

## Deep Learning-Based Low-Light Imaging Considering Image Signal Processing

Minsu Kwon\*

\*Researcher, ENERZAI, Seoul, Korea

### [Abstract]

In this paper, we propose a method for improving raw images captured in a low light condition based on deep learning considering the image signal processing. In the case of a smart phone camera, compared to a DSLR camera, the size of a lens or sensor is limited, so the noise increases and the reduces the quality of images in low light conditions. Existing deep learning-based low-light image processing methods create unnatural images in some cases since they do not consider the lens shading effect and white balance, which are major factors in the image signal processing. In this paper, pixel distances from the image center and channel average values are used to consider the lens shading effect and white balance with a deep learning model. Experiments with low-light images taken with a smart phone demonstrate that the proposed method achieves a higher peak signal to noise ratio and structural similarity index measure than the existing method by creating high-quality low-light images.

▶ **Key words:** Low Light, Image Sensor, Deep Learning, Smart Phone, Camera

### [요 약]

이 논문에서는 image signal processing 을 고려하여 저조도에서 촬영된 저품질의 raw 이미지를 딥러닝에 기반하여 개선하는 방법을 제안한다. 스마트폰 카메라의 경우 DSLR 카메라에 비해 렌즈나 센서의 확장에 제약이 있어 저조도 상황에서 이미지에 노이즈가 증가되고 품질이 저하되는 문제점을 보인다. 기존 딥러닝 기반 저조도 이미지 처리 방식은 image signal processing의 주요 요소인 렌즈 셰이딩 효과와 화이트 밸런스를 고려하지 못하여 부자연스러운 이미지를 생성하기도 한다. 본 논문에서는 렌즈 셰이딩 효과와 화이트 밸런스를 딥러닝 모델에 적용하기 위해 중심거리와 채널 평균을 활용한다. 스마트폰으로 촬영된 저조도 이미지를 통한 실험에서 제안하는 방법이 기존 방법에 비해 더 높은 peak signal to noise ratio 와 structural similarity index measure를 달성함과 동시에 높은 품질의 저조도 이미지를 생성함을 확인한다.

▶ **주제어:** 저조도, 이미지 센서, 딥러닝, 스마트폰, 카메라

---

• First Author: Minsu Kwon, Corresponding Author: Minsu Kwon  
\*Minsu Kwon (minsu.kwon@enerzai.com), ENERZAI  
• Received: 2023. 01. 02, Revised: 2023. 01. 25, Accepted: 2023. 02. 09.

## I. Introduction

저조도 환경에서 촬영된 이미지는 밝기가 충분한 환경에서 촬영된 이미지에 비해 낮은 품질을 보인다. 이는 어두운 환경에서 촬영시 이미지 센서로 들어오는 광자의 양이 충분하지 않아 노이즈의 영향을 많이 받기 때문이다. 기존 하드웨어에 내장되어있는 image signal processing (ISP) 는 이미지 센서에서 측정된 raw 이미지를 디스플레이 가능한 sRGB로 변환하는 작업을 수행한다. 이때 저조도에서 촬영된 raw 이미지에 대해서는 노이즈 제거 (noise reduction) 필터의 한계로 인해 노이즈 제거가 제대로 이루어지지 않거나 검정색에 가까운 낮은 RGB값을 출력한다.

물리적인 방법으로 저조도에서 고품질의 이미지를 얻는 방법에는 여러 가지가 있다. 먼저 렌즈나 이미지 센서의 크기를 증가시켜 받아들이는 빛의 양을 늘릴 수 있다. 동일한 시간에 더 많은 빛을 받게 되면 센서에 측정되는 값이 노이즈로부터 받는 영향이 적어져 기존 ISP만을 가지고 고도 고품질의 이미지를 얻을 수 있다. 하지만 스마트폰 카메라와 같이 렌즈나 센서 확장에 제약이 있는 경우 이러한 방법을 적용하기 어렵다.

