

일반논문 (Regular Paper)

방송공학회논문지 제26권 제4호, 2021년 7월 (JBE Vol. 26, No. 4, July 2021)

<https://doi.org/10.5909/JBE.2021.26.4.429>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

잔여 밀집 및 채널 집중 기법을 갖는 재귀적 경량 네트워크 기반의 단일 이미지 초해상도 기법

우희조^{a)}, 심지우^{a)}, 김응태^{a)†}

Single Image Super Resolution Based on Residual Dense Channel Attention Block-RecursiveSRNet

Hee-Jo Woo^{a)}, Ji-Woo Sim^{a)}, and Eung-Tae Kim^{a)†}

요 약

최근 심층 합성곱 신경망 학습의 발전에 따라 단일 이미지 초해상도에 적용되는 심층 학습 기법들은 좋은 성과를 보여주고 있다. 현존하는 딥러닝 기반 초해상도 기법들 중 하나로 잔여 밀집 블록을 이용하여 초기의 특징 정보를 마지막 계층에 전달하여 이후의 계층들이 이전의 계층들의 입력정보를 사용하여 복원하는 RDN(Residual Dense Network)이 있다. 하지만 계층적인 모든 특징을 연결하여 학습하고 다수의 잔여 밀집 블록을 쌓게 되면 좋은 성능에도 불구하고 많은 파라미터의 수와 연산량을 가지게 되어 느린 처리 속도와 네트워크를 학습하는데 많은 시간이 소요되고 모바일 시스템에 적용이 어렵다는 단점을 가지고 있다. 본 논문에서는 이전의 정보를 다시 사용하는 연속 메모리 구조인 잔여 밀집 구조와 이미지의 특징맵에 따라 중요도를 결정해주는 채널 집중 기법을 이용한 잔여 밀집 채널 집중 블록을 재귀적인 방식으로 사용하여 추가적인 파라미터 없이 네트워크의 깊이를 늘려 큰 수용 영역을 얻으며 동시에 간결한 모델을 유지할 수 있는 방식을 제안한다. 실험 결과 제안하는 네트워크는 RDN과 비교 하였을 때 4배 확대 배율에서 평균적으로 PSNR 0.205dB만큼 낮지만 약 1.8배 더 빠른 처리속도, 약 10배 더 적은 파라미터의 수와 약 1.74배 더 적은 연산량을 갖는 것을 실험을 통해 확인하였다.

Abstract

With the recent development of deep convolutional neural network learning, deep learning techniques applied to single image super-resolution are showing good results. One of the existing deep learning-based super-resolution techniques is RDN(Residual Dense Network), in which the initial feature information is transmitted to the last layer using residual dense blocks, and subsequent layers are restored using input information of previous layers. However, if all hierarchical features are connected and learned and a large number of residual dense blocks are stacked, despite good performance, a large number of parameters and huge computational load are needed, so it takes a lot of time to learn a network and a slow processing speed, and it is not applicable to a mobile system. In this paper, we use the residual dense structure, which is a continuous memory structure that reuses previous information, and the residual dense channel attention block using the channel attention method that determines the importance according to the feature map of the image. We propose a method that can increase the depth to obtain a large receptive field and maintain a concise model at the same time. As a result of the experiment, the proposed network obtained PSNR as low as 0.205dB on average at 4× magnification compared to RDN, but about 1.8 times faster processing speed, about 10 times less number of parameters and about 1.74 times less computation.

Keyword : Deep Learning, Super Resolution, Recursion, Residual Dense Block, Channel Attention

Copyright © 2021 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

I. 서론

단일 이미지 초해상도 기법은 저해상도 이미지로부터 고해상도 이미지를 생성하는 기법이다. Ultra-High Definition (UHD)과 같은 고해상도의 디스플레이가 등장하며 Full-High Definition(FHD)과 같은 저해상도 이미지를 고해상도 이미지로 변환할 수 있는 초해상도 알고리즘에 관심이 커지고 있고 초해상도는 디지털 TV, 모바일, 보안 카메라 등 다양한 분야에서 활용 되고 있다.

단일 이미지의 해상도를 높이는 대표적인 접근 방식으로는 크게 보간법^[1-4], 재구성법^[5-7] 및 심층 학습법^[8-10,12,17-19] 세 가지로 나눌 수 있다. 보간법은 주어진 주변의 화소값을 이용하여 정보가 없는 곳에 새로운 화소를 생성하는 기술로 대표적으로 최근접 화소 보간법, 선형 보간법, 양선형 보간법 및 고차원 보간법이 존재한다. 보간법은 계산량이 적다는 장점을 가지고 있지만 저주파수 필터 특성 때문에 이미지를 전반적으로 뭉롱화시켜 디테일한 부분의 화질 저하 및 열화가 존재한다. 재구성법은 보간법 보다 많이 개선되어 고주파수 성분을 잘 복원하지만 상대적으로 복잡한 계산량을 가지고 있다.

최근에 심층 학습법의 기술 발달로 딥러닝 기반의 초해상도 알고리즘이 기존의 이미지처리 방법들에 비해 높은 성능을 보여주고 있고 현재까지도 다양한 네트워크가 제안되며 활발히 연구되고 있다. 딥러닝을 적용한 초해상도의 발전은 선형^[17], 잔여학습^[18], 재귀^[9], 밀집^[8], 집중^[19]기법 등의 다양한 방식으로 발전됐으며 다양한 네트워크가 제안되었다.

밀집하게 연결된 네트워크 구조들 중 대표적인 Zhang et al^[8]등의 RDN(Residual Dense Network)는 그전의 모든 계

층의 정보를 활용하여 이미지의 특징 맵을 추출하는데 적합하였다. 이 방법은 학습된 다양한 특징 정보를 다시 사용하는 것을 통해 정보의 흐름을 유연하게 함으로써 성능 향상을 보여주었다. 하지만 다수의 잔여 밀집 블록을 쌓게 되면 좋은 성능에도 불구하고 많은 파라미터의 수와 연산량을 가지게 되어 네트워크를 학습하는데 많은 시간이 소요되고 큰 저장 공간이 필요하여 모바일 시스템에 적용이 어렵다는 문제점을 가지고 있다.

