상을 이루었으나, 신경망의 깊이가 증가할수록 정보 손실이 발생함으로써 성능증가에 한계가 발생하였다. 또한, 신경망의 구조가 복잡해짐에 따라 연산량이 증가하여 실제 문제에 적용하기가 어려워지거나, 생성된 초고해상도 영상은 성능지표 상의 높은 점수에도 불구하고 사람의 보기에 선명도가 떨어지는 등의 문제점이 발생하였다. 따라서, 이와 같은 문제점을 해결하기 위하여 다양한 시도들이 나타나면서 관련 연구가 발전하게 되었다.

초기 심층학습을 적용한 초해상화 연구들은 합성곱 층을 깊게 쌓을수록 성능이 좋아진다는 것이 일반적인 인식이었었는데, 이유는 네트워크가 깊어질수록 리셉티브 필드(receptive field), 즉 필터가 정보를 얻기 위해 살펴보는 이미지 영역이 넓어지면서 더 많은 정보를 추출할 수 있다는 생각 때문이었다<sup>[7]</sup>. 그러나 초해상화의 경우에는 무조건 깊이를 늘리다 보면 심층신경망의 낮은 층(low-level)의 정보 또는 영상의 낮은 주파수 성분이 점차로 소실되어 깊이가 증가해도 성능이 향상되지 않는다. 이에 따라, 심층신경망의 정보 손실을 최소화하기 위한 기법들이 도입되었으며, 대표적인 기법으로는 스킵 커넥션(skip connection)<sup>[8]</sup>이 있다. 스킵 커넥션은 합성곱 층을 통과하지 않은 입력 데이터를 합성곱 층의 출력에 더해줌으로써 입력 부분을 제외한 추가적인 정보만을 학습하므로 잔차 학습(residual learning) 혹은 잔차 커넥션(residual connection)으로도 불린다.

Kim et al. 의 매우 깊은 합성곱 신경망(Very Deep Convolutional Network, VDCN)<sup>[7]</sup>에서 처음 초해상화에 스킵 커넥션을 사용하였으며, 스킵 커넥션은 이후 심층 재귀 합성곱 신경망(Deply-recursive Convolutional Network, DRCN)<sup>[9]</sup>, 잔차 인코더-디코더 신경망(Residual Encoder-Decoder Network, RED-Net)<sup>[10]</sup>, 초해상화 잔차 신경망(Residual Network, ResNet)<sup>[11]</sup>에 사용되면서 그 효용성이 검증되었고 현재까지도 대부분의 심층학습 기반 초고해상화 신경망에 사용되고 있다. 또한, DenseNet<sup>[12]</sup>의 등장으로 초해상화에서도 텐스 연결(dense connection)이 사용되기 시작<sup>[13]</sup>하였다. 텐스 커넥션은 각 층에서 생성된 특징맵(feature map)을 모두 사용하여 심층신경망의 낮은 층과 높은 층의 모든 특징맵을 고려한다는 점에서 낮은 층 정보의 이용 또한 중요한 초해상화 문제에 적합하며, [14]와 같은 구조의 초고해상화 신경망에 사용되었다.

이후에는 특징맵 간에 상호의존을 고려하여 어떤 특징맵들을 어떻게 활용할 것인지에 대해 연구하기 시작하였다. 이러한 움직임의 핵심에 있는 것이 바로 어텐션 메커니즘(attention mechanism)<sup>[15]</sup>으로 특징맵 또는 특징 채널 간의 상호 관계를 분석하여 어떤 특징맵에 보다 가중치를 부여할지를 학습한다. [16]에서는 어텐션 메커니즘을 초해상화에 접목한 매우 깊은 잔차 채널 어텐션 신경망(Very Deep Residual Channel Attention Network, RCAN)을 제안하였다. Dai et al. 은 채널 간의 상호 관계 파악을 위하여 공분산(covariance) 행렬을 이용하는 방법<sup>[17]</sup>을 제안하였으며, Liu et al. 은 기존의 한 층 안에서 특징 채널 간 어텐션만을 고려하지 않고 층간에도 어텐션을 적용하는 방법<sup>[18]</sup>을 제안하였다. 또한, [18]에서는 순환(recurrent) 구조의 비지역(non-local) 연산을 사용하여 해당 특징 영역과 주변 영역 간의 상관관계 정보를 신경망 학습에 전파하는 구조를 사용하였다. 이와 유사하게, Mei et al. 은 입력 특징을 여러 스케일(scale)에서 비지역 연산을 수행하여 이를 활용하는 방식<sup>[19]</sup>을 제안하였다.

