

일반논문 (Regular Paper)

방송공학회논문지 제27권 제3호, 2022년 5월 (JBE Vol.27, No.3, May 2022)

<https://doi.org/10.5909/JBE.2022.27.3.378>

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

정보 증류 및 재귀적인 방식을 이용한 심층 학습법 기반 경량화된 초해상도 네트워크

우 희 조^{a)}, 심 지 우^{a)}, 김 응 태^{a)†}

Lightweight Super-Resolution Network Based on Deep Learning using Information Distillation and Recursive Methods

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

요 약

최근 심층 합성곱 신경망 학습의 발전에 따라 단일 영상 초해상도에 적용되는 심층 학습 기법들을 좋은 성과를 보여주고 있으며 깊은 네트워크의 강한 표현 능력으로 저해상도 영상과 고해상도 영상 사이의 복잡한 비선형 매핑이 가능해졌다. 하지만 과도한 합성곱 신경망의 사용으로 인해 증가하는 파라미터와 연산량으로 실시간 또는 저전력 장치에 적용하는데 제한이 있다. 본 논문은 정보 증류 방식을 이용하여 계층적인 특징을 조금씩 추출해내는 블록을 재귀적인 방식으로 사용하며 고주파수 잔여 정제 블록을 통해 더 정확한 고주파수 성분을 만들어 성능을 향상시키는 경량화된 네트워크인 Recursive Distillation Super Resolution Network (RDSRN)를 제안한다. 제안하는 네트워크는 RDN과 비교했을 때 비슷한 화질의 영상을 복원하며 약 32배 적은 파라미터와 약 10배 적은 연산량을 가지고 약 3.5배 더 빠르게 영상을 복원하며 기존 경량화 네트워크 CARN과 비교했을 때 약 2.2배 적은 파라미터와 약 1.8배 빠른 처리시간으로 평균 0.16dB 더 좋은 성능을 만들어 냈을 확인 하였다.

Abstract

With the recent development of deep composite multiplication neural network learning, deep learning techniques applied to single-image super-resolution have shown good results, and the strong expression ability of deep networks has enabled complex nonlinear mapping between low-resolution and high-resolution images. However, there are limitations in applying it to real-time or low-power devices with increasing parameters and computational amounts due to excessive use of composite multiplication neural networks. This paper uses blocks that extract hierarchical characteristics little by little using information distillation and suggests the Recursive Distillation Super Resolution Network (RDSRN), a lightweight network that improves performance by making more accurate high frequency components through high frequency residual purification blocks. It was confirmed that the proposed network restores images of similar quality compared to RDN, restores images 3.5 times faster with about 32 times fewer parameters and about 10 times less computation, and produces 0.16 dB better performance with about 2.2 times less parameters and 1.8 times faster processing time than the existing lightweight network CARN.

Keyword : Deep Learning, Super-Resolution, Recursion, Information Distillation, High-Frequency Refinement

Copyright © 2022 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. 고해상도 이미지는 감시 카메라 및 블랙박스 영상의 화질 개선 등에 응용되어 범죄나 교통에 큰 도움이 될 수 있고 스마트폰의 화질 개선과 반도체 및 부품 결함 검사의 정확성을 높여줄 수도 있다. 또한 인간과 밀접한 의료에도 응용되어 의료 영상을 개선해 인간의 생명 연장에도 도움이 되는 기술로 발전할 수 있으며 더 나아가 자율주행 자동차, 지능로봇, 드론, 군사, 위성영상 분석 등에 핵심적인 기술로 사용될 수 있다.

단일 이미지의 해상도를 높이는 대표적인 접근 방식으로 크게 보간법^[1-2], 재구성법^[3-4] 및 심층 학습법^[6-8,10,11-14,17] 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^[11], 잔여학습^[12], 재귀^[7], 밀집^[6], 집중^[13], 중속

^[14]기법 등의 다양한 방식으로 발전됐으며 다양한 네트워크가 제안되었다.

밀집하게 연결된 네트워크 구조들 중 대표적인 Zhang et al^[6] 등의 RDN (Residual Dense Network)는 그전의 모든 계층의 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 합성곱 신경망을 사용하는 잔여 밀집 블록을 다수 쌓게 되면 좋은 성능에도 불구하고 많은 파라미터 수와 연산량을 가지게 되어 네트워크를 학습하는데 많은 시간이 소요되고 큰 저장 공간이 필요하여 실시간 또는 저전력 컴퓨터에 적용하는데 비효율적이다.

본 논문에서는 이러한 문제를 해결하기 위해 정보 증류 (Information Distillation^[17] 및 재귀적인 방식 (Recursive Methods)^[7]을 이용한 심층 학습법 기반 경량화된 초해상도 네트워크 Recursive Distillation Super-Resolution Network (RDSRN)을 제안한다. 정보 증류 방식을 통해 합성곱에 들어가는 채널수를 조절하여 파라미터 수와 연산량을 감소시키며 동시에 Wider Channel (WC)을 적용하여 깊은 층으로 내려갈수록 활성화 함수로 인해 발생하는 특징 역제를 제어하는 블록을 설계한 뒤 추가적인 파라미터 없이 네트워크 공유를 통해 깊이를 늘려 큰 수용 영역을 얻으며 동시에 간결한 모델을 유지시키는 재귀적인 방식을 사용하여 적은 파라미터 수와 연산량을 가지며 정확한 초해상도 영상을 복원하도록 하였다.

본 논문은 다음과 같이 구성된다. 본 논문의 2장에는 관련 이론을 소개하고 3장에서는 제안된 네트워크인 Recursive Distillation Super Resolution Network (RDSRN)의 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^[11] 등의 SRCNN (Image Super-

a) 한국공학대학교 전자공학부(Department of Electronics Engineering, Tech University of KOREA)

‡ Corresponding Author : 김응태(Eung-Tae Kim)
E-mail: etkim@tukorea.ac.kr
Tel: +82-31-8041-0488

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

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

※ This work was partly supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea Government(MSIT) and Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (P0008458, HRD Program for Industrial Innovation).

· Manuscript April 7, 2022; Revised May 10, 2022; Accepted May 10, 2022.