본 논문에서는 이러한 문제를 해결하기 위해 Y.Tai^[9]등이 제안한 DRRN(Image Super-Resolution via Deep Recursive Residual Network)의 재귀적인 방식을 이용하여 잔여 밀집 및 Zhang et al^[19]등이 제안한 채널 집중 기법을 사용하는 RDCAB(Residual Dense Channel Attention Block)-RecursiveSRNet 단일 이미지 초해상도 네트워크를 제안한다. 제안된 방식은 초기의 특징 정보를 마지막 계층에 전달하고 이후의 계층들이 이전의 계층들의 입력 정보를 잘 사용하게 하기 위해 잔여 밀집 구조와 이미지의 전체 영역을 활용하여 이미지의 특징맵에 따라 중요도를 결정해주는 채널 집중 기법을 블록화 시켜 재귀적인 방식으로 사용하여 잔여 밀집 네트워크의 성능을 최대한 유지하면서 더 적은 파라미터의 수와 연산량으로 저해상도와 고해상도 사이의 연관성을 효율적으로 추론하도록 하였다.

본 논문은 다음과 같이 구성된다. 본 논문의 2장에는 관련 이론을 소개하고 3장에서는 제안된 네트워크인 RDCAB-RecursiveSRNet의 구조를 설명한다. 4장에서는 기존의 초해상도 알고리즘과 제안된 알고리즘을 비교하고 성능을 평가하며 5장에서 결론을 맺는다.

II. 기존 초해상도 관련 연구

초해상도 분야와 관련하여 합성곱 신경망을 사용하여 많은 알고리즘들이 제안되었다. 그 중 최초로 딥러닝을 초해상도에 적용한 Dong^[17]등의 SRCNN(Image Super-Resolution Using Deep Convolutional Networks)같은 경우 단 3개의 합성곱만 사용하여 딥러닝을 적용하지 않은 방법들에 비해 높은 성능을 만들어 냈고 초해상도에도 딥러닝을 적용할 수 있다는 가능성을 보여주었다. 하지만

a) 한국산업기술대학교 전자공학부(Department of Electronics Engineering, Korea Polytechnic University)

‡ Corresponding Author : 김응태(Eung-Tae Kim)

E-mail: etkim@kpu.ac.kr

Tel: +82-31-8041-0488

ORCID: <http://orcid.org/0000-0001-5984-0045>

※ This work was supported by the 2021 sabbatical year research grant of the Korea Polytechnic University.

※ This work was supported by the Technology development Program (S3025098) funded by the Ministry of SMEs and Startups(MSS, Korea).

· Manuscript received May 6, 2021; Revised June 14, 2021; Accepted June 14.

Dong의 방법은 사전 업 샘플링 구조를 가져 많은 양의 계산량이 필요하고 적은 수의 층으로 정확도 측면에서 제한적인 단점을 가지고 있다.

Kim^[18]등의 VDSR(Accurate Image Super-Resolution Using Very Deep Convolutional Networks)은 K.Simonyan^[19]등이 제안한 VGGNet에서 영감을 받아 잔여 학습을 사용하는 깊은 네트워크를 만들어 고주파수 정보만을 학습하여 성능을 개선하였지만 다른 계층에는 과거 특징 정보를 사용하지 않고 마지막 계층에만 과거 특징 정보를 사용하는 단점을 가지고 있다.

Y.Tai등이 제안한 DRRN은 깊은 네트워크가 좋은 성능을 만들 수 있지만 한 층씩 늘릴 때마다 파라미터의 수가 많이 늘어 하이퍼파라미터를 튜닝하기 어려운 점을 지적하였다. 전역 잔여 학습과 지역 잔여 학습을 사용하고 여러 개의 잔여 유닛으로 구성된 재귀 블록을 이용하여 추가적인 파라미터 없이 깊은 네트워크를 만들어 냈다.

모든 층에서의 정보를 취득하는 형태인 G.Huang^[10]등이 제안한 텐스넷(Dense Net)을 이용하여 Zhang et al^[8]등이 잔여 밀집 네트워크(Residual Dense Network, RDN)을 제안했다. 그림 1은 RDN의 구조를 나타내며 그림 2의 잔여 밀집 블록(Residual Dense Block)을 다수 쌓는 구조로 구성되어 있다.

RDB는 밀집하게 연결되어 있는 합성곱 층으로부터 지역 특징을 추출하고 이전의 RDB의 출력이 현재 RDB의 모

든 층의 입력으로 들어가게 설계되어 있다. 지역 특징 융합(Local Feature Fusion, LFF)을 통해 이전의 RDB의 출력과 지역 특징을 충분히 이용하여 학습 할 수 있도록 되어있고 밀집 지역 특징을 학습 한 후 전역 특징 융합을 통해 전역 계층 특징을 학습하게 된다.

여러 개의 RDB를 거쳐 지역 밀집 특징들을 얻은 후 밀집 특징 융합(Dense Feature Fusion, DFF)을 통해 계층적 특징들을 전역으로 이용을 하고 밀집 특징 융합은 전역 특징 융합(Global Feature Fusion, GFF)과 전역 잔여 학습(Global Residual Learning, GRL)으로 이루어져 있다. 전역 특징 융합은 각 RDB로부터 나온 출력 특징 맵을 연결한 다음 위치별 합성곱(Pointwise Convolution)인 1x1 합성곱과 3x3 합성곱을 거치도록 하여 전역 특징을 뽑아내며 전역 잔여 학습은 전역 특징 융합으로부터 나온 전역 특징과 네트워크의 첫 번째 합성곱의 출력을 더해준다. RDN의 각 층의 밀집 연결은 지역 특징들을 완전히 이용할 수 있도록 하며 지역 특징 융합은 깊은 네트워크의 학습을 안정화시키고 이전의 RDB의 출력과 현재 RDB의 지역 특징 정보를 보존할 수 있도록 해준다. 또한 지역 잔여 학습을 사용하여 정보가 잘 흐를 수 있도록 하며 전역 특징 융합을 통해 저해상도 공간의 계층적인 특징들을 추출하였다. 그러나 다수의 RDB를 쌓게 되면 네트워크의 깊이가 깊어져 과적합 문제와 많은 메모리 소비 문제가 생기며 급격하게 증가하는 계산량으로 네트워크를 학습하는데 많은 시간이 소요

