

## SRCNN과 VDSR의 구조와 방법 및 개선된 성능평가 함수

이광찬<sup>1</sup> · 왕광싱<sup>2</sup> · 신성윤<sup>3\*</sup>

### Structure, Method, and Improved Performance Evaluation Function of SRCNN and VDSR

Kwang-Chan Lee<sup>1</sup> · Guangxing Wang<sup>2</sup> · Seong-Yoon Shin<sup>3\*</sup>

<sup>1</sup>Graduate Student, School of Computer Information & Communication Engineering, Kunsan National University, Kunsan, 54150 Korea

<sup>2</sup>Associate Professor, Information Technology Center, Jiujiang University, Jiujiang, 332005 China

<sup>3\*</sup>Professor, School of Computer Information & Communication Engineering, Kunsan National University, Kunsan, 54150 Korea

#### 요약

이미지는 해상도가 높을수록 이미지를 시청하는 사람들의 만족도가 높아지며 초고해상도 이미지화는 컴퓨터 비전이나 영상처리 분야 중에서도 연구 가치가 꽤 높아지고 있다. 본 연구에서는 주로 딥 러닝 초 해상도 모델을 사용하여 저해상도 이미지 LR의 주요 특징을 추출한다. 추출된 특징을 학습 및 재구성하고, 고해상도 이미지 HR을 생성하는 재구성 기반 알고리즘에 중점을 둔다. 본 논문에서는 재구성에 기반을 둔 초 해상도 알고리즘 모델에서 SRCNN과 VDSR에 대하여 알아보도록 한다. SRCNN과 VDSR 모델의 구조 및 알고리즘 프로세스를 간략하게 소개하고 개선된 성능평가 함수에서도 다중 채널과 특수한 형태에 대하여 알아보도록 하며, 실험을 통하여 각 알고리즘의 성능을 이해하도록 한다. 실험에서는 SRCNN 및 VDSR 모델의 결과와 피크 신호 대 잡음 비 및 이미지 구조 유사도를 비교하는 실험을 수행하여 결과를 한눈에 볼 수 있도록 하였다.

#### ABSTRACT

The higher the resolution of the image, the higher the satisfaction of the viewers of the image, and the super-resolution imaging has a considerable increase in research value among the fields of computer vision and image processing. In this study, the main features of low-resolution image LR are extracted mainly using deep learning super-resolution models. It learns and reconstructs the extracted features, and focuses on reconstruction-based algorithms that generate high-resolution image HR. In this paper, we investigate SRCNN and VDSR in a super-resolution algorithm model based on reconstruction. The structure and algorithm process of the SRCNN and VDSR model are briefly introduced, and the multi-channel and special form are also examined in the improved performance evaluation function, and understand the performance of each algorithm through experiments. In the experiment, an experiment was performed to compare the results of the SRCNN and VDSR models with the peak signal-to-noise ratio and image structure similarity, so that the results can be easily judged.

**키워드** : 딥러닝, 저해상도, 초 해상도, SRCNN, VDSR

**Keywords** : Deep learning, Low-resolution, Super-resolution, SRCNN, VDSR

Received 17 February 2021, Revised 24 February 2021, Accepted 3 March 2021

\* Corresponding Author Seong-Yoon Shin(E-mail:s3397220@kunsan.ac.kr, Tel:+82-63-469-4860)

Professor, School of Computer Information & Communication Engineering, Kunsan National University, Kunsan, 54150 Korea

Open Access <http://doi.org/10.6109/jkiice.2021.25.4.543>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서론

하나의 이미지를 표현하는데 사용되는 기본 단위를 구성하는 것을 바로 해상도(Resolution)라 한다. 이미지의 해상도가 높으면 높을수록(픽셀의 숫자가 많으면 많을수록) 이미지의 세세한 부분까지 적나라하게 표현할 수 있다.

