

깊이별 분리 합성곱을 활용한 초해상도 모델의 경량화

김 대 희*, 강 주 희*, 이 재 구°

Lightweighting of Super-Resolution Model
Using Depth-Wise Separable Convolution

Daehee Kim*, Juhee Kang*, Jaekoo Lee°

요 약

초해상도는 저화질의 영상을 고화질의 영상으로 변환하는 과업으로, 무거운 연산을 활용하여 성능을 향상하는 연구에 비교적 집중되어 있다. 우리는 깊은 합성곱 신경망 기반 초해상도 모델을 다른 기술과 함께 사용하거나 모바일 환경에서 사용할 때, 모델의 경량화가 필요함을 인지하였다. 이에 따라 본 논문에서는 깊이별 분리 합성곱 (Depth-wise Separable Convolution)을 활용하여, 최신 초해상도 모델인 MSRN(Multi-Scale Residual Network)을 경량화한 SMSRN(Separable Convolution Based Multi-Scale Residual Network)의 구조를 제안한다. SMSRN의 매개변수는 MSRN의 14.64% 수준으로 감량되었다. 반면에 다양한 벤치마크 데이터 집합에 대한 정량적 실험을 진행한 결과로 성능은 98.53%를 유지함을 보였으며, 정성적 실험 결과로 성능 저하를 확인하기 힘든 결과를 보였다. SMSRN은 다양한 합성곱 필터 크기를 사용한 구조이므로, 다양한 구조의 초해상도 모델에도 적용하여 경량화할 수 있을 것으로 판단된다.

Key Words : Depth-wise Separable Convolution, Super-Resolution, Lightweighting, Deep Learning, Convolutional Neural Network

ABSTRACT

Super-resolution is the process of upscaling low-resolution images into high-resolution images and is relatively focused on research that improves system performance by utilizing heavy computation. We recognized that deep convolutional neural network based super-resolution models need to be lightweighted when they are used with other technologies or in mobile environments. In this paper, we propose a separable convolution-based multi-scale residual network(SMSRN) structure using depth-wise separable convolution, which is a lightweight version of the latest super-resolution model, that is, the multi-scale residual network (MSRN). Compare to the MSRN, the number of parameters in SMSRN is 14.64% of that in MSRN. On the other hand, quantitative experiments on a variety of benchmark datasets resulted in the remaining performance of 98.53%, and qualitative experiments showed that performance degradation was difficult to identify. Since the SMSRN is a structure that uses a variety of convolution filter sizes, it is expected that it can also be applied to various architectures of super-resolution models to make them lighter.

※ 본 연구는 2021년도 과학기술정보통신부의 재원으로 정보통신기획평가원의 지원을 받아 수행되었습니다. (No.2020-0-01826, AI 기반 선도적 실전문제해결 연구인재 양성)

• First Author : College of Computer Science, Kookmin University, frshield2@kookmin.ac.kr, 학생회원

° Corresponding Author : College of Computer Science, Kookmin University, jaekoo@kookmin.ac.kr, 정회원

* College of Computer Science, Kookmin University, kangjuhee18@kookmin.ac.kr

논문번호 : 202012-320-A-RN, Received December 22, 2020; Revised February 9, 2021; Accepted February 9, 2021

I. 서론

저화질의 영상을 고화질로 변환하는 초해상도(Super-Resolution) 과업은 깊은 합성곱 신경망(Deep Convolutional Neural Network)을 활용하여 급진적으로 발전하고 있다. 초해상도는 컴퓨터 비전(Computer Vision)의 필수적인 기술로서 광범위하게 연구되고 있는데, 현재 초해상도 연구는 무거운 연산을 활용하여 성능을 높이는 것에 비교적 집중되어 있다.

우리는 이전 연구에서 초해상도를 활용하여 깊은 신경망 기반 객체 인식 기술(Object Detection)의 근본적인 한계를 극복하고자 하였다¹⁾. 그러나 연구 과정에서 초해상도 모델을 객체 인식과 함께 사용하기 위해서는 실질적으로 모델의 경량화가 필요함을 인지하였다. 더불어 최근에는 고성능의 모바일 기기가 보편화되어 있으므로, 우리는 모바일 환경에서도 사용할 수 있도록 연산이 가벼우면서도 성능이 양호한 깊은 합성곱 신경망 기반 초해상도 모델 구조를 제안하고자 한다.

