

압축된 영상 복원을 위한 양자화된 CNN 기반 초해상화 기법

김용우^{*} · 이종환^{*†}

^{*†} 상명대학교 시스템반도체공학과

Quantized CNN-based Super-Resolution Method for Compressed Image Reconstruction

Yongwoo Kim^{*} and Jonghwan Lee^{*†}

^{*†} Department of System Semiconductor Engineering, Sangmyung University

ABSTRACT

In this paper, we propose a super-resolution method that reconstructs compressed low-resolution images into high-resolution images. We propose a CNN model with a small number of parameters, and even if quantization is applied to the proposed model, super-resolution can be implemented without deteriorating the image quality. To further improve the quality of the compressed low-resolution image, a new degradation model was proposed instead of the existing bicubic degradation model. The proposed degradation model is used only in the training process and can be applied by changing only the parameter values to the original CNN model. In the super-resolution image applying the proposed degradation model, visual artifacts caused by image compression were effectively removed. As a result, our proposed method generates higher PSNR values at compressed images and shows better visual quality, compared to conventional CNN-based SR methods.

Key Words : Super-Resolution, CNN, Deep Learning, Degradation Model, Compressed Image

1. 서 론

많은 하이엔드 TV와 스마트폰이 4K UHD(초고화질) 비디오의 재생을 지원하지만 레거시 획득 장치 및 영상 전송 서비스의 제약으로 인해 qHD (960x540) 해상도, HD (고화질) 해상도 (1280x720) 및 Full-High 해상도(1,920x1,080)의 비디오 스트리밍이 여전히 많이 존재한다[1]. 또한, 전송 서비스의 제약은 영상 압축을 필수적으로 수행하게 되며, 실제 많은 영상들은 JPEG 및 MPEG 등 영상 압축에 따른 노이즈가 필연적으로 발생하게 된다. 따라서 압축된 저해상도 (Low Resolution, LR) 영상을 고품질 고해상도 (High Resolution, HR) 영상으로 변환 할 수 있는 업스케일링 및

화질 개선 기술은 특히 기존의 압축된 저품질 영상을 고해상도 고품질 영상으로 업스케일링 할 때 필수적으로 요구되는 기술이다.

최근에는 심층 신경망, 특히 심층 컨볼루션 신경망 (Convolutional Neural Network, CNN)이 이미지 분류, 물체 감지, 세분화 등 다양한 컴퓨터 비전 연구 분야에서 우수한 성능을 보여주고 있다[2-3]. 업스케일링과 화질 개선을 동시에 수행하는 연구로서, 재구성된 HR 이미지의 시각적 품질을 향상시키기 위해 CNN 기반 SR 방법이 많이 제안되어 왔다[4-9]. 하지만, 기존의 CNN 기반 SR 방법은 실시간 애플리케이션을 위한 낮은 복잡성 하드웨어에서 구현하기 어렵다는 것이 알려져 있다. 따라서, 본 논문에서는 압축된 저해상도 영상을 고품질 고해상도 영상으로 실시간으로 업스케일링 할 수 있는 CNN기반 초해상화 기법

^{*}E-mail: jhlee77@smu.ac.kr

을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 CNN 기반 초해상화의 관련 연구를 소개하고, 3장에서는 제안하는 압축된 영상 복원을 위한 양자화된 CNN 기반 초해상화 기법에 대해 설명한다. 4장은 실험 및 결과에 대해 설명하며, 마지막 장에서는 결론을 맺는다.