두 번째로 같은 센서와 렌즈만을 가지고도 노출시간을 늘림으로서 받아들이는 빛의 양을 늘려 높은 품질의 저조도 이미지를 얻을 수 있다. 하지만 노출시간이 길어지게 되면 카메라의 떨림과 피사체의 순간적인 변화에 영향을 많이 받게 된다.

위 문제점 때문에 물리적으로 노이즈가 적은 값을 취득하기 보다는 이미지 처리 분야 [6-7][11] 에서 높은 성능을 보이는 딥러닝 모델 기반으로 저조도의 낮은 품질의 raw 이미지를 처리하려는 연구들이 진행되어왔다. [1][3] 이 중 딥러닝 학습기반의 ISP로 단노출의 저조도 raw 이미지를 처리하는 연구는 장노출 이미지와 유사한 결과를 달성하였다. [3] 하지만 이미지의 일부인 패치만을 가지고 지도 학습을 진행하기 때문에 전체 이미지를 통해 고려되는 화이트 밸런스 [2] 와 렌즈 왜곡 [10] 효과를 반영하지 못해 일부 영역에서 부자연스러운 결과를 낸다. 이러한 문제를 해결하기 위해 본 연구에서는 렌즈 왜곡과 화이트 밸런스를 고려한 딥러닝 기반 저조도 raw 이미지 처리 방법을 제안한다. 제안하는 모델은 렌즈 왜곡과 화이트 밸런스 추정에 필요한 정보를 UNet [8] 구조에 효과적으로 결합시킨다.

본 논문에서는 스마트폰을 이용하여 취득된 고해상도의 저조도 이미지 데이터셋에서의 실험을 통해 기존모델보다 향상된 PSNR (Peak Signal to Noise Ratio) 과 SSIM (Structural Similarity Index Measure) 을 다양한 색공간

(RGB, YUV, HSV) 에서 달성함을 확인한다. 또한 기존 모델의 출력 이미지와 개선된 모델의 출력 이미지를 비교하여 제안하는 방법이 가져오는 효과를 정성적으로 확인한다.

## II. Preliminaries

### 1. Background knowledge

#### 1.1 ISP

ISP는 이미지 센서에서 측정된 raw 이미지를 디스플레이 이등에 출력하여 인간이 인지할 수 있는 sRGB 형태로 변환하는 알고리즘이다. ISP의 종류는 여러 가지가 있지만 대표적으로 Fig. 1 과 같이 표현될 수 있다. ISP는 noise reduction, white balance, color filter array interpolation, color correction, gamma correction 과 같은 알고리즘들의 연속으로 이루어진다. 본 논문에서 제안하는 방법은 이러한 여러 개의 요소로 이루어진 ISP 를 하나의 딥러닝 기반 모델로 대체하여 저조도에서 발생하는 문제점을 해결한다.