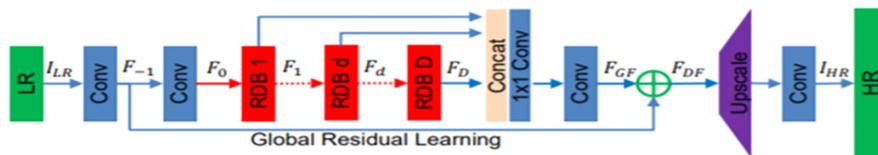


그림 1. Residual Dense Network 구조
 Fig. 1. The architecture of Residual Dense Network

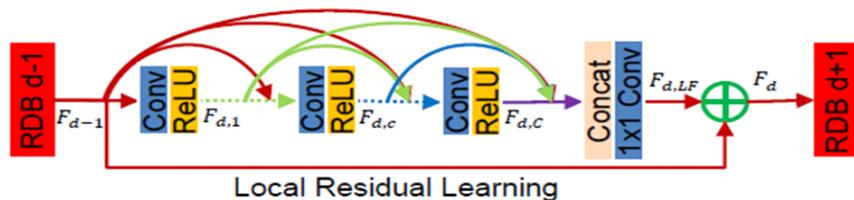


그림 2. Residual Dense Block 구조
 Fig. 2. The architecture of Residual Dense Block

되고 큰 저장 공간이 필요하여 모바일 시스템에 적용이 어려운 문제가 생긴다.

III. 제안된 RDCAB-RecursiveSRNet기반 초해상도 기법

본 논문의 그림 3은 RDCAB를 재귀적인 방식으로 사용하여 추가적인 파라미터 없이 네트워크의 깊이를 늘린 제안된 RDCAB-RecursiveSRNet의 구조를 보여준다. 제안하는 방법은 파라미터의 수와 연산량을 줄이기 위해 재귀적인 방식을 이용하였고 높은 성능을 만들기 위해 잔여 밀집 구조와 채널 집중 기법을 이용한다. 제안된 네트워크는 얇은 특징 추출 단계, 재귀 블록을 통한 특징 추출 단계, 업스케일 및 복원 단계로 구성되어 있다.

1. 얇은 특징 추출 단계

제안된 네트워크는 복잡한 계산량의 증대를 막기 위해 사후 업 샘플링 구조로 이루어져 있다. 저해상도의 이미지가 입력되게 되면 3x3 합성 곱을 이용하여 첫 번째와 두 번째 층에서 각각 64개의 특징 맵을 추출하고 첫 번째 합성 곱의 출력은 전역 잔여 학습에 사용이 되고 두 번째 합성 곱의 출력은 RDCAB의 지역 잔여 학습에 사용된다. 전역 잔여 학습과 지역 잔여 학습을 사용하여 추가적인 파라미터 없이 잔여 특징들을 더해주어 더 좋은 성능을 만들어 내었고 학습 시 기울기 소실(Gradient Vanishing)문제를 해결하였다.

입력으로 들어온 저해상도의 이미지가 I_{LR} 이고 네트워크

를 통과하여 만들어진 초해상도 이미지를 I_{SR} 이라 할 때 다음 식과 같이 나타낼 수 있으며 $F_{Recursive-SRNet}$ 은 전체 네트워크를 표현하는 함수이다.

$$I_{SR} = F_{Recursive-SRNet}(I_{LR}) \tag{1}$$

저해상도 이미지가 2단계로 구성되어 있는 얇은 특징 추출 단계를 통과하게 되면 다음 수식과 같이 나타낼 수 있다.

$$F_{-1} = H_{FE_1}(I_{LR}) \tag{2}$$

$$F_0 = H_{FE_2}(F_{-1}) \tag{3}$$

H_{FE_1} 과 H_{FE_2} 는 첫 번째와 두 번째 합성 곱을 나타내는 함수이며 F_{-1} 과 F_0 는 각각 합성 곱의 출력이고 전역 잔여 학습과 재귀 블록을 통한 특징 추출 단계의 입력과 지역 잔여 학습에 사용이 된다.

2. 재귀 블록을 통한 특징 추출 단계

제안된 네트워크와 RDN은 모두 표준 합성 곱(Standard Convolution)을 사용한다. 따라서 파라미터 수를 계산하는 식은 다음과 같다.

$$Params_{RDB} = k_x k_y m n \tag{4}$$

여기서 m 과 n 은 입력 및 출력 채널의 크기이고 k_x 과 k_y 는 합성 곱의 커널 크기이며 제안된 네트워크와 RDN에 사용된 RDB는 동일하게 3x3 커널을 사용한다.

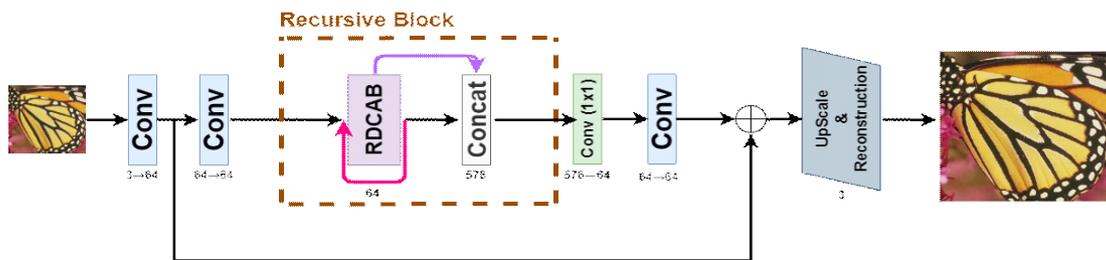


그림 3. 제안된 RDCAB-RecursiveSRNet 구조
Fig. 3. The architecture of proposed network RDCAB-RecursiveSRNet

표 1은 깊이에 따른 RDB의 파라미터 수를 식4를 통해 계산한 것이다. 하나의 RDB은 총 1,327,104개 파라미터 수를 가지고 있으며 n개의 RDB을 쌓게 되면 $1,327,104 \times n$ 개 파라미터 수가 필요하게 된다. 하지만 재귀적인 방식으로 사용하게 되면 네트워크를 공유하여 사용하기에 추가적인 파라미터 없이 수용 영역을 넓히며 사용하여 n번의 재귀에도 1,327,104개 파라미터 수만 필요하게 된다. 따라서 재귀적인 방식을 이용하면 깊은 네트워크를 만드는데 있어 중요한 저장 공간의 수요를 줄이며 동시에 깊이를 늘리는 간결한 모델을 유지할 수 있다.

