

## RDB 및 웨이블릿 예측 네트워크 기반 단일 영상을 위한 심층 학습기반 초해상도 기법

\*응우엔휴중 \*\*김응태

한국산업기술대학교

\*dungnh@kpu.ac.kr \*\*etkim@kpu.ac.kr

## Deep Learning-based SISR (Single Image Super Resolution) Method using RDB (Residual Dense Block) and Wavelet Prediction Network

\*Nguyen, Huu Dung \*\*Kim, Eung-Tae

Korea Polytechnic University

## 요약

단일 영상 초해상도 (Single Image Super-Resolution - SISR) 기법은 카메라로 획득된 저해상도 영상에 필터 기반의 연산을 적용하여 좋은 화질의 고해상도 영상을 복원하는 과정이다. 최근에 심층 합성곱 신경망 학습의 발전에 따라 단일 영상 초해상도에 적용되는 심층 학습 기법들은 좋은 성과를 보여 주고 있다. 본 논문은 단일 영상 초해상도 성능을 개선하기 위해 웨이블릿 예측 네트워크를 효율적으로 적용하는 방법에 대해 연구하였으며, 저해상도 입력 영상의 특징을 잘 추출해내기 위해 네트워크 내부에 RDB를 적용하여 기존 방식보다 효율적으로 고해상도 영상 복원하는 기법을 제안한다. 모의실험을 통해 제안하는 방법이 기존 방법보다 화질은 약 PSNR 0.18dB만큼 우수하며 속도는 1.17배 빠른 것을 확인하였다.

## 1. 서론

단일 영상 초해상도 (Single Image Super-Resolution - SISR)는 획득되는 저해상도 (Low Resolution - LR) 영상에 필터 기반의 연산을 적용하여 좋은 화질의 고해상도 (High Resolution - HR) 영상을 복원하는 기법이다. 고해상도 영상은 획득되는 저해상도 영상보다 화소 밀도 (Pixel Density)가 높고 더 많은 정보를 제공한다. 단일 영상 초해상도 기법은 디지털 TV, 모바일, 보안 카메라 등 다양한 분야에서 활용되고 있다. 게다가 최근 컴퓨터 성능 향상으로 인해 패턴인식 및 이미지 분석 기법, 영상 부호화/복호화 (Image Coding/Decoding), 영상 잡음 제거 (Image Denoising), 의료 영상 분석 등 다양한 기법들 초해상도 기법과 결합해서 성능을 향상시키는데 사용하고 있다.

단일 영상의 해상도를 높이는 기법은 보간법 (Interpolation) [1-4], 재구성법 (reconstruction) [6-8] 및 심층 학습법 (Deep Learning) [10,12,14,15] 3가지로 분류할 수 있다. 보간법 기반 초해상도는 해상도를 높이기 위해 기존 화소들 사이에 적절한 화소를 삽입하는 기술이며 간단하고 처리 속도가 비교적 빠르다는 장점이 있어 실시간 응용에 적합하다. 그러나 이러한 기법은 저주파수 대역필터 특성을 가지고 있으므로 계단 (Jagging) 현상 및 겹침 왜곡 (Aliasing), 파문 (Ringing) 현상에 의해 고주파수 성분이 효과 적으로 복원될 수 없다는 단점이 있다. 재구성법 기반 초해상도는 Multi-Scale Expected Patch Log Likelihood (EPLL) [7] 및 Adaptive High-Dimensional Non-Local Total Variation (AHNLTV) [8], Steering Kernel Regression (SKR)과 Nonlocal Means (NLM) [6] 등 다양한 기법들의 발달에 의해 많이 개선되어 보간법보다 계단 현상 및 겹침 왜곡, 파문 현상이 많이 감소되었다. 하지만, 재구성법 기반 초해상도 기법에 비해

상대적으로 복잡한 계산량을 가지는 단점이 있다.