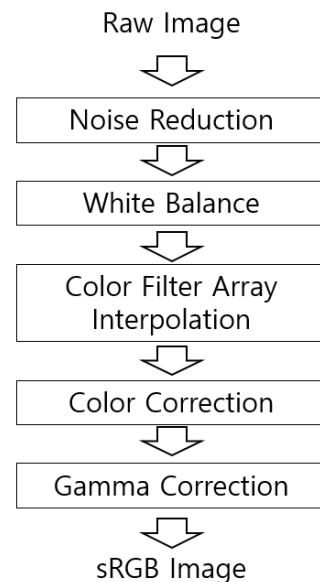


Fig. 1. Traditional Image Processing Pipeline

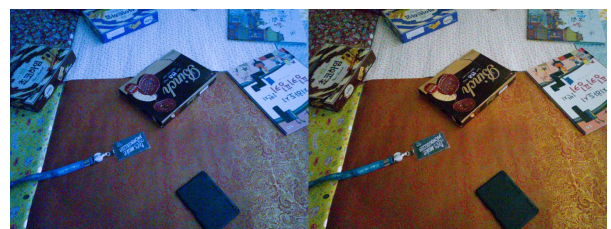


Fig. 2. Wrong and Correct White Balance

## 1.2 UNet

이 논문에서 사용되는 딥러닝 모델은 UNet 구조 [8] 를 따른다. UNet 구조는 피라미드 구조를 가지고 있어 이미지의 국소적인 특징 (local features) 과 광역적인 특징 (global feature)를 동시에 고려하여 의료, 제조업 등 다양한 분야에서 이미지 분석에 활용될 수 있다. 또한 저조도 이미지 개선에 UNet이 적합한 이유에는 먼저 광역적인 특징 추출을 통한 노이즈 제거 능력을 들 수 있다. 저조도 이미지 개선에서 UNet 은 점진적으로 특징맵을 줄여나가면서 노이즈 제거를 수행한다. 다음 점차적으로 특징맵을 확장해나가며 세부적인 특징들을 완성한다.

UNet의 경우 다른 이미지 처리 모델 구조에 비해 작은 크기의 특징맵을 다루므로 그에 비례하여 적은 계산량을 가진다. 이러한 UNet의 효율성은 ultra high definition 급의 고화질의 이미지를 처리하는 본 연구의 목적에 부합한다.

## 2. Related works

기존에 저조도 환경에서 전통적인 ISP가 갖는 한계점을 딥러닝을 통해 해결하려는 시도는 존재해왔다. 이미지의 밝기를 고려하여 픽셀값을 조절하는 연구 [4] 는 저조도의 노이즈를 고려하지 않아 고품질의 이미지를 합성하지 못한다. 저조도 환경의 노이즈를 처리하는 것에 초점을 둔 연구 [1] 는 raw 이미지 또는 RGB 이미지 상에서 denoising을 수행한다. raw 이미지에서 denoising을 하는 경우 추가작업을 통해 ISP를 최적화해야하며 RGB 상에서 denoising을 하는 경우 ISP에서 raw 이미지의 정보 손실이 있어 극저조도 환경에서는 효과적인 denoising이 어렵다.

반면 딥러닝 기반 ISP 방법 [3]은 모델의 입력은 단노출의 raw 이미지로 출력은 장노출의 RGB 이미지가 되도록 한다. 하나의 파이프라인으로 저조도의 raw 이미지를 고품질의 RGB 이미지로 처리할 수 있다는 장점이 있지만, 이 방법에서 모델은 이미지 패치단위로만 학습하여 픽셀의 위치에 따라 달라지는 렌즈 쉐이딩 [10] 효과와 이미지 전체의 화이트 밸런스 [2] 를 고려하지 못해 부자연스러운 이미지를 출력하는 문제점이 있다.

## III. The Proposed Scheme

### 1. Position Information

렌즈 쉐이딩 [10] 효과는 이미지 센서의 중심에서 멀어질수록 이미지 센서 픽셀당 받아들이는 빛의 양이 적어져 상대

적으로 중심부에 비해 어둡게 나오거나 더 많은 노이즈가 나타나는 현상이다 (Fig. 3). 이러한 렌즈 쉐이딩이 존재하는 이미지들을 이용하여 기존 딥러닝 기반 저조도 이미지 개선 모델을 학습하게 되면 이미지 가장자리 근처의 많은 노이즈들로 인해 학습된 노이즈 제거 경향에 따라 이미지 중심 근처의 패턴들 또한 노이즈 제거 작업을 거쳐 흐릿하고 단순한 형태의 이미지로 변환되는 문제점을 보일 수 있다.