초기에 초해상화 연구들은 신경망의 구조를 더욱 깊이 만들어서 성능을 향상하고자 하였으나, 학습에 많은 시간이 소요되고, 구조적 복잡성으로 인해 실제 문제 적용에 어려움이 있었다. 반면에, 실제 문제에 적용되는 장치들은 경량화되어 연산 효율성에 대한 이슈가 중요해지면서 신경망

구조를 효율적으로 만들기 위한 연구가 이루어졌다. 이와 같은 문제에 대한 해결 방안으로 저해상도 영상을 보간법을 통해 목표로 하는 크기의 고해상도 영상을 생성하고 이를 입력으로 사용하는 방식(pixel shuffling)이 제안되었으나 신경망의 크기가 커지는 문제점이 발생하였다. 따라서, 보간법으로 생성된 고해상도 영상을 입력으로 사용하면서도 효율적인 신경망을 설계하고자 [20]에서는 서브 픽셀 합성곱(sub-pixel convolution) 구조를 사용하였으며, [21]에서는 디콘볼루션(deconvolution)을 활용하는 방법, 점진적으로 영상 사이즈를 확대하기 위하여 네트워크 중간 단계마다 디콘볼루션을 사용하는 방식<sup>[22]</sup> 등이 제안되었다. 또한, 신경망의 매개변수 수를 줄여 학습 시간을 단축하는 회귀 구조를 활용하는 방식 [8], [18], [23], [24], [25], [26]도 제안되었다.

## 2. 로우 데이터 초해상화 연구

로우 센서 영상을 활용한 초해상화의 기존 연구는 매우 제한적이며, 연구에 활용 가능한 데이터셋(dataset)도 존재하지 않는다. 이에 따라, Zhang et al. 은 광학적인 확대를 이용해 직접 획득한 데이터셋인 SR-RAW를 이용하여 초고해상도 연구를 수행하였다<sup>[27]</sup>. 그러나 SR-RAW 데이터셋은 초점 거리(focal length)가 다른 렌즈를 사용하여 촬영한 영상 집합으로, 낮은 해상도를 가지는 로우 데이터 입력과 학습 목표가 되는 초고해상도 RGB 영상(ground-truth)이 완벽하게 정렬되지 않아, 이를 해결하기 위한 컨텍처 양방향 손실함수(contextual bilateral loss)를 제안하였다. Xu et al. 은 기존 초해상화의 학습을 위해 사용하는 합성(synthetic) 영상이 가지는 한계로 비선형성(non-linearity)과 단순한 가우시안 필터(Gaussian filter)를 사용한 쌍삼차 보간법(bicubic interpolation)을 논의하였으며, 이를 해결하

기 위한 카메라 영상신호처리(image signal processor, ISP) 과정을 대신할 컬러 보정(color correction)과 초고해상도 과정을 분리한 텐스 채널 어텐션 네트워크(dense channel-attention network)를 제안하였다<sup>[28]</sup>.

## III. 제안하는 방법

### 1. 로우 센서 데이터

컬러 필터 배열(color filter array, CFA)는 그림 1(a)의 예시와 같이 영상 센서의 한 화소에 R 또는 G 또는 B 요소가 일정하게 배치된 패턴이다. 이와 같은 필터의 배열에 따라 생성된 센서의 출력인 로우(raw 또는 CFA) 데이터는 3개의 컬러 채널 중 한 요소만 가지는 화소로 구성되어 그림 1(b)와 같이 회색조 영상(grayscale image)으로 보이며, 이는 화소마다 모든 컬러 요소를 가지는 RGB 영상의 컬러 채널 방향 서브샘플링(subsampling)으로 생각할 수 있다. 그림 1(b)의 예시에서도 볼 수 있듯이, 로우 센서 데이터의 각 화소는 빛의 세기(intensity)를 나타내며 컬러 필터가 결합되어 색 정보를 표현한다.

그림 1(a) 예시와 같은 컬러 필터 배열을 가지고 있는 영상 센서가 생성한 로우 데이터는 카메라 내 영상처리 프로세싱 파이프라인(camera processing pipeline)을 통해 RGB 영상으로 변환된다. 일반적인 카메라 프로세싱 파이프라인은 화소마다 완전하지 않은 컬러 정보를 주변 정보를 이용하여 완전한 정보로 구성하기 위한 디모자이킹(demosaicing), 색 보정(color correction), 화이트 밸런스(white balance), 감마 보정(gamma correction), 잡음 제거(noise suppression) 등과 같은 알고리즘으로 구성된다.