영상 분류 과업을 목적으로 구성된 MobileNet²⁾의 구조에서 깊이별 분리 합성곱(Depth-wise Separable Convolution)이 제안된 이후, CNN은 획기적인 모델

경량화가 가능하게 되었다. 우리는 이를 활용하여 깊은 합성곱 신경망 기반의 최신 초해상도 모델인 MSRN(Multi-Scale Residual Network, 다중 스케일 잔여 네트워크)³⁾을 경량화하였다. MSRN에 깊이별 분리 합성곱을 적용하여 SMSRN(Separable Convolution Based Multi-Scale Residual Network) 모델을 [그림 3]과 같이 작성하였으며, 기존의 MSRN처럼 다양한 합성곱 필터 크기를 사용하면서 성공적으로 경량화하였다.

[그림 1]은 Set5⁴⁾ 데이터 집합의 4배 초해상도 결과로 계산한 PSNR(Peak Signal-to-Noise Ratio, 최대 신호 대 잡음비) 성능지표와 매개변수의 수를 기준으로 사용하여, SOTA(State Of The Art) 초해상도 모델들과 SMSRN을 시각적으로 비교한 그래프이다. [그림 1]에서 볼 수 있듯이 SMSRN은 매개변수의 수가 MSRN의 14.64% 수준으로, 비약적인 감량에 성공하였다. 또한 다양한 벤치마크 데이터 집합을 기반으로 정량적인 실험을 진행하여 PSNR 성능지표를 비교하였을 때, SMSRN은 평균적으로 MSRN 대비 98.53%의 성능을 유지하였다.

II. 관련 연구

2.1 MSRN(Multi-Scale Residual Network)

MSRN의 전체적인 모델 구조는 [그림 3]에서 SMSRN의 구조인 (a)와 같다. 구체적으로는, 합성곱 필터 크기를 3×3과 5×5로 2가지를 사용하여, 수용장(Receptive Field)의 크기를 다양하게 적용하여 병렬적으로 쌓은 구조이다. 이러한 구조는 다양한 문맥적 정보(Contextual Information)을 취할 수 있다는 것이 장점이고, 많이 사용되는 3×3 필터보다 큰 사이즈의 필터를 사용한다. 이러한 문맥적 정보를 잘 활용하기 위해, [그림 3]에서 보는 것처럼, 각 블록에서 출력되는 특징 맵(Feature Map)을 접합(Concatenation)하고 1×1 합성곱으로 차원을 축소하여 활용하는 구조

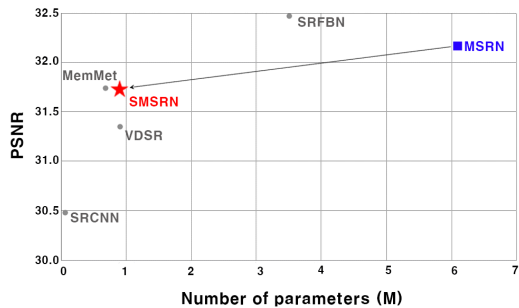


그림 1. SMSRN과 SOTA 초해상도 모델들의 매개변수와 PSNR 비교
Fig. 1. Comparison of parameters and PSNR for SMSRN and SOTA Super Resolution Models

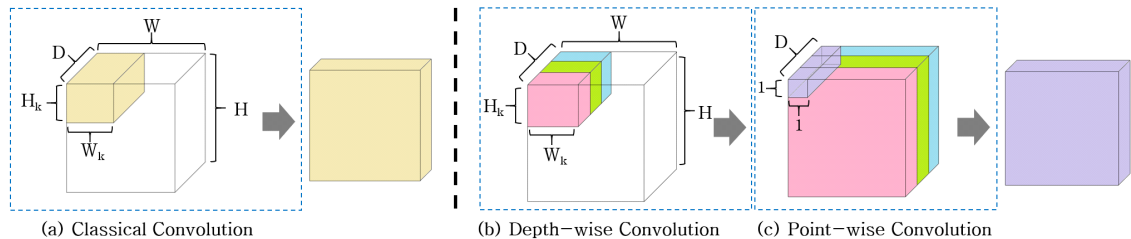


그림 2. 표준 합성곱 연산을 깊이별 합성곱과 점별 합성곱으로 분리하여 사용하는 과정
Fig. 2. The process of separating classical convolution into depth-wise and point-wise convolution