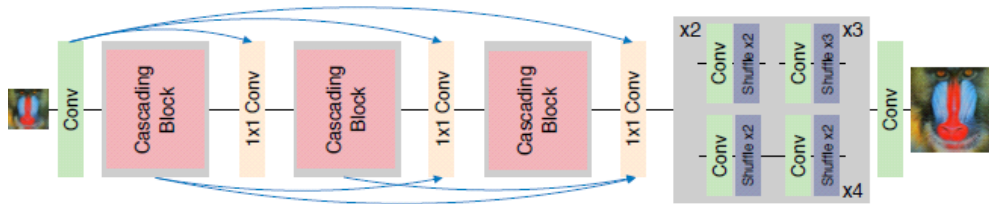


그림 1. Cascading Residual Network 구조
Fig. 1. The Architecture of Cascading Residual Network

Resolution Using Deep Convolutional Networks)같은 경우 단 3개의 합성 곱만 사용하여 딥러닝을 적용하지 않은 방법들에 비해 높은 성능을 만들어 냈고 초해상도에도 딥러닝을 적용할 수 있다는 가능성을 보여주었다. 하지만 Dong의 방법은 사전 업 샘플링 구조를 가져 많은 양의 계산량이 필요하고 적은 수의 층으로 정확도 측면에서 제한적인 단점을 가지고 있다.

모든 층에서의 정보를 취득하는 형태인 G.Huang^[8] 등이 제안한 텐스넷 (Dense Net)을 이용하여 Zhang et al^[6] 등이 잔여 밀집 네트워크 (Residual Dense Network, RDN)을 제안했다. RDN에 사용된 RDB는 밀집하게 연결되어 있는 합성 곱 층으로부터 지역 특징을 추출하고 이전의 RDB의 출력이 현재 RDB의 모든 층의 입력으로 들어가게 설계되어 있다. 지역 특징 융합 (Local Feature Fusion, LFF)을 통해 이전의 RDB의 출력과 지역 특징을 충분히 이용하여 학습할 수 있도록 되어있고 밀집 지역 특징을 학습 한 후 전역 특징 융합을 통해 전역 계층 특징을 학습하게 된다. 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 정보

들을 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 소요되고 큰 저장 공간이 필요하여 모바일 시스템에 적용이 어려운 문제가 생긴다.

이와 같은 문제를 해결하기 위해 경량화된 초해상도 네트워크들이 제안되었으며 그 중 종속 구조를 이용하여 Ahn et al^[14] 등이 종속 잔여 네트워크 (Cascading Residual Network, CARN)을 제안했다. CARN은 ResNet^[15]을 기반으로 만들어졌고 그림 1과 같이 구성되어 있으며 ResNet과는 다르게 전역 및 종속 모듈이 존재한다.

종속 블록끼리 사용되는 전역 종속 연결과 종속 블록 내 사용되는 지역 종속 연결의 사용은 다수의 층으로부터 특징을 통합시켜 계층적인 표현을 학습하여 하위 계층에서 상위 계층으로 정보를 빠르게 전파할 수 있다.

종속 블록에 사용된 그림 2의 개선된 잔여 블록은 MobileNet^[16]에서 사용된 접근 방식과 유사하지만 깊이별 합

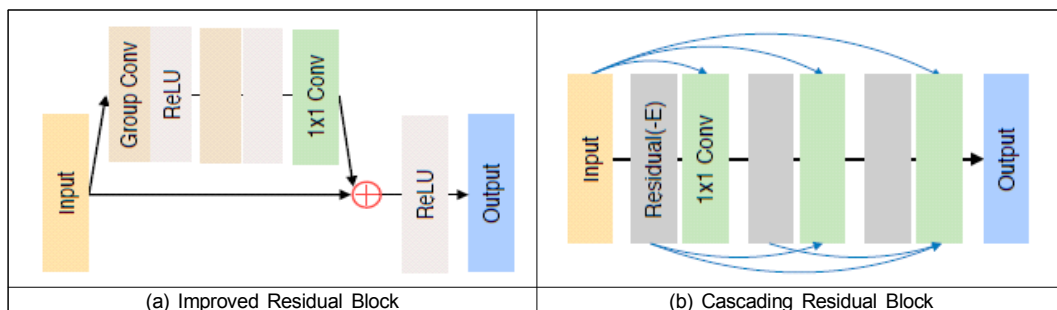


그림 2. 개선된 잔여 블록과 종속 잔여 블록 구조
Fig. 2. Improved Residual Block and Cascading Residual Block Structure

성곱 (Depthwise Convolution) 대신 2개의 그룹 합성곱 (Group Convolution)과 1개의 위치별 합성곱 (Pointwise Convolution)을 사용하여 모델의 효율성을 더 쉽게 조정할 수 있게 만들었다.

CARN은 전역 종속과 지역 종속을 사용하여 계층적인 특징을 통합한 뒤 사용하여 기울기 (Gradient) 값을 효과적으로 사용하였고, 개선된 잔여 블록을 이용하여 좋은 성능을 유지하며 기존의 잔여 블록대비 파라미터 수와 연산량을 감소시켰다. 하지만 CARN은 종속 모듈을 사용하여 파라미터 수와 연산량을 줄이긴 하였으나 성능 또한 다수 줄어들어 성능과 효율성 사이에서 균형을 이루고 있지 못하다.

크고 무거운 모델의 특징 정보를 작고 가벼운 모델로 전달하여 더 정확한 추론을 하도록 학습시키는 정보 증류 구조를 이용하여 Hun et al.^[17]등이 정보 증류 네트워크 (Information Multi Distillation Network, IMDN)을 제안했다. IMDN은 채널 분할을 통해 유용한 특징을 조금씩 추출하여 경량화를 하였다. 하지만 동일하지 않은 채널 분할로 인해 정보 손실 및 Identity Mapping에 한계점을 가지고 있다.

III. 제안된 RDSRN 초해상도 기법

본 연구는 깊은 네트워크가 갖는 많은 파라미터 수와 연산량으로 인해 발생하는 오랜 학습 시간과 추론 시간으로 개발 지연 등의 문제를 해결하기 위해 모델의 성능 손실을 기존 모델 대비 최소화하면서 모델 크기와 연산량을 크게 줄이며 요구 메모리와 에너지, 요구 연산량 등 여러 면에서

학습 및 추론의 효율성을 높일 수 있는 경량화된 초해상도 네트워크를 제안한다. 그림 3은 제안하는 경량화된 초해상도 네트워크인 Recursive Distillation Super Resolution Network (RDSRN)의 구조를 보여준다.

제안하는 방법은 파라미터 수와 연산량을 줄이기 위해 정보 증류 구조^[17]와 재귀적인 방식^[7]을 이용하며, 높은 성능을 만들기 위해 Wider Channel을 적용하여 학습 시 특징 억제를 시키고 고 주파수 잔여 정제 블록을 설계하여 정확한 고주파수 성분을 이용한다. 제안된 네트워크는 두 개의 합성곱을 이용하는 얇은 특징 추출 단계, 주어진 재귀 횟수 만큼 사용하는 재귀 블록을 통한 특징 추출 단계, 업 스케일 단계, 정확한 고 주파수 성분을 획득하기 위한 고 주파수 잔여 정제 및 복원 (High Frequency Residual Refinement & Reconstruction Block, HFRRB) 단계로 구성되어 있다.