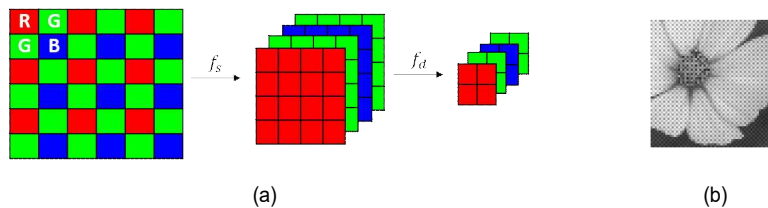
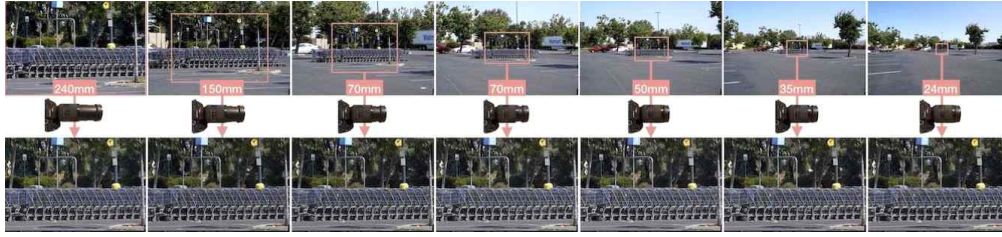


그림 1. (a) 베이어 컬러 필터 배열 패턴 (b) 로우 센서 데이터 예시 영상<sup>[29]</sup>  
 Fig. 1. (a) Bayer color filter array pattern (b) An example of raw sensor image<sup>[29]</sup>

그림 2. 로우 센서 영상 학습용 SR-RAW 쌍 예시<sup>[27]</sup>Fig. 2. Example sequence of SR-RAW dataset<sup>[27]</sup>

## 2. 학습 데이터

본 논문에서 제안하는 신경망 학습을 위해서는 II-2장에서 소개한 SR-RAW 데이터를 사용한다<sup>[27]</sup>. SR-RAW 데이터는 그림 2의 예시와 같이, 같은 영역을 24-240mm 줌렌즈의 초점 거리를 변경하여 촬영한 후, 짧은 초점 거리에서 촬영한 로우 센서 데이터를 저해상도 입력으로, 긴 초점 거리에서 촬영한 컬러 영상을 목표로 하는 고해상도 출력(ground-truth) 쌍으로 정의한다. 각 저해상도, 고해상도 영상 쌍은 그림 1과 같이 실내 혹은 야외에서 촬영한 7장의 영상으로 구성되어 있으며, 총 500개의 장면을 포함하고 있다.

위와 같은 방식으로 촬영한 영상은 그림 1의 예시에서 보는 바와 같이, 짧은 초점 거리에서 촬영한 영상이 긴 초점 거리에서 촬영한 영상보다 넓은 영역을 포함한다. 따라서, 본 논문에서는 저해상도 로우 센서 영상과 고해상도 컬러 영상을 학습 데이터 쌍으로 사용하기 위하여 짧은 초점 거리에서 촬영한 영상들을 가장 긴 초점 거리에서 촬영한 영상과 동일한 영역만을 포함하도록 잘라낸 후 사용한다. 그러나 영상을 잘라낸 후에도 학습 데이터 쌍이 기하학적으로 정확하게 정렬되지 않으므로, 영상 간의 유사도가 최대화되도록 변형 상관 계수 최대화(enhanced correlation coefficient maximization) 알고리즘<sup>[30]</sup>을 사용하여 비정렬(misalignment) 상태를 보상한다. 또한, 비정렬 상태 보완 후에도 입력 쌍이 완벽하게 정렬되지 않으므로 학습 과정 중에도 CoBi(Contextual Bilateral) 손실함수를 사용하여 화소 간 정렬을 보상한다. 그러나 해당 논문에서 사용한 비정렬 문제 해결 방안만으로는 대응쌍 간의 완벽한 정렬을 보장할 수 없으며, 데이터셋의 불완전성으로 인하여 기존 알고리즘과의 성능 비교를 어렵게 하는 한계점이 존재한다.