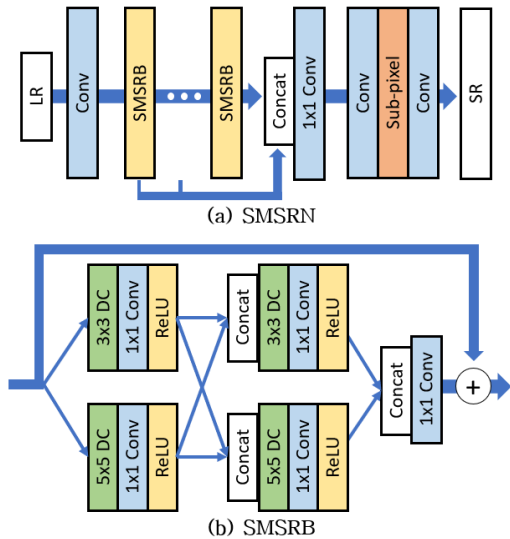


그림 3. SMSRN 모델 구조
Fig. 3. Architecture of SMSRN

를 가졌다. 또한, 큰 사이즈 필터를 사용하기 때문에 매개변수의 개수가 많아진다는 것과, 다양한 크기의 합성곱 필터를 병렬적으로 적용하므로 모델이 보다 무거워질 수 있는 점이 단점이다. 본 연구에서는 이러한 환경에서 깊이별 분리 합성곱을 활용하여 초해상도 모델을 경량화할 수 있는지 실험하였다.

기존의 MSRN은 YCbCr 색상 공간의 휘도 (Luminance) 채널만을 사용하여 학습하고, 추론하였다³⁾. 하지만 최근 초해상도 모델은 RGB 색상 공간의 3채널을 사용하는데, 이에 따라 MSRN을 RGB 3채널로 적용하니, 기존 MSRN보다 Set5 데이터 집합을 4 배 초해상화한 결과의 PSNR 지표 기준으로 0.3% 더 좋은 성능을 내었다.

2.2 깊이별 분리 합성곱(Depth-wise Separable Convolution)

[그림 2]는 MobileNet에서 제안된 깊이별 분리 합성곱을 묘사한 그림이다. 구체적으로, 기존의 표준 합성곱 (a)를 별개의 합성곱 연산인 (b)와 (c)로 분리하여 수행하는 과정이다. (b)는 깊이별 합성곱으로, 입력되는 특징 맵의 각 채널마다 독립적으로 합성곱을 수행하는 연산이며, (c)는 점별 합성곱(Point-wise Convolution)으로, 1×1 필터 크기의 합성곱 연산과 같은 과정이다. 이와 같이 (b)와 (c)를 연속하여 수행하면 (a)를 대체할 수 있으며, 약간의 성능 저하는 있지만 모델의 매개변수의 수를 혁신적으로 줄일 수 있다.

2.3 기존 초해상도 모델 경량화 연구

기존 초해상도 모델의 경량화 연구는 DRCN (Deeply Recursive Convolutional Network, 깊은 재귀 합성곱 신경망)⁵⁾, DRRN(Deep Recursive Residual Network, 깊은 재귀 잔여 신경망)⁶⁾ 등에서 이루어진 바 있다. 이 연구들은 동일한 합성곱 필터나 합성곱 블록 단위를 재귀적(Recursive)으로 사용하도록 구성되어 있다. 따라서 이러한 구조를 통해 매개변수의 수를 줄이는 것에는 성공하였다. 하지만 이는 초해상도 모델의 용량 자체는 경량화 할 수 있었으나, 결국 동일한 합성곱 필터를 반복적으로 사용하기 때문에, SMSRN과 달리 연산량 자체의 감소는 이루어지지 않았다.

III. SMSRN(Separable Convolution Based Multi-Scale Residual Network)

3.1 모델 구조

SMSRN은 MSRN을 바탕으로 깊이별 분리 합성곱을 적용하여 작성되었다. [그림 3]의 (b)와 같이 표준 합성곱 연산을 깊이별 합성곱과 점별 합성곱 연산으로 분리, 대체하여 SMSRB(Separable Multi-Scale Residual Block)를 구성하였다. [그림3]의 (b)에서 DC는 깊이별 합성곱을 의미한다. [그림3]의 (b)에서 앞단의 3×3과 5×5 합성곱 층에서는 각각 64개의 채널을 출력하며, 뒷단에서는 이를 접합하여 128개 채널의 특징 맵 연산하고 출력한다. 그리고 마지막 단에서는 총 256개 채널의 특징 맵을 접합하고, 1×1 합성곱을 통해 채널 수를 64개로 줄인 후, [그림 3]의 (b)에서의 입력값과 더하여 출력한다.