최근에는 CNN (Convolution Neural Networks) 기술의 발달에 힘입어 단일 영상 초해상도 분야에서도 CNN 기술을 적용한 방법들이 좋은 성능을 보여 주고 있다. 최초로 초해상도 분야에 CNN을 적용하는 SRCNN [13]로부터 최근에 제안된 방법 [14,15,17,18]들은 화소 단위 평균 자승 오차 (Mean Squared Error) 손실 함수 기반으로 학습을 수행한다. 학습을 통해 출력 영상 화소 단위로 점차 실측정보 (ground-truth) 고해상도 영상에 근접되고 있다. 그러나 이러한 방법은 몽롱화(blur) 현상과 과도한 평탄화 현상에 의해 출력 영상의 일부 정보, 특히 고주파수 정보 손실이 많이 발생된다. 게다가 이러한 방법은 한정된 확대율 (2x 혹은 4x) 일 때만 잘 수행되고 큰 확대율 (8x 혹은 16x) 일 경우에는 성능이 급격하게 저하된다.

웨이블릿 변환 (Wavelet transform)은 다중 해상도 (multi-resolution) 영상을 표현 및 복원할 때 효율적인 기법이다 [9]. 웨이블릿 변환은 영상의 정보를 다중 레벨 (different levels)에서 표현할 수 있어 높은 확대율 초해상도를 수행 가능하고 고주파수 성분을 잘 나타내는 장점이 있어 학습 기반 초해상도 분야에 적용 되었다 [10,11]. 그러나 이러한 방법은 복잡한 계산량 때문에 학습 진행에 어려움이 있고 저해상도 영상의 특징 (feature)을 잘 찾아 내지 못 하는 단점이 있다.

이러한 문제를 해결하기 위해 단일 영상 초해상도 네트워크인 RDB-WaveletSRNet를 제안한다. 제안된 방식은 저해상도 영상의 특징 맵을 잘 추출해내기 위해 특징 추출 네트워크에 Residual Dense Block(RDB)기법을 적용한다. RDB의 성장률(growth rate) 계수를 적절히 조절하여 복잡한 계산량 문제를 개선하도록 하였다. 또한 영상의

특징 맵 (feature map)으로 부터 웨이블릿 영역의 계수를 학습하고 확대율에 맞게 다중 레벨의 웨이블릿 계수를 구해서 초해상도 영상을 구현하게 된다.

본 논문의 구성은 다음과 같다. 2장에서 단일 영상 초해상도에 관련한 기존 연구들을 설명한다. 3장에서 제안하는 RDB-WaveletSRNet 네트워크 구조를 설명한다. 4장에서는 모의실험을 통해 기존의 기법들과 성능을 비교한다. 5장에서 결론은 맺는다.

## 2. 기존 초해상도 관련 연구

초해상도 분야와 관련하여 많은 웨이블릿 기반 기법들이 제안되었다. 그러나 저해상도 시퀀스 영상으로부터 초해상도 영상을 추정하거나 단일 영상 초해상도에서는 보간법과 재구성법 기반 초해상도를 보조해 주는 데만 웨이블릿이 적용된 기법이 대부분이다. Gao et al. [5]은 hybrid wavelet convolution network를 제안했다. 이 기법은 웨이블릿을 활용하여 sparse code 후보자 (candidates)를 생성하고 이를 CNN을 통해 고해상도 영상을 추출해내는 기법이다.

Huang et al. [10]은 그림 1에서 보는 바와 같이 저해상도 영상의 특징 맵으로부터 심층 학습으로 웨이블릿 계수를 학습시켜 고해상도 영상을 추출하는 WaveletSRNet 네트워크를 제안했다. 높은 확대율 (8x와 16x)에서도 다른 기법과 비교할 때 상대적으로 좋은 성능을 보여 주었다.