따라서, 위와 같은 데이터의 불완전성 문제를 해결하고자, 본 논문에서는 240mm 초점 거리에서 촬영한 높이가  $H$ , 너비가  $W$ 인 로우 센서 데이터  $X_r \in \mathbb{R}^{H \times W}$ 만을 활용하여 입력과 출력  $X_c \in \mathbb{R}^{H \times W \times 3}$  데이터 쌍을 생성한다. 다양한 크기의 저해상도 입력 영상을 생성하기 위하여, 로우 센서 데이터  $X_r$ 에 컬러 채널을 분리하기 위한 함수  $f_s(\cdot)$ 을 적용하여 입력 영상  $X_c \in \mathbb{R}^{H/2 \times W/2 \times 4}$ 을 생성하며,

$$X_c = f_s(X_r) \quad (1)$$

(1)에서 생성한  $X_c$ 에 다운샘플링(downsampling) 함수  $f_d(\cdot)$ 를 적용하여 크기를 4배, 8배 줄인 입력 데이터  $X_l$ 를 생성한다 (그림 1 참조).

$$X_l = f_d(X_c) \quad (2)$$

또한, 출력 데이터  $X_l$ 는 로우 센서 데이터  $X_r$ 을 RGB 영상으로 변환하여 생성한다.

## 3. 네트워크 구조

III-1장에서 기술한 바와 같이 센서에서 획득한 로우 센서 데이터는 카메라 내 영상처리 프로세싱 파이프라인을 통해 RGB 영상으로 변환된 후, 비선형적으로 색상, 명도, 채도 등을 보정하는 과정을 통해 시각적으로 보다 좋은 품질의 RGB 영상으로 렌더링된다. 이와 같은 비선형적 색상 보정 과정을 딥러닝으로 구현하기 위하여 기존의 연구들은 생성된 초고해상도 영상에 보정 상수를 곱해 컬러를 복원 하거나<sup>[27]</sup>, 해상도를 증가하는 과정과 분리하여 수행하는

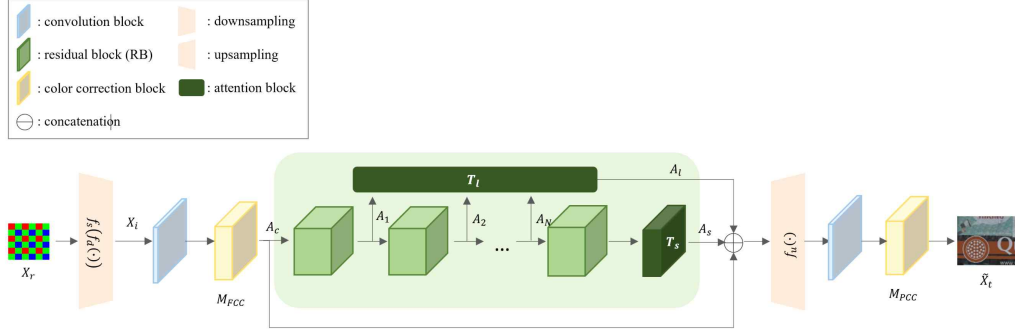


그림 3. 제안하는 로우 센서 데이터를 이용한 초고해상도 신경망 구조  
 Fig. 3. Proposed Corrector-Attention Network (CA-Net) architecture for raw sensor data

방식<sup>[28]</sup>을 사용한다. 그러나 본 논문에서는 저해상도 로우 센서 데이터를 종단간 학습(end-to-end learning) 방법을 통해 초고해상도 RGB 영상으로 변환하기 위하여 컬러 보정 모듈(corrector module)을 활용하는 그림 3과 같은 방식을 제안한다.

제안하는 로우 센서 데이터를 이용한 초고해상도 신경망 그림 3과 같은 구조로 이루어진다. 제안하는 신경망은 로우 센서 데이터를 입력으로 사용하여 컬러 채널이 분리된 (2)를 합성곱 층을 통과시켜 특징을 추출한 후, 추출된 특징 텐서(tensor)를 로우(raw) 레벨에서의 컬러 센서 별 신호 세기 균형을 위한 전단 컬러 보정 모듈(front color corrector module)  $M_{FCC}$ 의 입력으로 사용한다. 전단 컬러 보정 모듈은 주어진 입력값에 추정된 컬러 보정 행렬  $\theta_{FCC}$ 을 곱하여 식 (4)와 같이 다음 출력으로 반환하며, 이때,  $\theta_{FCC}$ 의 크기는 입력되는 미니 배치 크기에 이미지 또는 피쳐맵의 채널 수를 곱하여 결정한다.

