

Self-Calibrated Convolution을 이용한 단일 이미지 초해상화

강동균*, 이상국**

Single Image Super Resolution using Self-Calibrated Convolution

Dong-Gyun Kang*, Sang-Goog Lee**

요약

최근 합성곱 신경망(CNN) 학습 기술의 발전으로 단일 이미지 초해상화(SISR) 분야에서 다양한 심층 학습 기반 기법들이 우수한 성능을 보여주고 있다. 특히, 잔차 밀집 블록(Residual dense block)을 활용하여 초기 특징을 최종 계층까지 전달하고, 이전 계층의 출력을 후속 계층의 입력으로 사용하는 잔차 밀집 네트워크(RDN), 그리고 그와 유사한 기법들이 연구되어 왔다. 그러나 성능 향상을 위해 모델의 구조가 점점 복잡해지는 사례 역시 존재한다. 본 논문에서는 RDN의 기본 구조를 유지하면서도 성능 향상을 위해 기존 합성곱 연산을 자기 보정 합성곱으로 대체하는 방법을 제안한다. 제안한 모델은 구조가 복잡하지 않으면서 수용 영역을 확장하여 더 세밀한 복원 이미지를 생성하도록 한다. 실험 결과, 제안한 모델은 상당 수의 지표에서 기존 RDN 대비 우수한 성능을 보였으며, 이는 제안된 자기 보정 합성곱이 복원 품질 향상에 효과적임을 보여준다.

Abstract

Recent advances in Convolutional Neural Network (CNN)-based learning have led to the emergence of various deep learning-based techniques in the field of Single-Image Super-Resolution (SISR). In particular, Residual Dense Networks (RDN), which utilize residual dense blocks to propagate initial features to the final layer and use the output of previous layers as input for subsequent layers, and similar techniques have been actively studied. However, there are also cases where the model structure becomes increasingly complex to improve performance. In this paper, we propose a method that replace convolution with a self-calibrated convolution to improve performance while maintaining the basic structure of the RDN. The proposed model expands the receptive field without a complex structure, thereby generating more detailed restored images. Experimental results demonstrate that the proposed model outperforms conventional RDNs in a significant number of metrics, demonstrating the effectiveness of the proposed self-calibrated convolution in improving restoration quality.

Keywords

deep learning, computer vision, single image super resolution, self calibrated convolution

* 가톨릭대학교 대학원 디지털미디어학과
- ORCID: <http://orcid.org/0009-0002-6272-6598>
** 가톨릭대학교 미디어기술콘텐츠학과 교수(교신저자)
- ORCID: <http://orcid.org/0000-0003-3060-6882>

• Received: Aug. 06, 2025, Revised: Nov. 21, 2025, Accepted: Nov. 24, 2025
• Corresponding Author: Sang-Gook Lee
Dept. of Catholic University of Korea, 43 Jibong-ro, Wonmi-gu,
Bucheon-si, Gyeonggi-do, Korea
Tel.: +82-2-2164-4909, Email: gyun5164@catholic.ac.kr

1. 서 론

단일 이미지 초해상화(SISR, Single Image Super Resolution)는 해상도가 낮은 이미지를 시각적으로 품질이 높은 고해상도 이미지로 변환하는 것을 목표로 한다. 단일 이미지 초해상화는 보안[1], 의료 영상처리[2], 이미지 생성[3] 등 컴퓨터 비전 분야에서 다양하게 활용된다. 일반적으로 저해상도 이미지를 고해상도 이미지로 변환하는 문제는 부적절하게 정립된 역문제로 정의되기 때문에 저해상도 이미지에 대한 여러 고해상도 변환 이미지가 존재할 수 있다. 이를 해결하기 위한 초해상화 연구 기법으로는 세 가지가 있다. 첫 번째는 보간법[4]-[7]으로 주변의 픽셀값을 토대로 정보가 없는 곳에 새로운 픽셀을 생성하는 기술이다. 계산량이 적은 장점이 있지만 저주파수 필터 특성 상 고주파 정보를 충분히 복원하지 못해 품질이 좋지 않다. 두 번째는 재구성법[8]-[10]으로 보간법보다 많은 성능이 개선되어 고주파수 성분을 충분히 복원하지만 상대적으로 계산량이 많이 소요된다. 세 번째는 심층 학습법[11]-[13]으로 대량의 데이터에서 저해상도-고해상도 관계를 학습하여 초해상화를 수행하는 방식이다.

최근 심층 학습법의 발전 영향으로 초해상화를 위한 알고리즘은 합성곱 신경망을 기반으로 다양한 연구가 진행되고 있으며 높은 성능을 보여주고 있다.

C. Dong et al.[14]은 초해상화에 처음 딥러닝 기법을 활용한 네트워크인 SRCNN를 제안했다. 이는 상대적으로 단순한 3개의 계층을 가진 FCN(Fully Convolution Network) 구조의 네트워크로 바이큐빅 보간법으로 확대한 이미지의 화질을 높이는 알고리즘이다. 이 네트워크는 기존 초해상도 알고리즘과 비교하여 높은 성능을 보이며 초해상화를 위한 알

고리즘에 딥러닝 기반의 네트워크가 도움이 된다는 것을 입증하였다.