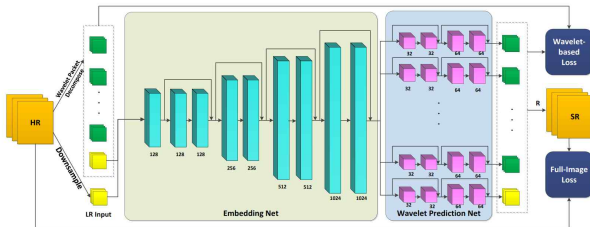


그림1. WaveletSRNet 네트워크 구조

WaveletSRNet에서는 저해상도 영상의 특징 맵을 추출할 때 제안된 Embedding Net을 사용한다. Embedding Net은 residual block [19]기반 심층 학습 네트워크이다. 다양한 분야에서 성능을 개선하기 위해 더 많은 계층의 심층 학습을 이용하는 경향이 있다. 그러나 전통적으로 많이 쓰는 활성화 함수의 기울기(gradient) 범위는 0에서 1까지이며 네트워크가 깊어질수록 기울기가 0으로 수렴하게 되는 소실 (vanishing) 문제가 발생한다. Residual block은 shortcut connection과 element-wise addition 기법으로 이 소실 문제를 잘 해결하였다. ILSVRC 2015에서, 3.57%의 가장 낮은 top-5 오차율 가진 ResNet[19]은 residual block의 좋은 성능을 보여 주었다. 그러나 residual block은 마지막 계층에만 과거 특징 정보만 도입되고 다른 계층에는 과거 특징 정보가 활용되지 못 하는 단점이 있다. 게다가 콘볼루션의 크기가 증가함에 따라 residual block의 계산량은 급격하게 증가되는 문제가 있다.

이 문제를 해결하기 위해 Zhang et al. [12]은 residual dense block (RDB) 제안했다. RDB은 residual block의 장점을 여전히 이어 받고 dense connection [20]로 각 계층은 그 전의 모든 계층의 정보를 활용하여 영상의 특징 맵을 추출하는데 적합한 기법이다.

## 3. 제안된 RDB-WaveletSRNet기반 초고해상도 기법

본 논문은 그림1의 WaveletSRNet [10]의 성능을 개선하기 위해 RDB를 이용하는 WaveletSRNet을 제안한다. 그림 2는 RDB-WaveletSRNet의 구조를 보여준다.

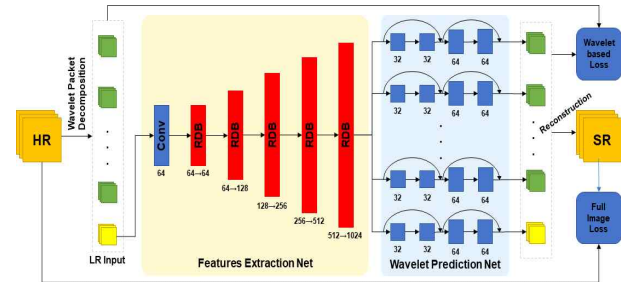


그림 2. 제안된 RDB-WaveletSRNet 네트워크의 구조

RDB-WaveletSRNet은 주어진 원본 고해상도 영상 (High Resolution)을 Wavelet Packet Decomposition을 통해 확대율에 맞는 웨이블릿 레벨의 웨이블릿 계수들과 저해상도 영상으로 분해한다. 저해상도 영상은 Feature Extraction Net의 입력이 되고 이 네트워크를 통해 저해상도 영상의 특징 맵을 추출하게 된다. Wavelet Prediction Net은 저해상도 영상의 특징 맵과 분해된 웨이블릿 계수기반 고해상도의 웨이블릿 계수를 학습시킨다. 복원단계에서 학습을 통해 얻어진 웨이블릿 계수를 저해상도 영상의 초해상도 영상으로 복원하는 단계이다.

### 3.1. Wavelet Packet Decomposition

제안된 방식에서 2D fast wavelet transform (FWT) [16]을 사용하여 입력 고해상도 영상의  $N_w$ 개의 wavelet 계수를 분해한다. 웨이블릿 변환은 적용되는 웨이블릿 원형함수 (wavelet function/ wavelet mother function)로 분류된다. 따라서 그 중에 가장 간단하고 계산량이 낮으며 에지 (edge) 성분에 아무 영향도 없이 완벽히 복원할 수 있는 특징을 가지는 Haar 웨이블릿을 선택하였다.

제안된 방식에서 확대율  $r=8$ 인 경우,  $l=\log_2(r)=3$ 개의 FWT 레벨로  $N_w$ 개의 웨이블릿 계수를 분해할 수 있다. 여기서,  $N_w=4^l=64$ 이다. 이전 웨이블릿 계수들을 그림 3에서 나타내는 FWT 3레벨까지 반복하여 통과시킨 후  $N_w$ 개의 웨이블릿 계수를 계산한다.

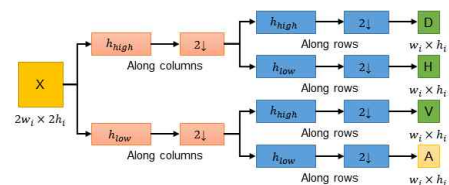


그림 3. i번째 레벨의 FWT

### 3.2. Features Extraction Net (FEN)

Wavelet Packet Decomposition을 통해서 얻어진 웨이블릿 계수들 중에 가장 낮은 주파수 성분 가진 웨이블릿 계수는 바로 주어진 고해상도 영상의 확대율  $r$ 만큼 down scale된 저해상도 영상이다. FEN을 통해 이 저해상도 영상의 features set을 추출한다.  $3 \times w \times h$  크기의 입력영상을 FEN을 통해  $1024 \times w \times h$  크기를 가지는 특징 맵을 추정한다.