3.2 깊이별 분리 합성곱을 활용한 연산량 감소

[그림 2]에서 D, W, H 는 각각 특징 맵의 깊이(채널 수), 너비, 높이를 의미하며, W_k, H_k 는 각각 합성곱의 필터 너비, 높이를 뜻한다. 연산량을 계산하면, (a)의 연산은 [수식1], (b)는 [수식2], (c)는 [수식3]과 같고, (b)와 (c)를 연속하여 수행하면 [수식4]와 같다. 이 때, N 은 합성곱 연산 후 출력할 채널 수이다. 합성곱 연산 후 출력되는 특징 맵의 크기는 두 연산에서 같다고 가정하므로 생략한다. 또한 MSRN의 조건에 맞추어 W_k, H_k 의 값은 각각 같고, W, H 도 각각 같다고 가정한다.

$$C_{classical} = W_k^2 \times D \times N \quad (1)$$

$$C_{depth} = W_k^2 \times D \quad (2)$$

$$C_{point} = 1 \times 1 \times D \times N \quad (3)$$

$$C_{separable} = C_{depth} + C_{point} = D(W_k^2 + N) \quad (4)$$

이에 따라 표준 합성곱과 깊이별 분리 합성곱과 점별 분리 합성곱을 연속한 분리 합성곱의 연산량 비율을 계산하면 [수식5]와 같다.

$$\frac{C_{separable}}{C_{classical}} = \frac{D(W_k^2 + N)}{W_k^2 \times N \times D} = \frac{1}{N} + \frac{1}{W_k^2} \quad (5)$$

[수식5]를 확인하면, 깊이별 분리 합성곱을 적용하였을 때, N 과 W_k 값에 따라 감량되는 연산량의 비율이 정해진다.

MSRN의 합성곱은 3×3 , 5×5 필터 크기로 구성되었으므로 이러한 조건을 대입해보면, N 을 128, W_k 를 3인 경우 약 0.1189가 계산되므로 기존 대비 약 88% 연산량이 감소되는 것을 알 수 있다. 또다른 조건인 N 을 128, W_k 를 5로 가정하면 약 0.0478이 계산되고, 기존 대비 약 95% 연산량이 감소된다. 이에 따라 SMSRN에서는 MSRN에 비해 평균적으로 약 92%에 해당하는 막대한 연산량이 감소하는 것을 알 수 있다. 뿐만 아니라, [표1]에서 확인할 수 있듯이 SMSRN의 매개변수 자체도 MSRN의 14.63% 수준으로 감소되었다.

IV. 실험

4.1 상세 구현

우리는 SMSRN과 MSRN을 비교하기 위해 같은 과정으로 학습하였다. 영상 복원 과업을 위한 고품질 영상 데이터 집합인 DIV2K^[7]을 기반으로 학습하였으며, 학습 과정에서 스케일링(Scaling), 회전(Rotation), 반전(Flipping) 방법을 활용하여 데이터 확대를 수행하였다. 목적 함수로는 평균 절대 오차(Mean Absolute Error)를 사용하고, 초기 학습률은 0.0001로 하여 Adam^[8] 최적화 기법을 사용하여 10개의 미니배치(Mini Batch) 단위를 적용한 500세대(Epochs)를 학습하였다. 초해상도 특성 상, 모델은 영상을 패치(Patch)로 조각내어 학습하는데, 이 패치 사이즈는

256으로 설정하였다.

SMSRN의 성능을 MSRN과 정량적으로 비교하기 위해 Set5, Set14^[9], B100^[10], Urban100^[11] 데이터 집합을 사용하여 실험하였다. [표 1]에서는 각 데이터 집합을 쌍입방 보간법(Bicubic Interpolation)으로 0.25배 다운샘플링(Down-sampling)하고, 그 영상을 각각의 모델로 4배 초해상화한 결과를 PSNR과 SSIM(Structural Similarity Index Measure, 구조 유사도 지표) 정량적 평가 지표로 기록하였으며, 각 모델의 매개변수 수도 포함하였다. PSNR은 영상의 원본과 픽셀 단위로 비교하여 비슷할수록 수치가 높아지는 성능지표이며, SSIM은 원본 영상과 비교 영상의 구조적 유사성을 측정한 지표이다.

