

## 두 장의 이미지를 활용한 이미지 화질 저하 커널 예측

조선우 조남익

서울대학교 전기정보공학부 뉴미디어통신연구소(INMC)

[etoo33@ispl.snu.ac.kr](mailto:etoo33@ispl.snu.ac.kr) [nicho@snu.ac.kr](mailto:nicho@snu.ac.kr)

## Blind Super-Resolution Kernel estimation using two images

Cho Sunwoo, Cho Nam Ik

Department of ECE, INMC, Seoul National University

## 요약

이미지 초해상도는 영상 취득 과정에서 센서와 렌즈의 물리적인 한계 등으로 인하여 의해 화질이 저하된 이미지를 더 높은 배율로 복원하는 문제이다. 이미지 초해상도는 딥러닝을 통해 놀라운 성능향상을 이루었지만, 카메라로 촬영된 실제 이미지에서는 좋은 성능을 내지 못하였다. 이는 딥러닝에서는 'bicubic' 커널로 down-sampling된 합성 이미지 데이터를 사용하였던 것과 달리 실제 이미지에서는 'bicubic' 커널을 통한 화질 저하와는 다른 화질 저하, 즉 다른 커널을 통한 화질 저하가 발생하기 때문이다. 따라서 실제 이미지에 대한 성능을 높이기 위해서는 이에 대한 정확한 커널 예측이 필요하다. 최근 주목받기 시작한 이미지 초해상도를 위한 커널 예측은 초해상도를 잘 시켜주는 커널을 직접 찾는 방법[10, 13]과 이미지의 분포와 커널을 통해 다운샘플링된 이미지에 대한 분포를 일치시켜주면서 커널을 예측하는 방법[14]으로 나누어져 있다. 그러나 두 방법 모두 ill-posed problem인 커널 예측 문제를 한 장의 이미지만으로 해결하려는 것이기 때문에 정확한 예측에는 어려움이 발생한다. 따라서 본 논문에서는 두 장의 이미지를 활용한 이미지 화질 저하 커널 예측 방법을 제안한다. 제안된 방법은 두 장의 이미지가 같은 카메라를 통해 촬영되었으며 이때 이미지 화질 저하는 카메라에 의해서만 영향을 받는다는 가정을 기반으로 한다. 즉, 두 장의 이미지는 같은 커널을 통해 저하된 이미지라는 가정을 한다. 제안된 방법은 [14]에서처럼 이미지 분포를 기반으로 한 커널 예측을 진행하며, 이미지 초해상도를 진행하고자 하는 이미지 외에 참고 이미지 또한 같은 커널에서 화질 저하를 시켰을 때 본래의 이미지와 같은 분포에 있도록 학습을 진행한다. 결과적으로 본 논문에서는 두 장의 이미지를 사용하였을 때 더욱 정확하게 커널을 찾을 수 있음을 보여준다. 두 장의 이미지를 활용하는 방식이 한 장의 이미지만을 활용하는 기존의 최고 수준의 방법에 비해 합성된 다양한 커널 데이터셋[14]에서 약 0.17dB 성능 향상이 있었다.

## 1. 서론

이미지 초해상도는 저해상도 이미지를 활용하여 고해상도 이미지를 복원하는 문제이다. 이미지 초해상도는 사용자들에게 고화질 영상 제공을 도울 뿐 아니라 다른 high-level vision 문제의 성능을 높여주기 위한 전처리 과정으로 많이 사용된다. 이에 따라 이미지 초해상도는 중요한 컴퓨터 비전 문제 중 하나로 여겨져 왔다.

노이즈가 없는 이미지 초해상도에서 고해상도 이미지와 저해상도 이미지 사이의 관계는 다음과 같은 수식으로 나타낼 수 있다.

$$y = (x \otimes k) \downarrow_s.$$

여기서  $x$ 는 고해상도 이미지를 나타내며  $y$ 는 저해상도 이미지를 나타낸다. 또한,  $K$ 는 이미지 화질 저하를 진행하는 커널을 나타내며  $\downarrow_s$  연산은 이미지를 배율  $s$ 로 다운 샘플링 하는 과정을 나타낸다.

