

Metrics for Low-Light Image Quality Assessment

Sangmin Kim*

*Researcher, ENERZAI, Seoul, Korea

[Abstract]

In this paper, it is confirmed that the metrics used to evaluate image quality can be applied to low-light images. Due to the nature of low-illumination images, factors related to light create various noise patterns, and the smaller the amount of light, the more severe the noise. Therefore, in situations where it is difficult to obtain a clean image without noise, the quality of a low-illumination image from which noise has been removed is often judged by the human eye. In this paper, noise in low-illumination images for which ground truth cannot be obtained is removed using Noise2Noise, and spatial resolution and radial resolution are evaluated using ISO 12233 charts and colorchecker as metrics such as MTF and SNR. It can be shown that the quality of the low-illumination image, which has been evaluated mainly for qualitative evaluation, can also be evaluated quantitatively.

▶ **Key words:** Low light enhancement, Denoising, Noise2Noise, Deep learning, MTF, SNR

[요약]

본 논문에서는 기존에 영상의 품질을 평가하는 데 사용되던 지표가 저조도 영상에 대해서도 적용될 수 있음을 확인한다. 저조도 영상의 특성상, 빛과 관련된 요인들이 다양한 잡음 패턴을 만들어내고 빛의 양이 적을수록 극심한 잡음을 가지고 있다. 그렇기 때문에, 잡음이 없는 깨끗한 영상을 구하기 힘든 상황에서 잡음이 제거된 저조도 영상의 품질을 사람의 눈으로 판단하는 경우가 많다. 본 논문에서는, ground truth를 구할 수 없는 저조도 영상의 잡음을 Noise2Noise를 이용해서 제거하고, MTF와 SNR 등의 지표로 공간 해상도와 방사 해상도를 ISO 12233 차트와 colorchecker를 대상으로 평가한다. 정성적 평가 위주로 평가되던 저조도 영상의 품질이 정량적으로도 평가될 수 있음을 보여줄 수 있다.

▶ **주제어:** 저조도 개선, 잡음 제거, Noise2Noise, 딥러닝, MTF, SNR

I. Introduction

사람들이 카메라나 컴퓨터를 통해 보고 있는 영상은 다양한 원인으로 인해 손상되어 원치 않는 잡음을 얻을 수 있다. 그 원인은 카메라 자체의 물리적인 원인일 수도 있고, 카메라나 컴퓨터 안의 소프트웨어로 인한 것일 수도 있다. 이와 같이, 영상의 잡음을 제거하여 품질을 개선하는 일은 하나의 연구 분야로서 자리 잡고 있다.

영상의 잡음을 제거하는 데에는 이전부터 Wiener filter [1] 등의 rule-based filter (규칙 기반 필터)를 많이 사용해왔지만, 최근에는 딥러닝 기반의 convolutional neural network (합성곱 신경망; CNN)이 많이 사용되고 있다. 영상의 품질이 저하되는 원인은 상당히 다양하고 이를 단순히 영상을 흐리게 하는 것 이상으로 없앨 수 있는 방법은 그다지 많지 않다. 수식이 아닌 데이터를 이용하는 CNN은 이를 해결하였으며, 품질을 개선하고 싶은 영상과 개선한 후의 영상을 데이터쌍으로 가지고 있다면 영상의 잡음을 쉽게 제거할 수 있다.

저조도 조건에서 촬영된 영상은 개선한 후의 영상을 얻기 힘든 경우에 속한다. 일상적인 조건에서는 빛이 적음에 따라 카메라의 조리개를 열거나 셔터 스피드와 센서의 감도를 늘리면서 빛의 양을 늘려 보상해준다. 하지만, 저조도 조건에서는 이러한 방법을 적용해도 어두운 영상이 촬영되며, 해당 방법들은 과도한 셔터 스피드로 인해 영상의 프레임 주사율을 낮추거나 센서의 감도로 인한 추가적인 잡음이 생길 수 있다.

또한, 영상의 품질을 평가하기 위한 측면에서는 잡음을 제거하기 전후 영상이 있으면 peak signal-to-noise ratio (최대 신호 대 잡음비; PSNR)나 structural similarity (구조적 유사성; SSIM)을 사용하여 데이터쌍의 유사성을 평가하는 것이 가능하다. 하지만, 저조도 조건에 의해 잡음이 발생한 영상이나 그러한 잡음을 제거한 영상은 비교할만한 ground truth (딥러닝 기반의 솔루션에서 학습하도록 설정하는 정답에 해당하는 데이터; GT)가 존재하지 않기 때문에 최대 신호 대 잡음비나 구조적 유사성을 사용하기 어렵다.

본 논문에서는 저조도 조건에서의 극심한 잡음을 제거하기 위해 GT를 필요로 하지 않는 Noise2Noise로 U-Net을 학습한다. 또한, 잡음이 있는 영상과 그러한 잡음이 제거된 영상을 평가하기 위한 지표로서 위성 영상 등에서 기존에 사용되고 있는 MTF와 SNR을 제안한다.