표 1. 입/출력 채널수에 따른 RDB의 파라미터 수 비교
 Table 1. Comparison of the number of RDB parameters according to the number of input/output channels

Number of Input(m) → Output(n) channels	Number of Parameters RDB - Residual Dense Block
64→64	36,864
128→64	73,728
192→64	110,592
256→64	147,456
320→64	184,320
384→64	221,184
448→64	258,048
512→64	294,912
Total	1,327,104

재귀 손실을 최소화하고 높은 성능을 유지하기 위해 잔여 밀집 구조와 채널 집중 기법으로 구성되어 있는 RDCAB를 사용하였고 파라미터의 수를 최대한 줄이기 위해 RDCAB를 재귀적으로 사용하였다. 잔여 밀집 구조는 연속 메모리

구조, 지역 특징결합, 지역 잔여학습 등을 통해 모든 계층의 정보를 활용하며 채널 집중 기법은 채널들 사이의 중요도에 따라 채널을 재정의해 주는 기능을 가지고 있다.

그림 4는 RDCAB구조를 나타내는 그림이며 8개의 3×3 합성 곱이 잔여 밀집 구조로 연결되어 있고 출력이 채널 집중 층을 통과한 뒤 1×1 위치별 합성 곱으로 채널을 조절하고 지역 잔여 학습을 진행하여 출력을 만들어 낸다. 각 층의 입력은 이전 층의 출력을 연결한 값을 사용하였고 64개의 특징 맵을 출력하였다. 최대한 높은 성능을 유지하기 위해 성장률(growth rate)은 64로 설정 하였다. RDCAB는 설정한 재귀 횟수만큼 반복하고 각 횟수마다 나온 RDCAB의 특징들을 연결하여 다시 한 번 1×1 위치별 합성 곱을 통과시켜 특징맵의 채널수를 64로 감소시켰다. 전체 영역의 정보를 활용하기 위해 각 채널들 사이의 중요도에 따라 가중치를 할당하는 채널 집중 기법을 전역 평균 풀링(Global Average Pooling)과 시그모이드(Sigmoid)를 사용하여 구성하였고 입력으로 들어온 채널을 감소율 16을 이용하여 나누고 다시 곱해주는 방식을 이용하였으며 네트워크의 첫 번째 합성 곱을 통과해서 나온 출력값과 전역 잔여 학습을 진행하였다.

F_{r-1} 와 F_r 을 r번째 재귀에서 RDCAB의 잔여 밀집 층의 입력과 출력이라고 하고 이 둘의 채널수는 G_0 로 동일하다고 가정하면 r번째 RDCAB에서 잔여 밀집 층의 c번째 합성 곱의 출력은 다음 수식과 같이 나타낼 수 있다.

$$F_{r,c} = \sigma(W_{r,c}[F_{r-1}, F_{r,1}, \dots, F_{r,c-1}]) \quad (5)$$

여기서 σ 는 정류 선형 유닛(Rectified Linear Unit,

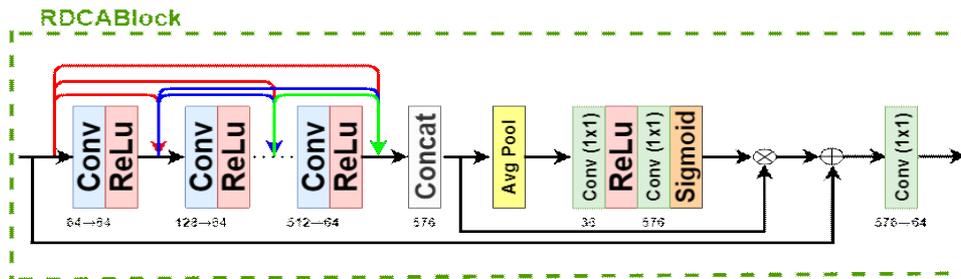


그림 4. 제안된 RDCABlock 구조
 Fig. 4. The architecture of proposed network RDCABlock

ReLU)을 의미하며 $W_{r,c}$ 는 c 번째 합성 곱 층의 가중치를 나타내고, \square 는 채널 연결을 나타낸다. 즉 이전의 RDCAB의 출력과 현재 RDCAB 내의 c 번째 이전 합성 곱 층의 출력들을 연결한 다음 합성 곱 시키고 정류 선형 유닛을 통과시켜 c 번째 합성 곱 층의 출력을 얻게 된다. 이러한 연속 메모리 구조는 순방향(Feedforward) 학습에 따른 성질을 유지할 수 있을 뿐만 아니라 지역 밀집 특징을 학습하는데도 용이하다.

RDACB의 잔여 밀집 층을 통과한 각각의 계층적인 특징들이 연결되어 채널 집중 층을 통과하게 되면 다음 수식과 같이 나타낼 수 있다.

$$F_{CA} = R_r(F_r) \cdot F_r \quad (6)$$

여기서 F_r 과 R_r 은 각각 r 번째 재귀에서 RDCAB에서 밀집하게 연결되어 있는 계층적인 특징과 전역 평균 풀링 및 시그모이드를 사용하는 채널 집중 층을 나타내며 F_{CA} 는 채널 집중 층을 통과한 출력을 의미한다.

채널 집중 층을 통과한 출력이 RDCAB의 입력으로 들어온 얇은 특징 추출단계의 두 번째 합성 곱의 출력과 지역 잔여 학습을 한 뒤 1×1 위치별 합성 곱을 통과하여 채널수를 조절하여 최종적으로 RDCAB의 출력을 만들어 내고 다음 수식과 같이 나타낼 수 있다.

$$F_{RDCAB} = H_{PW}(F_{CA} + F_0) \quad (7)$$

여기서 H_{PW} 는 1×1 위치별 합성 곱을 의미하며 F_0 는 얇은 특징 추출 단계의 두 번째 합성 곱의 출력을 의미한다.