해상도를 이해하는데 가장 중요한 dpi, ppi에 대하여 알아보도록 하자. 흔히 말하는 dpi(Dot per inch)와 ppi(Pixel per inch)라는 단어가 바로 해상도를 말한다.

dpi(Dot per inch)는 가로축과 세로축 1인치가 몇 개의 Pixel(점)로 구성되어있는가를 표현하며 대개 프린터의 성능 등의 인쇄 및 출력물에 대한 해상도를 나타내는 단위이다.

ppi(Pixel per inch)는 1인치당 몇 개의 픽셀로 구성되어있는가를 나타내는 화면 디스플레이 장치의 밀도를 나타내는 해상도의 단위이다. 모니터에서 작업하는 도중에 나타나는 이미지의 해상도를 말하여 인쇄출력물에서는 쓰지 않는 말이다. 컴퓨터에서 작업하고 있는 이미지에 대하여 모니터 상에서 해상도를 거론할 때는 ppi라는 용어가 보다 더 많이 쓰이고 정확하다.

이미지 해상도라 한 개의 이미지를 만들어 내는 전체 픽셀의 개수를 뜻하며 이미지의 크기를 표현하는 단위로 사용된다. Pixel 단위를 이용하면 파일의 크기를 쉽게 알아낼 수 있다.

영상 초 해상도 구현 방법은 주로 재구성 기반 초 해상도 기법과 학습 기반 초 해상도 기법이 있다. 재구성 기반의 초 해상도 방법은 주로 주파수 영역 알고리즘과 공간 영역 알고리즘을 포함한다. 학습 기반 초 해상도 방법에는 주로 분류 기반 방법, 재구성 기반 방법, 그리고 회귀 기반 방법이 포함된다.

영상 초 해상도 구현 방법에 관한 논문은 많다. 그중에서 딥러닝 기반의 초 해상도 구현 알고리즘에 관한 연구의 최신 기술 동향을 소개하고 딥러닝 기반의 초 해상도 구현 알고리즘 중 가장 많이 사용하는 네트워크들을 분석하고 여러 가지 데이터 세트에 대하여 알맞은 네트워크의 실험 결과에 따라 딥러닝 기반의 초 해상도 구현 알고리즘의 우수성을 확인하는 연구[1]가 있었다.

SRCNN[1], VDSR[2], EDSR[3]은 각각 딥러닝과 잔여 학습방법(residual learning)을 활용하여 더욱더 놀라운 성능의 향상을 가져왔다. 그리고 residual dense block

을 제시한 RDN [4]은 모델의 파라미터 수를 늘려 성능 향상을 초래했다. 또한 적은 파라미터로도 높은 성능을 보여준 RCAN[5], CARN[6]도 제시되었다.

본 논문에서는 SRCNN 모델과 VDSR 모델의 기본 구조와 성능평가 함수의 개선에 관하여 기술하고 전반적인 두 가지 방법의 실험을 수행하여 우수성을 입증하고자 한다.

## II. SRCNN 모델

중단 간 초고해상도 알고리즘이 바로 SRCNN(Super-Resolution Convolutional Neural Network) 알고리즘이다[2]. 이 알고리즘은 이미지 초고해상도 연구에 CNN(Convolutional Neural Network)을 적용하였다.

알고리즘의 흐름은, 처음에는 입력 전처리로서, 입력 저해상도 이미지를 확대하고 대상 크기로 확대하기 위해 bicubic 알고리즘을 사용했다. 알고리즘의 다음 과정은, 초고해상도 이미지를 얻기 위해 컨볼루션 네트워크를 통해 입력된 블러링된 저해상도 이미지를 처리하여 원본 고해상도 이미지와 최대한 비슷하게 만드는 것이다. 가령 G를 학습 모델의 목적 함수라고 하면, 저해상도 이미지 LI를 입력하고 이 함수를 통해 G(LI)를 처리하여 가능한 한 고해상도 원본 이미지 HOI와 유사한 결과를 얻는다. 이 과정은 다음과 같이 다음 세 단계로 구성된다.

첫 번째, 블록 추출 및 기능 표현 단계이다. 저해상도 이미지 LI에서 이미지 블록을 추출하고 각 이미지 블록을 고차원 벡터로 나타낸다. 모든 이미지 블록의 고차원 벡터는 일련의 특징 맵을 형성하며 특징 맵 채널의 수는 벡터의 차원과 같다. 다계층 컨볼루션 연산을 사용하여 생성된 특징 맵으로서 직접 이해할 수 있다.

두 번째, 비선형 매핑 단계로서, 고차원 벡터를 다른 고차원 벡터에 매핑한다. 다시 말해서, 첫 번째 컨볼루션 네트워크에서 출력된 고차원 특징 행렬은 다른 컨볼루션 커널 집합을 사용하여 컨볼루션 연산을 수행한다. 이것은 특징 매핑을 구현하는 것이다.