제안된 RDSRN은 추가적인 파라미터 없이 네트워크 공유를 통해 깊이를 늘려 큰 수용 영역을 얻으며 동시에 간결한 모델을 유지시키는 재귀적인 방식^[7]을 소수의 재귀 횟수로 사용하여 적은 파라미터 수와 연산량을 가지며, 재귀적으로 사용하는 공유 네트워크는 정보 증류 방식^[17]을 통해 계층적인 특징을 조금씩 추출하고 Wider Channel (WC)를 적용하여 깊은 층으로 내려갈수록 활성화 함수로 인해 발생하는 특징 억제를 막고 업 스케일을 통해 출력된 디테일이 부족한 고 주파수 성분에 대하여 정확한 고 주파수 성분을 만들어내기 위한 고 주파수 잔여 정제 블록 (High Frequency Residual Refinement Block, HFRRB)을 사용한다.

원본 영상에 대해 몽롱화, 잡음 등을 이용하여 획득한 저해상도 영상이 입력으로 들어오면 얇은 특징 추출 단계를 거쳐 영상의 특징 맵을 추출하고 재귀 블록을 통한 특징

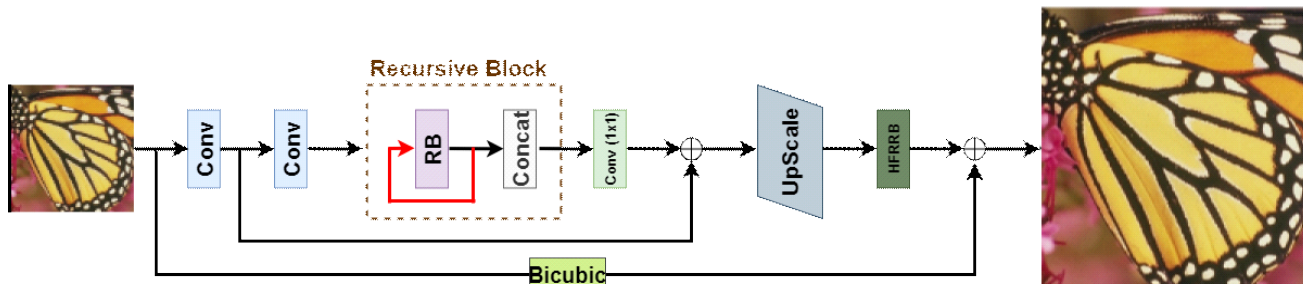


그림 3. 제안된 RDSRN 구조
Fig. 3. The Architecture of Proposed RDSRN

추출 단계로 들어간다. 재귀 블록을 통한 특징 추출 단계는 주어진 재귀 횟수만큼 공유된 네트워크를 통과한 뒤 각 재귀 횟수마다 나온 특징들은 하나로 연결하여 출력한다. 업스케일 단계는 ESPCN의 sub-pixel 합성 곱을 통해 원본 영상과 동일한 크기를 갖는 영상으로 특징을 확대 시키는 학습을 진행한다. 마지막 단계로 고 주파수 잔여 정제 및 복원 단계는 업스케일을 통해 출력된 디테일이 부족한 고 주파수 특징에 대해 추가 학습을 진행하고 RGB로 구성되어 있는 3채널 영상으로 복원한다.

1. 얇은 특징 추출 단계

입력으로 들어온 저해상도의 영상이 I_{LR} 이고 네트워크를 통과하여 만들어진 초해상도 영상을 I_{SR} 이라 할 때 다음 식과 같이 나타낼 수 있으며 F_{RDSRN} 은 전체 네트워크를 표현하는 함수이다.

$$I_{SR} = F_{RDSRN}(I_{LR}) \quad (1)$$

제안된 네트워크는 복잡한 연산량의 증대를 막기 위해 사후 업 샘플링 구조로 이루어져 있다. 원본 영상에 대해 모롱화, 잡음 등을 이용하여 획득한 저해상도 영상이 입력되게 되면 다음 식과 같이 첫 번째와 두 번째의 3×3 표준 합성 곱을 통해 각각 64개의 얇은 특징을 추출하게 된다. 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 출력이다. 첫 번째 합성 곱의 출력은 전역 잔여 학습에 사용이 되고 두 번째 합성 곱의 출력은 재귀 블록 내 지역 잔여 학습에 사용이 된다.

전역 잔여 학습과 지역 잔여 학습을 사용하여 추가적인 파라미터 없이 잔여 특징들을 더해주어 학습 시 기울기 소실 문제와 성능에 기여할 수 있도록 구성하였다.

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

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

2. 재귀 블록을 통한 특징 추출 단계

그림 4는 제안된 RDSRN에 사용된 재귀 블록을 보여준다. 적은 파라미터와 연산량을 가지며 빠른 처리시간으로 높은 성능을 만들어내기 위해 정보 증류를 통해 조금씩 계층적인 특징을 추출해내며, 깊은 층으로 내려갈수록 활성화 함수로 인해 발생하는 특징 억제를 막기 위해 Wider Channel (WC)을 적용한 Distillation Wider Channel Block (DWCB)와 채널들 사이의 중요도에 따라 채널을 재정의해주는 채널 집중^[13] 기법으로 구성되어 있는 재귀 블록을 설계하여 공통 네트워크로 설정한 뒤 재귀적으로 사용한다.

입력으로 들어온 64개 특징은 첫 번째 DWCB를 통해 64개 특징을 추출한 뒤 블록에 입력된 값과 연결되어 128개의 특징을 만들어 내고 얇은 특징 추출 단계에서 두 번째 합성 곱의 출력과 지역 잔여 학습을 진행하며 그 뒤 1×1 위치별 합성 곱 (Pointwise Convolution)을 통해 64개의 특징으로 축소한다. 두 번째 DWCB에 입력으로 들어온 특징에 대해서도 64개의 특징을 추출하고 이전 DWCB의 입력과 출력이 연결되었던 128개의 특징과 연결하여 192개의 특징을 만들어내며 지역 잔여 학습을 진행한 뒤 1×1 위치

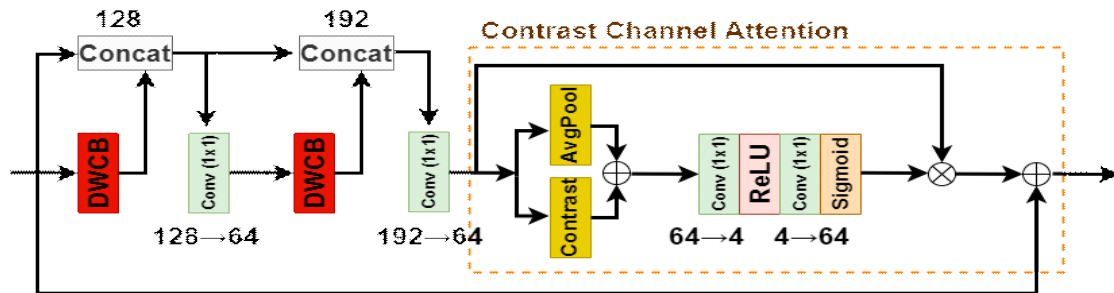


그림 4. 제안된 Recursive Block 구조

Fig. 4. The Architecture of Proposed Recursive Block

$$F_{DWCB} = H_{PW_2}((H_{DWCB_2}(H_{PW_1}(H_{DWCB_1}(F_{r-1}) + F_0)) + F_0)) \quad (4)$$

별 합성 곱을 통해 64개의 특징으로 축소한다. r 번째 재귀 블록에서 두 개의 채널 증류 블록을 통과하게 되면 다음 식과 같이 나타낼 수 있다.