기본적인 이미지 초해상도는 위 식에서 화질 저하를 진행하는 커널을 'bicubic' 커널로 가정하여 문제를 해결한다. 최근 딥러닝의 발전과 함께 이루어진 놀라운 성능향상도 'bicubic' 커널로 데이터를 합성하여 딥 뉴럴 네트워크(DNN) 학습시키며 진행되었다. SRCNN [1]은 최초로 이미지 초해상도에 딥러닝을 적용하여 고전적인 방법의 성능을 추월하

였다. 뒤이어 VDSR [2] 과 EDSR [3]은 이러한 모델 구조를 더욱 깊게 쌓고 잔차 학습방법(residual learning)을 활용한 구조를 도입하여 더욱 눈에 띄는 성능향상을 가져왔다. 또한, 네트워크의 레이어별로 어텐션을 활용하여 특징맵들을 모두 활용하는 residual dense block을 제안한 RDN [4]은 모델의 파라미터 수를 늘려 성능을 더욱 향상 시켰다. 이 후로, 모델의 채널 간 관심도를 예측하여 좀 더 낮은 파라미터 수로도 더욱 큰 성능을 보여주는 RCAN [5]과 네트워크를 효율적으로 사용하여 더욱 파라미터 수를 줄인 CARN[6]이 제안되었다.

그러나 이렇게 'bicubic' 커널에 대해서만 학습이 진행된 기존의 연구들은 실제 이미지에 대해서는 높은 성능을 보이지 못한다. 이는 실제 이미지들은 단순한 'bicubic' 커널 모델에 잘 맞지 않는 카메라의 렌즈에 따라서 화질 저하가 이루어졌기 때문이다. [7]은 카메라를 통해 카메라 렌즈의 배율을 바꾸어 화질 저하를 생성하며 실제 카메라를 통해 저하된 이미지가 'bicubic' 커널로 저하된 이미지와는 다른 형태를 갖고 있음을 보여주었다. 또한 [8]에서는 실제 촬영된 이미지에서 고전적인 방식으로 커널 예측을 하여 실제 이미지에서는 'bicubic'과는 달리 다양한 커널이 있음을 보여주었고 해당 커널들로 데이터를 합성하여 DNN을 학습시켰을 때 실제 이미지에서 더욱 높은 성능을 보여주었다.

이에 따라 최근에는 실제 이미지에서의 성능을 높이기 위해서 커널

을 고려하는 방향으로 연구가 진행되었다. 또한 이미지 초해상도 문제를 두 가지 관점에서 나누어 접근하기 시작하였다. 첫 번째는 이미지 커널이 주어졌을 때 해당 커널에 대해서 이미지 초해상도를 하는 문제이다. Kai Zhang et. al. 은 [9]에서 다양한 커널에 대한 모델링을 처음으로 제안하며 각 커널 별로 저화질 이미지와 커널을 함께 네트워크의 입력으로 주어 정답이미지를 예측하도록 네트워크를 학습시켰다. [10]의 논문에서는 이를 더욱 효율적으로 수행해줄 수 있도록 공간 피쳐 변환을 활용한 네트워크 구조를 제안하였다. 또한 [11]에서는 동적 컨볼루션을 활용한 네트워크 구조로 더욱 높은 성능을 보여주었다. 그러나 이러한 방법들은 기존의 높은 성능을 나타내는 'bicubic' 기반 네트워크를 사용하기 어렵고 새롭게 모든 커널에 대한 학습을 진행해야 한다는 단점이 있다. 따라서 [12]에서는 기존의 높은 성능을 보여주는 'bicubic' 기반으로 학습된 네트워크를 최적화 문제의 prior로 활용하여 최적화 문제를 푸는 방법으로 문제를 해결하였다. [13]에서는 특정 kernel로 합성된 저화질 이미지를 해당 이미지와 가장 유사한 'bicubic' 저화질 이미지로 변환시키는 최적화 문제에 대한 closed form으로 구하여 'bicubic' 저화질 이미지로 변환을 먼저 시킨 이후 'bicubic' 기반으로 학습된 네트워크에서 초해상도를 진행한다.