마지막은 복원 단계로서 이전에 계산된 고차원 기능이 융합되어 최종 초해상도 이미지를 생성하는 것이다. 예측된 초해상도 이미지가 원본 고해상도 이미지와 최대한 유사하게 만드는 것이 목표이다. SRCNN 모델의 구조는 그림 1[2]과 같다.

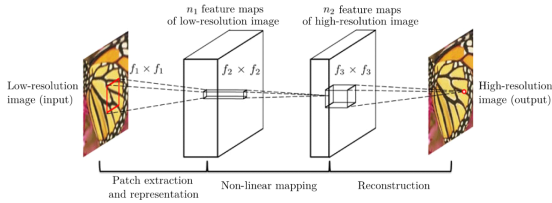


Fig. 1 SRCNN Model Structure

### III. VDSR 모델

VDSR (Very Deep Super-resolution) 모델은 이미지 초고해상도를 위한 대규모 컨볼루션 신경망이다[3]. 딥러닝 기반의 초고해상도 영상 알고리즘 중 입력 저해상도 영상과 고해상도 출력 영상은 대부분 유사하다. 즉, 저해상도 영상이 전달하는 저주파 정보는 고해상도 영상의 저주파 정보와 유사하다. 훈련 중에 이 부분을 가져오는 데 많은 시간이 걸린다. 실제로 우리는 고해상도 이미지와 저해상도 이미지 사이에서만 학습하면 된다. 이 고주파 정보의 획득은 잔여 학습 모듈의 장점이다[7]. 이 아이디어에 영감을 받아 연구진들은 잔여 네트워크를 이미지 초고해상도 모델에 도입했다. 실제 실험 결과를 살펴보면 잔여 네트워크 구조에 대한 아이디어가 특히 다음과 같은 딥러닝 초해상도 방법에 영향을 미친다고 할 수 있는 초해상도 문제를 해결하는 데 적합하다는 것을 알 수 있다.

VDSR은 보간 후 얻은 목표 크기의 저해상도 이미지를 네트워크의 입력으로 취한 다음이 이미지를 네트워크에서 학습한 잔여 학습에 추가하여 네트워크의 최종 출력을 얻는다. VDSR에는 네 가지 주요 기능이 있다.

첫째, VDSR 모델은 네트워크 구조(20계층)를 심화하여 네트워크 계층이 심오할수록 수용 필드가 더 커진다. VDSR 연구에서 3x3 컨볼루션 커널이 선택되고 깊이 D의 네트워크에는  $(2D + 1) \times (2D + 1)$  수용 필드가 있다.

둘째, VDSR 모델은 잔여 학습을 사용한다. 잔여 이미지는 상대적으로 희박하고 대부분의 값은 0이거나 상대적으로 작기 때문에 수렴 속도가 빠르다. 또한 VDSR은 적응형 그래디언트 클리핑(조정 가능한 그래디언트 클리핑)을 적용하여 그래디언트를 특정 범위로 제한하고 수렴 프로세스의 속도를 높일 수도 있다.

셋째, VDSR 모델은 각 컨볼루션 전에 이미지에 제로 채우기 작업을 수행하여 모든 기능 맵과 최종 출력 이미지의 크기가 일관되도록 한다. 연속적인 컨볼루션을 통해 이미지가 점점 작아지는 문제도 해결했다. 이 실험은 제로 필링 동작에 의한 경계 픽셀의 예측 결과도 향상될 수 있음을 증명한다.

마지막으로 VDSR 모델은 학습을 위해 여러 배수의 이미지를 혼합하여 학습된 모델이 여러 배수의 초고해상도 문제를 해결할 수 있도록 한다. VDSR 모델의 구조는 그림 2[3]에 나와 있다.