II. Preliminaries

1. Image Denoising

1.1 U-Net

CNN은 딥러닝 영상처리 분야에서 다양하게 연구되고 있다. 합성곱 층의 입력과 해당 층을 통과한 출력을 더하여 그 차이만 효율적으로 학습하는 ResNet [2] 이나 입력과 출력을 병합하여 다음 합성곱 층의 입력으로 사용하여 feature map (특징 맵) 의 활용도를 극대화하는 DenseNet [3] 등의 CNN은 영상에 찍힌 피사체를 구별하는 image classification (영상 분류) 분야에서 좋은 성능을 보였다.

U-Net [4] 은 image segmentation (영상 분할) 분야에서 좋은 성능을 보여, 지금은 딥러닝 영상처리 분야에서 많이 사용되고 있는 CNN이다. 영상 분할은 영상 단위가 아니라 영상을 이루고 있는 픽셀 단위로 피사체를 구별하는 분야이다. 그렇기 때문에, 영상을 전체적으로 보는 것보다는 작은 크기의 특징 맵을 많이 학습하는 것이 유리하다. U-Net 은 이 점에 착안하여, 작은 크기의 특징 맵을 많이 만들어내는 encoder (암호기)와 큰 크기의 특징 맵을 적게 만들어내는 decoder (복호기)를 결합한 encoder-decoder 구조로 구성되어있다. 암호기의 출력은 크기가 작기 때문에 합성곱 층을 많이 통과하여 특징 맵의 질을 높일 수 있다. 복호기에서는, DenseNet 과 같이 암호기의 입력과 복호기의 출력을 병합하는 과정을 반복해주기 때문에 원래의 데이터의 특징을 잃어버리지 않는다.

이러한 U-Net 의 장점은 영상의 잡음을 제거하는 분야에서도 각광받았으며, ResNet이나 DenseNet과 함께 대표적인 CNN으로 자리 잡아 잡음 제거 솔루션의 기반으로 사용되고 있다. 최근에는 ResNet, DenseNet, U-Net 을 이용한 RDUNet [5] 이나 ResNet과 질이 좋은 특징 맵에 가중치를 부여하는 attention mechanism [6] (어텐션 기법), U-Net을 이용한 RatUNet [7] 등 U-Net 기반의 CNN이 많이 연구되고 있다. 어텐션 기법은 번역을 포함한 자연어 처리 분야에서 처음 제안된 기법으로, 전체 문장에서 특정 단어 및 문자에 집중하여 학습하기 위해 고안되었다. 이외에도 U-Net 의 장점을 극대화하기 위해 encoder의 입력과 decoder의 출력을 병합하기 전에 합성곱 층을 추가하는 UNet++ [8], UNet3+ [9] 등의 CNN 역시 연구가 지속되고 있다.

1.2 Noise2Noise

딥러닝 기반으로 영상의 품질을 개선하는 Noise2Clean은, 잡음이 있는 입력 영상과 잡음이 없는 출력 영상의 데이터쌍을 학습하여 입력 영상이 출력 영상에 가까워지도록 하는 것이다. 본 논문에서는 해당 방법론을 기존의 Gaussian noise (가우시안 잡음)와 같은 영상 품질의 저하를 생성하기 쉬운 경우에는 깨끗한 임의의 영상을 모두 GT로 사용할 수 있다. 하지만, 저조도 조건의 영상을 촬영하기 위해서는 빛이 거의 존재하지 않는 환경에서 다양한 색상을 가진 사물이 필요하다. 즉, 저조도 조건은 Noise2Clean이 어려운 조건이라고 할 수 있다.

Noise2Noise [10]는 출력 영상에 잡음이 있어도 Noise2Clean과 비슷한 성능으로 영상 품질의 저하를 없애는 것이 가능하다는 방법론이다. 이는 딥러닝 기반 학습에서 사용하는 손실 함수와 관련이 있다. 이 중 mean square error loss (평균제곱오차 손실 함수)는 입력 영상과 출력 영상의 차이가 줄어드는 방향으로 학습하는데, 입력 영상은 모든 입력 영상의 평균으로 수렴하는 것으로 알려져 있다.

$$l_2 = \sum_{k=1}^n (x_k - y)^2 = ny^2 - (2 \sum_{k=1}^n x_k)y + \sum_{k=1}^n x_k^2$$

$$\frac{\delta l_2}{\delta y} = 2ny - 2 \sum_{k=1}^n x_k$$

$$y \Big|_{\frac{\delta l_2}{\delta y} = 0} = \frac{\sum_{k=1}^n x_k}{n} = \text{average}(x)$$

평균제곱오차 손실 함수는 위의 수식과 같이 표현할 수 있다. x 는 입력 영상, y 는 출력 영상, n 은 입력 영상의 개수이다. l_2 는 평균제곱오차 손실 함수이고, 출력 영상 y 에 관한 식으로 정리할 수가 있다. 이를 y 에 대해 미분을 했을 때 0이 되는 y 가 바로 딥러닝 기반의 솔루션에서 입력 영상이 학습되는 방향이라고 할 수가 있는데, 이는 잡음이 있는 입력 영상들의 평균과 일치한다. 즉, y 에 잡음이 있는지의 여부와는 상관없이 평균을 유지해준다면 잡음이 있는 영상들만으로도 딥러닝 기반의 학습을 통해 잡음을 제거하는 것이 가능하다. 잡음의 평균이 0이어야 한다는 의미이기도 하다.

