

인공지능 기반 영상 화질 개선 최신 기술 동향

□ 김원준 / 건국대학교

요약

최근 모바일 기기를 위한 카메라 관련 기술이 발전하면서 취득할 수 있는 영상의 화질 또한 크게 향상되고 있다. 그러나, 일상 생활에서 빈번히 발생하는 다양한 실내외 불규칙한 조명 조건 및 저조도 환경은 여전히 영상 화질 저하를 야기한다. 본 고에서는 이러한 문제를 해결하기 위해 최근 널리 연구되고 있는 심층신경망 기반 영상 화질 개선 연구의 최신 동향을 소개하고자 한다. 먼저, 다양한 최적화 기법을 바탕으로 영상 내 조명 성분을 추정하고, 이를 개선하는 방법들에 대해 간략히 설명한다. 또한, 영상 인식, 객체 검출 등에서 뛰어난 성능을 입증한 합성곱 신경망 구조를 기반으로 영상의 잠재적 특징을 효과적으로 검출한 후 이를 바탕으로 개선된 영상을 생성하는 방법에 대해 설명한다. 다양한 데이터셋에 대한 실험 결과를 통해 인공지능 기반 영상 화질 개선의 우수성을 보인다.

1. 서론

스마트폰 등 다양한 모바일 기기에 고성능 카메라

기능이 탑재되면서 많은 사용자가 하루에도 수많은 사진을 획득하고 가공하고 있다. 카메라 관련 기술이 꾸준히 발전하면서 고해상도의 영상을 손쉽게 획득할 수 있게 되었으며, 클라우드를 기반으로 한 대용량 영상 공유가 가능하기 때문에 뛰어난 화질의 영상 획득에 대한 관심도 크게 증가하고 있다. 특히, 스마트폰 제조사의 경우 제품의 우수성 기준으로 카메라의 성능을 제시하는 경우가 늘고 있다. 이러한 기술적 발전에도 불구하고, 불규칙한 실내외 조명 환경은 여전히 고화질 영상 획득에 어려움을 주고 있다. 이를 해결하기 위해 컴퓨터 비전 분야에서는 렌즈나 하드웨어 구성에 대한 개선이 아닌 소프트웨어적 솔루션에 대해 꾸준히 연구해오고 있다.

영상 화질 개선 방법은 다양한 최적화 기법을 이용한 조명 성분 예측 방법과 심층학습을 이용한 화질 개선 영상 생성 방법으로 크게 나눌 수 있다. 전

자의 경우, 영상 내 픽셀 값은 조명(Illumination) 성분과 반사(Reflectance) 성분의 곱으로 표현 가능하다는 Retinex 이론[1]을 기반으로 조명 성분을 추정하고자 노력하였다. 그러나 한 장의 영상만을 이용하여 조명과 반사 성분을 완벽하게 분리하기 매우 어려우며, 화질 개선 과정에서 색상 변형(Color Distortion) 및 예기치 못한 잡음이 발생되기도 한다. 한편, 영상 인식 분야에서 뛰어난 성과를 보여준 심층신경망(Deep Neural Network)을 영상 화질 개선에 적용하고자 하는 연구가 몇몇 연구자들에 의해 시작되고 있다. 이 방법의 경우, 주어진 장면 구조와 조명 간 복잡한 관계를 직접 추정하기보다는 신경망 학습을 통해 개선된 화질의 영상을 생성하고자 한다. 즉, 입력 영상과 이에 대한 화질 개선 버전을 이용한 지도 학습(Supervised Learning)을 통해 구축한 신경망이 그 차이를 효과적으로 학습할 수 있도록 하는 다양한 방법이 연구되고 있다. 이러한 생성 모델(Generative Model) 기반의 방법은 컬러 영상을 입력으로 하여 해당 화질 개선 영상을 출력하는 기본 구조를 가지고 있으며, 다양한 조명 조건을 대용량 학습을 통해 성공적으로 모델링 할 수 있다. 그러나, 영상 촬영 당시 조명 조건에 대한 개선 환경이 주어지지 않으며, 화질 개선 버전을 구축하기 위해서 현재는 전문가의 재조정(Re-touch)에 의존하고 있어 특정 개인의 주관적 화질 평가가 학습에 반영될 수 있는 문제점이 있다. 또한, 데이터셋 구축에 전문 인력과 많은 시간이 소요되기 때문에 큰 비용이 필요하며, 새로운 조명 환경이 추가될 때마다 리터치(Re-touch) 과정이 다시 수행되어야 하는 한계가 있다. 그럼에도 불구하고 뛰어난 정확도 및 확장성으로 인해 많은 연구자들이 다양한 심층신경망 구조를 기반으로 한 영상 화질 개선 연구를 진행하고 있다.