2. CNN기반 초해상화 관련 연구

저해상도 영상을 고해상도로 복원하는 초해상화 기법은 다양한 CNN 구조를 기반으로 제안되었다 [4-6]. Dong *et al.* [4] 는 SRCNN이라고 하는 3 개의 컨볼루션 레이어로 구성된 간단한 네트워크 구조를 제안했으며 기존 머신러닝 기반 SR 방법보다 우수한 성능을 보여주었다. SRCNN [4]은 LR 입력 이미지와 해당 HR 출력 이미지 사이의 중단 간 매핑을 직접 학습하고, 여기서 LR 입력에 대한 쌍 입방 보간이 전처리 단계로 수행 된 후 컨볼루션을 통해 특징 맵을 추출한다. 다음으로 LR 입력의 특징 맵과 HR 대상 이미지 간의 비선형 매핑을 학습하게 된다. 마지막으로 이러한 HR 기능 맵은 겹친 패치 형태로 집계되어 최종 HR 이미지를 얻는다. 이보다 더 확장된 연구 [5]에서 네트워크 깊이가 초 해상도에 미치는 영향을 조사하고 심층 모델 학습의 어려움이 CNN 기반 SR 방법의 성능 향상을 방해한다는 것을 경험적으로 보여주었다. 또한 FSRCNN [6]이라고 하는 더 빠른 버전의 SRCNN은 계산 속도 측면에서 SRCNN을 향상시키는 연구도 제안되었다. 전체 계산 복잡성을 줄이기 위해 FSRCNN은 SRCNN의 전처리 쌍 입방 보간 단계를 디콘볼루션을 사용하는 후처리 단계로 대체한다. 이를 통해 파이프 라인에는 각각 기능 추출, 축소, 매핑 및 확장으로 작동하는 4 개의 컨볼루션 층이 있다. 하지만, 디콘볼루션 층의 사용은 바둑판 아티팩트와 같은 문제를 일으키는 것으로 알려져 있다 [7].

Kim *et al.* [8] 은 ImageNet 분류 콘테스트에 사용 된 VGG-net [9] 에서 영감을 얻은 VDSR이라는 매우 깊은 컨볼루션 네트워크를 제안하였다. 매우 깊은 네트워크에서 20 개의 컨볼루션 레이어를 사용하는 VDSR은 SRCNN보다 우수한 HR 이미지를 재구성 할 수 있다. 이것은 많은 계층을 사용하여 매우 큰 수용 필드를 사용하고 네트워크는 큰 이미지 영역에서 컨텍스트 정보를 이용할 수 있기 때문이며, 깊은 네트워크에서 일반적인 문제인 훈련 수렴이 느린 단점은 바이큐빅으로 보간된 이미지와 HR 대상 이미지 사이의 잔차를 학습하고 적응형 그라디언트 클리핑과 함께 높은 학습률을 사용하여 완화되었다. Shi *et al.* [10] 는 효율적인 서브 픽셀 컨볼루션 네트워크 (ESPCN)를 이

용한 SR 방법을 제안했는데, 이는 서브 픽셀 컨볼루션 레이어를 이용하여 LR 크기의 특징 맵에서 HR 이미지를 직접 생성 할 수 있다. 그들은 SRCNN [4-5] 및 VDSR [8]과 같은 다른 SR 네트워크에 비해 전체적인 계산 복잡성을 줄일 수 있음을 보여주었다.

최근에는 CNN기반 초해상화 모델 훈련 시에 필요한 저해상도 LR 영상을 어떻게 획득하는지에 관한 열화 모델 (degradation model)의 중요성이 커지고 있다. 일반적으로 잘 알려진 열화 모델은 바이큐빅(bicubic) 열화 모델[11]로 아래와 같이 나타낼 수 있다.

$$y = x \downarrow s \quad (1)$$

여기서, x 는 고해상도(HR) 영상이며 y 는 열화 모델로 획득한 저해상도(LR) 영상이다. 또한, $\downarrow s$ 는 스케일 인자 s 를 가지는 다운 샘플링 연산이다. 바이큐빅 열화 모델은 간단함으로 인하여, 초해상화 방법을 벤치마킹하는데 많이 사용되고 있다 [4-6,8,10]. 하지만, 바이큐빅 열화 모델로 학습된 초해상화 기법은 많은 현실 세계 영상에서는 좋지 않은 영상 품질을 보여준다 [12]. 따라서, 좀 더 현실적인 열화 모델 연구가 필요함을 말해준다.