표 1. SMSRN과 MSRN[3]의 정량적 성능 비교 (PSNR / SSIM)

Table 1. Quantitative performance comparison between SMSRN and MSRN[3] (PSNR / SSIM)

Dataset	MSRN[3]	SMSRN
Set5[4]	32.16 / 0.8642	31.73 / 0.8558
Set14[9]	28.63 / 0.7460	28.30 / 0.7354
B100[10]	27.61 / 0.7139	27.39 / 0.7051
Urban100[11]	26.16 / 0.7679	25.45 / 0.7423
Average	28.64 / 0.7730	28.22 / 0.7597 (98.5%)/(98.3%)
Number of parameters	6,078,019	888,929 (14.6%)

4.2 실험 결과

4개의 데이터 집합을 실험한 결과, PSNR을 평균하였을 때 MSRN은 28.64, SMSRN은 28.22로 성능의 98.53%를 유지하였다. SSIM으로 비교하면 각각 0.7730, 0.7597로 98.27%를 보존하였다. 그와 반면에 매개변수의 수는 MSRN의 14.63%까지 감소시킬 수 있었다. [그림 4-7]에서 MSRN과 SMSRN의 초해상화 결과를 확인할 수 있듯이, 육안으로 확인하기 힘들 정도로 초해상도 성능을 유지하는 것을 볼 수 있다.

[표 1]의 정량적 비교와 [그림 4-7]의 정성적 비교에서 확인할 수 있듯이, 본 연구에서는 성공적으로 초해상도 모델을 경량화하였다. 특히, 깊이별 분리 합성곱을 활용하여 매개변수의 수와 함께 연산량도 감량하였다. 다만, 경량화된 모델의 성능은 기존 모델의 성능에 의존적이므로, 요구 조건에 맞게 경량화할 모델을 적절하게 선택하여야 할 것으로 예상된다.



그림 4. Set5[4] 데이터 집합의 baby 영상을 4배 초해상화한 결과(왼쪽부터 원본, 쌍입방 보간법, MSRN[3], SMSRN)
 Fig. 4. The 4-times super-resolved result of baby image of the set5[4] dataset (from left to original, bicubic interpolation, MSRN[3], SMSRN)



그림 5. Set14[9] 데이터 집합의 comic 영상을 4배 초해상화한 결과(왼쪽부터 원본, 쌍입방 보간법, MSRN[3], SMSRN)
 Fig. 5. The 4-times super-resolved result of comic image of the set14[9] dataset (from left to original, bicubic interpolation, MSRN[3], SMSRN)

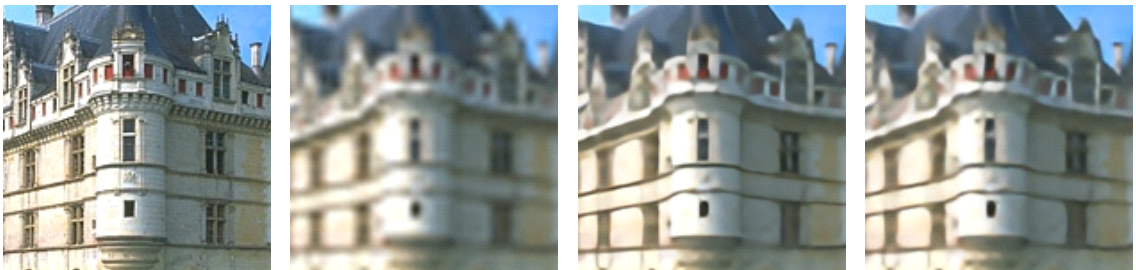


그림 6. B100[10] 데이터 집합의 102061 영상을 4배 초해상화한 결과(왼쪽부터 원본, 쌍입방 보간법, MSRN[3], SMSRN)
 Fig. 6. The 4-times super-resolved result of 102061 image of the B100[10] dataset (from left to original, bicubic interpolation, MSRN[3], SMSRN)