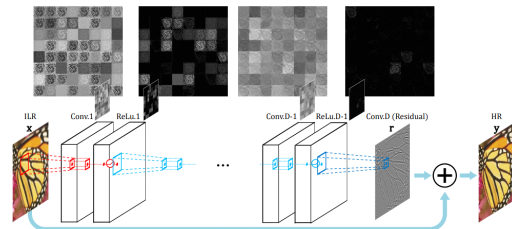


Fig. 2 VDSR Model Structure

### IV. 성능평가 함수

이미지 초고해상도에 대해 일반적으로 사용되는 평가 지수는 PSNR(Peak Signal to Noise Ratio) 및 구조적 유사성 SSIM(Structural Similarity) 이다[8]. 이 두 가지 평가 지표는 이미지 초고해상도의 필수 평가 지표이며 주로 압축된 재구성 이미지의 품질을 측정하는 데 사용된다.

#### 4.1. 피크 신호 대 잡음비

PSNR(Peak Signal-to-Noise Ratio)은 압축되어 재구성된 이미지의 품질을 측정하기 위해 신호의 최대 전력 대 신호의 잡음 전력의 비율이며 일반적으로 데시벨 (dB)로 표시된다. PSNR 지수가 높을수록 이미지 품질이 좋아진다. PSNR의 표현은 방정식 (1) 또는 방정식 (2)에 표시된다.

$$PSNR = 10 * \log_{10} \left( \frac{MAX_I^2}{MSE} \right) \quad (1)$$

$$PSNR = 20 * \log_{10} \left( \frac{MAX_I}{\sqrt{MSE}} \right) \quad (2)$$

여기서  $MAX_x$ 는 이미지 픽셀값의 최대값을 나타내고,  $MSE$ 는 현재 이미지  $X$ 와 참조 이미지  $Y$ 의 평균 제곱 오차( $MSE$ )를 나타내며,  $H$ 와  $W$ 는 각각 이미지의 높이와 너비를 나타낸다. 단일 채널 영상의  $MSE$  표현은 식 (3)에 나타나 있으며, 다중 채널 ( $C$  채널) 영상의  $MSE$  표현은 식 (4)과 같다.

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X(i,j) - Y(i,j))^2 \quad (3)$$

$$MSE = \frac{1}{C \times H \times W} \sum_{k=1}^C \sum_{i=1}^H \sum_{j=1}^W (X(i,j,k) - Y(i,j,k))^2 \quad (4)$$

#### 4.2. 구조적 유사성

$SSIM$ (Structural Similarity)은 두 이미지의 유사성을 측정하는 지표이며 값의 범위는  $[0,1]$ 이다.  $SSIM$  값이 클수록 이미지 왜곡 정도가 작아져 이미지 품질이 좋아진다.  $\mu_x, \mu_y$ 는 이미지  $X$ 와  $Y$ 의 평균값을 나타내고  $\sigma_x, \sigma_y$ 는 이미지  $X$ 와  $Y$ 의 분산을 나타내며  $\sigma_{xy}$ 는 이미지  $X$ 와  $Y$ 의 공분산을 나타낸다. 계산 공식은 방정식 (5), (6) 및 (7)과 같다

$$\mu_x = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X(i,j) \quad (5)$$

$$\sigma_x = \left( \frac{1}{H \times W - 1} \sum_{i=1}^H \sum_{j=1}^W (X(i,j) - \mu_x)^2 \right)^{\frac{1}{2}} \quad (6)$$

$$\sigma_{xy} = \frac{1}{H \times W - 1} \sum_{i=1}^H \sum_{j=1}^W (X(i,j) - \mu_x)(Y(i,j) - \mu_y) \quad (7)$$

$SSIM$ 은 밝기, 대비 및 구조의 세 가지 측면에서 이미지의 유사성을 측정한다. 밝기, 대비 및 구조적 유사성의 표현은 방정식 (8), (9) 및 (10)과 같다.

$$l(X, Y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (8)$$

$$c(X, Y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (9)$$

$$s(X, Y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (10)$$

그중  $C_1, C_2, C_3$ 는 상수이다. 0 분모를 피하려면 일반적으로  $C_1=(K_1*L)^2, C_2=(K_2*L)^2, C_3=C_2/2$ 를 취한다. 일

반적으로  $K_1=0.01, K_2=0.03, L=255$ 이다. 그러면  $SSIM$ 의 표현은 식 (11)과 같다.

$$SSIM(X, Y) = l(X, Y) * c(X, Y) * s(X, Y) \quad (11)$$

$SSIM$ 의 값 범위는  $[0,1]$ 이다.  $SSIM$  값이 클수록 이미지 왜곡이 작아지고 이미지  $X$ 가 참조 이미지에 더 가깝다.  $SSIM$ 의 특수한 개선된 형태는 방정식 (12)에 나와 있다.

$$SSIM(X, Y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (12)$$

성능평가 함수에서 피크 신호 대 잡음비는 식 (3-4)와 같이 개선되었고, 구조적 유사성은 식 (11-12)와 같이 개선되어 사용된다.