$$A_c = M_{FCC}(Conv(X_i)) \quad (3)$$

$$M_{FCC}(X) = \theta_{FCC} \cdot X \quad (4)$$

전단 컬러 보정 모듈  $M_{FCC}$ 은 그림 4의 구조와 같이 특징맵  $conv(X_i)$ 를 입력으로 받아 합성곱 층, PReLU(Parametric Rectified Linear Unit), 평균 풀링(average pooling)을 통해 정보를 압축한 후, 트랜스포즈드 합성곱(transposed convolution)을 통해 추정된 컬러  $N$ 보정 행렬이 곱해진 특징맵을 생성한다. 이후, 컬러 보정 모듈을 통해 생성된 특징

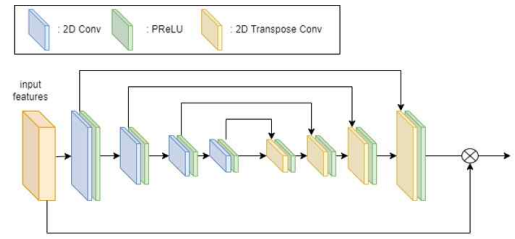


그림 4. 제안하는 컬러 보정 블록의 구조  
 Fig. 4. Proposed color corrector module

맵은 식 (5)와 같이 순차적으로  $N$ 개의 잔차 블록(residual block)을 통과한 후,

$$A_i = RB(A_{i-1}), \quad i = 1, 2, \dots, N \quad (5)$$

어텐션 블록(attention block)으로 구성된 심층 합성곱 층을 통하여 층간 기여도를 반영한 특징맵  $A_l$ 과 공간간, 채널간 기여도를 반영한 특징맵  $A_s$ 으로 생성된다<sup>[16][18]</sup>.

$$A_l = T_l(A_1, A_2, \dots, A_N) \quad (6)$$

$$A_s = T_s(A_N) \quad (7)$$

(7)에서 생성된 특징맵은 고해상도 영상 생성을 위한 샘플링  $f_u$  모듈을 통과한 후 화이트 밸런스와 같은 컬러 보정을 위하여 후단 컬러 보정 모듈(post color corrector module)  $M_{PCC}$ 을 통과한 후 최종적인 초고해상도 영상  $X_t$ 으로 생성된다.

$$X_t = M_{PCC}(f_u(concat(A_c + A_l + A_s))) \quad (8)$$

이와 같이, 제안하는 로우 센서 데이터를 사용한 초고해상도 신경망은 배치 크기가  $B$ 일 때, 아래와 같은 손실 함수  $L$ 을 사용하여 학습한다.

$$L = \frac{1}{B} \sum_{b=1}^B \| \widetilde{X}_{t,b} - X_{t,b} \|_1 \quad (9)$$

#### IV. 연구 결과

##### 1. 실험 설정

기본적으로 네트워크의 hyper-parameter 값들은 SR연구 중 SOTA 성능을 내는 HAN(Niu et al., 2020)의 실험에서 사용한 설정을 따라서 20개 residual group에 10개, residual block으로 네트워크 깊이를 맞췄다. 학습을 위한 입력 patch size는 96으로 하였고 미니배치크기(minibatch size)는 12개로 설정하였으며, x2, x3, x4, x8 배율로 SR-RAW 데이터를 학습용과 평가용으로 나눠서 준비하였다. 로우 영상을 읽고 후처리(post-processing)하기 위해서는 rawpy 라이브러리를 사용하였으며, 네트워크 입력을 위한 데이터

로더(dataloader)에서 해당 라이브러리를 이용해서 로우 영상을 읽고 처리하면 모델의 학습속도가 떨어지기 때문에 이를 npz 파일로 미리 데이터 세트를 준비하여 1 에폭(epoch)당 2927개 학습 데이터를 입력하였으며, 평가용으로는 337개 영상을 사용하여 학습과 평가를 하였다. Adam optimizer를 0.9, 0.999로 각각  $\beta_1, \beta_2$ 로 설정하였고 학습률(learning rate)를  $10^{-4}$ 로 시작하여  $10^{-8}$ 까지 80, 150, 200, 230 epoch 마다  $10^{-1}$ 씩 감소시켜가면서 총 250epoch를 학습하였다. 학습 하드웨어는 Intel i9 CPU에 RTX 3090 GPU를 사용하여 1.19sec/it의 속도로 초고해상도 스케일마다 10시간 정도의 학습시간이 들었다.