2. Metrics

2.1 Modulation Transfer Function (MTF)

영상을 평가하기 위한 해상도 중 하나는 spatial resolution (공간 해상도)이 있다. 같은 조건에서 촬영된 두 사물을 구분할 수 있는 가장 짧은 거리, 즉 resolution (분해능)이 해당 영상의 공간 해상도를 판단할 수 있는 척도가 될 수 있다. 실제로, 위성 영상에서는 하나의 픽셀이 투영하는 거리를 ground sample distance (GSD)라 칭하고 위성 영상의 공간 해상도를 표현하고 있다.

공간 해상도를 정의하기 위한 두 가지 개념이 있는데, 하나가 위에서 언급한 분해능이라면 다른 하나는 신호의 값과 관련된 modulation (변조)이다. 우리가 사물을 구분할 때, 아무리 짧은 거리를 구분할 수 있다고 하더라도 두 사물이 주는 신호의 값이 같으면 구분할 수 없다. 변조가 높을수록 가까이 있는 두 사물을 뚜렷하게 구분할 수 있다. 변조는 본래 통신 분야에서 아날로그 신호를 디지털 신호로 변환함을 의미하지만, 본 논문에서 변조 M 은 아래의 수식과 같이 정의한다. Fig. 1에서는 square wave (사각파)에서의 예시를 보여준다.

$$M = \frac{I_{Max} - I_{Min}}{I_{Max} + I_{Min}}$$

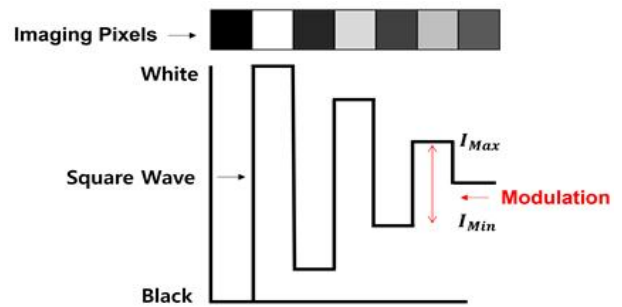


Fig. 1. Modulation in a square wave

영상의 공간 해상도를 표현할 수 있는 가장 좋은 지표 중 하나는 modulation transfer function (변조 전달 함수; MTF)이며, 이는 영상의 패턴이 단위 거리마다 반복되는 정도를 주파수로 나타내고 해당 주파수에 따라 변조의 값을 차트로 시각화한 것이다. 높은 주파수에서 높은 변조를 가질수록 해당 영상은 주파수가 높은, 즉 디테일에 해당하는 부분을 온전히 가지고 있다고 여겨진다.

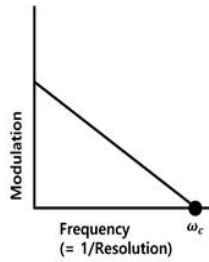


Fig. 2. Simple MTF plot (ω_c : cutoff frequency)

단순한 광학계 또는 영상의 MTF는 Fig. 2와 같이 주파수가 cutoff frequency (차단 주파수)일 때 0을 가지게 된다. 실제로 지표로 사용할 때에는 Fig. 2와 같은 plot의 대푯값을 사용하게 되는데, MTF가 50% (0.5)가 되는 주파수인 MTF50과 주파수가 Nyquist frequency (나이키스트 주파수)일 때의 MTF ($MTF@\omega_{Nyquist}$)를 주로 사용한다. 나이키스트 주파수는 차단 주파수의 절반보다 조금 작은 값으로, 차단 주파수가 나이키스트 주파수의 2배를 넘는 값이 되어야 aliasing (위신호 현상)없이 신호를 전달할 수 있다는 Nyquist theorem (나이키스트 정리)에 근거하여 값을 정한다. 실제로 적용할 때에는 차단 주파수에 1보다 작은 Kell factor (켈 계수) K 를 곱하여 아래의 수식과 같이 사용한다.

$$\omega_{Nyquist} = \frac{1}{2}K\omega_c (K \approx 0.7)$$

MTF를 측정하는 대표적인 방법은 slanted-edge method 이다. Fig. 1의 Imaging Pixels과 같이 사물 또는 경계가 구분되는 edge profile을 미분하면 가우시안 함수 형태의 line spread function (LSF)를 얻을 수 있는데, LSF를 푸리에 변환하면 주파수에 따른 변조를 plot한 MTF를 얻을 수 있다.