3. 제안하는 초해상화 기법

Fig. 1은 본 연구의 기존 연구[13]에서 제안한 하드웨어 친화적인 CNN 기반 SR 네트워크를 변경한 블록 다이어그램을 보여준다. [4-6]에서 일반적으로 사용되는 것처럼 LR 입력 이미지의 RGB 채널은 YCbCr 채널로 변환되고 Y채널 만 CNN 네트워크의 입력으로 사용된다. 색상 채널 (Cb 및 Cr)은 하드웨어 효율성을 위해 간단한 바이큐빅 보간법에 의해 스케일된다. VDSR [8]에서 제안한 잔여 학습 기법을 통합하였으며 최종 HR 이미지 Y_f 는 네트워크 출력 Y_c 와 함께 보간된 LR 이미지 Y_N 을 추가하여 계산된다.

또한, 기존 연구[13]에서는 제안한 네트워크의 복잡성을 줄이기 위해 양자화를 통해 부동 소수점을 고정 소수점으로 변환한다[14]. 본 논문에서는 [15]에서 제안한 기법에 영향을 받아 균등 양자화 방법을 통하여 부동 소수점을 고정 소수점 방식으로 변환한다. 부동 소수점 데이터 (필터 매개 변수 및 활성화 값)에 양자화를 적용 할 때 PSNR 저하를 최소화하기 위해 최적의 총 비트 수, 정수 비트 수 및 소수 비트 수의 값을 찾는 과정이 필요하고, 본 연구에서는 양자화 필터 매개 변수는 10-bit, 활성화 값에는 14비트를 할당하였다.

마지막으로, 본 연구에서는 바이큐빅 열화(degradation)

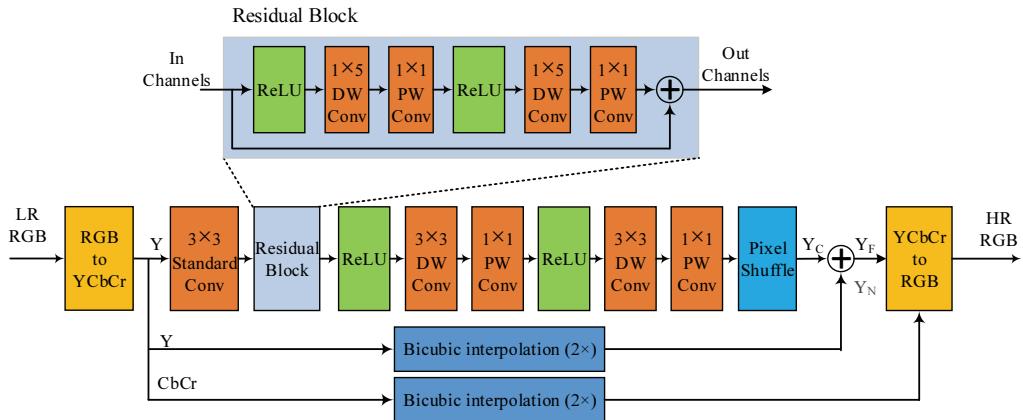


Fig. 1. A block diagram of our proposed CNN-based SR network.