F_{r-1} 은 r 번째 재귀블록에 입력으로 들어온 특징을 의미하며 H_{DWCB_1} 과 H_{DWCB_2} 는 각각 첫 번째와 두 번째 채널 증류 블록을 나타내는 함수이다. H_{PW_1} 과 H_{PW_2} 는 1×1 위치별 합성 곱을 의미하며 F_0 는 얇은 특징 추출 단계의 두 번째 합성 곱의 출력을 의미하고 F_{DWCB} 는 두 개의 채널 증류 블록과 두 개의 1×1 위치별 합성 곱을 통과한 출력을 의미한다.

전체 영역의 정보를 활용하기 위해 각 채널들 사이의 중요도에 따라 가중치를 할당하는 채널 집중 기법^[13]을 전역 평균 풀링 (Global Average Pooling, GAP), 영상의 화소 분포가 더 넓은 영역에 걸쳐서 퍼지도록 하는 Contrast와 Sigmoid를 사용하여 구성하였다. 입력으로 들어온 채널을 감소율 16을 이용하여 나누고 다시 곱해주는 방식을 이용하였으며 재귀 블록의 최종 출력은 채널 집중 층을 통과한 출력과 재귀 블록의 입력 값이 지역 잔여 학습을 진행한 뒤 출력 된다. 두 개의 채널 증류 블록을 통과한 F_{DWCB} 특징이 채널 집중 층으로 입력된 뒤 얇은 특징 추출 단계의 두 번째 합성 곱 F_0 와 지역 잔여 학습을 하여 최종 재귀 블록의 출력을 만들어 내게 되면 다음 식과 같이 나타낼 수 있다.

$$F_r = (R_{CCA}(F_{DWCB}) \times F_{DWCB}) + F_0 \quad (5)$$

여기서 R_{CCA} 는 전역 평균 풀링, Contrast, Sigmoid를 사용하는 채널 집중 층을 나타내며 F_r 은 재귀 블록의 최종 출력을 나타낸다.

표준 합성 곱 신경망과 정보 증류 방식을 이용한 합성 곱의 연산량을 비교하게 되면 k_x 와 k_y 를 합성 곱 커널의 크기, i 번째 합성 곱 층에서 m_i 와 n_i 를 입력과 출력의 채널의 수, h_i 와 w_i 를 특징 맵의 높이와 너비라고 할 때 표준 합성 곱 신경망의 연산량은 다음 식과 같이 나타낼 수 있다.

정보 증류 방식을 이용한 합성 곱의 연산량은 0에서

$$Operations = k_x k_y m_i n_i h_i w_i \quad (6)$$

1 사이의 값을 갖는 채널 분할 비율 (Channel Split Ratio, CSR)이 주어졌을 때 식 7과 같이 나타낼 수 있다. 따라서 표준 합성 곱 신경망과 정보 증류 방식을 이용한 합성 곱의 연산량을 비교하게 되면 식 8과 같이 주어진 만큼 연산량을 줄일 수 있음을 확인할 수 있다.

$$\begin{aligned} Refine &= m_i \times CSR \\ Retain &= m - (Refine) \\ Operations_{Distillation} &= k_x k_y (Refine) n_i h_i w_i \end{aligned} \quad (7)$$

$$\frac{k_x k_y (m_i \times CSR) n_i h_i w_i}{k_x k_y m_i n_i h_i w_i} = CSR \quad (8)$$

그림 5는 Distillation Wider Channel Block (DWCB) 구조를 나타낸다. DWCB로 입력된 64개의 특징은 3×3 표준 합성 곱을 통해 특징을 추출한 뒤 입력으로 들어온 특징과 잔여 학습을 진행한다. 정보 증류를 통해 주어진 초매개변수만큼 정제 특징과 유지 특징으로 나눈 후 정제 특징에 대해서 Wider Channel (WC)을 적용한다. 첫 번째 소 블록에서는 3배로 특징을 키워 활성화 함수를 통과한 뒤 다시 3배로 특징을 줄이며 활성화 함수로 인해 발생하는 특징 역제를 막는다. 그 뒤 정제 특징의 입력과 잔여 학습을 진행하고 유지 특징과 연결시켜 이전 특징에서 일부 특징을 유지하며 정제된 특징을 동시에 얻을 수 있도록 구성하였다.

두 번째 소 블록은 4배로 WC을 적용하고 첫 번째 소 블록과 동일하게 동작을 한다. DWCB는 이후의 계층들이 이전의 계층들의 입력 정보를 잘 사용하기 위해 밀집 연결로 두 개의 소 블록이 연결되어 있으며 정보 증류를 통해 정제 특징과 유지 특징을 나눠 이전의 특징을 유지하며 동시에 WC을 적용해 정제 특징에 대해 더 정확한 학습을 진행하였다. DWCB는 다수의 합성 곱 층을 쌓는 것이 아닌 정보 증류를 통해 일부 특징을 유지하며 학습시킨 뒤 연결하여 사용함으로 적은 파라미터와 연산량을 가질 수 있다.

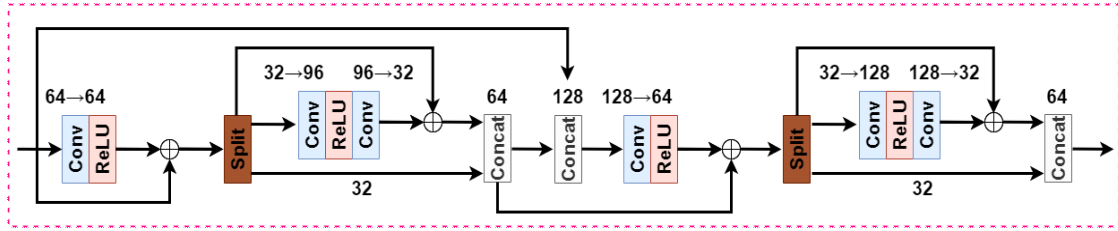


그림 5. 제안된 Distillation Wider Channel Block 구조

Fig. 5. The Architecture of Proposed Distillation Wider Channel Block

$$I_{SR} = H_{Reconstruction} (R(H_3(R(H_3(R(H_3(R(H_3(F_{upscale})))))))) \quad (9)$$

3. 업 스케일 단계

W. Shi et al^[10] 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 변형시켜 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.