본 고에서는 이와 같이 단일 영상에 대한 화질 개선을 수행하는 최신 기술 동향을 살펴보고자 한다. II장에서는 최적화 기법을 이용한 영상 화질 개선 방법에 대해 간략히 살펴본 후, 심층신경망을 이용한 화질 개선 최신 연구를 소개한다. III장에서는 다양한 영상에 대한 화질 개선 결과를 비교 제시하고, 마지막으로 V장에서 결론을 맺는다.

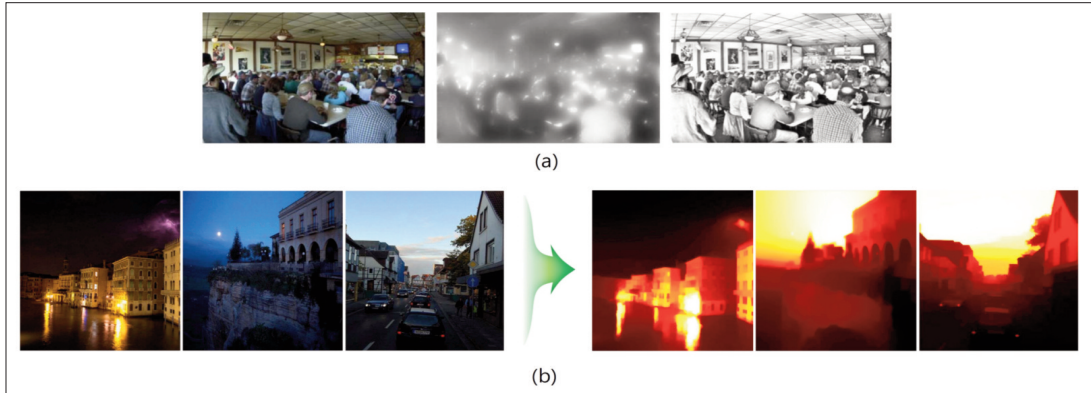
II. 영상 화질 개선 기술 동향

이번 장에서는 영상 화질 개선 기술을 두 가지 방법으로 나누어 자세히 살펴보고자 한다. 먼저 영상 내 조명 성분을 최적화 기법을 통해 추정하고, 해당 결과를 감마 보정 등을 통해 간단히 개선한 후 반사 성분을 이용하여 화질 개선 영상을 복원하는 방법을 간략히 살펴본다. 이후 심층학습 기술을 기반으로 하는 화질 개선 영상 생성 방법에 대해 자세히 살펴본다.

1. 다양한 최적화 기법을 이용한 영상 화질 개선 방법

영상 내 픽셀 값은 장면 내 객체의 반사 성분과 해당 조명 성분의 곱(Multiplication)으로 표현할 수 있다는 Retinex 이론[1]을 기반으로 많은 연구자들이 영상 내 조명 성분을 추정에 초점을 맞추어 연구를 진행해왔다.

먼저, Wang[2] 등은 영상 내 조명 성분은 RGB 컬러 채널 값 중 가장 큰 값으로 나타난다는 관찰 결과를 바탕으로 초기 조명 성분을 예측하고, 각 픽셀 위치에서의 주변 픽셀 밝기 값과의 유사도를 이용하여 예측한 조명 성분을 정제한다. 이 후 [2]에서는 Bi-log 변환을 이용하여 예측한 조명 성분



〈그림 1〉 조명 성분 추정 예. (a) Fu[3] 등의 결과(왼쪽: 원본 영상, 중앙: 조명 성분, 오른쪽: 반사 성분) (b) Guo[4] 등의 결과(왼쪽: 원본 영상, 오른쪽: 조명 성분)