Y. Tai et al.은 DRRN[15]에서 재귀 블록을 MemNet메모리 블록[16]으로 대체하여 깊은 네트워크 구조에서도 학습 안정성을 유지할 수 있도록 설계했다. 그러나 이러한 방법은 네트워크에 적용하기 전에 원본 저해상도 이미지를 원하는 크기로 보간해야 하는 단점이 있다. 이 전처리 단계는 계산 복잡도를 증가시키고[17] 원본 저해상도 이미지를 과도하게 평활화하여 세부 정보 손실을 초래한다. 그 결과 이러한 방법은 보간된 저해상도 이미지로부터 특징을 얻게 되어 원본 저해상도 이미지로부터 고해상도 이미지로의 매핑을 제대로 수행하지 못한다.

G. Huang et al.[18]은 동일한 밀집 블록 내에서 임의의 두 계층 간 직접 연결을 허용하는 DenseNet을 제안했다. 이러한 지역 밀집 연결을 통해 각 계층은 동일한 밀집 블록 내에서 모든 이전 층으로부터 정보를 읽을 수 있다. 이러한 DenseNet 구조를 초해상화에 적용하는 SRDenseNet[19]이 제안되었다.

G. Huang et al.[18]이 제안한 DenseNet을 이용해서 Y. Zhang et al.[11] 등은 잔차 밀집 네트워크를 제안했다. 그림 1은 잔차 밀집 네트워크(RDN, Residual Dense Network)의 구조를 나타내며 그림 2의 잔차 밀집 블록을 다수 쌓는 구조로 구성된다. 잔차 밀집 블록(RDB, Residual Dense Block)은 밀집하게 연결된 합성곱 층으로부터 지역 특징을 얻고 이전 RDB의 출력이 현재 RDB의 모든 층의 입력으로 가도록 설계되어 있다. 지역 특징 융합(LFF, Local Feature Fusion)을 통해 이전 RDB의 출력, 지역 특징을 충분히 학습할 수 있고 밀집 지역 특징 학습 후 전역 특징 융합(Global feature fusion)을 통해 전역 계층 특징을 학습한다.

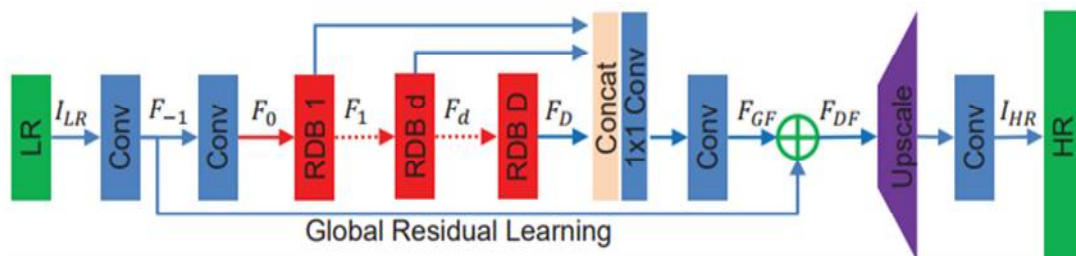


그림 1. Residual Dense Network[11] 구조
Fig. 1. Architecture of Residual Dense Network[11]

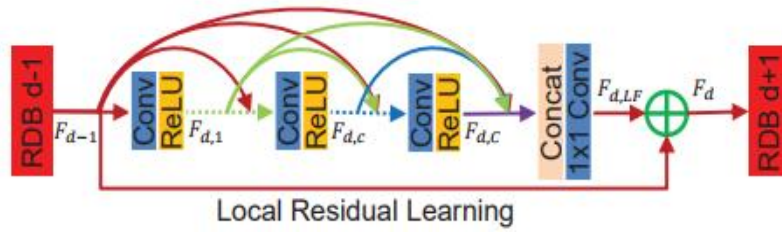


그림 2. Residual Dense Block[11] 구조
 Fig. 2. Architecture of Residual Dense Block[11]

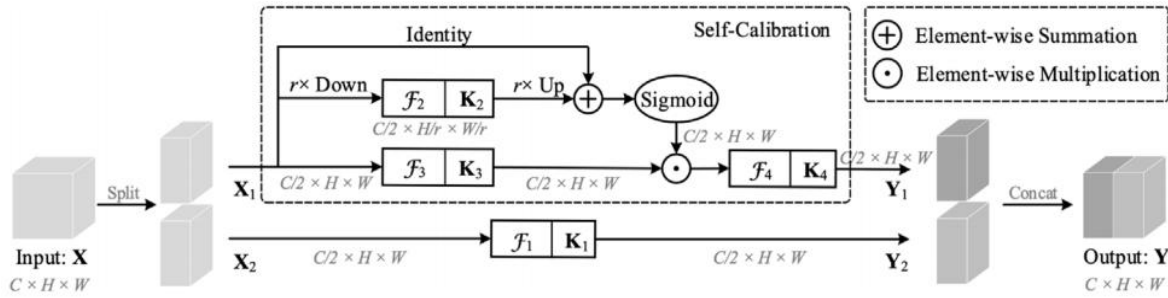


그림 3. Self-Calibrated Convolution[20] 구조
 Fig. 3. Architecture of Self-Calibrated Convolution[20]

합성곱 신경망뿐만 아니라 자연어 처리에서 많이 사용되는 Transformer를 Vision Task에 적용한 Swin Transformer[20]를 이용한 이미지 초해상화 연구 역시 진행되었다. J. Liang et al.[21]은 Swin Transformer를 기반으로 이미지 복원을 위한 강력한 베이스라인 모델인 SwinIR을 제안했다. CONDE[22]은 Swin TransformerV2[23]를 기반으로 한 이미지 초해상도 네트워크를 제안했다.