RDCAB의 밀집하게 연결되어 있는 계층적인 특징들을 채널 집중 층을 통과시켜 지역 잔여 학습을 수행하는 RDCAB를 주어진 재귀 횟수 r 만큼 네트워크를 공유하며 사용할 때 다음 수식과 같이 나타낼 수 있으며 g 는 재귀 계층에서 단일 재귀로 모델링된 함수를 나타낸다.

$$F_{recursive} = g_r(g_{r-1}(\dots g_1(F_{RDCAB}))) \quad (8)$$

주어진 재귀 횟수마다 나온 특징 값들이 전역 특징 결합

을 하며 RDCAB와 동일하게 1×1 위치별 합성 곱을 통과하여 채널수를 조절한 뒤 전역 잔여 학습을 하여 재귀 블록을 통한 특징 추출단계의 출력을 만들어 낸다. 최종적으로 저해상도 이미지가 재귀 블록을 통한 특징 추출단계를 통과하게 되면 다음 수식과 같이 나타낼 수 있다.

$$F_{GFF} = H_{GFF}(F_{recursive_1}, F_{recursive_2}, \dots, F_{recursive_r}) + F_{-1} \quad (9)$$

여기서 H_{GFF} 는 재귀를 통해 나온 RDCAB 특징들의 전역 특징 결합을 의미 하고 F_{-1} 은 전역 잔여 학습으로 사용되는 얇은 특징 추출 단계에서 첫 번째 합성 곱의 출력을 의미한다.

3. 업 스케일 및 복원 단계

W. Shi et al^[12]등의 ESPCN(Efficient Sub-Pixel Convolutional Neural Network)에서 제안한 sub-pixel 합성 곱을 통해 채널 별 서플을 이용하여 $H \times W \times (C \times r \times r)$ 을 $rH \times rW \times C$ 로 변형 시켜 업 스케일을 진행 하였다. 여기서 r 은 확대 배율을 의미한다. 최종 복원 단계에서는 업 스케일 단계를 거쳐 나온 특징들을 3×3 합성 곱을 통해 RGB로 구성되어 있는 3채널 이미지로 복원한다.

IV. 모의실험 결과

제안된 네트워크의 성능을 확인하기 위해 학습을 위한 데이터 셋으로 DIV2K 800장 영상을 이용하여 영상 당 서로 다른 위치에서 동일한 크기로 5장을 크롭(Crop)하여 총 4,000장의 영상을 만들었고, 원본 고해상도 영상인 4,000장에 대해 쌍입방 보간법을 통해 다운 스케일 하여 저해상도 영상을 획득한 뒤 학습시키는데 사용하였다. 검증 셋으로 5장으로 이루어져있는 Set5를 이용하였다. 제안된 네트워크에 입력과 출력으로 3채널 RGB 이미지를 사용하였고 수평, 수직 뒤집기 및 90° 회전을 이용하는 데이터 증강 기술을 사용하였다. 입력으로 들어온 저해상도 이미지는 32×32 크기의 패치(Patch)로 무작위로 잘린 다음 배치 크기인 16

만큼 모델에 들어가게 된다. 사용된 활성화 함수로는 정류 선형 유닛(Rectified Linear Unit, ReLU)을 사용 하였다. 학습 횟수는 800에폭(epoch)으로 학습하였고 학습률(Learning Rate)은 0.0001을 초기 값으로 사용하였고 640에폭 이후에는 0.00001로 감소시켜 사용하였다. $\beta_1=0.9$, $\beta_2=0.999$, $\epsilon=10^{-8}$ 로 설정한 Adam 최적화 방식을 사용 하였고, 손실 함수는 L1 Loss를 사용 하였다. 실험은 파이토치(PyTorch)를 이용하였고 재귀 블록의 수를 1개, 재귀 블록 내 사용되는 RDCAB를 9번 재귀적으로 사용하는 네트워크를 기본 모델로 사용하였다. 사용된 하드웨어 환경은 Intel Xeon

Gold 6240R 2.40GHz CPU와 RTX 3090 GPU 24G를 사용 하였다.

표 2, 표 3은 각 테스트 셋을 이용하여 기존의 네트워크들과 제안된 네트워크의 신호 대 잡음비(Peak signal to noise, PSNR), 구조적 유사도(Structural Similarity, SSIM) 및 각 네트워크의 파라미터의 수와 연산량을 비교한 표이다.

제안된 방법이 RDN을 제외한 기존의 방법들보다 더 좋은 성능을 만들었고 RDN과 비교하였을 때 표 2에서 보이는바와 같이 2배 확대 배율 BSD100 테스트 셋에서 RDN은

표 2. 테스트 셋에 대한 최대 신호 대 잡음비, 구조적 유사도 비교
 Table 2. Comparison of PSNR, SSIM for the test set

Model	scale	Set5 ^[13] (PSNR,SSIM)	Set14 ^[14] (PSNR,SSIM)	BSD100 ^[15] (PSNR,SSIM)	Urban100 ^[16] (PSNR,SSIM)
SRCNN	×2	36.66 / 0.9542	32.45 / 0.9067	31.36 / 0.8435	29.51 / 0.8405
	×3	32.75 / 0.9090	29.29 / 0.8215	28.41 / 0.7863	26.24 / 0.7991
	×4	30.48 / 0.8628	27.50 / 0.7513	26.90 / 0.7103	24.52 / 0.7226
VDSR	×2	37.53 / 0.9587	33.05 / 0.9127	31.90 / 0.8960	30.77 / 0.9141
	×3	33.66 / 0.9213	29.78 / 0.8318	28.83 / 0.7976	27.14 / 0.8279
	×4	31.35 / 0.8838	28.02 / 0.7678	27.29 / 0.7252	25.18 / 0.7525
DRRN	×2	37.63 / 0.9591	32.23 / 0.9136	32.05 / 0.8973	31.23 / 0.9188
	×3	34.03 / 0.9244	29.60 / 0.8349	28.95 / 0.8004	27.53 / 0.8377
	×4	31.68 / 0.8888	28.21 / 0.7720	27.38 / 0.7284	25.44 / 0.7638
SRDenseNet	×2	-	-	-	-
	×3	-	-	-	-
	×4	32.02 / 0.8934	28.50 / 0.7782	27.53 / 0.7337	26.05 / 0.7819
RDN	×2	38.24 / 0.9614	34.01 / 0.9212	32.34 / 0.9017	32.89 / 0.9353
	×3	34.71 / 0.9296	30.57 / 0.8468	29.26 / 0.8093	28.80 / 0.8653
	×4	32.47 / 0.8990	28.81 / 0.7871	27.72 / 0.7419	26.61 / 0.8028
CARN	×2	37.76 / 0.9590	33.52 / 0.9166	32.09 / 0.8978	31.92 / 0.9256
	×3	34.29 / 0.9255	30.29 / 0.8407	29.06 / 0.8034	28.06 / 0.8493
	×4	32.13 / 0.8937	28.60 / 0.7806	27.58 / 0.7349	26.07 / 0.7837
CARN-M	×2	37.53 / 0.9583	33.26 / 0.9141	31.92 / 0.8960	31.23 / 0.9193
	×3	33.99 / 0.9236	30.08 / 0.8367	28.91 / 0.8000	27.55 / 0.8385
	×4	31.92 / 0.8903	28.42 / 0.7762	27.44 / 0.7304	25.62 / 0.7694
OURS	×2	38.05 / 0.9610	33.61 / 0.9179	32.23 / 0.9005	32.18 / 0.9284
	×3	34.51 / 0.9278	30.40 / 0.8438	29.15 / 0.8067	28.26 / 0.8544
	×4	32.32 / 0.8951	28.66 / 0.7837	27.63 / 0.7382	26.18 / 0.7888