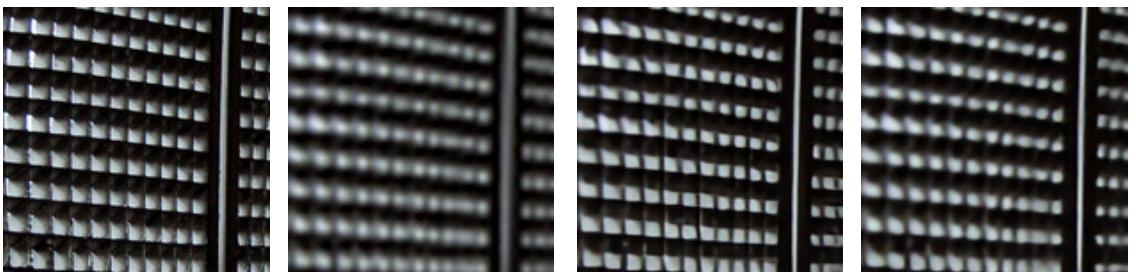


그림 7. Urban100[11] 데이터 집합의 img004 영상을 4배 초해상화한 결과(왼쪽부터 원본, 쌍입방 보간법, MSRN[3], SMSRN)
 Fig. 7. The 4-times super-resolved result of img004 image of the Urban100[11] dataset (from left to original, bicubic interpolation, MSRN[3], SMSRN)

V. 결 론

본 연구에서는 깊은 합성곱 신경망 기반 최신 초해상도 모델인 MSRN에 깊이별 분리 합성곱을 적용하여 경량화하였으며, 그 결과로 SMSRN 모델을 제안하였다. SMSRN은 MSRN의 14.63% 수준의 매개변수로 구성되었으며, 이에 반해 PSNR 성능지표 기준으로 98.53%를 유지하였다. SMSRN은 다양한 합성곱 필터 크기를 사용하는 모델임에도, 깊이별 분리 합성곱을 통해 성공적으로 경량화하였다. 따라서 여러 다양한 구조의 초해상도 모델에도 깊이별 분리 합성곱을 적용하여 경량화할 수 있을 것으로 판단된다.

References

[1] D. Kim, Y. Kim, Y. Yoo, and J. Lee, "Image super-resolution based on deep neural network for improving performance of face detection," in *Proc. KICS Winter Conf.*, pp. 42-43, Kangwon, Korea, Feb. 2020.

[2] A. G. Howard, et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[3] J. Li, et al., "Multi-scale residual network for image super-resolution," in *Proc. Eur. Conf. Computer Vision*, pp. 517-532, Munich, Germany, Sep. 2018.

[4] M. Bevilacqua, et al., "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. British Mach. Vision Conf.*, Surrey, UK, Sep. 2012.

[5] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution" in *Proc. IEEE Conf. CVPR*, pp. 1637-1645, Nevada, USA, Jun. 2016.

[6] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. CVPR*, pp. 3147-3155, Hawaii, USA, Jul. 2017.

[7] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf.*

CVPR Wkshp, pp. 126-135, Hawaii, USA, Jul. 2017.

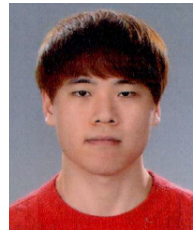
[8] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[9] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Int. Conf. Curves and Surfaces*, pp. 711-730, Avignon, France, Jun. 2010.

[10] D. Martin, et al., "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. Eighth IEEE Int. Conf. Computer Vision*, Vancouver, Canada, Jul. 2001.

[11] J.-B. Huang, et al., "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. CVPR*, pp. 5197-5206, Boston, USA, Jun. 2015.

김 대 희 (Daehee Kim)



2020년 2월 : 국민대학교 소프트웨어학부 졸업
 2020년 3월~현재 : 국민대학교 컴퓨터공학과 석사
 <관심분야> 인공지능, 초해상도 (Super-Resolution)
 [ORCID:0000-0001-9676-9604]

강 주 희 (Juhee Kang)



2018년 3월~현재 : 국민대학교 소프트웨어학부 학사
 <관심분야> 인공지능
 [ORCID:0000-0003-0520-1672]

이 재 구 (Jaekoo Lee)



2011년~2013년 : LG전자 CTO

부문 (주임연구원)

2018년 : 서울대학교 전기컴퓨

터공학부 박사

2018년 : SK 텔레콤 ICT기술원

매니저

2018년~현재 : 국민대학교 소프

트웨어학부 조교수

<관심분야> 인공지능, 기계학습, Deep Learning,
Data Science

[ORCID:0000-0002-5947-5487]