2.2 Signal-to-Noise Ratio (SNR)

영상을 평가하기 위한 해상도 중 하나에는 radiometric resolution (방사 해상도)이 있다. 영상의 한 픽셀이 표현할 수 있는 dynamic range (동적 범위)를 의미하는데 이는 곧 몇 비트 영상인지에 따라 달라지기 때문에 bit depth (비트 깊이)로 표현되기도 한다. 우리가 보통 카메라나 컴퓨터를 통해 보는 sRGB 영상의 각 픽셀은 0부터 255 사이의 값을 가지고 있는 8비트 영상이다. 8비트 영상의 비트 깊이는 8로 표현한다.

비트 깊이는 잡음과 어느 정도 연관이 있다고 할 수 있다. 잡음이 아무리 많아도, 비트 깊이보다 한참 작은 값이면 그 잡음은 강하다고 할 수 없다. 반대로, 비트 깊이와 비슷한 값이 조금만 있으면 해당 픽셀은 포화되고, 제대로 된 값을 표시할 수 없다. 즉, signal-to-noise ratio (신호 대 잡음비; SNR)가 중요한 것이다. 최대 신호 대 잡음비나 구조적 유사성과 다르게 SNR은 데이터쌍이 필요하지 않고, 하나의 영상 내에서 계산할 수 있기 때문에 영상의 잡음 정도를 평가하기 유용하다.

SNR은 정의 그대로 ‘신호’를 ‘잡음’으로 나누어 계산해 줄 수 있다. 영상 내에서 region of interest (ROI)를 잡았을 때 신호의 대푯값은 해당 ROI의 평균이다. 잡음의 대푯값은 ROI의 표준편차를 사용하지만 그 값은 잡음의 종류에 따라서 달라진다. Fig. 3 [11]과 같이, 광자가 충분히 많으면 우리가 흔히 다루는 Gaussian distribution (가우시안 분포)를 따르게 되지만, 저조도 조건에서는 광자가 충분할 때에는 보이지 않는 미세한 잡음과 함께, shot noise (광자 잡음)가 발생한다. 광자 잡음은 제한적인 광자의 수로 인해 Poisson distribution (푸아송 분포)를 따른다. 가우시안 분포와는 다르게 평균과 분산이 같은 분포이기 때문에, 표준편차는 평균의 제곱근이 된다.

광자 잡음만 존재한다는 가정 하에서는 평균의 제곱근을 SNR로 대체할 수 있지만, 가장 좋은 방법은 촬영한 영상에서 신호가 일정한 ROI를 찾는 것이 가장 좋다. 신호가 일정하다는 가정 하에, 즉 잡음을 제외한 신호의 표준편차가 0이라는 가정 하에 신호의 평균을 표준편차로 나눈 값은 SNR과 정확히 일치하게 된다. 표준편차 자체가 온전히 잡음의 대푯값이 되기 때문이다.

$$SNR [dB] = 10\log_{10} \frac{Signal}{Noise} = 10\log_{10} \frac{\lambda}{\sigma}$$

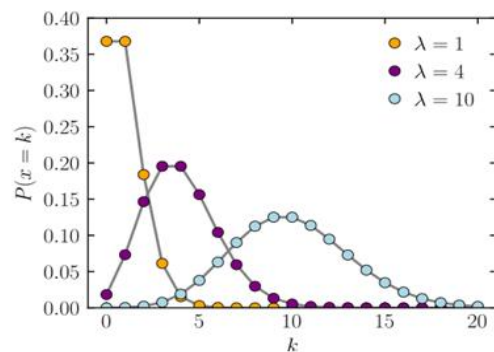


Fig. 3. Poisson distribution ($\lambda = 1, 4$) and Gaussian distribution ($\lambda = 10$)

III. The Proposed Scheme

본 논문에서는 저조도 조건에서 촬영된 Raw 영역의 영상을 입력하여 Noise2Noise 방법으로 학습된 U-Net으로 잡음을 제거하고, 연산량을 줄이기 위해 간소화된 software image signal processor (SW ISP)로 sRGB 영역의 영상을 출력하는 파이프라인을 제안한다. Raw 영역의 영상은 카메라의 ISP를 거치기 이전의 Bayer pattern (베이어 격자)인데, 카메라의 센서가 sRGB 영역과는 다르게 R, G1, G2, B 값들이 격자 패턴으로 배열되어있기 때문에 ISP에서 sRGB 영역으로 변환해주지 않으면 사람이 눈으로 보는 영상을 얻을 수 없다.