표 3. 각 네트워크의 파라미터 수와 연산량 비교
 Table 3. Comparison of the number of parameters and the computational complexity of each network

×4	SRCNN	VDSR	DRRN	SRDenseNet	RDN	CARN	CARN-M	OURS
Parameter	0.057M	0.66M	0.29M	5M	22M	1.5M	0.4M	2.1M
Multi-Adds	52.7G	612.6G	6,796.9G	389.9G	1,309G	90.9G	32.5G	750G

32.34dB를 얻었고 제안된 방법은 32.23dB를 얻어 0.11dB 만큼 더 낮은 성능을 만들었고 4배 확대 배율 Urban100 테스트 셋에서 RDN은 26.61dB를 얻었고 제안된 방법은 26.18dB를 얻어 0.43dB 만큼 낮은 성능을 만들었지만 표 3에서 보이는바와 같이 RDN은 22M 파라미터 수와 1,309G 연산량을 갖고 제안된 네트워크는 2.1M 파라미터 수와 750G 연산량을 가져 제안된 네트워크가 약 10배 더 적은 파라미터의 수와 약 1.74배 더 적은 연산량을 가짐을 알 수 있다.

그림 5는 표 2의 4배 확대 배율에 따른 Set5 테스트 셋의 성능을 파라미터의 수로 기존의 네트워크들과 비교한 그림으로 SRCNN, VDSR, DRRN은 사용하는 네트워크 또는 블록의 층이 얇아 적은 수의 파라미터의 수를 갖지만 얇은 층의 한계로 정확한 특징을 학습하지 못해 제안하는 네트워크보다 낮은 성능을 만들었고 1,327,104개의 파라미터 수를 갖는 잔여 밀집 블록을 16개 쌓아 약 22M의 파라미터의 수를 갖는 RDN보다 약 10배 더 적은 파라미터 수로 비슷한 성능을 만들어 낸 것을 확인 할 수 있으며 경량화에 초점을 맞춘 CARN과 CARN-M은 중속 모듈을 사용하여 파라미터의 수와 연산량을 줄이긴 하였으나 성능 또한 다수 줄어들어 성능과 효율성 사이에서 균형을 이루고 있지 못하며 각각 제안된 방법보다 1.4배, 5.2배 적은 파라미터 수를 갖지만 0.19dB, 0.4dB 낮은 성능을 갖는 것을 확인할 수 있다.

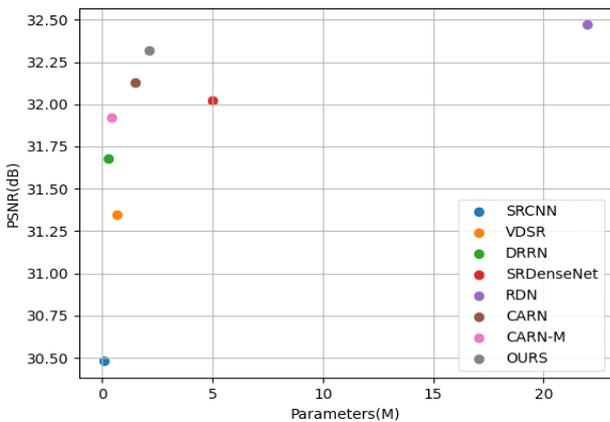


그림 5. 기존 네트워크들과 제안된 네트워크의 파라미터 수 비교
 Fig. 5. Comparison of the number of parameters of the existing networks and the proposed network

표 4는 네트워크를 학습하는데 총 걸린 시간과 Urban100 테스트 셋에서 'img007' 이미지를 4배 확대 배율로 초해상화 시킬 때 처리시간에 대한 성능을 비교한 표로 RDN은 학습하는데 약 8시간이 소요 됐으며 제안된 네트워크는 약 5시간이 소요되어 RDN에 비해 1.6배 더 적은 학습 시간을 필요로 하였다. 또한 RDN은 0.033초 처리 시간으로 29.89dB 성능을 만들어 냈는데 제안된 네트워크는 0.019초 처리 시간으로 29.66dB 성능으로 만들어 내어 0.23dB 더 적은 성능을 만들었지만 처리속도는 RDN 대비 약 1.74배 더 빠른 결과를 얻었고 평균적으로 4배 확대배율에 대한 모든 테스트 셋에서 약 1.8배 더 빠른 결과를 얻었다.

표 4. 학습 시간 및 처리시간에 대한 성능 비교
 Table 4. Performance comparison for training time and processing time

Urban100, x4, 'img007'	RDN	OURS
Training Time	8h 23m 24s	4h 45m 6s
Processing Time(sec)	0.033	0.019
PSNR(dB)	29.89	29.66

표 5는 재귀 블록의 수와 재귀 횟수에 따른 파라미터와 성능을 Urban100 테스트 셋을 이용하여 비교한 표이다. RB는 재귀 블록의 수를 의미하고 RDB는 재귀 블록 내 RDCAB를 재귀적으로 사용한 횟수를 의미한다. 재귀 블록의 수를 늘려 사용하게 되면 RDN과 동일하게 블록을 다수 쌓은 형태가 되어 재귀 횟수를 1로 고정하고 재귀 블록의 수를 1개씩 증가시켰을 때 재귀 블록의 수가 증가할수록 미세한 성능 향상을 만들어 내지만 약 140만개 이상의 파라미터의 수 증가와 느린 처리속도를 가져 경량화의 목적을 얻을 수 없다. 따라서 성능과 효율성 사이에 균형을 이루기 위해 재귀 블록의 수 1개와 재귀 횟수 9번을 사용하는 것이 적절하다. 재귀 횟수를 3으로 설정한 결과는 2,087,207개 파라미터 수로 25.97dB 성능을 만들어 냈고 재귀 횟수를 9로 설정한 결과는 2,111,783개 파라미터 수로 26.18dB 성능을 만들어 내어 재귀 횟수를 9로 설정한 결과가 재귀 횟수를 3으로 설정한 결과보다 파라미터의 수는 24,576개 정도 많았지만 성능은 0.21dB 더 높은 것을 확인할 수 있다. 따라서 재귀횟수를 늘려 사용하게 되면 각 재귀마다 나온 특징들을 더 많이 사용할 수 있어 성능이 향상됨을 볼 수 있다.