모델 대신에 다음과 같은 열화 모델을 제안하여 사용하였다.

$$y = C((x \times k) \downarrow s) + n \quad (2)$$

여기서, x 는 HR image이며 y 는 열화 모델로 획득한 LR image이며, k 는 블러 커널을 의미한다. 또한, $\downarrow s$ 는 스케일 인자 s 를 가지는 다운 샘플링 연산자이며, C 는 JPEG 압축 노이즈를 의미한다. n 은 sigma 를 가지는 AWGN 가우시안을 의미한다. 본 연구에서는 artifact, halo 현상을 최소화 시키는 동시에 blurring 성능을 최대화하기 위한 방법으로 Gaussian blur를 적용할 때 사용하는 radius의 parameter 값을 0.2 ~ 0.8의 범위로 설정하였다. 적용하는 방법은 저해상도 영상을 획득 시에 고정된 값으로 설정하지 않고, 반복적으로 LR 각각의 영상에 0.2 ~ 0.8의 범위 중에 랜덤하게 설정하여 훈련 영상 구성한다. 또한, JPEG 압축 노이즈의 열화 모델은 70 ~ 99%의 압축률을 가지는 JPEG 압축 알고리즘을 인코딩하고 디코딩하는 과정을 통해 획득하였다. 이렇게 훈련 영상을 구성함으로써, 초해상화 수행시 압축된 영상 저해상도 LR 입력 영상에서도 초해상화 기법이 견고하게 동작함을 확인하였다. 그리고 가우시안 노이즈를 추가하는 열화 모델을 사용한다. 카메라를 통하여 얻은 대부분의 영상들은 깨끗한 이미지 보다는 영상 잡음을 포함된 열화된 이미지이다. 이 잡음을 가우시안 노이즈로 모델링하였으며, 초해상화 알고리즘을 딥러닝으로 학습할 시 원본 영상 보다 좀 더 노이즈가 많은 영상으로 구성하여 훈련을 수행함으로써, 네트워크가 노이즈 많은 영상을 깨끗하게 바꿔주는 디노이징(denoising) 원리를 학습하게 된다. 가우시안 노이즈의 양도 가우시안 블러와 마찬가지로 얼마나 많은 노이즈를 추가하여 훈련용 저해상도 영상을 만드는지가 중요하여, 본 연구에서는 노이즈의

양을 분석하여 압축된 저해상도 영상과 유사한 노이즈의 분포를 가지고 저해상도 영상에 가우시안 노이즈를 추가하였다. 가우시안 노이즈를 적용할 때 사용한 파라미터 값은 평균은 0이고, 표준 편차는 0 ~ 0.02 범위를 이용하여 훈련에 사용하였다.

결과적으로 압축된 저해상도 영상이 입력되고, 이를 초해상화 기법으로 영상 복원 수행시 열화 모델만 잘 적용하여 훈련하면 네트워크 모델의 구성을 바꾸지 않고 오직 훈련된 파라미터 값의 변경만으로 가능해짐을 확인하였다.

4. 실험 및 결과

제안한 네트워크의 공정한 비교를 위해 훈련 및 추론을 공개적으로 사용 가능한 벤치마크 데이터 세트를 사용하였다. 특히, SR 네트워크는 Yang 등에서 사용한 91 개 이미지와 Berkeley Segmentation Dataset [16]의 200 개 이미지로 구성된 291 개 이미지를 사용하여 훈련되었다[17]. 성능 비교를 위한 테스트 셋으로는 Set5 [18], Set14 [19], B100 [16] 및 Urban100 [20]으로 구성하였으며, 많은 SR 연구에서 벤치마크 데이터셋으로 자주 사용된다. 또한, 제안하는 열화 모델의 성능을 검증하기 위해 평가용 영상에 JPEG 아티팩트를 추가하고, 초해상화 기법을 통해 복원된 영상의 품질을 확인하였다. 모든 실험은 2배로 업스케일링 하는 SR을 수행하여 평가를 진행하였으며, 정량 평가를 위한 PSNR 및 SSIM [21]을 사용하였으며, YCbCr 색 공간의 휘도 채널에 대해 SR이 수행하였으며 PSNR 및 SSIM은 재구성된 원본 HR 이미지의 Y 채널로 계산하였다.