저조도 조건의 Raw 영상에서 잡음 제거를 하는 방법은 영역 관점에서 다양하게 나뉘는데, Raw \rightarrow Raw (Raw2Raw), Raw \rightarrow sRGB (Raw2RGB), sRGB \rightarrow sRGB (RGB2RGB) 등으로 구분할 수 있다. Raw2Raw 는 ISP를 거치기 전에 잡음 제거를 하는 방법으로 카메라에 촬영된 영상이 바로 딥러닝 학습에 사용된다. Raw2RGB는 딥러닝 기반의 솔루션이 ISP의 역할을 대신해줄 수 있기 때문에 ISP 의 튜닝이 불필요한 꽤나 간편한 솔루션이지만, 딥러닝을 사용해서 영역을 변환하는 만큼 디테일을 온전히 살리기는 힘들다는 단점이 있다. RGB2RGB는 ISP를 거친 이후의 영상을 잡음을 제거하는 방법으로 잡음의 형태가 ISP에 의존할 수 있다는 단점이 있다. 본 논문에서는 Raw2Raw 방법을 택하고, Noise2Noise와 ISP를 차례대로 사용하는 파이프라인을 Fig.4 와 같이 제안한다.

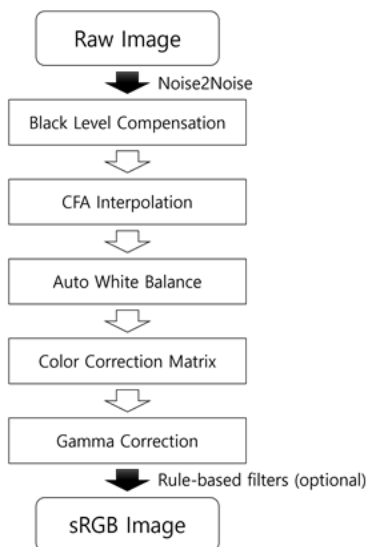


Fig. 4. Pipeline of Raw \rightarrow Raw denoising and SW ISP

우선, Raw 영역의 영상은 비트 깊이에 따라 동적 범위가 다양하다. 딥러닝으로 학습을 진행하기 위해서는 0과 1 사이의 값으로 normalization (정규화)를 하여 입출력 데이터쌍의 분포를 유사하게 만들어주는 것이 필수적이다. 본 논문에서는 ISP에서 black level compensation (BLC)를 포함하기 때문에, Noise2Noise 이전엔 입출력 데이터쌍을 모두 최댓값으로 나누어 정규화를 해주었고, BLC에서 영상의 최솟값에 해당하는 black level을 빼주었다. 10비트 영상 기준 64, 12비트 기준 256이 적절한 값이라고 알려져 있다. 0과 1 사이에서 BLC를 진행한다면, $64/1024 = 0.0625$ 정도의 값이 적절한 것이다.

Color filter array interpolation (CFA)는 Raw 영역의 영상을 sRGB 영역의 영상을 변환해주는 과정이다. 베이어 격자 특유의 모자이크 패턴을 없애주기 때문에 demosaicing (모자이크 제거)이라고도 불리고 있다. CFA는 우선 격자로 배열되어있는 R, G1, G2, B를 각각의 채널로 분리한다. 각각의 채널은 하나의 평면이 정보의 양을 유지한 상태로 넷으로 분리되었기 때문에, 크기가 절반으로 줄어들어 있다. Bilinear, malvar [12] 등의 방법으로 interpolation (보간)을 하면 베이어 격자와 동일한 크기의 R, G, B 채널을 얻을 수 있으며, 베이어 격자가 밝기에 민감한 사람의 눈을 고려하여 G1, G2 두 개의 G 채널을 가지고 있는 반면에 sRGB 영역의 영상은 두 G 채널을 이용하여 하나의 G 채널을 만들어 사용한다.

Auto white balance (AWB)는 영상의 색온도를 조절하는 과정이다. R, G, B 세 채널 중 어느 채널이 강하냐에 따라서 영상의 색깔이 다른 빛을 띠 수 있는데, 흰색이 흰색으로 보일 수 있는 색온도가 기준이 된다. 5100K, 6500K 등이 사용되고, 이보다 낮을수록 영상이 전체적으로 노란빛을 띠게 되고 높을수록 파란빛을 띠게 된다. 본 논문에서는 gray world [13] 기법을 사용하여 R, G, B 채널의 평균을 모두 G 채널에 맞게 변환한다. 이에 따라, 세 채널의 값은 균형을 맞추게 되고 밝기가 충분했을 때 흰색이 흰색으로 보일 수 있다.