4. 고 주파수 잔여 정제 및 복원 단계

고 주파수와 저 주파수 특성을 결합하여 사용하게 되면 고화질의 영상으로 복원하는데 효과적이다. 따라서 업 스케일을 통해 출력된 디테일이 부족한 고 주파수 성분을 더 정확하게 만들기 위해 그림 6과 같이 추가적인 학습을 진행 하였다. 고 주파수 잔여 정제 및 복원 단계는 입력으로 들어온 특징에 대해 4개의 3×3 표준 합성 곱과 4개의 활성화 함수를 통과시킨 뒤 잔여 학습을 통해 고 주파수 성분에 대해 추가적인 학습을 진행함으로써 초해상도를 위한 더 정확한 고 주파수 성분을 만들어 내었고 복원 단계는 하나

의 합성 곱으로 구성되어 있으며 고 주파수 정제 단계를 거쳐 나온 특징들을 3×3 표준 합성 곱을 통해 RGB로 구성되어 있는 3채널 영상으로 복원한다.

고 주파수 잔여 정제 및 복원 단계는 식 9와 같이 나타낼 수 있다.

$F_{upscale}$ 은 업 스케일 단계를 통해 출력된 특징을 의미하며 H_3 와 R 은 각각 3×3 표준 합성 곱과 정류 선형 유닛 (Rectified Linear Unit, ReLU)을 의미하고 $H_{Reconstruction}$ 은 초해상도 영상으로 복원하는 함수를 나타낸다.

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 뭉뚱화, 잡음 등을 통해 다운 스케일 하여 저해상도 영상을 획득한 뒤 학습시키는데 사용하였다. 검증 셋으로 5장으로 이루어져있는 Set5를 이용하였다. 제안된 네트워크에 입력과 출력으로 3채널 RGB 이미지를 사용하였고 수

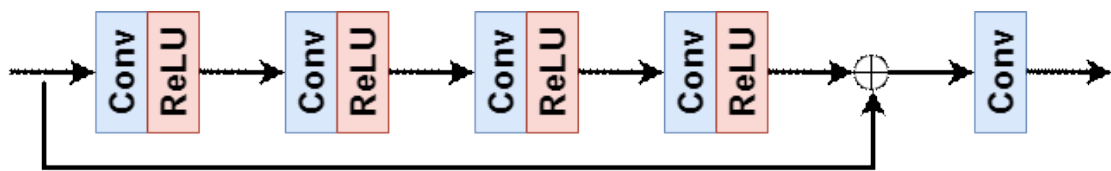


그림 6. 고 주파수 잔여 정제 및 복원 블록 구조

Fig. 6. The Architecture of Proposed High Frequency Residual Refinement & Reconstruction Block

평, 수직 뒤집기 및 90° 회전을 이용하는 데이터 증강 기술을 사용하였다. 입력으로 들어온 저해상도 이미지는 64×64 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 0.001을 초기 값으로 사용하였고 640에폭 이후에는 0.0001로 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 이용하였으며 그림 3의 제안된 RDSRN 구조 에서 Recursive Block (RB)를 3번 재귀적으로 사용하는 네트워크를 기본 모델로 사용하였다. 사용된 하드웨어 환경 은 Intel Xeon Gold 6240R 2.40GHz CPU와 RTX 3090 GPU 24G를 사용하였다.

표 1은 재귀 횟수와 채널 분할 비율에 따른 정량적 비교를 나타낸다. 재귀 손실을 막기 위해 다수의 재귀 횟수를 사용하게 되면 급격하게 증가하는 연산량으로 네트워크가 무거워지고 느린 학습 속도와 처리시간의 문제가 생기며 특징을 나누는 채널 분할 비율 (Channel Split Ratio, CSR)

또한 합성 곱으로 들어가는 채널수를 조절하기 때문에 경량화를 판단하는 요소와 성능에 많은 영향을 끼쳐 적절한 재귀 횟수와 CSR을 사용하는 것이 중요하다. 3번의 재귀 횟수와 0.5 채널 분할 비율이 다른 재귀 횟수와 다른 채널 분할 비율에 비해 가장 적게 연산량이 증가하며 빠른 처리 속도로 높은 성능 향상을 만들어 내는 것을 확인할 수 있다. 반면 9번의 재귀 횟수와 0.25 채널 분할 비율을 사용하게 되면 3번의 재귀 횟수와 0.25 채널 분할 비율 대비 각각 0.17dB, 0.07dB 성능 향상을 만들어내지만 채널 수 증가로 인해 파라미터 수와 연산량이 증가하여 약 2.8배, 약 1.1배 느린 처리시간을 갖게 된다.

그림 7의 (a), (b)는 확대율 $r=4$ 로 Set5 테스트 셋의 “baby” 영상을 복원할 때 3번째 재귀에서 DWCB의 두 번째 소블럭에 적용된 Wider Channel (WC)의 사용 여부에 따른 학습된 32개의 특징을 비교한 실험 결과이다. WC를 적용한 그림 7의 (a)는 WC를 적용하지 않은 그림 7의 (b)보다 더 넓은 채널로 학습을 진행하여 눈, 코, 입과 가장자리 성분을 더 뚜렷하게 학습한 것을 확인할 수 있으며 반면 WC를 적용하지 않은 영상은 깊은 네트워크로 내려갈수록 생기는 활성화 함수로 인한 특징 억제로 인해 뚜렷한 특징

표 1. 재귀 횟수와 채널 분할 비율에 따른 정량적 비교
Table 1. Comparison of PSNR, SSIM for the test set

Urban100, $\times 4$	재귀 횟수			채널 분할 비율		
	1	3	9	0.25	0.5	0.75
Parameters(M)	0.68	0.68	0.71	1.01	0.68	0.49
Multi-Adds(G)	77	136	223	192	136	102
Processing Time(sec)	0.0085	0.0112	0.0320	0.0123	0.0112	0.0126
PSNR(dB)	25.95	26.29	26.46	26.36	26.29	26.04