두 번째는 이미지 커널이 주어지지 않았을 때 이미지 초해상도를 진행하는 문제이다. 두 번째 문제는 대체로 한 장의 이미지로 커널을 먼저 예측한 이후 첫 번째 문제의 방법을 활용하여 초해상도를 진행한다. 따라서 두 번째 문제는 커널을 예측하는 문제로 귀결될 수 있다. 초해상도를 위한 커널 예측은 초해상도를 잘 시켜주는 커널을 직접적으로 찾는 방법[10, 13]과 이미지의 분포와 커널을 통해 다운샘플 된 이미지에 대한 분포를 일치시켜주면서 커널을 예측하는 방법[14]으로 나누어져있다. [11]에서는 최초로 딥러닝을 활용하여 커널을 예측하기 시작하였다. 다양한 커널로 학습된 네트워크에 예측된 커널의 정보가 들어가는데, 이러한 커널 예측에 오차에 따라 초해상도 이미지에 어떠한 왜곡이 발생하는 보여주었다. 따라서 초기에 예측된 커널을 통해 도출한 초해상도 이미지를 통해 예측된 커널을 수정하는 방식으로 커널 예측을 진행한다. [13]에서는 앞에서 언급했던 바와 같이 'bicubic' 저화질 이미지로의 근사를 진행하여 기존 'bicubic' 이미지로 학습을 진행시킨 네트워크에서 초해상도 이미지를 예측한다. 여기서도 마찬가지로 예측된 초해상도 이미지에서의 저화질 이미지와 입력 이미지와 가장 유사하게 만들어주는 커널을 DNN을 통해 예측한다. 위 두 방법에서 초해상도 이미지의 정보를 활용한 것과 달리 Bell-Kligler et. al.은 입력 이미지와 그의 특정 커널로의 저화질 이미지가 같은 커널로 이미지 저하가 이루어졌기 때문에 같은 이미지 분포에 있다는 가정을 활용하여 커널을 예측한다. 그러나 커널 예측 문제는 한 장의 저화질 이미지는 무수히 많은 고품질 이미지로부터 올 수 있기 때문에 커널 또한 무수히 많이 존재할 수 있는 ill-posed problem 이다. 따라서 한 장의 이미지로만 커널을 예측하는 이러한 방법들은 정확한 커널을 예측하는 데에 한계가 있다.

본 논문에서는 두 장의 이미지를 활용한 화질 저하 커널 예측 방법을 제안한다. 본 논문에서는 이미지 화질 저하가 주로 이미지를 촬영하는 카메라의 특성에 국한된 것, 즉 이미지를 촬영하는 촬영자의 손떨림이나 이미지 촬영 순간의 주변 환경의 변화와는 무관하다고 가정한다. 따라서, 같은 카메라로 서로 다른 이미지에 대한 촬영을 진행한다면 동일한 화질 저하 커널의 특성을 갖고 있을 것이며 이에 따라 두 장의 이미지는 같은 커널을 통해 화질 저하가 이루어졌다고 가정한다.

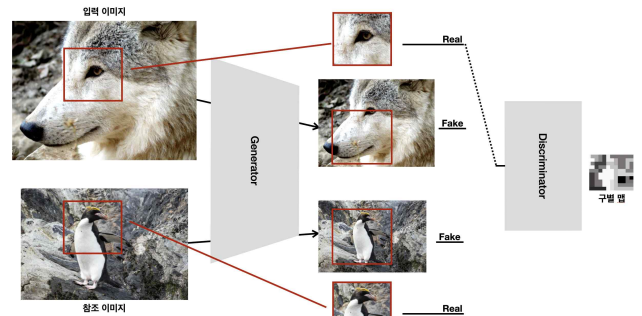


그림1 두 장의 이미지를 활용한 커널 예측 방법

위의 입력 이미지와 아래의 참조 이미지가 모두 Generator 네트워크에 들어가 특정 커널로 down-sampling이 이루어진 후 Discriminator에서 down-sampling 된 이미지와 원본 이미지를 구분하며 서로 적대적인 학습을 진행한다.

즉, 본 논문에서 제시하는 커널 예측 문제에서는 초해상도를 진행하고자 하는 한 장의 입력 이미지 이외에 같은 커널을 통해 화질이 저하된 참고 이미지 한 장이 추가로 제공된다. 본 논문은 [14]와 같은 실험 환경과 네트워크 구조를 사용하였다. 그러나 참조 이미지를 함께 사용함으로써 한 장의 이미지를 사용하던 [14]에서 보다 높은 정확도로 커널을 예측함을 보여주었다. 또한 합성된 다양한 커널 데이터셋[14]에서 0.17dB의 성능 향상을 이루었다.

## 2. 실험 방법

### 2.1 네트워크 구조

본 논문에서 사용한 전체적인 네트워크 구조는 그림 1에 잘 나타나 있다. 총 2개의 네트워크로 구성되어 있는데 하나는 입력 이미지가 들어갔을 때 특정한 커널로 down-sampling을 수행해주는 generator 네트워크이다. 다른 하나는 generator를 통해 나온 이미지를 임의로 자른 이미지와 입력 이미지를 임의로 자른 이미지가 들어갔을 때 둘을 잘 구분하도록 학습시키는 discriminator 네트워크이다.