Color correction matrix (CCM)은 영상의 색공간을 보정하는 과정이다. 색상은 3차원 공간으로 정의될 수 있다. 영상의 색공간이 sRGB와 유사한 분포를 가지지 않으면, AWB 이후에도 색상이 다르게 보일 수 있다. 3×3 행렬을 곱하여 영상의 색공간을 sRGB와 가깝게 보정해준다. 본 논문에서 사용한 식은 아래의 수식과 같다. r, g, b 는 CCM 전 채널, R, G, B 는 CCM 후 채널을 의미한다.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.8 & -0.8 & 0 \\ -0.3 & 1.5 & -0.2 \\ 0 & -0.8 & 1.8 \end{bmatrix} \begin{bmatrix} r \\ g \\ b \end{bmatrix}$$

Gamma correction (GAC) 는 선형적인 영상의 분포를 비선형 분포로 바꾸어주는 과정이다. 사람의 눈은 미세한 신호에 민감하기 때문에 어두울수록 변화를 더욱 크게 느끼게 된다. 즉, 기존의 선형적인 신호를 이차곡선 형태로 변환하여 어두울 때는 변화가 완만하지만, 밝을 때는 가파르게 변할 수 있도록 해야 한다. γ 는 이차곡선의 곡률을 정의하는 요인이며, 일반적인 카메라의 ISP나 TV에는 2.2의 값을 이용한다고 알려져 있다. GAC의 식은 아래와 같다. I 는 R, G, B 세 채널을 포함한 영상이다.

$$I = I^{1/\gamma} = I^{1/2.2}$$

IV. Experimental Results

본 논문에서는 Sony IMX327 Sensor 와 Jetson Xavier NX 기기를 결합하여 촬영한 저조도 조건의 영상을 대상으로 Noise2Noise 학습 기반의 Raw2Raw 잡음 제거와 간단한 ISP를 거쳐 저조도 조건의 영상의 잡음을 제거하기 전후의 영상을 비교한다. 영상은 MTF와 SNR 측정을 위한 각각의 차트마다 50lux, 10lux, 1lux, 0.1lux 에서 촬영되었으며, 잡음이 제거된 영상은 빛이 거의 존재하지 않는 극한의 저조도 조건인 1lux와 0.1lux에서 비교한다.

Noise2Noise 학습에는 U-Net을 사용하였으며, 학습 데이터로는 Galaxy S22 로 촬영한 4000x3000 크기의 저조도 Raw 및 sRGB 영상을 이용하였다. 512x512 크기의 패치를 학습시켰으며, 손실 함수로는 Noise2Noise 학습의 효과를 위해 평균제곱오차 손실 함수를 Adam optimizer 를 learning rate 0.0001로 설정하여 학습하였다. Inference 영상의 크기는 1920x1080이다.

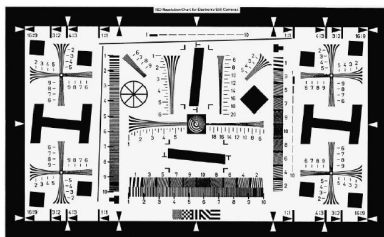


Fig. 5. ISO 12233 chart for MTF measurement



Fig. 6. Colorchecker for SNR measurement

잡음이 제거된 영상 품질의 정성적인 평가뿐 아니라 MTF, SNR 등의 지표를 통한 정량적인 평가까지 포함한다. MTF의 경우엔 Fig. 5와 같은 ISO 12233 [14] 차트를 촬영한 영상의 잡음을 제거하였다. 차트의 줄의 간격을 관찰하여 정성적인 평가를 하고, 차트의 edge 부분을 참고하여 정량적인 MTF 결과를 확인할 수 있다. SNR의 경우엔 Fig. 6 과 같은 colorchecker [15] 를 촬영한 영상의 잡음을 제거하여 영상의 평탄한 ROI를 대상으로 다양한 색상에서의 SNR을 얻을 수 있다. 본 논문에서는 24개의 ROI에서 계산한 SNR을 평균한 값으로 표시한다.

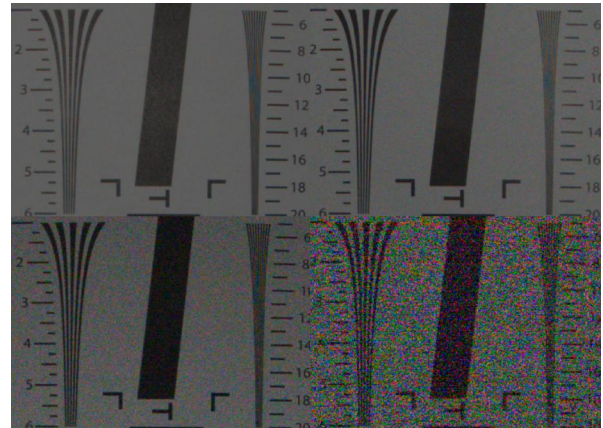


Fig. 7. ROI of Raw ISO 12233 image visualized by ISP (50lux, 10lux, 1lux, 0.1lux from left top to right bottom)



Fig. 8. Raw colorchecker image visualized by ISP (50lux, 10lux, 1lux, 0.1lux from left top to right bottom)

Fig. 7과 Fig. 8은 본 논문 3장에서 제안한 파이프라인의 입력 영상에 해당한다. 실제 Raw 영상은 사람의 눈으로 보기 힘들기 때문에 Fig. 7과 Fig. 8은 Raw2Raw 잡음 제거 과정을 생략하고 ISP 만 거친 입력 영상이다. 조도가 낮을수록 저조도로 인한 잡음이 심해짐을 알 수 있고, 그에 따라 ISO 12233 차트에서 줄의 간격을 구분하기 힘들음을 알 수 있다.