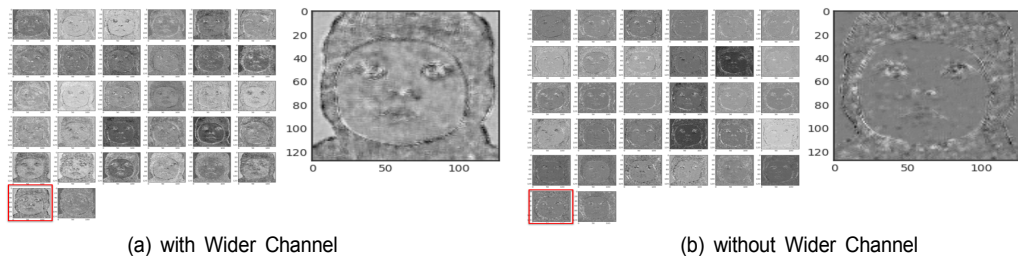


그림 7. Wider Channel의 사용 여부에 따른 학습된 특징 비교
Fig. 7. Comparison of learned features with or without Wider Channel

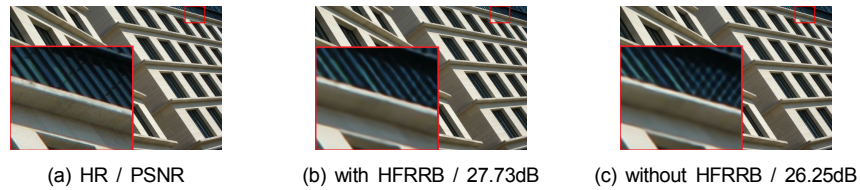


그림 8. HFRRB의 사용 여부에 따른 정성적, 정량적 비교
Fig. 8. Qualitative and quantitative comparison with or without HFRRB

들을 학습하지 못한 것을 확인 할 수 있다.

그림 8은 고 주파수 잔여 정제 블록 (High Frequency Residual Refinement Block, HFRRB)의 사용 여부에 따른 비교 실험 결과이다. 그림 8의 (a), (b), (c)는 각각 정답 값인 고해상도 영상, HFRRB를 사용한 영상, HFRRB를 사용하지 않은 영상이며 섬세하게 비교하기 위해 각 영상에 나타나 있는 빨간색 박스를 크롭하여 확대 후 삽입하였다. 확대된 크롭 영상을 시각적으로 비교하게 되면 HFRRB를 사용한 (b)는 정답 값인 (a)와 유사한 형태의 무늬를 복원하였지만 HFRRB를 사용하지 않은 (c)는 (a)에는 없는 격자무늬가 다수 나타나 있는 것을 확인 할 수 있으며, 정량적으로 비교하였을 때 HFRRB는 사용한 (b)가 사용하지 않은 (c)보다 1.48dB 높은 성능을 만들어 내었다. 따라서 고 주파수 잔여 정제 블록은 업 스케일을 통해 나온 디테일이 부족한 고 주파수 성분을 추가 학습 시켜 더 정확한 고 주파수 특징들을 만들어 낼 수 있음을 확인 할 수 있다.

표 2는 Set5, Set14, BSD100, Urban100 테스트 셋에 대하여 기존의 네트워크들과 제안된 네트워크의 최대 신호 대 잡음비, 구조유사도를 비교한 표이며 표 3은 각 네트워크의 파라미터와 1280 x 720 HD급 영상을 복원하는데 필요한 연산량을 나타내었고, 확대율 r=4로 Urban100 테스트

표 3. 각 네트워크의 파라미터, 연산량, 처리시간 비교

Table 3. Comparison of the number of parameters, multi-adds, and processing time of each network

Method	Parameters(M)	Multi-Adds(G)	Processing Time(sec)
SRCNN	0.057	52.7	0.0101
DRRN	0.29	6,796.9	0.0907
RDN	22	1,309	0.0401
RCAN	16	919.9	0.1841
CARN	1.5	90.9	0.0207
CBPN	1.1	97.9	0.0189
RDSRN(ours)	0.68	136	0.0114

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 (PSNR,SSIM)	Set14 (PSNR,SSIM)	BSD100 (PSNR,SSIM)	Urban100 (PSNR,SSIM)
SRCNN	×2	36.66 / 0.9542	32.45 / 0.9067	31.36 / 0.8435	29.51 / 0.8405
DRRN		37.74 / 0.9591	33.23 / 0.9136	32.05 / 0.8973	31.23 / 0.9188
RDN		38.24 / 0.9614	34.01 / 0.9212	32.34 / 0.9017	32.89 / 0.9353
RCAN		38.27 / 0.9614	34.12 / 0.9216	32.41 / 0.9027	33.34 / 0.9384
CARN		37.76 / 0.9590	33.52 / 0.9166	32.09 / 0.8978	31.92 / 0.9256
CBPN		37.90 / 0.9590	33.60 / 0.9171	32.17 / 0.8989	32.14 / 0.9279
RDSRN(ours)		38.08 / 0.9610	33.64 / 0.9181	32.33 / 0.9004	32.25 / 0.9294
SRCNN	×4	30.48 / 0.8628	27.50 / 0.7513	26.90 / 0.7103	24.52 / 0.7226
DRRN		31.68 / 0.8888	28.21 / 0.7720	27.38 / 0.7284	25.44 / 0.7638
RDN		32.47 / 0.8990	28.81 / 0.7871	27.72 / 0.7419	26.61 / 0.8028
RCAN		32.63 / 0.9002	28.87 / 0.7889	27.77 / 0.7436	26.82 / 0.8087
CARN		32.13 / 0.8937	28.60 / 0.7806	27.58 / 0.7349	26.07 / 0.7837
CBPN		32.21 / 0.8944	28.63 / 0.7813	27.58 / 0.7356	26.14 / 0.7869
RDSRN(ours)		32.34 / 0.8959	28.72 / 0.7848	27.66 / 0.7390	26.29 / 0.7918

트 셋을 복원할 때 Pytorch에서 제공하는 torch.cuda.Event library를 이용하여 영상을 복원하는데 걸리는 평균 처리시간을 나타내었다.

표 2, 3을 보게 되면 제안된 네트워크는 확대율 $r=4$ 일 때 Urban100 테스트 셋에서 약 0.68M 파라미터와 약 136G 연산량으로 26.29dB 성능을 0.0114sec 소요하여 영상을 복원하였다. 깊은 네트워크인 RDN, RCAN과 비교하였을 때 RDN은 약 22M 파라미터와 약 1,309G 연산량으로 26.61dB 성능을 0.0401sec 소요하여 영상을 복원하였으며 RCAN은 약 16M 파라미터와 약 919.9G 연산량으로 26.82dB 성능을 0.1841sec 소요하여 영상을 복원하였다. 제안된 네트워크가 RDN보다 0.32dB만큼 낮은 성능을 만들었지만 약 32배 적은 파라미터와 약 10배 적은 연산량을 갖으며 약 3.5배 더 빠르게 영상을 복원하였고 RCAN보다 0.53dB 만큼 낮은 성능을 만들었지만 약 24배 적은 파라미터와 약 7배 적은 연산량을 가지며 약 16배 더 빠르게 영상을 복원하였다.