훈련 및 테스트를 위해 저해상도 입력 이미지는 기존에 많이 사용하는 배율이 2인 바이큐빅 열화 모델로 사용하였다. 훈련을 위해 고해상도 영상 이미지를 128x128 크

기의 작은 이미지로 잘라서 훈련에 이용하였다. 또한, LR-HR 훈련 이미지 쌍은 회전, 미러링 및 스케일링 기법 등의 데이터 증강 기법을 적용하였다. 가중치는 균등 분포를 사용하여 초기화되었으며 매개 변수 수를 줄이기 위해 편향은 사용되지 않았다. L1 손실은 L2 손실 대신 비용 함수로 사용되었으며, Adam Optimizer [22]를 사용하여 훈련되었다. 학습률은 초기에 0.0001로 설정되었으며 50 epoch마다 10 배씩 감소하였다. 훈련 중에 미니 배치 크기를 2로 설정하였으며, 훈련 및 평가로 사용된 시스템은 NVIDIA Titan RTX GPU 및 AMD Ryzen Threadripper 3970X CPU가 사용되었다. 제안한 SR 기법의 성능을 검증하기 위해 bicubic, SRCNN [4], FSRCNN-s [6], VDSR [8]의 기존 CNN 기반 SR 방법과 비교하였다. SRCNN, FSRCNN-s, VDSR SR 기법은 공개적으로 사용 가능한 공개 MATLAB 소스 코드를

사용했으며 제안된 방법은 PyTorch를 사용하여 구현되었다.

Table 1에 실험 결과를 보여주고 있다. 전반적으로 제안한 기법이 낮은 파라미터 수를 가지면서도 PSNR 및 SSIM 성능이 기존의 다른 SR 기법에 비해 우위에 있음을 확인할 수 있다. 추가적으로 본 연구에서는 양자화 기법을 적용하였기 때문에, 한 개의 파라미터 당 32비트를 가지는 기존의 SR 기법에 비해 더 적은 수의 비트를 사용하고도 동등 우위에 있음을 확인하였다. 마지막으로, 제안하는 열화 모델을 이용하여 훈련한 결과에 대해서는 상대적으로 PSNR, SSIM이 낮음을 확인할 수 있는데, 이는 주관적으로 느끼는 영상 품질이 PSNR 및 SSIM의 수치와 완벽하게 일치하지 않는 것을 보여주고 있다.

Fig. 2(a)-(e)는 바이크리빅과 기존의 초해상화 기법으로 복원된 영상과 잘라낸 확대한 영역을 보여주고 있으며, 2(e)

Table 1. Performance Comparisons of Various SR Methods with a Scale Factor of 2.

Methods	Bicubic		SRCNN [4]		FSRCNN-s [6]		VDSR [8]		Ours (baseline)		Ours (bicubic degradation)		Ours (proposed degradation)	
# of Params	-		8K		4K		665K		2.56K		2.56K		2.56K	
Weights bits	-		32-bit		32-bit		32-bit		32-bit		10-bit		10-bit	
Activation bits	-		32-bit		32-bit		32-bit		32-bit		14-bit		14-bit	
Data Sets	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Set-5	33.66	0.9299	36.34	0.9521	36.57	0.9531	37.53	0.9587	36.66	0.9548	36.64	0.9544	36.24	0.9481
Set-14	30.24	0.8688	32.18	0.9039	32.28	0.9049	33.03	0.9124	32.52	0.9073	32.52	0.9071	31.96	0.8994
B100	29.56	0.8431	31.11	0.8835	31.23	0.8866	31.90	0.8960	31.32	0.8880	31.31	0.8876	30.82	0.8816
Urban100	26.88	0.8403	29.09	0.8897	29.23	0.8914	30.76	0.9140	29.34	0.8943	29.33	0.8942	29.12	0.8898