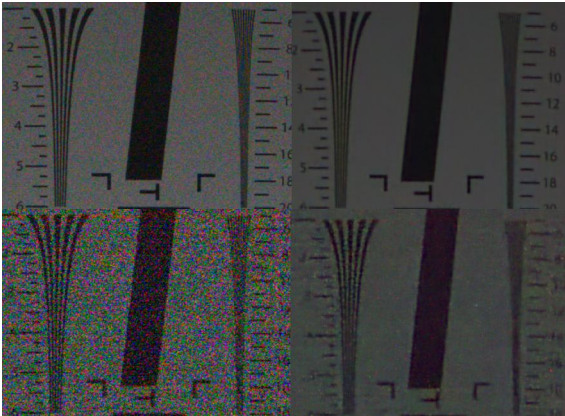


Fig. 9. ROI of Raw and sRGB ISO 12233 image visualized by ISP (1lux noisy, 1lux denoised, 0.1lux noisy, 0.1lux denoised from left top to right bottom)



Fig. 10. Raw and sRGB colorchecker image visualized by ISP (1lux noisy, 1lux denoised, 0.1lux noisy, 0.1lux denoised from left top to right bottom)

Fig. 9와 Fig. 10은 빛이 거의 없는 1lux 및 0.1lux 환경에서 잡음이 제거되기 전후의 영상을 비교한 것이다. ISO 12233 차트의 잡음 제거 결과를 봤을 때, 1lux에서는 색상 잡음이 완전히 제거된 결과를 보여주었다. 하지만, 줄의 간격의 선명도가 이전보다 낮기 때문에 줄의 간격이 이전보다 뚜렷하게 잘 보인다고 말하기 어렵고 이는 잡음 제거가 곧 공간 해상도의 상승으로 이어지지 못했음을 시사한다. 0.1lux에서는 차트의 디테일이 온전히 복원되지 못하였고, white balance 역시 크게 어긋났다.

Colorchecker의 잡음 제거 결과를 봤을 때, 1lux에서는 배경이 전체적으로 어두워진 것을 제외하고는 색상 잡음이 완전히 제거되었다. 하지만, 0.1lux에서는 ISO 12233에서의 결과와 마찬가지로 white balance가 전체적으로 G 채널 쪽으로 치우쳤고, 영상 곳곳에서 특정 채널이 포화된 hot pixel이 많아져서 영상의 품질이 저하된 상태이다.

Table 1. $MTF@w_{Nyquist}$ [%] results in Fig. 9

Lux	noisy	denoised
1lux	5.07	2.96
0.1lux	4.36	2.97

Table 2. SNR [dB] results in Fig. 10

Lux	noisy	denoised
1lux	16.86	31.20
0.1lux	8.58	23.59

Table 1과 Table 2는 MTF와 SNR을 활용한 저조도 조건의 영상의 정량적인 평가 결과이다. 0.1lux에서의 MTF가 1lux에서의 MTF보다 낮은 것을 보아, 디테일이나 글씨를 알아보기 더욱 힘든 조건이다. 잡음 제거 이후에도 MTF가 크게 나아지지 않는 것을 보아, 현재의 Noise2Noise는 저조도 조건에서의 영상의 디테일을 향상시켜주지 못하고 있으며, 2장의 설명과 같이, 실제로 데이터의 평균에 가까워지도록 학습하고 있기 때문에 평균 필터를 쓴 것과 같이 흐려질 수밖에 없다. 색상 잡음에 의해 인간의 눈에는 잡음이 있는 영상이 글씨를 보기 더 힘들 수도 있지만, 색상 잡음의 투명도가 높기 때문에 영상의 디테일을 예상보다는 덜 해친 것으로 볼 수 있다.

반면에, SNR은 잡음 제거 이후에 크게 상승하였다. 0.1lux에서의 SNR이 1lux에서의 그것보다 낮은 것을 보아, 저조도 조건으로 인한 잡음이 더욱 심하다고 볼 수 있다. 잡음 제거 이후에 0.1lux 조건에서의 SNR은 잡음 제거 이전의 1lux 조건에서의 SNR보다 높게 나왔으며, 현재의 Noise2Noise가 GT 없이도 잡음 제거의 성능이 훌륭하다고 평가할 수 있다. 다만, 밝기가 어두워지거나, white balance가 맞지 않는 부분을 개선할 필요가 있고, 이는 ISP에도 해당되는 부분이다.