표 5. 재귀 블록의 수와 재귀 횟수에 따른 파라미터 수, 처리시간, 성능 비교

Table 5. Comparison of number of parameters, processing time, and performance according to the number of recursive blocks and the number of recursion

Urban100, ×4	RB1,RDB1	RB2,RDB1	RB2,RDB2	RB3,RDB1	RB1,RDB3	RB1,RDB6	RB1,RDB9
Parameter	2,079,015	3,489,739	3,497,931	4,900,463	2,087,207	2,099,495	2,111,783
Processing Time(sec)	0.0048	0.0117	0.0199	0.0236	0.0081	0.0131	0.0195
PSNR(dB)	25.60	26.08	26.22	26.26	25.97	26.14	26.18

표 6은 제안된 네트워크에 사용된 채널 집중 기법과 전역 잔여 학습의 사용 여부에 따른 성능 비교를 나타낸다. 4배 확대 배율 Set5 테스트 셋으로 부터 채널 집중 기법을 사용하지 않았을 때 32.26dB 성능을 만들어 냈고 전역 잔여 학습을 사용하지 않았을 때 32.28dB 성능을 만들어 냈으며

모든 기법을 적용한 경우 32.32dB 성능을 만들어 각 기법을 사용하지 않았을 때보다 향상된 성능을 보여주었다. 채널 집중 기법 같은 경우 잔여 밀집 구조에 있어 계층적인 특징들에 대해 중요도를 재정의 해줘 중요 특징들을 이용하여 성능이 향상 될 수 있음을 보여주고 전역 잔여 학습 같은 경우 깊은 네트워크를 학습하는데 있어 잔여 특징들을 더해줘 학습 시 기울기 소실(Gradient Vanishing)문제를 뿐만 아니라 성능도 향상 될 수 있음을 보여준다.

표 6. 기법에 따른 성능 비교

Table 6. Comparison of performance according to technique

Set5, ×4	Original	without Channel Attention	without Global Residual Learning
PSNR(dB)	32.32	32.26	32.28

그림 6, 그림 7, 그림 8은 각각 확대 배율 2배, 3배, 4배에 대해서 기존의 네트워크들과 결과를 비교한 그림으로 주관적으로 비교하기 위해 확대한 영상을 바탕으로 화질과 디



그림 6. 확대율 r=2에서 기존 네트워크들과 제안된 네트워크 비교

Fig. 6. Comparison of existing networks and proposed networks at scaling factor r=2



그림 7. 확대율 r=3에서 기존 네트워크들과 제안된 네트워크 비교

Fig. 7. Comparison of existing networks and proposed networks at scaling factor r=3

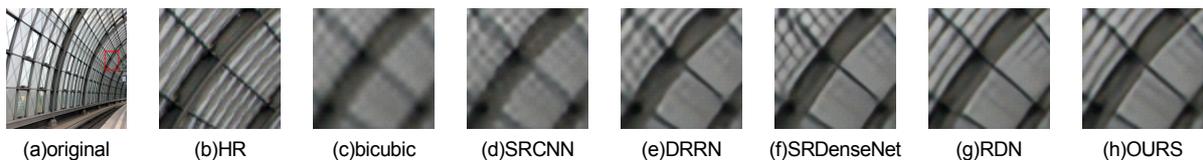


그림 8. 확대율 r=4에서 기존 네트워크들과 제안된 네트워크 비교

Fig. 8. Comparison of existing networks and proposed networks at scaling factor r=4

테일을 비교하였다. HR은 네트워크를 통과하여 나온 영상과 비교하기 위해 original 영상에서 빨간색 박스로 표시한 부분을 크롭(Crop)을 통해 획득한 확대 영상이다. 16개의 화소 값을 이용하여 결과 영상의 화소의 값을 계산하는 bi-cubic^[4]같은 경우 모든 확대 배율에서 저주파수 특성 때문에 뭉롱화한 것을 확인 할 수 있다. 그림 6에서 보는 바와 같이 2배 확대 배율에서 기존의 네트워크는 'y'와 'o' 사이가 선으로 연결되어있고 열화가 존재하는데 제안된 네트워크는 'y'와 'o' 사이가 연결되어 있지 않은 것을 확인 할 수 있다. 그림 7에서 보는 바와 같이 3배 확대 배율에서 기존의 네트워크는 왼쪽 상단의 창문을 나누는 창틀을 휘어져 있거나 알아볼 수 없는 창틀로 복원하였는데 제안된 네트워크는 5개의 창문을 나누는 창문의 창틀을 복원한 것을 확인 할 수 있다. 그림 8에서 보는 바와 같이 4배 확대 배율에서 기존의 네트워크는 사선 방향과 반대의 사선 무늬로 복원을 하여 체커보드 형태를 띄우고 있지만 제안된 네트워크는 사선 방향의 선만 복원 한 것을 확인 할 수 있다. 전체적으로 제안된 방식은 RDN과 비교하여 적은 파라미터와 빠른 처리속도에서 주관적으로 비슷한 화질을 얻을 수 있음을 확인할 수 있었다.

V. 결론

국내외 시장에서는 고해상도의 스마트폰 및 모니터 제품의 점유율이 증가하고 있으며 저해상도 이미지를 고해상도 이미지로 변환하여 얻은 고품질의 이미지를 사용자는 즐길 수 있게 되었다. 또한 고해상도의 이미지는 반도체 및 부품 결합 검사의 정확성을 높여 줄 수 있고 인간과 밀접한 의료에도 응용되어 의료 이미지를 개선해 인간의 생명 연장에 도 도움이 되는 기술로 발전할 수 있으며 더 나아가 자율주행 자동차, 지능로봇, 드론, 군사, 위성 이미지 분석 등에 핵심적인 기술로 사용이 될 수 있다.