본 논문에서는 RDN[11]으로부터 영감을 받아, 기존의 RDN 구조를 유지하면서도 성능을 향상시키기 위해 기존 합성곱을 그림 3의 자기 보정 합성곱[24]으로 대체하였다. 기존 필터는 K1, K2, K3, K4 네 부분으로 나뉘고 각 부분은 서로 다른 기능을 담당한다. 먼저, 입력 특징맵 X를 두 X1과 X2로 균등하게 나눈다. 그 다음, 특징 X1에 대해 평균 풀링과 함께 다운샘플링 r-폴드를 사용하고, 이전 출력에 대해 업샘플링(쌍선형 보간)을 수행한다. 이는 아래 식 (1)에 나타나있다

$$X'_1 = Up(F_2(AvgPool_r(X_1))) \quad (1)$$

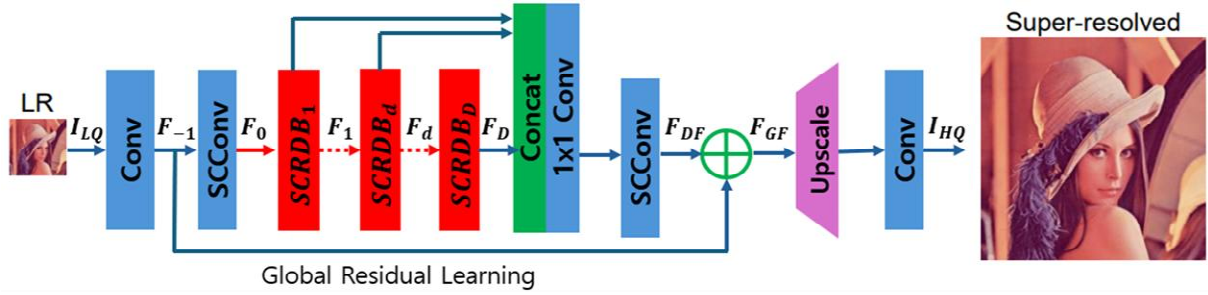
가중치 행렬 W는 아래 식 (2)로 얻는다.

$$W = \sigma(X_1 + X'_1) \quad (2)$$

식 (2)에서 σ 는 시그모이드 함수를 나타낸다. 가중치 행렬은 필터 K3에서 추출된 특징맵을 보정하여 중간 출력 특징을 얻는 데 사용된다. 중간 출력은 필터 K4를 사용하여 추출하여 특징맵 Y1을 얻는다. 또한, 입력 특징맵 X2는 필터 K1을 사용하여 추출하여 특징맵 Y2를 얻는다. 마지막으로, 서로 다른 스케일 공간에서 출력 특징 Y1과 Y2를 이어 붙여 최종 출력 특징 Y를 얻는다. 그림 4, 5는 각각 그림 1의 RDN과 그림 2의 RDB 구조에서 합성곱을 자기 보정 합성곱으로 대체한 것을 나타낸다.

이렇게 함으로써 자기 보정 잔차 밀집 네트워크를 통해 각 합성곱 계층의 시야를 확장하고 출력 특징을 더 풍부하게 할 수 있다. 또 기존 네트워크의 합성곱 층에 쉽게 적용 가능하며, 연산 복잡도를 증가시키지 않는다. 결과적으로 구조를 복잡하지 않게 하면서 네트워크의 성능을 향상시킬 수 있다.

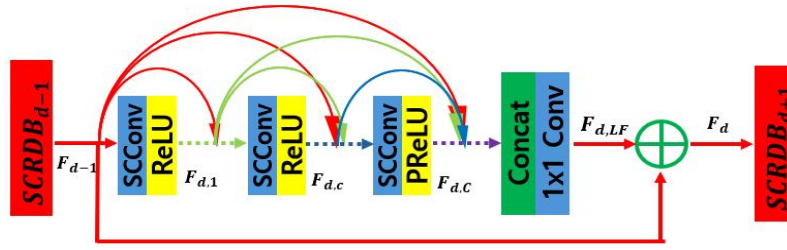
본 논문의 구성은 다음과 같다. 2장에서는 자기 보정 합성곱 기반의 잔차 밀집 네트워크를 통한 이미지 초해상화를 제안한다. 3장에서는 실험을 통해서 제안한 방법의 성능을 선형연구와 비교, 분석하며 4장에서는 결론과 향후 과제를 기술한다.



Global Residual Learning

그림 4. Self-Calibrated Residual Dense Network 구조

Fig 4. Architecture of Self-Calibrated Residual Dense Network



Local Residual Learning

그림 5. Self-Calibrated Residual Dense Block 구조

Fig. 5. Architecture of Self-Calibrated Residual Dense Block

II. 제안된 Self-Calibrated Convolution 기반 초해상화 기법

본 논문의 SCRDN은 그림 4와 같이 얇은 특징 추출(Shallow feature extraction net), 자기 보정 잔차 밀집 블록(SCRDB, Self-Calibrated Residual Dense Block), 밀집 특징 융합(DFF, Dense Feature Fusion), 그리고 마지막으로 업샘플링 네트워크(UPNet) 등 4 가지 파트로 구성되어 있다.