Generator 네트워크의 경우 특정한 커널로 이미지를 down-sampling 해주는 연산을 대신 수행하는 역할을 한다. 이 때 이 연산은 선형 연산이기 때문에 generator 네트워크의 구조에는 non-linear 활성화 함수 요소가 없이 convolution layer를 5개 쌓아서 구성하였다. 중간 hidden layer는 모두 64 channel을 사용하였고 마지막 convolution에서만 stride를 2로 주어 downscaling이 수행되도록 구성하였다. Convolution layer의 kernel width는 순서대로 (5, 3, 1, 1, 1)으로 설정하였고 이렇게 하였을 때 13x13의 kernel을 통해 convolution을 한 후 downsampling을 수행하는 연산과 같아진다. 이러한 generator의 구조는 [14]에서 제안되었는데, 이는 단순히 한 레이어의 convolution으로 kernel을 구성하는 것보다 많은 전역 최적점을 갖고 있기 때문에 최적점을 갖고 있어 gradient 기반의 최적화 방법에서는 깊은 linear network가 얇은 것보다 빠르게 수렴한다 [15].

Discriminator는 [16]에서 소개된 patch 기반의 구조를 사용한다. 구조는 7개의 convolution 레이어로만 이루어져 있고 첫 번째를 제외한 나머지 레이어에서는 1x1 convolution을 수행하고 첫 번째 레이어에서는 7x7 convolution을 수행한다. 각 convolution layer는

spectral normalization과 batch normalization과 ReLU가 따라오며 마지막 layer에서는 이들 대신 sigmoid가 따라온다. Discriminator는 최종적으로 각 pixel 별로 해당 픽셀이 실제 이미지와 유사한지 가짜 이미지와 유사한지를 예측하는 구별 맵을 내보낸다.

## 2.2 학습 방법

본 논문에서 제안하는 학습 방법은 입력 이미지를 generator 네트워크를 통과하여 down-sample된 이미지에서 임의로 한 부분을 잘라낸 것과 원래의 입력 이미지에서 임의로 한 부분을 잘라낸 것이 같은 분포를 따르도록 학습을 진행한다. 참조 이미지 또한 같은 촬영 환경에서 찍었다는 가정이 있으므로 같은 down-sample 된 이미지와 같은 위치에서 generator 네트워크와 discriminator 네트워크가 서로 적대적인 관계에서 학습을 진행하는 GAN[17]의 adversarial training을 사용한다. Generator는 최대한 입력 이미지와 유사한 이미지를 만들어 discriminator를 속이게 학습을 진행한다. 이러한 방법 중에서도 LSGAN[19]을 활용하여 학습을 진행한다. 이에 대한 식은 다음과 같다.

$$\min_G \max_D E_{x \sim patches(I_{LR})} [ |D(x) - 1| + |D(G(x))| ]$$

본 논문에서는 위의 손실 함수 이외에도 학습이 잘 진행되기 위한 커널의 조건에 대한 4가지의 추가적인 손실함수를 부수적으로 사용하였다. 이에 대한 식은 다음과 같다. 이 때, 커널의 조건을 적용하기 위해서 커널은 generator에 delta 입력을 넣어준 결과로 도출할 수 있다.

$$\alpha L_{one} + \beta L_{boundary} + \gamma L_{sparse} + \delta L_{center}$$

$L_{one} = \left| 1 - \sum_{i,j} k_{i,j} \right|$ 은 커널의 전체의 합이 1이 되도록 하는

요소이다.  $L_{boundary} = \sum_{i,j} |k_{i,j} * m_{i,j}|$ 는 커널의 가장자리에 갈

수록 0에 가까워지도록 하는 요소이다.  $L_{sparse} = \sum_{i,j} |k_{i,j}|^{1/2}$ 는

커널이 과도한 스무딩을 하지 않도록 희소성을 유지 시키는 요소이다.

마지막으로  $L_{center} = \left\| (x_0, y_0) - \frac{\sum_{i,j} k_{i,j} * (i, j)}{\sum_{i,j} k_{i,j}} \right\|$ 는 커널

의 무게중심이 커널의 중심에 있도록 하는 요소이다.