## V. 실험

실험은 각 방법을 테스트하는 유일한 기준이다. 이 장에서는 실험 데이터에 대한 설명, 실험 환경의 구성 및 실험 결과를 소개한다. 실험을 통해 우리는 이미지 초고 해상도 솔루션에서 SRCNN과 VDSR에 대한 우수성을 비교 테스트한다. 데이터 세트의 이미지 세부 정보는 표 1에 나타나 있다.

**Table. 1** Details of experimental data set of image super-resolution

Dataset	Amount	Avg. Resolution	Avg. Pixels	Format	Category Keywords
Set5	5	313*336	113,491	BMP	Baby, Bird, Butterfly, Head, Woman
Set14	14	492*446	230,203	BMP	Humans, Animals, Insects, Flowers, etc.
BSD100	100	435*367	154,401	JPG	Animal, Building, Food, Landscape, etc.
Urban100	100	984*797	774,314	JPG	Architecture, City, Structure, Urban, etc.

실험 환경은 하드웨어 환경과 소프트웨어 환경으로 구분된다. 하드웨어 환경은 Inter Core i7-4790칩과 그래픽 카드로 GeForce GTX960을 사용했다. 소프트웨어

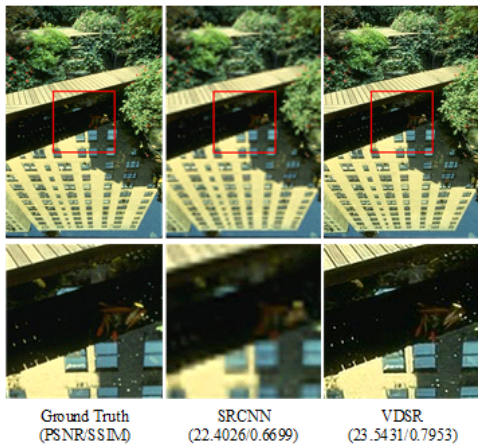
환경은 Python 프로그래밍 언어를 사용했고 모델 구성과 학습 및 테스트를 위하여 Pytorch 딥 러닝 프레임워크 및 동적 링크 라이브러리를 설치하여 사용하였다. 실험 단계는 주로 데이터 전처리, 모델 학습 및 모델 테스트의 세 부분으로 나뉩니다.

이 실험에서는 SRCNN 및 VDSR 모델의 결과와 피크 신호 대 잡음비 및 이미지 구조 유사도를 비교하는 실험을 수행했다. 비교 실험이 수행되었으며 실험 결과는 표 2에 나와 있다. 표에서는 서로 다른 초고해상도 방법으로 얻은 개선된 평균 PSNR 및 SSIM 값이 데이터 세트 Set5, Set14, BSD100 및 Urban100에서 x4의 배울을 갖는 것으로 표시된다.

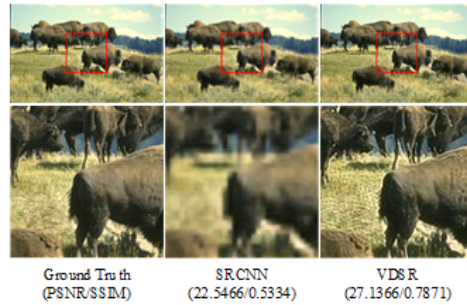
**Table. 2** Results of PSNR and SSIM values with different models

Method	Training Dataset	Set5	Set14	BSD100	Urban100
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
SRCNN	291	30.48/0.8628	27.50/0.7513	26.90/0.7101	24.52/0.7221
VDSR	291	25.64/0.8830	28.01/0.7674	27.29/0.7251	25.18/0.7524

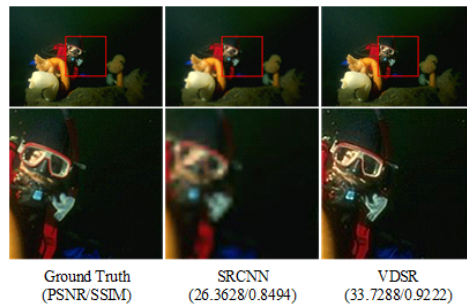
그림 3 ~ 5는 서로 다른 초고해상도 모델의 세 가지 이미지에서 얻은 SR 이미지의 실험 결과를 비교한 것으로 서로 다른 SR 이미지의 PSNR 및 SSIM 평가 지표를 계산하여 나타난다.