##### 2. 실험 결과

일반적인 초고해상도 연구에서 사용하고 있는 최대 신호 대잡음비(PeakSignal-to-noise ratio, PSNR), 영상 구조 유사성 지수(structural similarity index measure, SSIM)를 기준으로 성능을 측정하였으며, 색상 보정 모듈이 있을 때와 없는 경우를 실험하여 결과를 비교하였다. 결과는 표 1과 같다.

표 1. 성능 결과 지표의 비교. 본 연구에서 SR-RAW 데이터를 대상으로 하는 연구의 성능 결과 지표를 PSNR, SSIM 지표에 대해 비교한 결과 Table 1. Performance matric table shows precedent raw sensor images SR results. And comparing performance metric between with and without color collector module.

Model	Test data	4X		8X	
		PSNR	SSIM	PSNR	SSIM
without_color_collector	SR-RAW	22.06	0.8433	21.99	0.7939
with_color_collector	SR-RAW	23.05	0.8352	22.61	0.7768

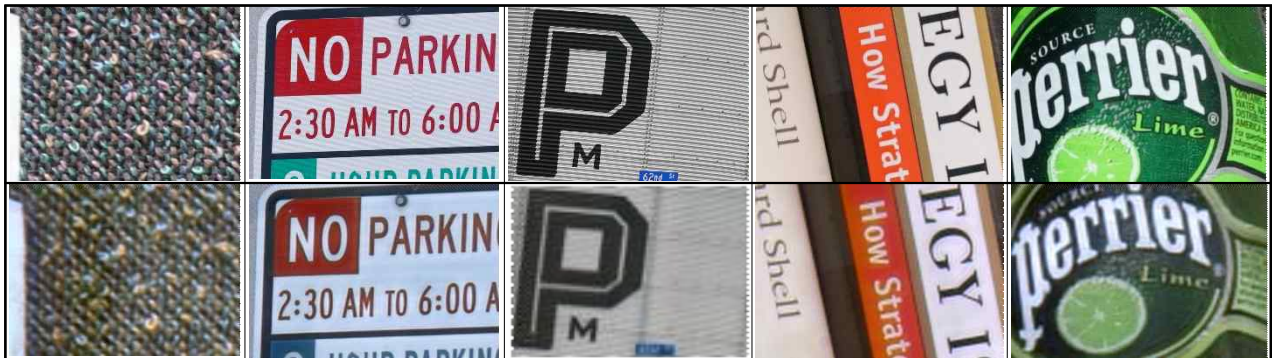


그림 5. 위열은 ground truth인 RGB 영상에서 특정 영역 확대한 패치, 아래열은 본 연구에서 만들어낸 4배 SR RGB 영상 Fig. 5. First line is ground truth HR RGB images, below is output x4 SR RGB images from out network

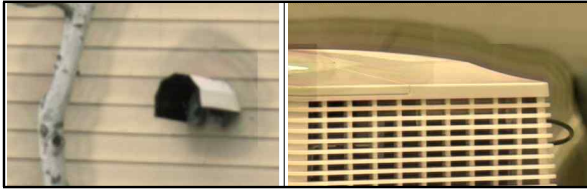


그림 6. 의도하지 않게 영상 내 사물 주변에 음영 형태의 아티팩트  
 Fig. 6. Unintended shadowing artifact around of objects on our SR RGB image

표 1에서 보는 것처럼 색상 보정 모듈을 사용하는 경우, PSNR에서 4X, 8X 스케일 모두 높은 성능을 나타내었다. 반면, SSIM은 오히려 색상 보정 모듈이 있는 경우에 색상 보정 모듈을 사용하지 않는 경우보다 지표가 낮게 나오는 것을 확인하였다. 이런 결과가 나타나는 이유는 색상 보정 모듈을 거치면서 밝기(brightness)가 조정되거나 네트워크가 깊어지면서 산출 이미지가 흐려지는(blurring)되는 이유인 것으로 추정된다. 하지만 SSIM의 성능 차이가 미미하여 출력된 결과 이미지만을 비교하여서는 정확한 원인을 찾을 수 없었다.