경량화 네트워크인 CARN, CBPN을 제안된 방법과 비교하였을 때 CARN은 약 1.5M 파라미터와 약 90.9G 연산량으로 26.07dB 성능을 0.0207sec 소요하여 영상을 복원하였으며 CBPN은 약 1.1M 파라미터와 약 97.9G 연산량으로 26.14dB 성능을 0.0189sec 소요하여 영상을 복원하였다. 제안된 네트워크가 CARN보다 약 2.2배 적은 파라미터, 약 1.5배 많은 연산량을 가지며 0.22dB 더 높은 성능을 약 1.8배 빠르게 복원하였고 CBPN보다 약 1.6배 적은 파라미터, 약 1.4배 많은 연산량을 가지며 0.15dB 더 높은 성능을 약 1.7배 빠르게 복원하였다. CARN은 종속 모듈을 사용하여 파라미터 수와 연산량을 줄이긴 하였으나 성능 또한 다수 줄어들어 성능과 효율성 사이에서 균형을 이루고 있지 못하며 CBPN은 저 해상도와 고 해상도 사이의 정보에 대해 피드백 구조를 적용하여 사용하기에 적은 파라미터 수와 연산량 대비 느린 처리 속도를 갖는 것을 확인할 수 있다.

그림 9, 10, 11, 12의 (a)는 원본 고해상도 영상이며 (b)는 원본 고해상도 영상으로부터 각 확대배율에 맞게 bicubic

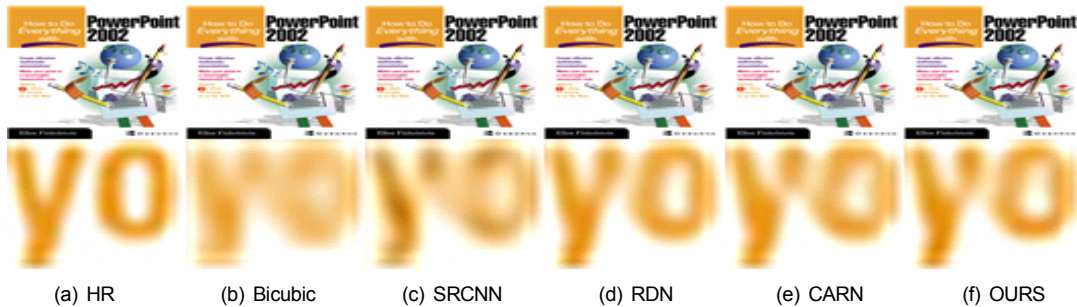


그림 9. Set 14 테스트 셋 "ppt3"에 대한 확대율 $r=2$ 에서 다른 네트워크와 시각적 비교

Fig. 9. Visual comparison with other networks at magnification $r=2$ for the Set 14 test set "ppt3"

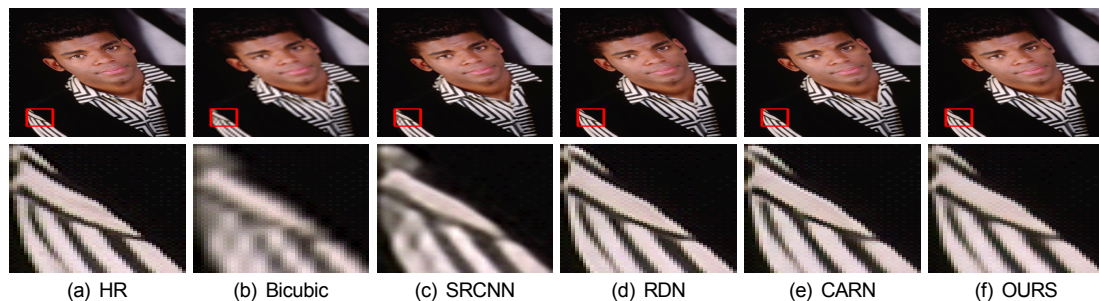


그림 10. BSD 100 테스트 셋 "302008"에 대한 확대율 $r=2$ 에서 다른 네트워크와 시각적 비교

Fig. 10. Visual comparison with other networks at magnification $r=2$ for the BSD 100 test set "302008"

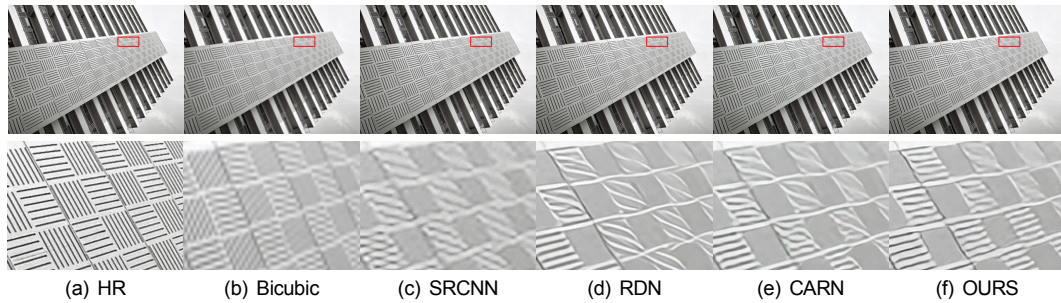


그림 11. Urban 100 테스트 셋 “img092”에 대한 확대율 $r=4$ 에서 다른 네트워크와 시각적 비교
Fig. 11. Visual comparison with other networks at magnification $r=4$ for the Urban 100 test set “img092”

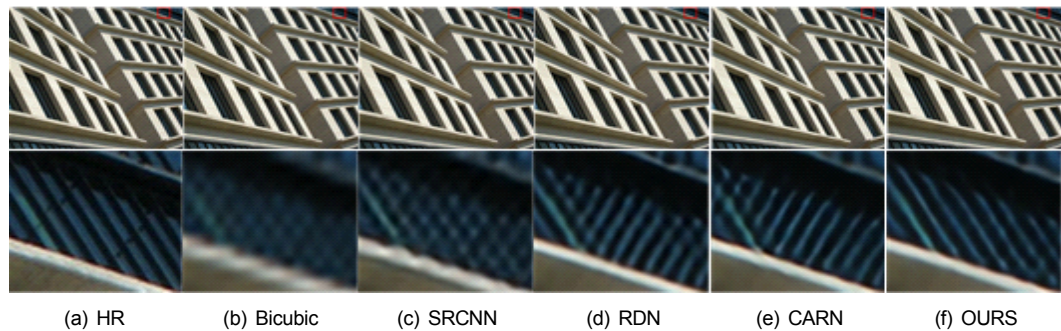


그림 12. Urban 100 테스트 셋 “img025”에 대한 확대율 $r=4$ 에서 다른 네트워크와 시각적 비교
Fig. 12. Visual comparison with other networks at magnification $r=4$ for the Urban 100 test set “img025”

보간법을 통해 다운 샘플 된 저해상도 영상이고, (c), (d), (e), (f)는 각각 SRCNN, RDN, CARN과 본 논문에서 제안하는 RDSRN을 통해 각 확대배율에 맞게 복원된 결과를 보여준다.