2.1 얇은 특징 추출 단계

저해상도 이미지가 입력되면 먼저 합성곱과 자기 보정 합성곱을 거쳐 64개의 특징맵을 추출한다. 첫 번째 합성곱은 저해상도 이미지로부터 특징 F_{-1} 을 추출한다[11]. 첫 번째 얇은 특징 추출은 다음 식(3)으로 나타낼 수 있다.

$$F_{-1} = H_{SFE1}(I_{LQ}) \quad (3)$$

식 (3)에서 $H_{SFE1}(\cdot)$ 은 표준 합성곱(Standard convolution) 연산자를 나타낸다. F_{-1} 은 추가적인 얇

은 특징 추출과 전역 잔차 학습(GRL, Global Residual Learning)에 사용된다. 두 번째 얇은 특징 F_0 추출은 다음 식(4)로 나타낸다.

$$F_0 = H_{SFE2}(F_{-1}) \quad (4)$$

식 (4)에서 $H_{SFE2}(\cdot)$ 은 두 번째 얇은 특징 추출을 위한 자기 보정 합성곱 연산자로 구조는 그림 3에 나타나 있다. 아래 식 (5)의 출력은 자기 보정 잔차 밀집 블록의 입력으로 사용된다.

$$F_d = H_{SCRDB,d}(F_{d-1}) = H_{SCRDB,d}(H_{SCRDB,d-1}(\dots(H_{SCRDB,d-1}(F_0))\dots)) \quad (5)$$

식 (5)에서 $H_{SCRDB,d}$ 은 d 번째 SCRDB 연산자를 나타낸다. 이때 $H_{SCRDB,d}$ 는 자기 보정 합성곱, ReLU[25]와 같은 연산의 합성 함수가 될 수 있다. d 번째 SCRDB에 의해 F_d 가 출력되면서 F_d 는 지역 특징으로 볼 수 있다.

SCRDB 세트를 이용하여 계층적 특징을 추출한 후, 전역 특징 융합(GFF)과 전역 잔차 학습(GRL)을

포함하는 밀집 특징 융합(DFE)을 추가로 수행한다. 밀집 특징 융합(DFE)은 모든 이전 단계의 특징을 활용할 수 있는데 이는 다음 식 (6)으로 나타낸다.

$$F_{DF} = H_{DFE}(F_{-1}, F_0, F_1, \dots, F_D) \quad (6)$$

식 (6)에서 F_{DF} 는 합성 함수 H_{DFE} 를 활용하여 얻은 밀집 특징 융합(DFE)의 출력 특징이다. 저해상도 공간에서 지역 및 전역 특징을 추출한 후, 고해상도 공간에서 업샘플링 네트워크를 적용한다. 업샘플링 네트워크에서 ESPCN[26]을 적용한다. RDN의 출력 I_{SR} 은 다음 식 (7)로 얻어진다.

$$I_{SR} = H_{SCRDN}(I_{LR}) \quad (7)$$

식 (7)에서 H_{SCRDN} 은 SCRDN 함수를 나타낸다.

2.2 Self-Calibrated Residual Dense Block

그림 5의 자기 보정 잔차 밀집 블록은 지역 밀집 연결과 전역 잔차 학습(GRL)을 결합함으로써 더 많은 지역 정보를 효과적으로 학습하고 전달할 수 있다. 본 논문의 SCRDB는 밀집연결층, 지역 특징 융합(LFE), 그리고 지역 잔차 학습(LRL, Local Residual Learning)을 포함하며 연속 메모리로 이어진다.

연속 메모리 메커니즘은 이전 SCRDB의 상태를 현재 SCRDB의 각 계층에 전달하여 실현된다. 각 SCRDB의 출력은 아래 식 (8)로 나타낸다.

$$F_{d,c} = \sigma(W_{d,c}[F_{d-1}, F_{d,1}, \dots, F_{d,c-1}]) \quad (8)$$

식 (8)에서 σ 는 ReLU[25] 활성화 함수를 가리킨다. $W_{d,c}$ 는 c 번째 합성곱 레이어의 가중치이다. $F_{d,c}$ 는 G(growth rate) 특징맵들로 구성되어 있다. $[F_{d-1}, F_{d,1}, \dots, F_{d,c-1}]$ 은 $d-1$ 번째 SCRDB에 의해 만들어진 특징맵의 연결 연산자를 가리킨다. 이전 RDB와 각 계층의 출력은 이후의 모든 계층과 직접 연결되어 있어 피드-포워드 특성을 유지할 뿐만 아니라 지역 밀집 특징을 추출한다.

지역 특징 융합(LFE)은 이전 SCRDB의 상태와 현재 SCRDB의 전체 합성곱 계층을 적응적으로 융

합하는데 적용된다. 위에서 분석한 바와 같이 $d-1$ 번째 SCRDB의 특징맵은 연결 방식으로 d 번째 SCRDB에 직접 도입되므로 특징 수를 줄이는 것이 필수적이다. 한편, MemNet[16]에서 영감을 받아 출력 정보를 적응적으로 제어하기 위해 1×1 합성곱 계층을 도입한다. 이를 지역 특징 융합(LFE)이라고 하며 아래 식 (9)로 나타낸다.

$$F_{d,LF} = H_{LFE}^d([F_{d-1}, F_{d,1}, \dots, F_{d,c}, \dots, F_{d,c}]) \quad (9)$$

식 (9)에서 H_{LFE}^d 는 d 번째 SCRDB 안에서의 1×1 합성곱 레이어의 함수를 가리킨다.