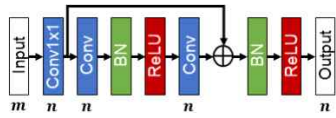


그림 4. Residual Block 구조

FEN은 서로 크기 다른 병렬로 연결된 5개의 RDB로 구성되어 있다. 앞서 언급한 바와 같이 RDB는 모든 과거 관측값을 사용하여 완벽 연결 구조로 그림 4에서 나타내는 Residual Block의 단점을 개선하는 기법이다. 그림 5는 본 논문에서 사용하는 RDB 구조이다. 각 RDB를 통해 저해상도 영상의 feature 개수를 확대시키므로 RDB의 입력 채널 개수  $m$ 과 출력 채널 개수  $n$ 은 서로 다르다. RDB의 addition 연산을 하기 위해  $m$ 개 채널 입력을  $n$ 개로 확대해 주는 expand convolution(Conv\_ex)을 사용하는 것을 제안한다.  $g$ 는 RDB의 성장률(growth rate)이며 residual block의 복잡한 계산량 문제를 해결하기 위해 적절한  $g = 32$ 를 사용한다.

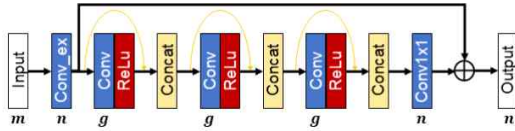


그림 5. 제안된 방식에서 사용되는 개선된 RDB 구조

Residual block (RB) 구조로부터 RB의 파라미터 수를 계산하는 식은 다음과 같다.

$$Params_{RB} = mn + 2nkn_xk_y \quad (1)$$

본 논문에서 사용하는 RDB의 파라미터 수를 계산하는 식은

$$Params_{RDB} = mnk_xk_y + (Dg + n)n + \sum_{i=0}^{D-1} (n + ig)k_xk_y \quad (2)$$

이다. 여기서,  $k_x$ 과  $k_y$ 는 convolution (Conv)의 커널 크기이며 RB와 RDB는 동일하게 3x3 커널을 쓴다.  $D$ 는 RDB의 dense 계층 수이며, 본 논문에서  $D = 3$ 을 적용한다. 표 1은 식 (1)과 (2)로부터 채널 크기에 따라 RB와 RDB의 파라미터수를 나타낸다. 표 1을 통해 낮은 계산량으로 RDB는 폭이 넓고 깊은 학습에 적합함을 보여주었다.

표 1. 입/출력 채널 수에 따라 RB와 RDB의 파라미터 수

입력→출력 채널 수 ( $m \rightarrow n$ )	Number of Parameter	
	Residual Block	Residual Dense Block
3→64	39,040	94,912
64→128	229,888	240,640
128→256	918,528	633,856
256→512	3,672,064	1,960,960
512→1024	14,684,160	6,777,856
Total	19,543,680	9,708,224