그림 9에서 보이는 바와 같이 (b) bicubic과 (c) SRCNN은 (a) 원본 고해상도 영상과 다르게 ‘y’와 ‘o’ 사이가 선으로 연결되어 있고 열화가 존재하는데 비해 제안된 네트워크는 ‘y’와 ‘o’ 사이가 연결되어 있지 않고 뚜렷하게 영상을 복원한 것을 확인할 수 있다.

그림 10에서 보이는 바와 같이 (b) bicubic과 (c) SRCNN은 (a) 원본 고해상도 영상과 다르게 옷 무늬에 반대 방향의 선과 많은 열화가 존재하는데 제안된 네트워크는 (a) 원본 고해상도와 유사한 뚜렷한 옷 무늬 영상을 복원한 것을 확인할 수 있다.

그림 11, 12에서 보이는 바와 같이 (d) RDN과 (e) CARN은 (a) 원본 고해상도 영상과 다른 반대 또는 물결무늬의 사선으로 다수 복원하였지만 제안된 네트워크는 원본 고해

상도 영상과 가장 비슷한 사선무늬로 영상을 복원한 것을 확인할 수 있다.

V. 결 론

최근 심층 합성곱 신경망 학습의 발전에 따라 심층학습 기법을 사용한 초해상도기법들이 기존 필터링기반 초해상도기법들 보다 좋은 성과를 보여주고 있으며, RDN 또는 RCAN과 같은 깊은 네트워크는 강한 표현 능력으로 저해상도 영상과 고해상도 영상 사이의 복잡한 비선형 매핑이 가능해졌다. 하지만 좋은 성능에도 불구하고 과도한 합성곱 신경망의 사용으로 인해 증가하는 파라미터 및 연산량과 느린 처리시간으로 저장 공간의 제약이 있는 실시간 또는 저전력 장치에 적용하는데 제한이 있다.

위와 같은 문제를 해결하기 위해 본 논문은 정보 증류 및 재귀적인 방식을 이용한 심층 학습법 기반 경량화된 단

일 영상 초해상도 기법을 제안하였다. 정보 증류 방식을 통해 유지 특징과 정제 특징으로 나눠 학습시켜 파라미터와 연산량을 감소시켰고 또한 정제 특징에 대해 Wider Channel을 적용함으로써 특징 억제를 제어하는 Distillation Wider Channel Block을 설계한 뒤 재귀적으로 사용하여 추가적인 파라미터 없이 네트워크의 깊이를 늘려 큰 수용 영역을 얻는 Recursive Distillation Super Resolution Network (RDSRN)를 제안한다.

실험 결과 RDN과 비교하였을 때 비슷한 화질의 영상을 복원하며 RDN보다 약 32배 적은 파라미터와 약 10배 적은 연산량을 가지고 기존대비 약 3.5배 더 빠른 처리시간으로 영상을 복원하는 것을 확인하였으며 기존 경량화 네트워크 CARN과 비교하였을 때 약 2.2배 적은 파라미터와 약 1.8배 빠른 처리시간으로 모든 테스트 셋에서 평균적으로 0.16dB 더 좋은 성능을 만들어 냄을 확인하였다.

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
doi: <https://doi.org/10.1109/TIP.2008.924279>
- [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
doi: <https://doi.org/10.1109/TIP.2014.2325774>
- [3] 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
doi: <https://doi.org/10.1109/TIP.2012.2208977>
- [4] V. Pappayan and M. Elad, "Multi-Scale Patch-Based Image Restoration," in IEEE Transactions on Image Processing, pp. 249-261, 2016
doi: <https://doi.org/10.1109/TIP.2015.2499698>
- [5] 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
doi: <https://doi.org/10.1109/TIP.2016.2619265>
- [6] Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in CVPR, pp. 2472-2481, 2018.
doi: <https://doi.org/10.1109/CVPR.2018.00262>
- [7] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in CVPR, pp. 2790-2798, 2017.
doi: <https://doi.org/10.1109/CVPR.2017.298>
- [8] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," in CVPR, pp. 2261-2269, 2017
doi: <https://doi.org/10.1109/CVPR.2017.243>
- [9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," In ICLR, 2015
- [10] 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.
doi: <https://doi.org/10.1109/CVPR.2016.207>
- [11] 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.
- [12] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super resolution using very deep convolutional networks," in CVPR, pp. 1646-1654, 2016.
doi: <https://doi.org/10.1109/CVPR.2016.182>
- [13] 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.
doi: https://doi.org/10.1007/978-3-030-01234-2_18
- [14] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and, lightweight super-resolution with cascading residual network," ECCV, 2018.
doi: https://doi.org/10.1007/978-3-030-01249-6_16
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in CVPR, 2016.
doi: <https://doi.org/10.1109/CVPR.2016.90>
- [16] Andrew G. Howard, Bo Chen, D. Kalenichenko, W. Wang, T. Weand, M. Andreetto, H. Adam, "MobileNets: Efficient Convolutional Neural Networks for mobile vision applications," in CVPR, 2017.
- [17] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," IEEE Conference on Computer Vision and Pattern Recognition, pp. 723 - 731, 2018.
doi: <https://doi.org/10.1109/CVPR.2018.00082>

— 저 자 소 개 —



우 희 조

- 2019년 : 한국공학대학교 전자공학부 공학사
- 2022년 : 한국공학대학교 전자공학부 공학석사
- 2022년 3월 ~ 현재 : 한화시스템 지능형 SW팀 연구원
- ORCID : <https://orcid.org/0000-0003-4956-6809>
- 주관심분야 : 딥러닝기반 영상처리, 경량화, 컴퓨터 비전, 임베디드 시스템



심 지 우

- 2019년 : 한국공학대학교 전자공학부 공학사
- 2022년 : 한국공학대학교 전자공학부 공학석사
- 2022년 3월 ~ 현재 : ㈜메디쌍류 기술연구소 연구원
- ORCID : <https://orcid.org/0000-0003-1964-3615>
- 주관심분야 : 딥러닝기반 영상처리, 컴퓨터 비전, 임베디드 시스템



김 응 태

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