

비지역 희소 어텐션 메커니즘을 활용한 초해상화

김소원¹, 박한훈^{1*}

¹부경대학교 전자공학과

Super-Resolution Using NLSA Mechanism

Sowon Kim¹, Hanhoon Park^{1*}

¹Department of Electronic Engineering, Pukyong National University

요약 딥러닝이 발전하면서 초해상화 기술은 단순 보간법(Interpolation)에서 벗어나 딥러닝을 활용해 발전하고 있다. 딥러닝을 사용한 초해상화 기술은 합성곱 신경망(Convolutional Neural Network, CNN) 기반의 연구가 일반적이지만, 최근에는 어텐션(Attention) 메커니즘을 활용한 초해상화 연구가 활발히 진행되고 있다. 본 논문에서는 어텐션 메커니즘 중 하나인 비지역 희소 어텐션(Non-Local Sparse Attention, NLSA)을 활용한 초해상화 성능 향상 방법을 제안한다. 실험을 통해 NLSA를 함께 활용하면 기존 초해상화 신경망 모델인 IMDN, CARN, OISR-LF-s의 성능이 향상되는 것을 확인할 수 있었다.

• 주제어 : 초해상화, 딥러닝, 어텐션 메커니즘, NLSA, IMDN, CARN, OISR-LF-s

Abstract With the development of deep learning, super-resolution (SR) methods have tried to use deep learning mechanism, instead of using simple interpolation. SR methods using deep learning is generally based on convolutional neural networks (CNN), but recently, SR researches using attention mechanism have been actively conducted. In this paper, we propose an approach of improving SR performance using one of the attention mechanisms, non-local sparse attention (NLSA). Through experiments, we confirmed that the performance of the existing SR models, IMDN, CARN, and OISR-LF-s can be improved by using NLSA.

• Key Words : Super-resolution, Deep learning, Attention mechanism, NLSA, IMDN, CARN, OISR-LF-s

Received 16 March 2022, Revised 26 March 2022, Accepted 28 March 2022

* Corresponding Author Hanhoon Park, Department of Electronic Engineering, Pukyong National University, 45, Yongso-ro, Nam-gu, Busan, Korea. E-mail: hanhoon.park@pknu.ac.kr

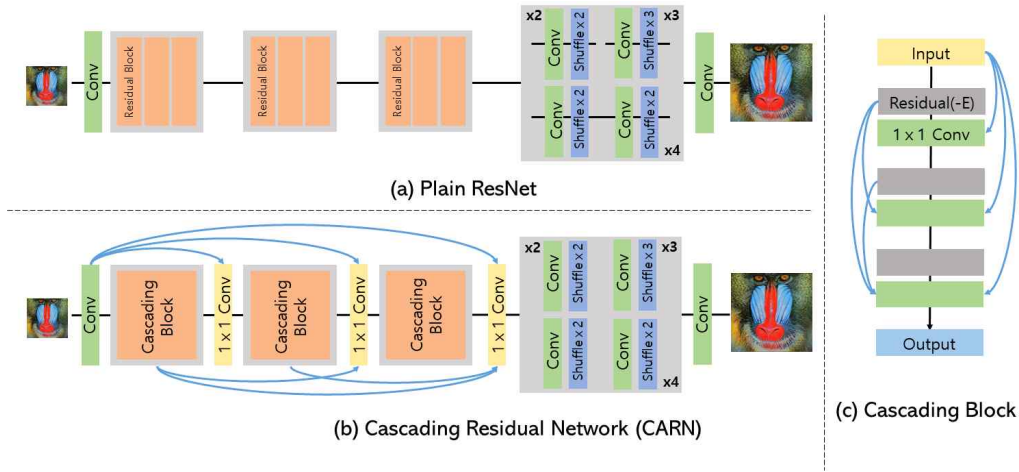


Fig. 1. Network architectures of plain ResNet and CARN

I. 서론

초해상화(Super-Resolution)란 저해상도 영상을 고해상도의 영상으로 변환하는 영상처리 기술로서, 미디어, 인터넷, 보안 등 다양한 분야에서 활용되고 있다. 일반적으로 초해상화 기술은 부적절하게 정립된 역 문제(ill-posed inverse problem)로 정의되기 때문에 이를 해결하기 위한 다양한 초해상화 기법들이 제안되어 왔다. 초기의 초해상화 기술은 쌍선형 보간법(bilinear interpolation), 쌍삼차 보간법(bicubic interpolation) 등의 보간법을 이용하였다. 하지만 최근에는 딥러닝의 발전으로 초해상화 기술 또한 딥러닝을 이용하여 연구되고 있다[1-2].