심층 학습법 기반 초해상도 기법은 최근 합성곱 신경망의 기술 발달로 고해상도의 타겟 이미지를 실측 정보로 정의하고 뭉롱화, 잡음 등을 이용해 저해상도 이미지로 만든 뒤 최적화 방법 및 손실 함수 기반으로 학습되는 모종의 네트워크를 통해 저해상도 이미지를 실측 정보로 복원 시

키도록 네트워크를 학습시키는 구조를 따르게 된다.

본 논문에서는 재귀를 이용한 잔여 밀집 및 채널 집중 기법을 사용한 경량화된 네트워크 기반의 단일 이미지 초해상도 기법을 제안하였다. 기존의 잔여 밀집 블록을 사용한 RDN은 이전의 특징 정보를 다시 사용하는 것을 통해 정보의 흐름을 유연하게 하여 성능을 개선하였다. 하지만 다수의 RDB를 쌓게 되면 좋은 성능에도 불구하고 많은 파라미터의 수와 연산량을 가져 느린 처리 속도와 학습하는데 많은 시간이 소요되고 큰 저장 공간이 필요하여 모바일 시스템에 적용하는데 있어 단점을 가지고 있다. 따라서 이전의 정보를 연결하여 다시 사용하는 연속 메모리 구조인 잔여 밀집 구조와 이미지의 특징맵에 따라 중요도를 결정해주는 채널 집중 기법을 이용한 잔여 밀집 채널 집중 블록을 재귀적인 방식으로 사용하여 추가적인 파라미터 없이 네트워크의 깊이를 늘려 큰 수용 영역을 얻었고 동시에 간결한 모델을 유지할 수 있는 방식을 제안하였다. 제안된 네트워크를 다양한 이미지를 통해 실험하였고 기존의 RDN과 비교했을 때 주관적으로 비슷한 화질의 영상을 복원하였지만 약 10배 정도 적은 파라미터의 수와 약 1.74배 더 적은 연산량을 가지고 기존대비 약 1.8배 더 빠른 처리 속도를 갖는 것을 확인 하였다.

본 논문에서 제안된 RDCAB-RecursiveSRNet 기법은 디지털 TV, 보안 카메라, 이미지 편집 분야, 의료 이미지 등에 널리 이용 될 수 있다.

참 고 문 헌 (References)

- [1] X. Zhang and X. Wu, "Image Interpolation by Adaptive 2-D Autoregressive Modeling and Soft-Decision Estimation," in IEEE Transactions on Image Processing, pp. 887-896, 2008
- [2] Y. Romano, M. Protter and M. Elad, "Single Image Interpolation Via Adaptive Nonlocal Sparsity-Based Modeling," in IEEE Transactions on Image Processing, pp. 3085-3098, 2014
- [3] W. Ye and K. Ma, "Convolutional Edge Diffusion for Fast Contrast-guided Image Interpolation," in IEEE Signal Processing Letters, pp. 1260-1264, 2016
- [4] H. A. Aly and E. Dubois, "Image up-sampling using total-variation regularization with a new observation model," in IEEE Transactions on Image Processing, pp. 1647-1659, 2005
- [5] K. Zhang, X. Gao, D. Tao and X. Li, "Single Image Super-Resolution With Non-Local Means and Steering Kernel Regression," in IEEE

- Transactions on Image Processing, pp. 4544-4556, 2012
- [6] V. Pappas and M. Elad, "Multi-Scale Patch-Based Image Restoration," in IEEE Transactions on Image Processing, pp. 249-261, 2016
- [7] 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 Transactions on Image Processing, pp. 90-106, 2017
- [8] Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in CVPR, pp. 2472-2481, 2018.
- [9] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in CVPR, pp. 2790-2798, 2017.
- [10] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," in CVPR, pp. 2261-2269, 2017
- [11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," In ICLR, 2015
- [12] 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," in CVPR, pp. 1874-1883, 2016.
- [13] Bevilacqua, A. Roumy, C. Guillemot and M.-L. A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," In Proceedings British Machine Vision Conference, pp. 135.1-135.10, 2012.
- [14] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," In International conference on Curves and Surfaces, pp. 711-730, 2012.
- [15] 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," In International Conference on Computer Vision, pp. 416-423, 2001.
- [16] J.-B. Huang, A. Singh, and N. Ahuja, "Single Image Super-Resolution From Transformed Self -Exemplars," In Conference on Computer Vision and Pattern Recognition, pp. 5197-5206, 2015.
- [17] C. Dong, C. C. Loy, K. He, and X. Tang. "Image Super-Resolution Using Deep Convolutional Networks," In Transactions on Pattern Analysis and Machine Intelligence, pp. 295-307, 2016.
- [18] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super resolution using very deep convolutional networks," in CVPR, pp. 1646-1654, 2016.
- [19] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," ECCV, pp. 286-301, 2018.

저 자 소 개



우 희 조

- 2019년 : 한국산업기술대학교 전자공학부 공학사
- 2020 ~ 현재 : 한국산업기술대학교 전자공학부 공학석사
- ORCID : <https://orcid.org/0000-0003-4956-6809>
- 주관심분야 : 딥러닝기반 영상처리, 경량화, 컴퓨터 비전, 임베디드 시스템



심 지 우

- 2019년 : 한국산업기술대학교 전자공학부 공학사
- 2020 ~ 현재 : 한국산업기술대학교 전자공학부 공학석사
- ORCID : <https://orcid.org/0000-0003-1964-3615>
- 주관심분야 : 딥러닝기반 영상처리, 컴퓨터 비전, 임베디드 시스템

저 자 소 개



김응태

- 1991년 : 인하대학교 전자공학과 공학사
- 1993년 : KAIST 전기및전자공학과 공학석사
- 1999년 : KAIST 전기및전자고학과 공학박사
- 1998년 3월 ~ 2004년 2월 : ㈜LG전자 DTV연구소 책임연구원
- 2004년 3월 ~ 현재 : 한국산업기술대학교 전자공학부 교수
- ORCID : <https://orcid.org/0000-0001-5984-0045>
- 주관심분야 : 멀티미디어 신호처리, DTV SOC, 지능형 영상감시 시스템, 딥러닝기반 영상처리