## 3. 실험 결과

실험은 [14]에서 제안한 임의의 커널을 통해 합성된 저화질 이미지 데이터셋 DIV2K를 통해 진행되었다. DIV2K는 DIV2K 데이터셋의 validation set 100장을 임의로 생성된 커널로 blur 시킨 이후 down-sample을 통해 생성된 것이다. 이 때 임의로 생성된 커널은 11x11의 이방적인 가우시안으로  $\lambda_1, \lambda_2 \sim U(0.5, 5)$ 을 각 축별로 독립적으로 따르고,  $\theta \sim U(-\pi, \pi)$ 의 각도로 회전시켰다. 커널을 예측하는 과정에서 학습은 Adam optimizer를 활용하였고 learning rate

는 0.0002를 사용하였으며 750번의 iteration에 한번씩 0.1배 하였다. 이미지당 3000번의 iteration을 통해 학습을 진행하였고 NVIDIA GEFORCE 2080 TI에서 한 장의 이미지 당 약 1분 30초의 시간이 소요되었다.

제안한 방법을 이용한 학습을 통하여 표 1과 같은 성능을 얻을 수 있었다. 이 때, 다른 방법들에 대한 수치는 [14]를 참고하였다. 커널 예측을 통한 초해상도 성능을 측정하기 위해 종류1에서는 커널의 정보를 주지 않고 기존의 'bicubic' 커널이라고 가정한 상태에서의 초해상도 성능을 보여주었고, 종류 2, 3에서는 한 장의 이미지를 통한 커널 예측 방법들과 이렇게 예측된 커널 정보가 있을 때 초해상도를 진행할 수 있는 초해상도 방법을 함께 사용하여 성능을 측정하였다. 표에 나타난 바와 같이 기존의 'bicubic' 커널에서 학습되어 높은 성능을 보여주었던 방법들은 다양한 일반화 된 커널 데이터셋으로 테스트를 했을 때에는 단순히 bicubic으로 interpolation 한 이미지에서 0.5dB 정도의 성능 향상만 있음을 알 수 있다. 한 장의 이미지를 활용하여 커널을 예측하는 방법인 [14]와 [13]은 한 장의 이미지로도 비교적 정확하게 커널을 예측하여 해당 정보를 입력에 넣어주어 더욱 높은 성능을 보여줄 수 있다. 제안하는 방법인 두 장의 이미지를 활용하여 커널 예측을 하는 방법은 많은 정보량으로 더욱 정확하게 커널을 예측하여 [14]에 비해 약 0.17dB 가량의 PSNR 성능 향상이 있음을 알 수 있다. 이는 정답 커널을 넣어주어 같은 초해상도 방법을 사용한 상한에 좀 더 가까워진 결과라고 할 수 있다.

### 표 1 다른 모델들과의 성능 비교

제안하는 방법을 제외한 나머지 방법들의 수치는 [14]를 참고하였다. 기존의 'bicubic' kernel로 학습된 우수한 성능을 보였던 방법들도 다양한 커널의 상황에서는 낮은 성능을 보여준다. 제안하는 방법은 KernelGAN을 활용한 방법보다 약 0.17dB 높은 성능을 보여준다.

	방법	2배율(PSNR)
종류 1 'bicubic' kernel로 학습된 성능 좋은 기존 방법들	Bicubic	28.731
	Interpolation	28.731
	Bicubic kernel + ZSSR[18]	29.102
	EDSR [3] RCAN [5]	29.172 29.198
종류 2 한 장의 이미지를 활용하여 kernel estimation을 하는 방법들	KernelGAN[14] + SRMD[9]	29.565
	KernelGAN + ZSSR[9]	30.363
	Correctionfilter[13] + DBPN[20]	29.389
종류 3 두 장의 이미지를 활용하여 kernel estimation을 하는 방법	제안하는 방법 + ZSSR	30.536
종류 4 정답 커널을 활용한 방법 (상한)	정답 커널 + SRMD	31.962
	정답 커널 + ZSSR	32.436

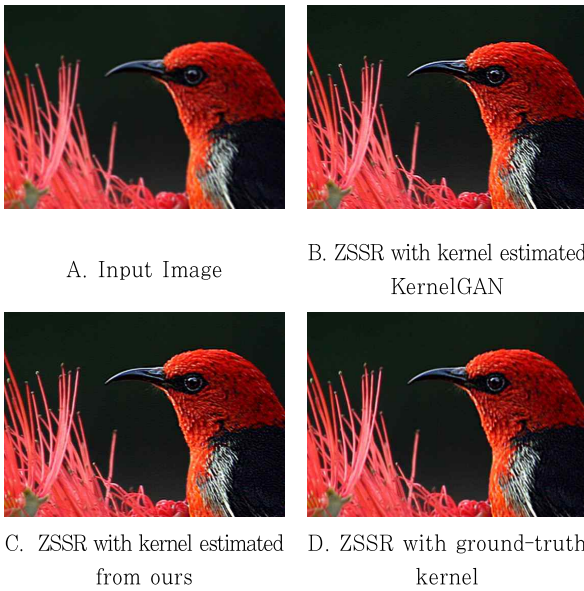


그림2 두 장의 이미지를 활용한 커널 예측에 대한 결과 이미지.