딥러닝 기반의 초해상화의 알고리즘 중 가장 일반화된 것은 합성곱 신경망(Convolutional Neural Network, CNN)을 활용하는 것이다. CNN 기반 초해상화 알고리즘의 초기 연구인 SRCNN(Super-Resolution CNN)이 제안된 이후 다양한 CNN 기반의 초해상화 기술이 연구되었다. 하지만 최근에는 CNN만 이용한 초해상화에는 한계가 있음을 지적하며 이를 극복하기 위해 자연어 처리에 쓰였던 어텐션(attention) 메커니즘을 활용한 초해상화 기술들이 제안되고 있다. 어텐션을 활용한 대표적인 초해상화 신경망으로는 RCAN, SAN 등이 있으며, 기존 CNN 기반으로만 만들어진 초해상화보다 우수한 성능을 보여주었다[3-4].

본 논문은 어텐션 메커니즘을 활용하여 초해상화 성능을 개선하는 것을 목표로 한다. 이를 위해, 다양한

구조의 어텐션 모듈을 초해상화 모델에 적용했을 때의 성능 변화를 분석한 후, 어텐션 모듈의 성능 개선 효과를 극대화하기 위해 비지역 희소 어텐션(Non-Local Sparse Attention, NLSA) 모듈을 함께 사용하는 것을 제안한다. 기존 어텐션 모듈과 NLSA 모듈 결합의 유효성을 실험적으로 확인한다.

II. 관련연구

2.1 초해상화 신경망

이 절에서는 본 논문의 실험에서 사용된 초해상화 신경망 모델들을 간략히 설명한다.

초기 신경망 모델인 SRCNN 이후, 다양한 구조와 형태를 가진 신경망들이 제안되어왔다. 몇몇 연구들은 초해상화 성능을 향상시키기 위해 신경망을 더 깊고 넓게 만들었지만 이러한 구조적인 확장은 느리고 무거워진다는 문제점이 있다. 최근에는 이런 문제점을 해결하기 위해 빠르고 가벼운 초해상화 신경망 모델들이 제안되고 있으며 그 중 CARN(Cascading Residual Network)은 모바일에서 사용하기 위해 경량화된 초해상화 모델이다[5]. CARN은 ResNet(Residual Neural Network)을 기반으로 만든 모델이지만 ResNet과의 차이점은 지역적(local), 전역적(global) 연쇄 구조(cascade structure) 모듈을 사용했다는 것이다. 전역적 연쇄 모듈은 중간층(layer)의 출력이 상위층으로 연쇄적으로

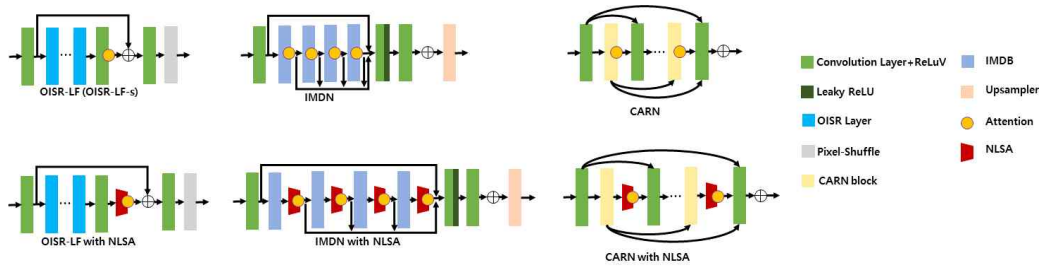


Fig. 2. Addition of NLSA to SR models

연결되며 마지막으로 단일 1×1 합성곱 층으로 수렴하게 된다(그림 1의 파란색 선 참조). 지역적 연쇄 모듈 또한 전역적 연쇄 모듈과 같은 원리이지만 행해지는 구간이 연쇄 블록 내부이다(그림 1(c) 참조). 이 두 가지의 연쇄 구조 모듈을 통해 CARN은 여러 계층 정보를 더 효율적으로 얻을 수 있다.