V. Conclusion

본 논문에서 저조도 조건에서의 잡음 제거에서는 GT를 통한 학습이나 평가 지표의 사용이 불가능하기 때문에 주로 정성적 평가 위주임을 문제로 제기했다. U-Net 구조의 CNN을 Noise2Noise 방법으로 학습하여 저조도 조건으로 인한 잡음을 제거할 수 있음을 확인함과 동시에, MTF 및 SNR 과 같은 지표가 영상의 공간 해상도 및 방사 해상도의 정량적인 평가에 도움이 될 수 있음을 설명하고 이를 실험으로 증명하였다. 실험으로 검증된 다양한 지표들은 저조도 영상 품질 개선 필요로 하는 카메라 및 CCTV 업체나 저조도 영상의 품질 개선 알고리즘을 제공하는 업체 간의 객관적인 평가 기준으로서 활용될 수 있다.

본 논문에서 Noise2Noise 로 학습한 U-Net 기반의 잡음 제거 솔루션은 저조도 조건 특유의 색상 잡음을 상당히 제거하였음을 SNR 지표의 상승을 통해 확인하였다. 하지만, 영상의 디테일을 유지하거나 작은 글씨 등을 쉽게 보일 수 있도록 하는 능력의 부족함 역시 MTF 지표를 통해 확인했다. 향후 연구에서는 영상의 공간 해상도와 방사 해상도 간의 상충관계가 있음을 인지하고, 잡음을 줄임과 동시에 영상의 디테일한 부분의 손상도 줄일 수 있는 솔루션을 개발하는 것이 중요한 방향으로 설정될 것이다.

ACKNOWLEDGEMENT

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning(KETEP) and the Ministry of Trade, Industry & Energy(MOTIE) of the Republic of Korea (No. 20216110100060).

This work was supported by the Technology Innovation Program (20019466, Development of Main computing system dedicated to integrate video codec device, autonomous flight, object detect/recognition, communication system, FC) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea).

REFERENCES

- [1] R. Gonzalez, and R. Woods, "Digital Image Processing," Pearson, pp.358-363, 2018.
- [2] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 2016, pp. 770-778, doi: <https://doi.org/10.1109/CVPR.2016.90>.
- [3] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, July 2017, pp. 2261-2269, doi: <https://doi.org/10.1109/CVPR.2017.243>.
- [4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 2015, pp. 234-241, doi: https://doi.org/10.1007/978-3-319-24574-4_28.
- [5] J. Gurrola-Ramos, O. Dalmau, and T. Alarcón, "A residual dense u-net neural network for image denoising," IEEE Access, Volume 9, pp. 31742-31754, February 2021, doi: <https://doi.org/10.1109/ACCESS.2021.3061062>.
- [6] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, September 2014, doi: <https://doi.org/10.48550/arXiv.1409.0473>.
- [7] H. Zhang, Q. Lian, J. Zhao, Y. Wang, Y. Yang, and S. Feng, "RatUNet: residual U-Net based on attention mechanism for image denoising," PeerJ Computer Science 8:e970, pp. 234-241, May 2022, doi: <https://doi.org/10.7717/peerj-cs.970>.
- [8] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh and J. Liang, "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," IEEE Transactions on Medical Imaging, Volume 39, No. 6, pp. 1856-1867, June 2020, doi: <https://doi.org/10.1109/TMI.2019.2959609>.
- [9] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y. Chen, and J. Wu, "UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, May 2020, pp. 1055-1059, doi: <https://doi.org/10.1109/ICASSP40776.2020.9053405>.
- [10] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and Timo Aila, "Noise2Noise: Learning Image Restoration without Clean Data," arXiv preprint arXiv:1803.04189, March 2018, doi: <https://doi.org/10.48550/arXiv.1803.04189>.
- [11] Wikipedia, "Poisson distribution," https://en.wikipedia.org/wiki/Poisson_distribution.
- [12] H. S. Malvar, Li-wei He and R. Cutler, "High-quality linear

- interpolation for demosaicing of Bayer-patterned color images," 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, QC, Canada, 2004, pp. iii-485, doi: <https://doi.org/10.1109/ICASSP.2004.1326587>.
- [13] G. Buchsbaum, "A spatial processor model for object colour perception," *Journal of the Franklin Institute*, Volume 310, Issue 1, pp. 1-26, July 1980, doi: [https://doi.org/10.1016/0016-0032\(80\)90058-7](https://doi.org/10.1016/0016-0032(80)90058-7).
- [14] Imatest, "Electronic Still Picture Camera Resolution Test Chart (ISO-12233) Product Specifications," https://www.imatest.com/wp-content/uploads/2015/05/ISO-12233-Digital-Still-Camera-Resolution-Chart-QA-72_spec_v1-01.pdf.
- [15] Calibrite, "ColorChecker Classic," <https://calibrite.com/us/product/colorchecker-classic/>.

Authors



Sangmin Kim received the B.S. degree in Electronic Engineering from Hanyang University, Korea, in 2019 and the M.S. degree in Electronic Engineering from Hanyang University, Korea, in 2021.

Mr. Kim researched remote sensing and developed web crawler as an Associate Engineer in SatreCI, until 2022. He is currently a ML Engineer in ENERZAI. He is interested in low light imaging and neural ISP.