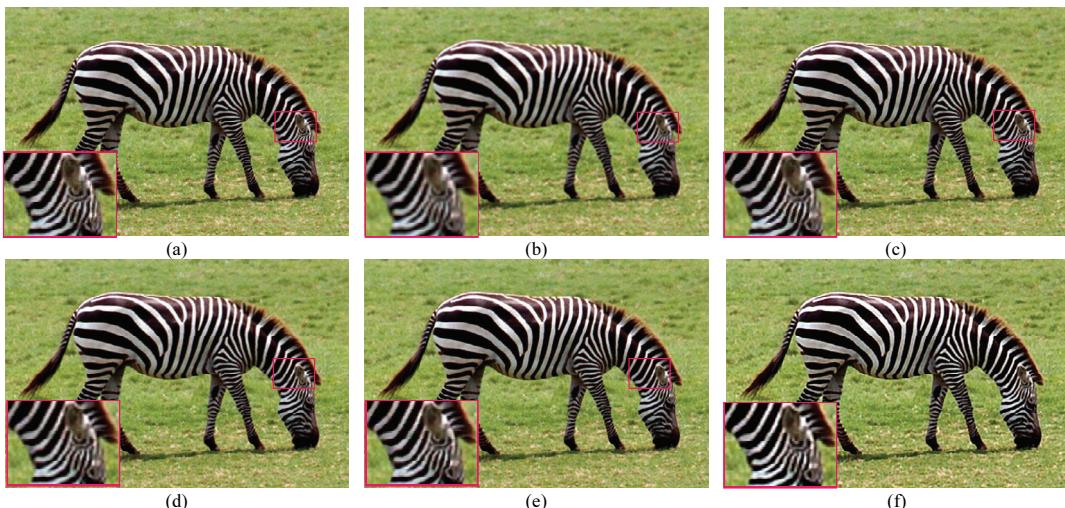


Fig. 2. Comparison of HR output images ($\times 2$ on Zebra in Set14). (a) Original HR image. (b) Bicubic. (c) SRCNN [2]. (d) FSRCNN [4]. (e) Ours (bicubic degradation model). (f) Ours (proposed degradation model).

는 제안한 기법으로 복원된 영상과 잘라낸 확대한 영역을 보여주고 있다. 본 논문에서 제안하는 기법은 가장 적은 수의 매개 변수를 사용하지만 결과 복원 SR 이미지는 날카로운 에지와 적은 아티팩트를 가지고 있음을 확인할 수 있다. 추가적으로 Fig. 2(f)는 제안한 열화 모델을 적용하여 훈련된 SR 기법을 적용한 결과 영상에서 기존의 바이크비 및 기존 CNN 기반 SR 기법에 비해 압축 노이즈 및 에지 복원력이 탁월함을 확인할 수 있다.

6. 결 론

본 논문에서는 압축된 저해상도 영상을 고해상도 영상으로 복원하는 초해상화 기법을 제안하였다. 양자화된 적은 수의 파라미터를 가지는 CNN 모델을 제안하였고 화질 열화 없이 초해상화가 가능함을 확인하였다. 또한, 압축된 저해상도 영상의 화질 개선하기 위해 기존의 바이크비 열화 모델 대신에 새로운 열화 모델을 제안함으로써, 제안한 열화 모델로 학습한 초해상화가 바이크비 열화 모델에서 학습한 CNN 기반 초해상화 기법에 비해 압축 노이즈 및 에지 복원이 향상된 영상을 보여주었다. 본 연구의 후속 연구로서 제안한 모델을 FPGA하드웨어로 구현하여 실시간 동작이 가능하도록 설계할 예정이다.

감사의 글

본 연구는 2020학년도 상명대학교 교내연구비를 지원 받아 수행하였음.