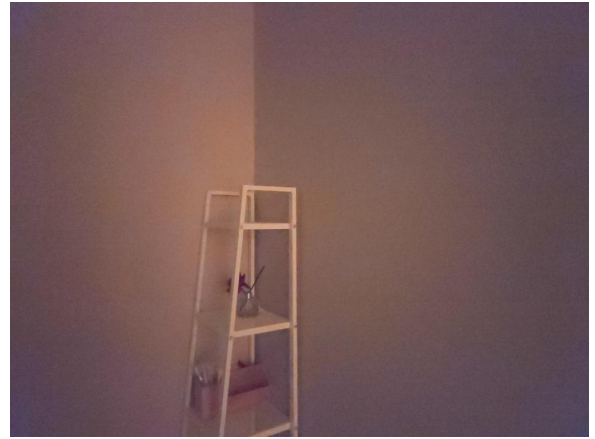


Fig. 3. Lens shading Effect (Vignetting)

본 연구에서는 UNet의 입력에 각 픽셀마다 픽셀의 위치정보를 주어 각 위치에 따라 렌즈 쉐이딩 효과를 고려하도록 한다. 본 논문에서는 직접적으로 위치 정보를 주기보다는 이미지 중심으로 부터의 거리에 따라 렌즈 쉐이딩 효과가 커지는 점에 착안하여 이미지 중심으로 부터의 거리를 모델의 입력에 주도록 한다. 거리는 이미지의 가로가  $H$ , 세로가  $W$ 로 정의되어 있을 때 아래와 같이 각 픽셀  $i, j$  마다 계산된다.

$$D_{i,j} = \sqrt{\left(1 - \frac{i}{0.5 \times H}\right)^2 + \left(1 - \frac{j}{0.5 \times W}\right)^2}$$

이 위치정보를 통하여 모델은 같은 패턴의 raw 데이터가 입력이 되더라도 픽셀의 위치에 따라 다르게 처리할 수 있다. 예를 들어 특정 고주파 패턴이 raw에서 관측되었을 때 이것이 이미지 중심부에 있다면 노이즈가 아닌 실제 이미지의 패턴으로 인식되며 가장자리에서 관측된다면 렌즈 쉐이딩 효과에 의한 노이즈로 인식하여 노이즈 제거 작업을 수행한다.

### 2. Channel Average

ISP에서 화이트 밸런스 [2] 는 이미지 전체의 색분포를 참고하여 피사체의 색온도를 적절한 값으로 조절한다. 하

지만 학습기반의 딥러닝 모델은 이미지 전체가 아닌 이미지 일부의 패치로 학습이 되기 때문에 학습에 전체 색분포가 고려되기 어렵다. 우리는 딥러닝 모델의 입력에 패치 말고도 전체 raw 이미지에서 각 RGBG 채널별로 평균값을 계산한 다음 모델의 입력값으로 활용한다. 이미지의 가로, 세로가  $H, W$  이고 각 채널  $c$ 에 대해 평균 값  $A^c$ 는 다음과 같이 계산된다.

$$A^c = \frac{1}{H \times W} \sum I_{i,j}^c$$

### 3. Loss Function and Training

제안하는 저조도 이미지 개선 모델( $f$ )에 대해 저조도에서 짧은 노출에서 촬영된 raw 이미지  $I$  와 이에 대한 픽셀 위치정보  $D$ , 채널 평균  $A$  가 주어졌을 때 합성된 높은 품질의 이미지 ( $\hat{Y}$ )는 다음과 같이 정의되며 모델 구조는 Fig. 4와 같이 표현된다.

$$\hat{Y} = f(I, D, A)$$

이때 최소화해야 하는 손실함수는 다음과 같이 정의한다.

$$\|Y - \hat{Y}\|_1 + \alpha(1 - MSSSIM(Y, \hat{Y}))$$

위 식에서 첫 번째 항은 합성한 이미지  $\hat{Y}$  와 실제 장노출 이미지  $Y$  의 픽셀값 차이가 최소화 되도록 하며 MSSSIM (multi-scale structural similarity index measure) [9] 는 합성된 이미지가 고품질의 이미지와 사람이 시각적으로 비슷하게 인지하도록 한다.