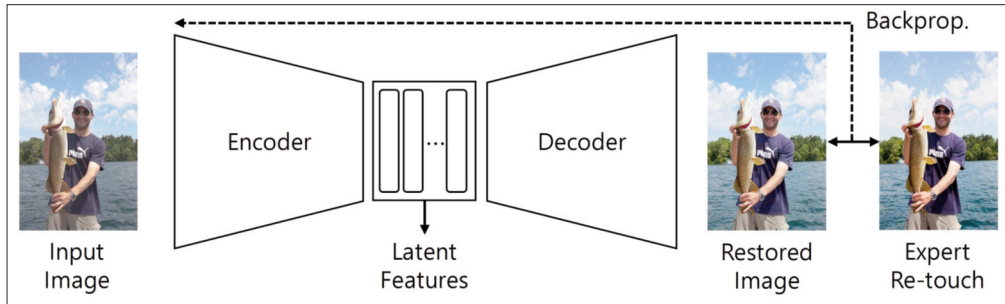
대해 개선 과정을 수행하고, 반사 성분과의 곱을 통해 최종적으로 화질 개선된 영상을 생성한다. Fu[3] 등은 조명 성분과 반사 성분을 동시에 추정할 수 있는 최적화 방법을 제안하여 화질 개선 성능을 향상 시켰다. 특히, 기존 최적화 과정에서 발생하는 텍스처(Texture) 성분에 의한 불필요한 잡음 증폭 현상을 효과적으로 개선할 수 있는 방법을 제안하여 해당 방법은 최근까지도 널리 인용되고 있다. Guo[4] 등은 [2]의 방법과 마찬가지로 초기 조명 성분을 추정하여 이를 기반으로 최적화를 고속으로 수행할 수 있는 효율적 개선 방법을 제안하였다. 해당 방법은 주파수 공간에서의 고속 동작이 가능한 구현 방법론을 제시하여 기존 방법 대비 월등히 빠른 수행 속도를 보이며 화질 개선 효과도 뛰어나다 (〈그림 1〉 참조). [5]에서는 영상 내 지역적 영역의 주성분(Principal Component)은 조명이라는 가정 하에 각 픽셀 위치에서 하위 공간 분석(Subspace Analysis)을 수행하여 조명 성분을 추정하는 방법을 제안하였다. 또한, Kim[6] 등은 영상 확산 과정에서 조명 성분은 가장 밝은 값을 유지한다는 관찰 결과를 바탕으로 조명 성분을 효과적으로 검출하는

방법을 제안하였다.

이와 같이 다양한 최적화 기법을 이용하여 영상 화질을 개선하는 방법이 최근까지도 꾸준히 연구되어 오고 있으나, 영상의 크기가 커질수록 대부분의 방법에서 수행 속도가 현저히 느려지는 문제가 있다. 또한, 실외 환경에서의 과도한 복원(Over-enhancement)과 이로 인한 잡음 증폭 현상이 종종 발생하기도 한다.

2. 심층신경망을 이용한 화질 개선 영상 생성 방법

최근 영상 인식 분야에서 뛰어난 성능 향상을 입증한 심층학습 기술을 영상 화질 개선에 적용하려는 시도가 늘고 있다. 기본적으로 원본 영상을 신경망에 입력으로 하여 개선된 영상이 출력되는 인코더-디코더(Encoder-Decoder) 구조를 기반으로 적용이 시작되고 있으며(〈그림 2〉 참조), 조명 성분을 별도로 생성하는 방법도 소개되고 있다. 인코더 구조에서는 원본 영상 내 의미있는(Underlying) 특징을 효과적으로 추출하여 잠재 특징(Latent



〈그림 2〉 심층신경망 기반 영상 화질 개선 방법의 기본 구조. 인코더-디코더 구조를 기반으로 개선된 화질의 영상을 생성

Features)을 생성하고, 디코더에서는 정의된 손실 함수 값을 최소화하는 방향으로 잠재 특징으로부터 영상을 복원한다. 이러한 디코더 과정에서 영상 해상도 손실로 인한 흐려짐 현상을 방지하기 위해 Skip Connection[7]이나 다양한 업스케이링(예를 들어, Bilateral Grid-based Upsampling [8]) 방법을 적용한다.