## V. 결론

본 연구에서는 로우 센서 영상을 입력 영상으로 활용할 수 있는 초해상화 네트워크를 제시하고 이를 실험하였다. 로우 센서 영상을 일반 RGB 영상처럼 네트워크 안에서 초해상화가 가능하도록 하기 위하여 색상 보정 모듈을 제안 하였으면, 이를 토대로 색상 보정 모듈이 유무에 따른 성능 변화를 확인하였다. 다만, 본 실험에서 모델을 통해 출력된 영상 결과물이 ground truth인 초고해상도 RGB영상과 비교하여 색상 보정 모듈의 사용에도 불구하고 일부 밝기(brightness)가 달라서 전체적인 색감이 다르게 보이는 문제가 발생하였으나 이에 대한 원인을 찾아내고 보완하지는 못하였다. 또한 색상 보정 모듈 사용 시에 PSNR 성능 지표가 향상됨에도 SSIM 지표가 더 낮게 나오는 이유를 명확히 밝히지 못하였다. 추후 연구에서 이러한 원인을 찾는 방향으로 연구를 진행하고자 한다.

## 참고 문헌 (References)

- [1] (2022) NITRE. [Online]. Available: <https://data.vision.ee.ethz.ch/cvl/ntire22/>
- [2] Q. Yuan, L. Zhang, and H. Shen, "Regional Spatially Adaptive Total Variation Super-Resolution With Spatial Information Filtering and Clustering," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp.2327-2342, 2013.  
doi: <https://doi.org/10.1109/TIP.2013.2251648>
- [3] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," *IEEE International Conference on Computer Vision (ICCV)*, pp.2272-2279, 2009.  
doi: <https://doi.org/10.1109/ICCV.2009.5459452>
- [4] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep Networks for Image Super-Resolution with Sparse Prior," *IEEE International Conference on Computer Vision (ICCV)*, pp.370-378, 2015.  
doi: <https://doi.org/10.1109/ICCV.2015.50>
- [5] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," *European Conference on Computer Vision (ECCV)*, pp.184-199, 2014.  
doi: [https://doi.org/10.1007/978-3-319-10593-2\\_13](https://doi.org/10.1007/978-3-319-10593-2_13)
- [6] 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.  
doi: <https://doi.org/10.48550/arXiv.1501.00092>
- [7] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1646-1654, 2016.  
doi: <https://doi.org/10.48550/arXiv.1511.04587>
- [8] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.770-778, 2016.  
doi: <https://doi.org/10.48550/arXiv.1512.03385>
- [9] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-Recursive Convolutional Network for Image Super-Resolution," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1637-1645, 2016.  
doi: <https://doi.org/10.48550/arXiv.1511.04491>
- [10] X.-J. Mao, C. Shen, and Y. Yang, "Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections," *A journal extension of the version at arXiv:1603.09056*, 2016.  
doi: <https://doi.org/10.48550/arXiv.1606.08921>
- [11] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.105-114, 2017.  
doi: <https://doi.org/10.48550/arXiv.1609.04802>
- [12] G. Huang, Z. Liu, L. Maaten and K. Weinberger, "Densely Connected Convolutional Networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.4700-4708, 2018.  
doi: <https://doi.org/10.48550/arXiv.1608.06993>
- [13] T. Tong, G. Li, X. Liu, and Q. Gao, "Image Super-Resolution Using