### 3.3. Wavelet Prediction Network

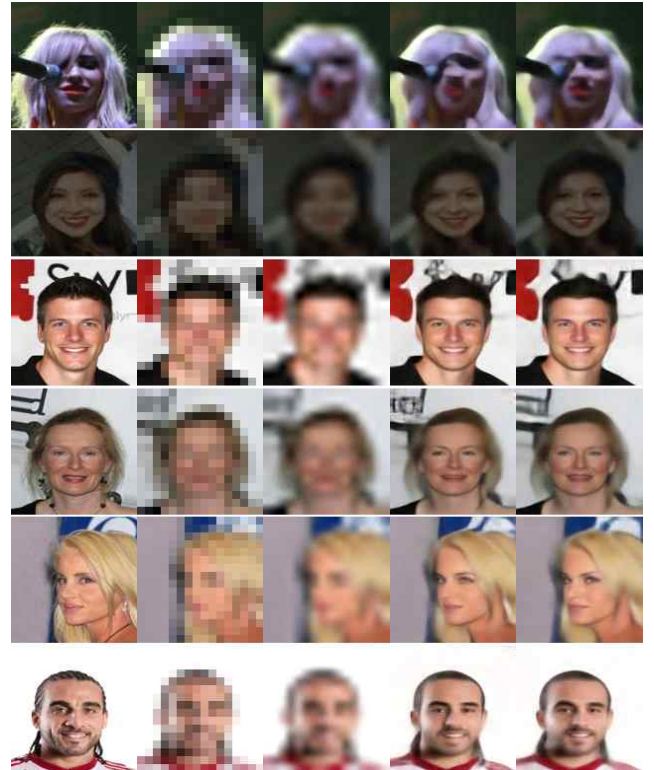
복구할 초해상도 영상의  $N_w$ 개의 웨이블릿 계수를  $N_w$ 개의 독립적인 subnet으로 Wavelet Prediction Network (WPN)으로 학습한다. 각 subnet은 2개의 residual block으로 구성되며  $1024 \times w \times h$  feature map 입력으로 해당되는 웨이블릿 계수를 생성한다. 저해상도 입력 영상과 모든  $N_w$ 개의 웨이블릿 계수의 크기는 동일하게  $3 \times w \times h$ 로 유지하기 위해 본 논문에서 WPN에 3x3 커널, stride 1,

padding 1 가지는 convolution을 사용한다. Haar 웨이블릿의 웨이블릿 계수들은 서로 독립적이기 때문에  $N_w$ 개의 subnet을 병렬로 독립적으로 설계한다.

### 3.4. Reconstruction

Reconstruction은 단순히 inverse-FWT를 사용하여  $N_w$ 개의  $3 \times w \times h$  웨이블릿 계수를 원래  $3 \times (r \times w) \times (r \times h)$  영상으로 복원하는 단계이다.

## 4. 실험 결과 및 고찰



(a) HR (b) LR (c) Bicubic (d) [10] (e) Ours

그림 6.  $16 \times 16$  입력 저해상도, 확대율  $r = 8$  경우

기존 방식들과 비교

모의실험에서 CelebA [21] 데이터셋을 기반으로 모의실험하였다. CelebA 데이터셋의 202,599개의 영상 중에 162,700개 학습 영상과 19,867개 validation영상을 사용했다. 150개의 epoch (905k iterations)을 학습한 후 19,962개의 영상으로 bicubic 보간법, WaveletSRNet[10]과 PSNR 및 SSIM을 비교하여 제안된 RDB-WaveletSRNet의 성능을 검토하였다. 그림 6에서 보는 바와 같이 CelebA 데이터셋의 얼굴 가운데 정렬된  $178 \times 218$ 크기인 데이터셋으로부터  $128 \times 128$ 의 고해상도 영상을 얻었다. 그림 6에서 (b)  $16 \times 16$  입력 저해상도를 확대율  $r = 8$  경우 (c) bicubic 보간법, (d) WaveletSRNet [10]과 제안된 RDB-WaveletSRNet의 결과를 보여 준다.

그림 7은 150개의 epoch (905k iterations)을 통해 1000 iterations마다 WaveletSRNet과 제안된 RDB-WaveletSRNet의 PSNR을 측정 한 학습 결과이다. 200k iteration부터 WaveletSRNet은 overfitting되는 반면에 제안하는 RDB-WaveletSRNet은 계속 잘 학습되는 것을 볼 수 있다.

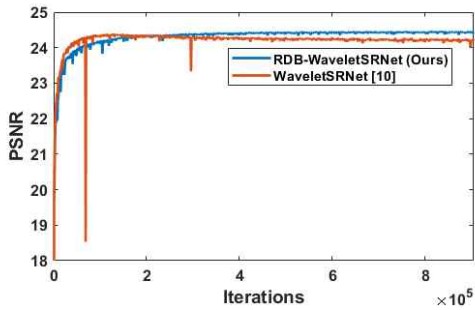


그림 7. 학습 반복에 따른 PSNR 변환

표 2. CelebA 데이터 세트 기반 PSNR, SSIM, 파라미터 수와 학습시간 비교

Method	PSNR	SSIM	Number of Parameter	Training Time (hours)
Bicubic	21.6975	0.7787	-	-
WaveletSRNet	24.6090	0.8496	50M	96
Ours	<b>24.7903</b>	<b>0.8521</b>	<b>40M</b>	<b>82</b>