지역 잔차 학습은 여러 개의 합성곱 계층이 하나의 SCRDB에 존재하기 때문에, 정보 흐름을 더욱 개선하기 위해 존재한다. d 번째 SCRDB의 마지막 출력은 다음 식 (10)으로 나타낸다.

$$F_d = F_{d-1} + F_{d,LF} \quad (10)$$

지역 잔차 학습은 네트워크 표현 능력을 더욱 향상시켜 더 좋은 성능을 제공할 수 있다.

2.3 밀집 특징 융합(DFE)

SCRDB 세트를 이용하여 지역 밀집 특징을 추출한 후, 계층적 특징을 전역적으로 활용하기 위해 밀집 특징 융합(DFE)을 활용한다. DFE는 전역 특징 융합(GFE)과 전역 잔차 학습(GRL)으로 구성된다. 전역 특징 융합은 모든 SCRDB의 지역 특징들을 융합하여 전역 특징 F_{GF} 를 추출하는 것을 의미한다. 전역 특징 F_{GF} 는 이미지 전체 구조나 고수준의 의미를 반영할 수 있다. 전역 특징 F_{GF} 는 아래 식 (11)로 얻는다.

$$F_{GF} = H_{GFE}([F_1, \dots, F_D]) \quad (11)$$

$[F_1, \dots, F_D]$ 는 SCRDB 1, ..., D 에 의해 생성된 특징맵의 연결을 나타낸다. H_{GFE} 는 1×1 , 3×3 합성곱의 합성 함수를 가리킨다. 1×1 합성곱 계층은 다양한 수준의 특징 범위를 적응적으로 융합하

는 데 사용된다. 다음 3×3 합성곱 계층은 전역 잔차 학습을 위해 특징을 추가로 추출하기 위해 사용된다.

그런 다음 아래 식 (12)로 전역 잔차 학습(GRL)을 사용하여 업스케일링을 수행하기 전에 특징맵을 얻는다.

$$F_{DF} = F_{-1} + F_{GF} \quad (12)$$

F_{-1} 은 얇은 특징맵을 가리킨다. 전역 특징 융합 이전의 모든 다른 층들은 전적으로 SCRDB와 함께 활용된다. SCRDB들은 F_{GF} 에 적응적으로 융합되는 다중 지역 밀집 특징을 생산한다. 전역 잔차 학습(GRL) 후에 밀집 특징인 F_{DF} 를 얻는다.

2.4 기타 세부사항

얇은 특징 추출 계층, 지역 및 전역 특징 융합 계층에는 G0=64 필터를 사용했다. 각 SCRDB의 다른 계층에는 G 필터가 있으며 ReLU[25]가 이어져 있다. ESPCN[26]을 사용하여 UPNet의 거친 해상도 특징을 세밀하게 조정한다. 최종 합성곱 계층에는 색상 HR 이미지를 출력하므로 출력 채널이 3개 있다. 그러나 네트워크는 회색 이미지도 처리할 수 있다.

III. 실험 및 고찰

제안한 모델의 성능을 확인하기 위해서 학습을 위한 데이터셋으로 800장의 DIV2K를 이용하여 이미지 당 서로 다른 위치에서 동일 크기로 5장을 크롭(Crop)해서 총 4,000장의 이미지를 만들었다. 검증 데이터셋으로 Set5, Set14, Urban100 세 개를 이용했다. 제안된 네트워크의 입력과 출력으로 3채널 RGB 이미지를 사용했고 수평, 수직 뒤집기 및 90° 회전을 이용하는 데이터 증강 기술을 사용하였다.

활성화 함수로는 ReLU[25] 함수를 사용하였다. 학습 횟수는 800 에포크(epoch)로 학습하였고 학습률(Learnig rate)은 0.0001로 설정했다. 실험은 파이토치(PyTorch) 코드로 진행하였고 사용된 하드웨어 환경은 Intel(R) Core(TM) i9-10900X CPU와 Geforce RTX 3090 GPU 24G를 사용하였다.

표 1은 각 테스트셋을 사용하여 기존의 초해상화 네트워크들과 제안된 네트워크의 신호 대 잡음비(PSNR, Peak Signal to Noise), 구조적 유사도(SSIM, Structural Similarity)[27]를 비교한 표이다.