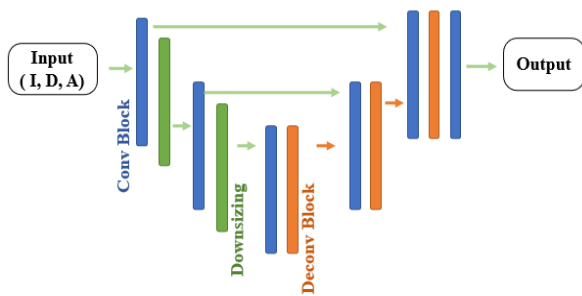


Fig. 4. Model architecture

## IV. Experiments

### 1. Data

우리는 Sony IMX 563 CMOS 이미지 센서가 탑재되어 있는 삼성 갤럭시 Z Flip3 스마트폰을 이용하여 데이터셋을 취득하였다. 한 장면에 대하여 0.1초의 노출과 400의

ISO로 촬영을 한 다음 같은 고정된 초점에서 4초의 노출로 높은 품질의 이미지를 얻었다. 총 260 장면들에 대하여 위와 같이 촬영을 진행하였으며 이 중 230장을 학습 30장을 테스트에 활용한다. 한편 촬영된 이미지의 세로는 3000 가로는 4000의 사이즈를 가진다.

### 2. Training

딥러닝 모델의 학습은 Intel(R) Xeon(R) Silver 4210R CPU @ 2.40GHz CPU, 256GB RAM 의 기기의 NVIDIA RTX 3090 그래픽카드를 이용하여 진행하였다. 본 실험에서 사용된 하이퍼파라미터는 다음과 같다. 600 epochs 동안 학습을 진행하였으며 learning rate 는 0.0001 로 설정 되었다. 학습 raw 이미지 패치의 사이즈는 512이며 학습데이터에서 무작위로 추출되었다. 배치 사이즈는 1로 설정되었다. 입력과 타겟출력은 Table 1과 같이 정리된다.

Table 1. Model Input and Target

	Input	Target
Image Type	Raw	sRGB
ISO	400	400
Exposure Time	0.1 sec.	4.0 sec.

Table 2. RGB Space Score

	Baseline [1]	Proposed
PSNR	24.96	25.67
SSIM	0.7860	0.8008

### 3. Results

#### 3.1 RGB Space

이 절에서는 UNet을 가지고 기본적인 학습방법만을 이용하여 학습한 baseline [3] 과 position information 과 channel average 정보를 추가하여 개선한 모델 (proposed model) 의 성능과 결과를 비교한다.

먼저 4초 노출로 촬영된 고품질의 이미지와 0.1초 노출의 raw 이미지를 처리하여 생성된 결과의 유사도를 PSNR 과 SSIM으로 측정하였다. 그 결과 Table 2 와 같이 두 지표 모두에서 기존 모델보다 높은 유사도를 달성함을 확인하였다.



Fig. 5. Zoomed view of output samples of baseline model (top) and proposed model (bottom) show the effectiveness of the position information

또한 테스트셋의 이미지에서 이미지 중심부근의 그물 구조의 패턴을 기존 모델은 노이즈로 인식하여 흐릿하게 처리하였지만 제안하는 모델은 위치 정보를 통해 노이즈가 아닌 패턴으로 인식하고 선명한 이미지를 생성하였다. (Fig. 5)

또한 Fig. 6 과 같이 기존 모델의 결과에서는 하늘의 색이 부자연스럽게 출력되었지만 제안하는 모델에서는 고품질의 이미지와 유사한 색으로 자연스러운 이미지를 합성하였다.

### 3.2 YUV Space

학습된 모델의 결과에서 대체적으로 baseline 과 제안하는 방법 모두 밝기에 대한 학습은 효과적으로 이루어졌지만 색상에 표현력 차이를 확인하기 위해 YUV 색공간에서 비교를 진행하였다. Table 3 의 U채널 PSNR과 V채널 PSNR 에서의 baseline 과 제안하는 모델의 성능 차이가 밝기 정보인 Y채널의 PSNR보다 높은 것을 확인할 수 있다. 이는 Fig. 7 의 UV채널 에러맵과 같이 기존 baseline 모델이 화이트 밸런스를 맞추지 못했던 영역에서 제안하는 모델은 channel average를 통해 적절한 화이트 밸런스를 찾기 때문이다.