표 2에서 19,962개의 테스트 영상을 대상으로 bicubic 보간법과 WaveletSRNet, RDB-WaveletSRNet을 통해 측정된 PSNR, SSIM, 파라미터 수와 수행시간을 보여준다. 파라미터 수는 식 (1)을 사용하여 WPN 네트워크의 파라미터 수를 계산한 후 3.2절에 계산된 결과와 더하여 얻어진 WaveletSRNet과 RDB-WaveletSRNet의 전체 파라미터 수량이다. GTX 1070 GPU를 사용하여 학습을 진행한 결과, 기존 WaveletSRNet보다 PSNR이나 시간측면에서 개선된 효과를 얻을 수 있었다.

### 5. 결론

본 논문에서는 RDB와 웨이블릿 예측 네트워크를 사용하여 단일 영상 초해상도에 적용되는 RDB-WaveletSRNet 기법을 제안하였다. RDB의 장점을 활용하여 저해상도의 특징 맵을 잘 추출해내게 구현함에 따라서 단일 영상 초해상도 성능이 좋아지고, 적절한 성장률 계수를 설정하여 복잡한 계산량 문제까지 해결할 수 있었다. 제안하는 방법은 기존 방법보다 화질은 PSNR 0.18dB만큼 우수하며 속도는 1.17배 빠른 것을 모의실험을 통해 확인하였다. 향후 RDB-WaveletSRNet 기법을 다양한 응용분야에 적용하여 좋은 화질의 초해상도 영상을 얻으리라 기대된다.

### References

[1] X. Zhang and X. Wu, "Image Interpolation by Adaptive 2-D Autoregressive Modeling and Soft-Decision Estimation," in IEEE Trans. Image Process., 2008.

[2] Y. Romano, M. Protter and M. Elad, "Single Image Interpolation Via Adaptive Nonlocal Sparsity-Based Modeling," in IEEE Trans. Image Process., 2014.

[3] W. Ye and K. Ma, "Convolutional Edge Diffusion for Fast Contrast-guided Image Interpolation," in IEEE Signal Process. Lett., 2016.

[4] C.Ji, X.He, Y.Fu, Z.Liang and L.Qing, "An edge directed interpolation algorithm based on regularization," in J.Electron.

Inf. Technol., 2014.

[5] X. Gao and H. Xiong, "A hybrid wavelet convolution network with sparse-coding for image super-resolution," in ICIP, 2016.

[6] K. Zhang, X. Gao, D. Tao and X. Li, "Single Image Super-Resolution With Non-Local Means and Steering Kernel Regression," in IEEE Trans. Image Process., 2012.

[7] V. Papyan and M. Elad, "Multi-Scale Patch-Based Image Restoration," in IEEE Trans. Image Process., 2016.

[8] C. Ren, X. He and T. Q. Nguyen, "Single Image Super-Resolution via Adaptive High-Dimensional Non-Local Total Variation and Adaptive Geometric Feature," in IEEE Trans. Image Process., 2017.

[9] S. Mallat, "Wavelets for a vision," in Proc. IEEE, 1996.

[10] H. Huang, R. He, Z. Sun and T. Tan, "Wavelet-SRNet: A Wavelet-Based CNN for Multi-Scale Face Super Resolution," in ICCV, 2017.

[11] Y. Matsuo and S. Sakaida, "Super-resolution for 2K/8K television using wavelet-based image registration," in IEEE GlobalSIP, 2017.

[12] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in CVPR, 2018.

[13] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," in ECCV, 2014

[14] J. Bruna, P. Sprechmann, and Y. LeCun, "Super-Resolution with Deep Convolutional Sufficient Statistics," in ICLR, 2016

[15] C. Dong, C. C. Loy, K. He and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," in TPAMI, 2015.

[16] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," in TPAMI, 1989.

[17] H. Chang, D.-Y. Yeung and Y. Xiong, "Super-resolution through neighbor embedding," in CVPR, 2004.

[18] A. Singh, F. Porikli and N. Ahuja, "Super-resolving Noisy Images," in CVPR, 2014.

[19] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in CVPR, 2015.

[20] T. Tong, G. Li, X. Liu and Q. Gao, "Image Super-Resolution Using Dense Skip Connections," in ICCV, 2017.

[21] Z. Liu, P. Luo, X. Wang and X. Tang, "Deep Learning Face Attributes in the Wild," in ICCV, 2015.