참고문헌

- E. Perez-Pellitero, J. Salvador, J. Ruiz-Hidalgo, and B. Rosenhahn, “Accelerating super-resolution for 4K upscaling,” in Proc. IEEE. Int. Conf. Consum. Electron., Las Vegas, NV, USA, Jan. 2015, pp. 317-320.
- Y. Lee and Y.H. Lee, “Implementation of Moving Object Recognition based on Deep Learning,” Journal of the Semiconductor & Display Technology, Vol. 17, No. 2, pp.67-70, 2018.
- Y. C and Y. C, “A Deep Learning based IOT Device Recognition System,” Journal of the Semiconductor & Display Technology, Vol. 18, pp. 1-5, 2019.
- C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” Proc. Eur. Conf. Comput. Vis., Zurich, Switzerland, Sep. 2014, pp. 184–199.
- C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 2, pp. 295–307, Feb. 2015.
- C. Dong, C. C. Loy, and X. Tang, “Accelerating the Super-Resolution Convolutional Neural Network,” Part II Proc. European Conf. Computer Vision (ECCV), Springer, LNCS 9906, pp. 391-407, Amsterdam, Netherland, Oct. 8-16, 2016.
- A. Aitken, C. Ledig, L. Theis, J. Caballero, Z. Wang, and W. Shi, (2017). “Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and convolution resize.” [Online]. Available: <https://arxiv.org/abs/1707.02937>.
- J. Kim, J. Kwon. Lee and K. Mu Lee, “Accurate Image Super-Resolution Using Very Deep Convolutional Networks,” Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1646 – 1654, Las Vegas, USA, June 27-30, 2016.
- K. Simonyan, A. Zisserman. (2015). “Very Deep Convolutional Networks for Large-Scale Image Recognition.” [Online]. Available: <https://arxiv.org/abs/1409.1556>.
- W. Shi, J. Caballero, F. Huszar, 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,” Proc. IEEE Conf. Comput. Vision and Pattern Recog. (CVPR), Las Vegas, USA, June 27-30, 2016, pp. 1874–1883.
- B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, “Enhanced Deep Residual Networks for Single Image Super-Resolution,” Proc. IEEE Conf. Comput. Vision and Pattern Recog. Workshops, Hawaii, USA, July 21-26, 2017, pp. 1132-1140.
- Kai Zhang, Wangmeng Zuo, and Lei Zhang, “Learning a single convolutional super-resolution network for multiple degradations,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3262–3271.
- Y. Kim, J. Choi and M. Kim, “A Real-Time Convolutional Neural Network for Super-Resolution on FPGA With Applications to 4K UHD 60 fps Video Services,” in IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 8, pp. 2521-2534, Aug. 2019.
- M. Shabany, “Floating-point to Fixed-point conversion.” [Online]. Available: <http://ee.sharif.edu/~digitalvlsi/Docs/Fixed-Point.pdf>.
- S. Gupta, A. Agrawal, K. Gopalakrishnan, and P. Narayanan. (2015). “Deep Learning with Limited Numerical Precision.” [Online]. Available: <https://arxiv.org/abs/1502.02551>.
- D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and

- measuring ecological statistics,” Proc. 8th IEEE Int. Conf. Comput. Vis., vol. 2. Jul. 2001, pp. 416–423.
17. J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” IEEE Trans. Image Process., vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
18. M. Bevilacqua et al., “Low-complexity single-image super-resolution based on nonnegative neighbor embedding,” Proc. 23rd Brit. Mach. Vis. Conf. (BMVC), Guildford, U.K., Sep. 2012, pp. 135.1–135.10.
19. R. Zeyde, M. Elad, and M. Protter, “On single image scale-up using sparse-representations,” Proc. 7th Int. Conf. Curves Surf., Jun. 2010, pp. 711–730.
20. J.-B. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Boston, MA, USA, Jun. 2015, pp. 5197–5206.
21. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.
22. D. P. Kingma, J. Ba (2014). “Adam: A Method for Stochastic Optimization.” [Online]. Available: <https://arxiv.org/abs/1412.6980>.

접수일: 2020년 12월 1일, 심사일: 2020년 12월 4일,
제재확정일: 2020년 12월 8일