Fig. 6. Output samples with different white balance : 0.1s exposed image (1st), 0.1s exposed image scaled up with Adobe Lightroom (2nd), baseline model (3rd), proposed model (4th), 4s exposed image (5th)

Table 3. YUV Space Score

		Baseline [1]	Proposed
PSNR	Y	31.22	31.53
	U	35.88	36.86
	V	34.34	35.57
SSIM	Y	0.7836	0.7983
	U	0.9753	0.9759
	V	0.9787	0.9790

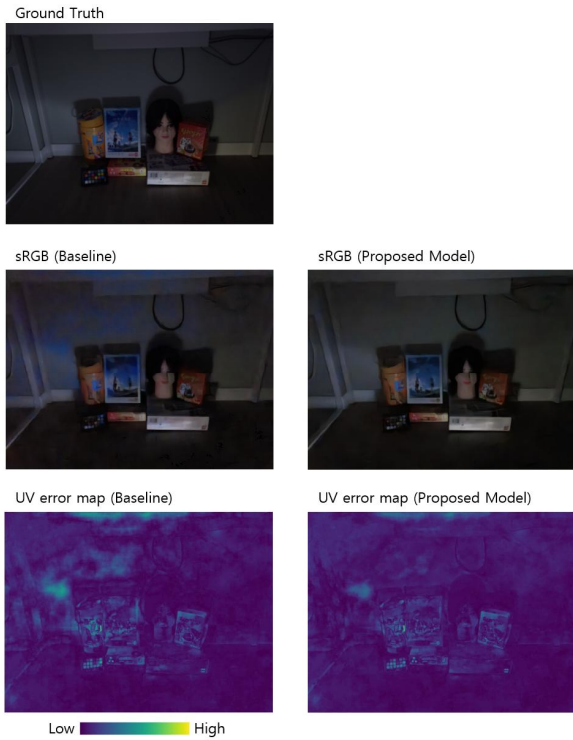


Fig. 7. UV channel error maps

### 3.3 HSV Space

HSV 색공간은 색상(Hue), 채도(Saturation), 명도(Value) 로 이루어져 있어 보다 인간의 시각과 비슷한 측면을 지닌다. HSV 색공간에서 PSNR과 SSIM은 Table 4 와 같이 정리되며 특히 채도 (S) 의 경우 SSIM에서 baseline과 제안하는 모델의 성능차이가 컸다. Fig. 8의 채도 에러맵에서 기존 baseline의 채도값들이 고품질의 ground truth 채도값보다 높아 높은 에러를 발생시키고 부자연스러운 색상을 나타낸다. 반면 제안하는 모델의 결

Table 4. HSV Space Score

		Baseline [1]	Proposed
PSNR	H	30.16	30.42
	S	28.36	28.54
	V	30.09	30.65
SSIM	H	0.6316	0.6506
	S	0.5208	0.5435
	V	0.7786	0.7928

과는 적절한 채도로 자연스러운 색상을 보이며 낮은 채도 에러를 보인다.