**Fig. 3** Super-resolution results of "148026"(scale factor x4)



**Fig. 4** Super-resolution results of "38092"(scale factor x4)



**Fig. 5** Super-resolution results of "45096"(scale factor x4)

실험에서는 그리 커다란 변화를 느끼기에는 부족한 것으로 알고 있지만, 평가함수를 조금 다르게 변화시켜도 원래의 평가함수가 갖는 결과보다도 특정 부분에서 더 좋은 이미지를 얻을 수 있다는 것이 핵심이다.

## VI. 결론

본 논문에서는 저해상도 이미지에서 딥 러닝 초해상도 모델을 이용하여 주요 특징을 추출하고 이를 학습을 수행하고 재구성하여 고해상도 이미지를 생성하는 두 가지 방법에 초점을 두었다. 두 가지 방법은 재구성에 기반을 둔 초해상도 알고리즘 모델에서 SRCNN과 VDSR 모델이다. 본 논문에서는 SRCNN과 VDSR 모델의 구조 및 알고리즘 프로세스를 간략하게 알아보았고 개선된 성능평가 함수에서도 다중 채널과 특수한 형태에 대해서도 살펴보았다. 또한, 실험을 통하여 각 알고리즘의 성능과 우수성도 알아보았다. 그리고 성능평가 함수를 개선하여 이미지에서의 미세한 변화와 특정 부분에 대한 변화도 알아보았으며, 이는 이미지 처리에서 하나의 중요 자료가 될 것이다.

## References

- [ 1 ] Y. L. Seo and S. J. Kang, "Current status and latest trend of deep learning-based super resolution technology," *Broadcasting and Media Magazine*, vol. 25, no. 2, pp. 7-16, Apr. 2020.
- [ 2 ] C. Dong, C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," In *Proceedings of the European conference on computer vision*, Glasgow, United Kingdom, Springer, Cham, pp. 184-199, Sep. 2014.
- [ 3 ] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, pp. 1646-1654, 2016.
- [ 4 ] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, Salt Lake City, UT, USA, pp. 2472-2481, 2018.
- [ 5 ] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," In *Proceedings of the European Conference on Computer Vision(ECCV)*, Munich, Germany, pp. 286-301, 2018.
- [ 6 ] N. Ahn, B. Kang, and K. A. Sohn, "Fast, Accurate, and Lightweight Super-Resolution with Cascading Residual Network," In *Proceedings of the European Conference on Computer Vision(ECCV)*, Munich, Germany, pp. 256-272, 2018.
- [ 7 ] D. Vint, G. Di Caterina, J. J. Soraghan, R. A. Lamb, and D. Humphreys, "Evaluation of performance of VDSR super resolution on real and synthetic images," In *2019 Sensor Signal Processing for Defence Conference(SSPD)*, Brighton, United Kingdom, pp. 1-5, May. 2019.
- [ 8 ] 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.



**이광찬(Kwang-Chan Lee)**

한국외국어대학교 경영정보학과 석사  
2017.9~현재 군산대학교 컴퓨터정보공학과 박사과정  
※ 관심분야: 영상처리, 빅데이터



**왕광싱(Gwangxing Wang)**

중국화중과기대학교 컴퓨터응용기술 석사  
2018.3~2021.2 군산대학교 컴퓨터정보공학과 박사  
※ 관심분야: 인공지능, 정보시스템



**신성윤(Seong-Yoon Shin)**

2003.2: 군산대학교 컴퓨터학과 이학박사  
2006.3~현재 : 군산대학교 컴퓨터정보통신공학부 교수  
2018.1~2019.12 : 한국정보통신학회 총무부회장  
2020.1~현재 : 한국정보통신학회 수석부회장  
2020.3~현재 : 군산대학교 정보전산원장  
※ 관심분야: 비디오처리, 가상현실, 멀티미디어