자세히 살펴보면, Ignatov[9] 등은 ResBlock[7]을 뼈대 구조(Backbone Network)로 하여 원본 영상에서 특징을 추출하고, 다양한 손실 함수(예를 들어, 색상 유사도, 콘텐츠 유사도 등)를 바탕으로 자연스러운 화질 개선 버전의 영상을 복원하는 구조를 제안하였다. Park[10] 등은 강화 학습을 이용하여 색상의 톤이나 밝기를 적응적으로 학습할 수 있는 새로운 신경망 구조를 제시하여 생동감 있는 영상을 복원하였다. Chen[11] 등은 순환 기반 적대적 학습 모델(Generative Adversarial Network, GAN)[12]을 적용하여 화질 개선된 영상을 복원하고 이를 기반으로 다시 원본 영상을 복원하여 픽셀 값의 유사도를 측정하는 방식을 제안하였다. 두 번의 일관성 유지를 통해 성공적인 화질 개선을 수행할 수 있으며, 순환 기반 GAN 특성으로부터 다양한 도메인로의 영상 생성 또한 가능하다. 가장 최근에서는 심

층신경망을 이용하여 화질 개선된 영상을 직접 복원하지 않고 조명 성분을 추정하는 방법이 제안되었다[8]. 원본 영상과 화질 개선된 버전의 관계는 매우 비선형적이기 때문에 이러한 관계를 직접 학습하는 것은 심층신경망을 이용하더라도 여전히 어려운 문제이다. 해당 방법은 조명 성분을 먼저 추정함으로써 기존 방법 대비 다양한 조명 환경에 대한 학습을 효과적으로 진행할 수 있는 장점이 있다.

이와 같은 심층신경망 기반의 방법은 직접적인 조명 성분 모델링 없이 학습을 통해 성공적인 화질 개선을 수행할 수 있지만, 학습 위해 원본 영상에 대한 개선된 화질 영상이 필요한 문제점이 있다. 즉, 원본 영상이 획득될 당시 조명 환경에 대한 화질 개선 버전이 동시에 촬영되어야 하나 이는 불가능하기 때문에 현재는 촬영된 원본 영상에 대한 전문가 리터치(Re-touch) 버전을 학습에 사용하고 있다. 따라서, 현재 가장 널리 사용되고 있는 MIT Adobe 5K 데이터셋[13]의 경우 다섯 명의 전문가가 리터치한 결과 영상을 학습을 위해 제공하지만, 특정 개인의 화질에 대한 주관적 평가가 학습 과정에 반영되는 문제점을 야기할 수 있다. 이러한 한계를 극복하기 위해서 한 가지 고려할 수 있는 방안은 영상 화질 평가 메트릭(Visual Quality Assessment,

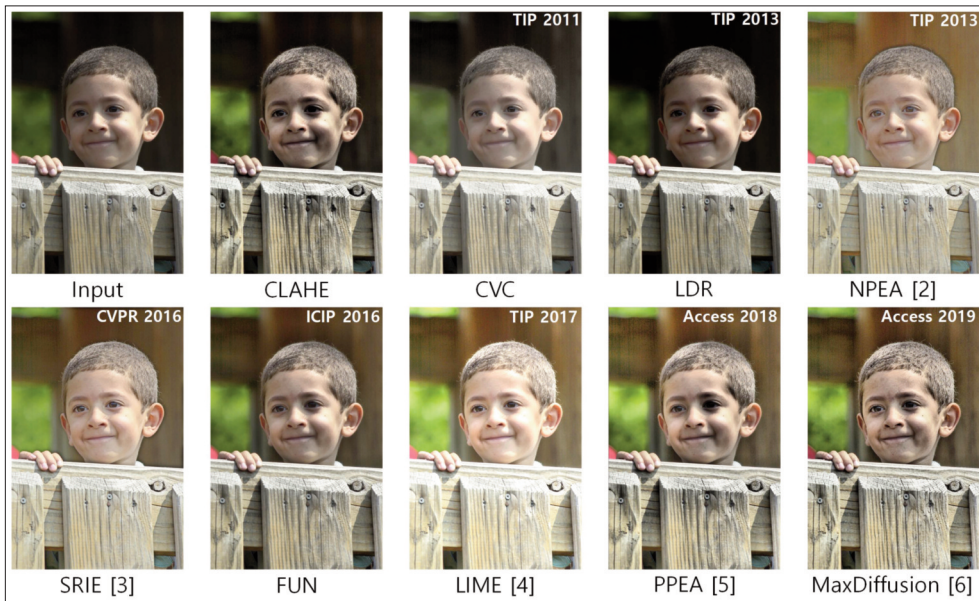
VQA)을 손실 함수에 반영하는 것이다. 이를 통해 객관적 화질 평가 결과가 복원된 영상에 반영되어 보다 효과적인 학습이 가능할 것으로 예상된다.