표 1. 테스트셋에 대한 최대 신호 대 잡음비, 구조적 유사도 비교

Table 1. Comparison of PSNR, SSIM for the test set

Model	scale	Set5	Set14	Urban100
		(PSNR, SSIM)	(PSNR, SSIM)	(PSNR, SSIM)
Bicubic	x2	33.66 / 0.9542	30.24 / 0.8688	26.88 / 0.8403
	x3	30.39 / 0.8682	27.55 / 0.7742	24.46 / 0.7349
	x4	28.42 / 0.8104	26.00 / 0.7027	23.14 / 0.6577
DRRN	x2	37.74 / 0.9591	33.23 / 0.9136	31.23 / 0.9188
	x3	34.03 / 0.9244	29.96 / 0.8349	27.53 / 0.8378
	x4	31.68 / 0.8888	28.21 / 0.7721	25.44 / 0.7638
SRDenseNet	x2	-	-	-
	x3	-	-	-
	x4	32.02 / 0.8934	28.50 / 0.7782	26.05 / 0.7819
MemNet	x2	37.78 / 0.9597	33.28 / 0.9142	31.31 / 0.9195
	x3	34.09 / 0.9248	30.00 / 0.8350	27.56 / 0.8376
	x4	31.74 / 0.8893	28.26 / 0.7723	25.50 / 0.7630
SRCNN	x2	36.66 / 0.9542	32.45 / 0.9067	29.51 / 0.8405
	x3	32.75 / 0.9090	29.29 / 0.8215	26.24 / 0.7991
	x4	30.48 / 0.8628	27.50 / 0.7513	24.52 / 0.7226
RDN	x2	38.24 / 0.9614	34.01 / 0.9212	32.89 / 0.9353
	x3	34.71 / 0.9296	30.57 / 0.8468	28.80 / 0.8653
	x4	32.47 / 0.8990	28.81 / 0.7871	26.61 / 0.8028
SCRDN	x2	38.22 / 0.9620	33.92 / 0.9220	32.96 / 0.9364
	x3	34.76 / 0.9306	30.52 / 0.8476	28.85 / 0.8679
	x4	32.54 / 0.8999	28.78 / 0.7876	26.66 / 0.8053

PSNR은 픽셀값의 차이를 통해서 노이즈 수준을 측정하며, 값이 높을수록 품질이 우수한 것을 의미한다. SSIM은 두 이미지 간 구조적으로 유사한 정도를 평가하는 지표로, 마찬가지로 높은 값일수록 품질이 우수함을 의미한다. 제안된 모델은 상당 부분에서 기존의 방법들보다 더 좋은 성능을 나타냈다.

Urban100 데이터셋의 경우 모든 확대 배율의 지표에서 기존의 네트워크들보다 성능이 우수했다. Set5 데이터셋의 경우, 확대 배율 2의 PSNR을 제외한 모든 부분에서 기존 네트워크보다 성능이 더 좋았다. Set14의 경우 모든 확대 배율의 PSNR 지표는 RDN 다음으로 높았고 SSIM은 모든 네트워크보다 성능이 좋았다. SCRDN은 복잡한 도시 이미지 데이터셋인 Urban100에서 특히 성능 향상이 뚜렷했다. 이는 제안한 네트워크가 복잡한 텍스처나 구조를 처리하는데 강점이 있음을 알 수 있다. 또 확대 배율이 높아질수록 RDN 대비 성능 차이가 안정적으로 유지되거나 커졌는데 이는 모델이 고배율 초해상화에서도 일관된 성능을 유지한다는 점에서 유의

미한 결과이다.

또 본 논문에서 제안한 네트워크가 모든 SSIM 지표에서 가장 우수한 점을 확인할 수 있다. SSIM은 이미지의 구조, 명암도, 대비를 종합적으로 고려한 지표로써 왜곡, 압축 시 시각 품질 평가에 유리하고, 이미지의 품질을 사람 눈의 인식에 더 가깝게 판단할 수 있는 지표이다. 이러한 점으로 봤을 때 제안한 네트워크가 기존의 네트워크들보다 사람 눈의 인식에 더 부합하는 우수한 품질을 생산할 수 있다는 점에서 유의미하다.

그림 6, 7, 8은 Urban100의 2번 이미지를 확대 배율 2, 3, 4에 각각 RDN과 SCRDN으로 초해상화를 이룬 결과를 비교한 것이다. 위 표 1의 PSNR과 SSIM 결과와 마찬가지로 모든 확대 배율 부분의 이미지에서 SCRDN을 적용하여 초해상화 시킨 이미지가 RDN을 적용시킨 이미지보다 철골 구조물의 윤곽을 더 뚜렷하게 표현하는 것을 알 수 있다. 즉, SCRDN을 적용한 이미지가 디테일 보존과 경계 표현에서 상대적으로 우수한 것을 알 수 있다.



그림 6. 확대 배율 2에서 기존 네트워크와 제안된 네트워크 비교
 Fig. 6. Comparison of existing network and proposed network at scaling factor 2



그림 7. 확대 배율 3에서 기존 네트워크와 제안된 네트워크 비교
 Fig. 7. Comparison of existing network and proposed network at scaling factor 3