- Dense Skip Connections," IEEE International Conference on Computer Vision (ICCV), pp.4809-4817, 2017.  
doi: <https://doi.org/10.1109/ICCV.2017.514>
- [14] S. Anwar, and N. Barnes, "Densely Residual Laplacian Super-Resolution," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 3, pp.1192 - 1204, 2022.  
doi: <https://doi.org/10.48550/arXiv.1906.12021>
- [15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, Ł. Kaiser, and I. Polosukhin. "Attention is all you need," Advances in Neural Information Processing Systems(NeurlPS), 2017.
- [16] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," European conference on computer vision (ECCV), pp.286-301, 2018.  
doi: <https://doi.org/10.48550/arXiv.1807.02758>
- [17] T. Dai, J. Cai, Y. Zhang, S. Xia, and L. Zhang, "Second-Order Attention Network for Single Image Super-Resolution," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.11057-11066, 2019.  
doi: <https://doi.org/10.1109/CVPR.2019.01132>
- [18] B. Niu, W. Wen, W. Ren, X. Zhang, L. Yang, S. Wang, K. Zhang, X. Cao, and H. Shen, "Single Image Super-Resolution via a Holistic Attention Network," European Conference on Computer Vision (ECCV), pp.191-207, 2020.  
doi: <https://doi.org/10.48550/arXiv.2008.08767>
- [19] Y. Mei, Y. Fan, Y. Zhou, L. Huang, T. Huang, and H. Shi, "Image Super-Resolution with Cross-Scale Non-Local Attention and Exhaustive Self-Exemplars Mining," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.5690-5699, 2020.  
doi: <https://doi.org/10.48550/arXiv.2006.01424>
- [20] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1874-1883, 2016.  
doi: <https://doi.org/10.48550/arXiv.1609.05158>
- [21] C. Dong, C. C. Loy, and X. Tang, "Accelerating the Super-Resolution Convolutional Neural Network," European Conference on Computer Vision (ECCV) pp.391-407, 2016.  
doi: <https://doi.org/10.48550/arXiv.1608.00367>
- [22] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.624-632, 2017.  
doi: <https://doi.org/10.48550/arXiv.1704.03915>
- [23] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A Persistent Memory Network for Image Restoration," IEEE International Conference on Computer Vision (ICCV), pp.4549-4557, 2017.  
doi: <https://doi.org/10.48550/arXiv.1708.02209>
- [24] W. Han, S. Chang, D. Liu, M. Yu, M. Witbrock, and T. S. Huang, "Image Super-Resolution via Dual-State Recurrent Networks," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1654-1663, 2018.  
doi: <https://doi.org/10.48550/arXiv.1805.02704>
- [25] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep Back-Projection Networks for Super-Resolution," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1654-1663, 2018.  
doi: <https://doi.org/10.48550/arXiv.1803.02735>
- [26] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback Network for Image Super-Resolution," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.3862-3871, 2019.  
doi: <https://doi.org/10.48550/arXiv.1903.09814>
- [27] X. C. Zhang, Q. Chen, R. Ng, and V. Koltun, "Zoom to Learn, Learn to Zoom," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.3757-3765, 2019.  
doi: <https://doi.org/10.48550/arXiv.1905.05169>
- [28] X. Xu, Y. Ma, W. Sun, and M.-H. Yang, "Exploiting Raw Images for Real-Scene Super-Resolution," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 1, pp.1905-1921, 2022.  
doi: <https://doi.org/10.48550/arXiv.2102.01579>
- [29] Y. Lee and K. Hiraakawa, "Lossless White Balance for Improved Lossless CFA Image and Video Compression," IEEE Transactions on Image Processing, vol. 31, pp.3309-3321, 2022.  
doi: <https://doi.org/10.48550/arXiv.2009.09137>
- [30] G. D. Evangelidis, and E. Z. Psarakis, "Parametric Image Alignment Using Enhanced Correlation Coefficient Maximization," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 10, pp. 1858-1865, 2008.  
doi: <https://doi.org/10.1109/TPAMI.2008.113>
- [31] Xu, Xuan, and Xin Li. "Scan: Spatial color attention networks for real single image super-resolution." IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW), 2019.  
doi: <https://doi.org/10.1109/CVPRW.2019.00254>

---

저 자 소 개

---

신 바 울



- 2020년 3월 ~ 현재 : 서울과학기술대학교 스마트에너지시스템공학과 석사과정
- 2021년 11월 ~ 현재 : ㈜크레이지알파카 소프트웨어 엔지니어
- ORCID : <https://orcid.org/0000-0002-0009-8832>
- 주관심분야 : Computer Vision컴퓨터 비전, Recommender System추천 시스템

김 태 하



- 2020년 2월 : 서울과학기술대학교 전기정보공학과 공학사
- 2020년 2월 ~ 2022년 8월 : 서울과학기술대학교 전기정보공학과 석사과정
- 2020년 5월 ~ 현재 : LG전자
- 주관심분야 : 컴퓨터 비전, 영상 처리, 기계 학습:

이 의 진



- 2017년 9월 : 미국 University of California, San Diego 전기컴퓨터공학과 공학박사
- 2017년 ~ 2018년 : 미국 University of California, Los Angeles Postdoctoral Fellow
- 2018년 ~ 현재 : 서울과학기술대학교 조교수
- ORCID : <https://orcid.org/0000-0002-3439-5042>
- 주관심분야 : 컴퓨터 비전, 기계 학습