III. 영상 화질 개선 성능 비교

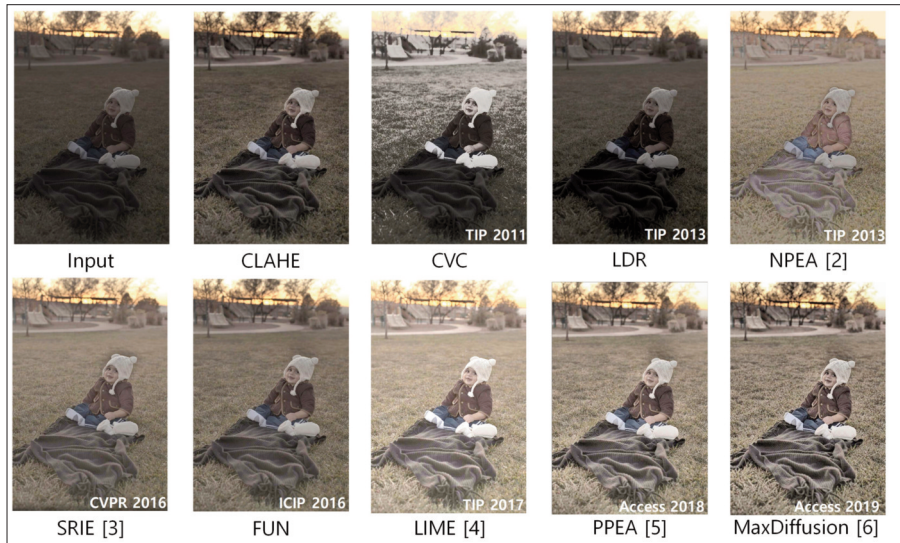
이번 장에서는 다양한 영상에 대한 화질 개선 결과를 살펴보고자 한다. 화질 개선 성능 평가를 위해 널리 사용되는 데이터셋으로는 NASA[14], HDR[15], 그리고 앞 장에서 소개한 MIT Adobe 5K[13] 데이터셋이 있다. NASA 데이터셋과 HDR 데이터셋은 원본 영상에 대한 화질 개선 버전은 포함하고 있지 않으나 심층학습에 널리 사용되고 있는 MIT Adobe 5K 데이터셋은 전문가 리터치 버전이 포함되어 있다. NASA 데이터셋과 HDR 데이터셋은 화질 개선 버전을 포함하고 있지 않기 때문에 성능 평

가를 위해 정답 영상을 사용하지 않는 객관적 화질 평가 메트릭(예를 들어, NIQE[16], BTMQUI[17], NIQMC[18], C-PCQI[19] 등)을 사용한다. 각각의 메트릭은 영상 내 콘텐츠의 왜곡 정도, 색상의 자연스러움 등을 평가할 수 있다. 한편, 전문가 리터치 결과를 포함한 MIT Adobe 5K를 기반으로 성능 평가를 수행할 시 복원된 영상과의 차이를 계산할 수 있기 때문에 PSNR이나 SSIM과 같은 전형적인 평가 메트릭을 사용할 수 있다.

〈그림 3〉과 〈그림 4〉는 NASA 데이터셋 영상과 HDR 데이터셋 영상을 기반으로 다양한 최적화 기법을 이용한 방법들 간 영상 화질 개선 결과를 비교하여 보여주고 있다. 최적화 기반의 대부분의 기존 방법은 효과적으로 영상 화질 개선을 수행할 수 있음을 볼 수 있다. 최근에 출판된 논문의 경우 저조도 환경이나 복잡한 장면에서 좀 더 개선된 성능을 보여주고 있다. 〈그림 5〉는 심층학습 기반 방법의



〈그림 3〉 최적화 기반의 영상 화질 개선 결과의 예(NASA 데이터셋 영상)



〈그림 4〉 최적화 기반의 영상 화질 개선 결과의 예(HDR 데이터셋 영상)



〈그림 5〉 심층학습 기반의 영상 화질 개선 결과의 예(MIT Adobe 5K 데이터셋 영상)

영상 화질 개선 결과를 보여주고 있다. 대부분의 방법들이 전문가 리터치 결과를 충실히 학습하여 유사한 복원 결과를 보여주고 있다. 복원 과정에서 원본 영상의 색상(Color Attribute)을 최대한 유지하는 방향으로 결과가 도출되고 있음을 확인할 수 있다. 또한, 최적화 기반 방법 대비 불필요한 잡음 증폭 등의 왜곡 현상을 효과적으로 줄일 수 있다. 그러나, 전문가 리터치가 항상 개선된 화질의 정답이라고 할 수 없는 문제는 여전히 해결해야 할 문제로 남아 있다.