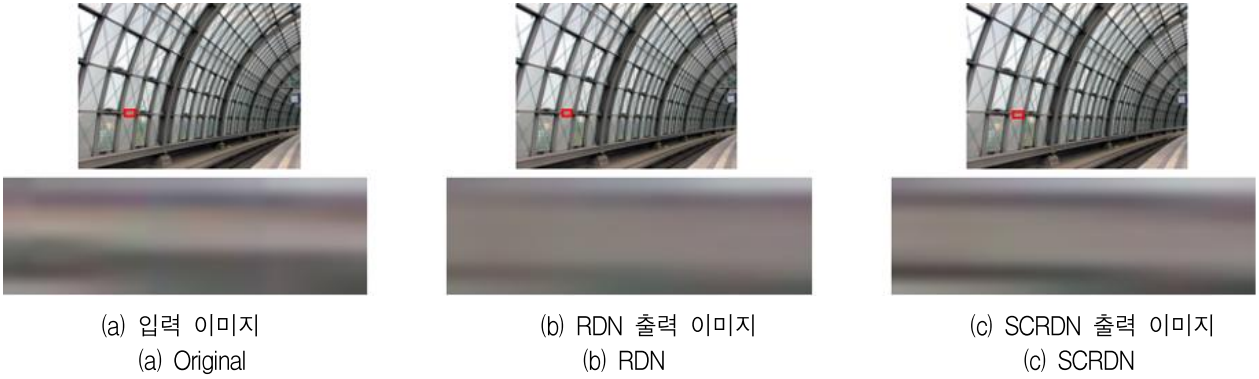


그림 8. 확대 배율 4에서 기존 네트워크와 제안된 네트워크 비교

Fig. 8. Comparison of existing network and proposed network at scaling factor 4



그림 9. 확대 배율 3에서 기존 네트워크와 제안된 네트워크 비교

Fig. 9. Comparison of existing network and proposed network at scaling factor 3



그림 10. 확대 배율 4에서 기존 네트워크와 제안된 네트워크 비교

Fig. 10. Comparison of existing network and proposed network at scaling factor 4

그림 9, 10은 Set5 데이터셋의 이미지를 확대 배율 3, 4에 각각 RDN과 SCRDN으로 초해상화를 적용한 결과를 비교한 것이다. 해당 확대 배율 부분의 이미지에서 SCRDN을 적용하여 초해상화 시킨 이미지가 RDN을 적용시킨 이미지보다 눈썹 부분을 더 뚜렷하게 표현한 걸 알 수 있다.

다만 표 2와 같이 SCRDN은 RDN의 합성곱과 달리 채널 분할과 보정 경로가 있는 자기 보정 합성곱을 사용했기 때문에 합성곱을 사용한 RDN보다 파라미터 수와 FLOPs가 소폭 증가한다는 점을 확인했다.

표 2. 네트워크의 파라미터 수

Table 2. Parameter number of network

	Params(M)		FLOPs(G)	
	RDN[11]	SCRDN	RDN[11]	SCRDN
x2	22M	25M	90.71	92.95
x3	22M	26M	91.5	95.32
x4	29M	26M	93.22	96.48

IV. 결론 및 향후 과제

본 논문에서는 RDN의 합성곱을 자기 보정 합성곱(Self-calibrated convolution)으로 대체하여 복잡하지

않은 구조를 유지하면서 각 합성곱 계층의 시야를 확장하고 출력 특징을 더 풍부하게 만듦으로써 단일 이미지 초해상화의 품질을 향상시키는 연구를 진행했다. 연구 결과 PSNR의 경우 데이터셋 마다 일부 차이가 있었지만 모든 데이터셋의 SSIM, 즉 이미지의 구조, 명암도, 대비를 종합적으로 고려한 지표가 가장 우수하게 나오므로써 제안된 네트워크가 사람 눈의 인식에 더 부합하는 우수한 품질을 생산할 수 있다는 가능성을 보였다. 그러나 합성곱을 자기 보정 합성곱으로 대체했기 때문에 모델의 파라미터 수, 연산량 등이 소폭 증가하는 한계점 역시 확인했다. 추후 연구에서는 이러한 문제를 해결하고 보완하는 방향으로 실험을 진행할 계획이다.

References

- [1] W. W. W. Zou and P. C. Yuen, "Very low-resolution face recognition problem", *IEEE Transactions on image processing*, Vol. 21, No. 1, pp. 327-340, Jan. 2012. <https://doi.org/10.1109/TIP.2011.2162423>.
- [2] W. Shi, et al., "Cardiac Image Super-Resolution with Global Correspondence Using Multi-Atlas PatchMatch", *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2013*, Vol. 8151, pp. 9-16, Sep. 2013. https://doi.org/10.1007/978-3-642-40760-4_2.
- [3] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation", *arXiv preprint arXiv:1710.10196*, Oct. 2017. <https://doi.org/10.48550/arXiv.1710.10196>.
- [4] X. Zhang and X. Wu, "Image Interpolation by Adaptive 2-D Autoregressive Modeling and Soft-Decision Estimation", *IEEE Transactions on Image Processing*, Vol. 17, No. 6, pp. 887-896, Jun. 2008. <https://doi.org/10.1109/TIP.2008.924279>.
- [5] Y. Romano, M. Protter, and M. Elad, "Single Image Interpolation Via Adaptive Nonlocal Sparsity-Based Modeling", *IEEE Transactions on Image Processing*, Vol. 23, No. 7, pp. 3085-3098, Jul. 2014. <https://doi.org/10.1109/TIP.2014.2325774>.
- [6] W. Ye and K. Ma, "Convolutional Edge Diffusion for Fast Contrast-guided Image Interpolation", *IEEE Signal Processing Letters*, Vol. 23, No. 9, pp. 1260-1264, Sep. 2016. <https://doi.org/10.1109/LSP.2016.2571738>.
- [7] H. A. Aly and E. Dubois, "Image up-sampling using total-variation regularization with a new observation model", *IEEE Transactions on Image Processing*, Vol. 14, No. 10, pp. 1647-1659, Oct. 2005. <https://doi.org/10.1109/TIP.2005.851684>.
- [8] K. Zhang, X. Gao, D. Tao, and X. Li, "Single Image Super-Resolution With Non-Local Means and Steering Kernel Regression", *IEEE Transactions on Image Processing*, Vol. 21, No. 11, pp. 4544-4556, Nov. 2012. <https://doi.org/10.1109/TIP.2012.2208977>.
- [9] V. Papyan and M. Elad, "Multi-Scale Patch-Based Image Restoration", *IEEE Transactions on Image Processing*, Vol. 25, No. 1, pp. 249-261, Jan. 2016. <https://doi.org/10.1109/TIP.2015.2499698>.
- [10] C. Ren, X. He, and T. Q. Nguyen, "Single Image Super-Resolution via Adaptive High-Dimensional Non-Local Total Variation and Adaptive Geometric Feature", *IEEE Transactions on Image Processing*, Vol. 26, No. 1, pp. 90-106, Jan. 2017. <https://doi.org/10.1109/TIP.2016.2619265>.
- [11] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution", *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 2472-2481, Jun. 2018. <https://doi.org/10.1109/CVPR.2018.00262>.
- [12] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network", *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 2790-2798, Jul. 2017. <https://doi.org/10.1109/CVPR.2017.298>.