상좌측부터 A 입력 이미지, B KernelGAN을 활용하여 예측된 커널과 한 장의 이미지를 활용한 이미지 초해상도, C 제안된 방법을 활용하여 예측된 커널과 한 장의 이미지를 활용한 이미지 초해상도 D 정답 커널을 활용하여 한 장의 이미지를 활용한 이미지 초해상도

#### 4. 결론

본 논문에서는 두 장의 영상을 활용하여 이미지 초해상도 문제에서 정의되는 화질 저하 커널을 예측하는 방법을 제안하였다. 실험 결과를 통해 본 논문에서 제안하는 방식으로 두 장의 이미지를 활용하여 커널을 예측하였을 때 한 장의 이미지만을 사용한 것보다 커널을 더욱 정확하게 예측하여 실제 이미지에서의 이미지 초해상도 성능을 향상시켜 주었다. 본 연구는 두 장의 이미지를 활용하여 비교적 정확하게 커널을 예측하지만 한 장의 이미지마다 약 1분 30초 정도의 적지 않은 시간이 소요된다는 한계를 갖고 있어 추후 이를 보완하는 연구가 따라야 할 것으로 보인다.

#### 감사의 글

이 논문은 삼성전자의 지원과 2021년도 산업통상자원부와 한국산업기술진흥원의 “지역혁신클러스터육성사업(R&D, P0002072)”으로 수행된 연구결과입니다.

#### 참고 문헌

[1] Dong Chao, et al. "Learning a deep convolutional network for image super-resolution." *In ECCV 2014*  
 [2] Jiwon Kim et al. "Accurate image super-resolution using very deep convolutional networks." *In CVPR 2016*  
 [3] Bee Lim et al "Enhanced Deep Residual Networks for Single Image Super-Resolution". *In CVPR 2017*

#### Workshops.

[4] Yulun Zhang, et al. "Residual dense network for image super-resolution." *In CVPR 2018*  
 [5] Yulun Zhang, et al. "Image super-resolution using very deep residual channel attention networks." *In ECCV 2018*  
 [6] Namhyuk Ahn, et al. "Fast, Accurate, and Lightweight Super-Resolution with Cascading Residual Network" *In ECCV 2018*  
 [7] Chang Chen et. al. "Camera lens super-resolution," *In CVPR 2019.*  
 [8] Ruofan Zhou et. al. "Kernel Modeling Super-Resolution on Real Low-Resolution Images." *In ICCV 2019*  
 [9] Kai Zhang et. al. "Learning a single convolutional super-resolution network for multiple degradations." *In CVPR 2018.*  
 [10] Jinjin Gu et. al. "Blind super-resolution with iterative kernel correction." *In CVPR 2019.*  
 [11] Yu-Syuan Xu et. al. "Unified Dynamic Convolutional Network for Super-Resolution with Variational Degradations." *In CVPR 2020.*  
 [12] Kai Zhang et. al. "Deep Plug-and-Play Super-Resolution for Arbitrary Blur Kernels." *In CVPR 2019.*  
 [13] Shady Abu Hussein et. al. "Correction Filter for Single Image Super-Resolution: Robustifying Off-the-Shelf Deep Super-Resolvers" *In CVPR 2020.*  
 [14] Sefi Bell-Kligler et. al. "Blind Super-Resolution Kernel Estimation using an Internal-GAN." *In NeurIPS 2019.*  
 [15] Sanjeev Arora et. al. "On the optimization of deep networks: Implicit acceleration by overparameterization." *In ICML 2018.*  
 [16] Phillip Isola et. al. "Image-to-image translation with conditional adversarial networks." *In CVPR 2017.*  
 [17] Ian Goodfellow et. al. "Generative adversarial nets." *In NeurIPS 2014.*  
 [18] Assaf Shocher et. al. "Zero-shot super-resolution using deep internal learning." *In CVPR 2018.*  
 [19] Xudong Mao et. al. "Least squares generative adversarial networks." *In ICCV 2017.*  
 [20] Muhammad Haris et. al. "Deep back-projection networks for super-resolution." *In CVPR 2018.*