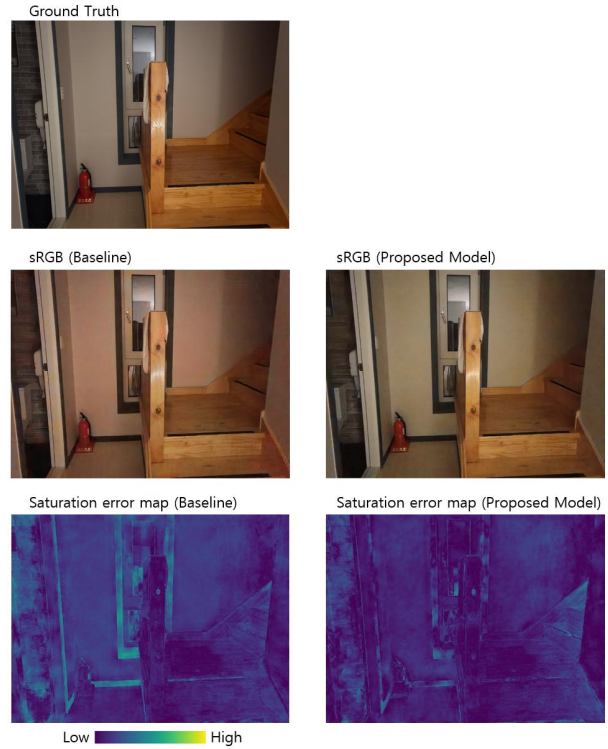


Fig. 8. Saturation channel error maps

## V. Conclusions

이 논문을 통해 이미지 센서의 특징을 고려하여 저조도에서 촬영된 이미지를 개선하는 방법을 제안하였다. 기존 딥러닝 기반 저조도 이미지 처리 방식인 렌즈 웨이딩 효과와 광역 화이트 밸런스를 고려하지 못하는 문제점을 본 논문에서는 중심거리와 채널 평균을 모델의 입력에 추가해 주었다. 스마트폰 이미지를 통한 실험으로 위 해결방법이 가져오는 효과를 확인하였다.

## ACKNOWLEDGEMENT

This work was supported by The Ministry of Trade, Industry and Energy(MOTIE, Korea). [Project Name: Development and Substantiation of Real-time Acquisition and Artificial Intelligence based Analysis Technology of Deposit Information for Rapid Exploration of Rare Metals Resources / Project Number: 20216110100060]

## REFERENCES

- [1] Abdelhamed, A., Lin, S., and Brown, M. S. A high-quality denoising dataset for smartphone cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1692-1700, June 2018. DOI: 10.1109/cvpr.2018.00182
- [2] Cepeda-Negrete, J., and Sanchez-Yanez, R. E. Gray-world assumption on perceptual color spaces. In Pacific-Rim Symposium on Image and Video Technology, pp. 493-504. 2014. DOI: 10.1007/978-3-642-53842-1\_42
- [3] Chen, C., Chen, Q., Xu, J., and Koltun, V. Learning to see in the dark. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3291-3300. June 2018. DOI: 10.1109/cvpr.2018.00347
- [4] Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., and Cong, R. Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1780-1789. June 2020. DOI: 10.1109/cvpr42600.2020.00185
- [5] Ignatov, A., Byeoung-Su, K., Timofte, R., and Pouget, A. Fast camera image denoising on mobile gpus with deep learning, mobile ai 2021 challenge: Report. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2515-2524, June 2021. DOI: 10.1109/cvprw53098.2021.00285
- [6] Lim, B., Son, S., Kim, H., Nah, S., and Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 136-144, July 2017. DOI: 10.1109/cvprw.2017.151
- [7] Lugmayr, A., Danelljan, M., Gool, L. V., and Timofte, R. SrfLOW: Learning the super-resolution space with normalizing flow. In European conference on computer vision, pp. 715-732, Aug. 2020. Springer, Cham. DOI: 10.1007/978-3-030-58558-7\_42
- [8] Ronneberger, O., Fischer, P., and Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham. DOI: 10.1007/978-3-319-24574-4\_28
- [9] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), pp. 600-612, Apr. 2004. DOI: 10.1109/tip.2003.819861
- [10] Young, I. T. Shading correction: compensation for illumination and sensor inhomogeneities. Current Protocols in Cytometry, 14(1), pp. 2-11, Oct. 2000. DOI: 10.1002/0471142956.cy0211s14
- [11] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European conference on computer vision (ECCV) pp. 286-301, Sep. 2018. DOI: 10.1007/978-3-030-01234-2\_18

## Authors



Minsu Kwon received the B.S. degree in Electrical and Computer Engineering from Seoul National University, Korea, in 2018 and M.S. degree in Computer Science from KAIST, Korea, in 2020.

Mr. Kwon is currently a Researcher in ENERZAi, Korea. He is interested in low light imaging, neural ISP and neural network compression.