IV. 결론

본 고에서는 인공지능 기술, 특히, 심층학습을 이용한 영상 화질 개선 기술 동향에 대해 살펴보았다. 모바일 기기를 위한 카메라 기술이 꾸준히 발전하

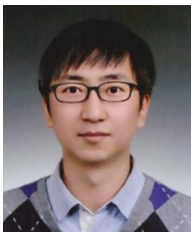
고 있지만, 일상 생활에서 발생하는 다양한 조명 환경에서 획득되는 영상에는 여전히 열화 현상이 포함되어 있다. 컴퓨터 비전 분야에서는 이를 해결하기 위해 최적화 기법을 이용하여 영상 내 조명 성분을 추정하여 개선하는 방법과 최근 영상 인식 분야에서 뛰어난 성능을 보여준 심층학습 기술을 이용하여 화질 개선을 수행하는 방법에 대하여 활발히 연구가 진행되고 있다. 특히, 심층신경망 기반 영상 화질 개선은 복잡한 모델링 없이 대용량 학습을 통해 효과적으로 영상 화질을 개선할 수 있기 때문에 최근 연구가 활발히 진행되고 있다. 그러나 여전히 학습을 위한 전문가 리터치(Re-touch) 과정이 요구되기 때문에 이를 극복할 수 있는 방안이 필요한 실정이다. 이를 효과적으로 극복한다면 인공지능 기술 기반 영상 화질 개선 방법이 모바일 기기에도 성공적으로 적용될 수 있을 것으로 기대된다.

참고 문헌

- [1] E. H. Land, "The Retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108-128, Dec. 1977.
- [2] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538-3548, Sep. 2013.
- [3] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2782-2790.
- [4] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982-993, Feb. 2017.
- [5] W. Kim, "Image enhancement using patch-based principal energy analysis," *IEEE Access*, vol. 6, pp. 72620-72628, 2018.
- [6] W. Kim, R. Lee, M. Park, and S-H. Lee, "Low-light image enhancement based on maximal diffusion values," *IEEE Access*, vol. 7, pp. 129150-129163, Dec. 2019.
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Int'l. Conf. Comput. Vis. Patt. Recognit.*, Jun. 2016, pp. 770-778.
- [8] R. Wang, Q. Zhang, C-W. Fu, X. Shen, W-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 6849-6857.
- [9] A. Ignatov, N. Kobyshev, K. Vanhoey, R. Timofte, and L. Van Gool, "DSLR-quality photos on mobile devices with deep convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 3297-3305.
- [10] J. Park, J-Y. Lee, D. Yoo, and I. S. Kweon "Distort-and-recover: color enhancement using deep reinforcement learning," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5928-5936.

- [11] Y-S. Chen, Y-C. Wang, M-H. Kao, and Y-Y. Chuang, "Deep photo enhancer: unpaired learning for image enhancement from photographs with GANs," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6306-6314.
- [12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Proc. Neural Inf. Process. Syst., Dec. 2014, pp. 1-9.
- [13] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input / output image pairs," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2011, pp. 97-104.
- [14] Retinex Theory of Color Vision, NASA, Washington, DC, USA, 2001.
- [15] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust patch-based HDR reconstruction of dynamic scenes," ACM Trans. Graph., vol. 31, no. 6, 2012, Art. no. 203.
- [16] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," IEEE Signal Process. Lett., vol. 20, no. 3, pp. 209-212, Mar. 2013.
- [17] K. Gu, S. Wang, G. Zhai, S. Ma, X. Yang, W. Lin, W. Zhang, and W. Gao, "Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure," IEEE Trans. Multimedia, vol. 18, no. 3, pp. 432-443, Mar. 2016.
- [18] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," IEEE Trans. Cybern., vol. 47, no. 12, pp. 4559-4565, Dec. 2017.
- [19] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 4, pp. 1301-1313, Apr. 2018.

필자소개



김원준

- 2012년 8월 : 한국과학기술원(KAIST) 박사
- 2012년 9월 ~ 2016년 2월 : 삼성종합기술원 전문연구원
- 2016년 3월 ~ 현재 : 건국대학교 전기전자공학부 조교수
- 주관심분야 : 컴퓨터 비전, 영상처리, 기계학습