- [13] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super resolution using very deep convolutional networks", CVPR, Las Vegas Nevada, pp. 1646-1654, Jun.-Jul. 2016.
- [14] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution", Computer Vision – ECCV 2014, Zurich, Switzerland, Vol. 8692, pp. 184-199, Sep. 2014. https://doi.org/10.1007/978-3-319-10593-2_13.
- [15] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network", CVPR, Honolulu, Hawaii, USA, pp. 2790-2798, Jul. 2017.
- [16] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration", ICCV, Venice, Italy, Oct. 2017.
- [17] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super resolution convolutional neural network", Computer Vision – ECCV 2016, Amsterdam, Netherlands, Vol. 9906, pp. 391-407, Oct. 2016. https://doi.org/10.1007/978-3-319-46475-6_25.
- [18] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks", CVPR, Honolulu, HI, USA, pp. 2261-2269, Jul. 2017. <https://doi.org/10.1109/CVPR.2017.243>.
- [19] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections", ICCV, Venice, Italy, Oct. 2017. <https://doi.org/10.1109/ICCV.2017.514>.
- [20] Z. Liu, et al., "Swin transformer: Hierarchical vision transformer using shifted windows", Proc. of the IEEE/CVF international conference on computer vision, Montreal, QC, Canada, pp. 10012-10022, Oct. 2021.
- [21] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte, "Swinir: Image restoration using swin transformer", Proc. of the IEEE/CVF international conference on computer vision, Montreal, BC, Canada, pp. 1833-1844, Oct. 2021. <https://doi.org/10.1109/ICCVW54120.2021.00210>.
- [22] M. V. Conde, et al., "Swin2sr: Swinv2 transformer for compressed image super-resolution and restoration", Computer Vision – ECCV 2022 Workshops, Tel Aviv, Israel, Vol. 13802, pp. 669-687, Oct. 2022. https://doi.org/10.1007/978-3-031-25063-7_42.
- [23] Z. Liu, et al., "Swin transformer v2: Scaling up capacity and resolution", Proc. of the IEEE/CVF conference on computer vision and pattern recognition, New Orleans, LA, USA, pp. 12009-12019, Jun. 2022. <https://doi.org/10.1109/CVPR52688.2022.01170>.
- [24] J.-J. Liu, et al., "Improving convolutional networks with self-calibrated convolutions", CVPR, Seattle, WA, USA, pp. 10096-10105, Jun. 2020. <https://doi.org/10.1109/CVPR42600.2020.01011>.
- [25] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks", AISTATS, Fort Lauderdale, USA, Vol. 15, pp. 315-323, Apr. 2011.
- [26] 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", CVPR, Las Vegas Nevada, pp. 1874-1883, Jun.-Jul. 2016.
- [27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", IEEE transactions on image processing, Vol. 13, No. 4, pp. 600-612, Apr. 2004. <https://doi.org/10.1109/TIP.2003.819861>.

저자소개

강 동 균 (Dong-Gyun Kang)



2019년 8월 : 가톨릭대학교
미디어기술콘텐츠학과(공학사)
2019년 8월 ~ 현재 : 가톨릭대학교
디지털미디어학과(공학석박사
통합과정)
관심분야 : 인공지능, 딥러닝,
컴퓨터 비전

이 상 국 (Sang-Goog Lee)



1988년 2월 : 인하대학교
전기공학과(학사)
1990년 2월 : 인하대학교
전기공학과(석사)
1994년 10월 : 프랑스 국립
응용과학원(Institut National des
Sciences Appliquees) 전자공학과

(공학박사)

1995년 9월 ~ 1999년 8월 : 프랑스 국립 루앙(ROUEN Univ.) 대학교 전기전자공학부 조교수
1990년 9월 ~ 1999년 8월 : 프랑스 국립응용과학원
PSI(Pertception System & Information)연구소 연구원
1999년 9월 ~ 2001년 5월 : 삼성전자 중앙연구소
Wearable Computer Project 팀장, 수석연구원
2001년 6월 ~ 2006년 8월 : 삼성종합기술원
Master(기술명인) Mobile Augmented Reality Project
팀장, Ubicomp Interactive Device 그룹장
2006년 9월 ~ 현재 : 가톨릭대학교 미디어기술콘텐츠학과
교수/ 가톨릭대학교 대학원 디지털미디어학과 교수
관심분야 : 스마트센서와 시스템, 인간-컴퓨터 상호작용,
유비쿼터스/웨어러블컴퓨팅, 가상 및 증강 현실,
인공지능