IDN(Information Distillation Network)은 지식 증류를 통해 경량화된 신경망으로, 증류된 특징(feature)을 얻기 위해 채널 분리 연산(channel split operation)을 사용한다[6]. 신경망 중간의 특징을 두 부분으로 나누어 하나는 유지하고 나머지는 합성곱 층을 통과 시켜 프로세스를 진행한다. 이러한 과정을 통해 가벼운 모델에서도 좋은 성능을 낼 수 있다. 이후, IDN을 개선하여 IMDN(Information Multi-Distillation Network)이 나왔으며, IMDB(Information Multi-Distillation Blocks)가 반복되는 구조를 가진다[7]. IMDB 안에서는 채널 분리 연산이 여러 번 반복된다. 초해상화 성능을 향상시키기 위해 IMDB 뒤에 CCA(Contrast-aware Channel Attention)가 사용된다(그림 2의 IMDN 구조 참조).

OISR(ODE-Inspired SR)은 상미분방정식(ODE)에 영감을 받아 만든 초해상화 모델이다[8]. OISR에서는 상미분방정식을 활용해 새로운 잔차 모듈과 세 개의 특별한 블록(LF-block, RK2-block, RK3-block)을 설계하였고, 각 블록을 사용한 서로 다른 모델이 제안되었다. 본 논문의 실험에서는 LF-block을 갖는 OISR-LF 모델을 사용한다. 또한 OISR의 크기에 따라 작은 크기의 OISR-LF-s와 원래 크기의 OISR-LF 모델로 나누어 지는데 본 논문의 실험에서는 작은 크기의 OISR-LF-s를 사용한다.

2.2 NLSA 메커니즘

초해상화에서 비지역(non-local) 연산과 희소 표현

(sparse representation) 모두 중요한 역할을 수행할 수 있으며, 최근 이를 결합한 새로운 어텐션 모듈인 NLSA 모듈이 제안되었다[9]. NLSA는 비지역 연산으로 영상 내 넓은 영역의 정보를 얻으며, 희소 표현을 통해 견고성과 높은 효율성을 얻을 수 있도록 하였다.

2.2.1 희소 표현

희소 표현은 관측된 신호를 적은 수의 기저(basis) 신호의 선형 결합으로 구성한다. 식 (1)에서 y 는 관측된 신호, a 는 희소 계수, x 는 기저 신호를 나타낸다. a 의 희소성은 a 가 대부분 0이거나 0에 가까운 특성으로 정의된다.

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n. \quad (1)$$

식 (1)은 식 (2)로 변환할 수도 있다. 이때 D 는 사전(dictionary)이고, α_i 는 x_i 의 희소 계수이며, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]$ 으로 나타내어진다.

$$y = D\alpha, \quad s.t. \|\alpha\|_0 \leq k. \quad (2)$$

희소 표현은 영상 재구축에 효과적이라고 입증되어 NLSA에서는 희소 표현을 비지역 연산에 포함하였다.

2.2.2 비지역 연산

CNN에서 넓은 수용 영역(receptive field)을 가지기 위해서는 합성곱 층을 더 쌓을 수밖에 없는데 이때 연산량이 커지면서 비효율적이게 된다. 비지역 연산(non-local operation)은 이러한 문제점을 해결하기 위한 것으로, 자연(natural) 영상에서 비슷한 패치가 여러 번 발생하는 것을 이용해 넓은 수용 영역을 효과적으로 받아들인다. NLSA는 이러한 비지역 연산의 특성을 이용하여 복잡성을 감소시키면서 영상에서 더 넓은 범위의 정보를 받아들일도록 하였다.

2.2.3 비지역 희소 어텐션(NLSA)

비지역 어텐션(non-local attention)은 여전히 지역적이며 연산량의 문제가 있다. 이를 보완하기 위해 희소성을 사용하여 비지역적이며 연산량을 줄이는 NLSA가 제안되었다[9]. 식 (3)은 식 (2)의 희소성을 이용한 NLSA 표현식이다.

$$y_i = D\alpha_i = \sum_{j \in \delta_i} \frac{f(x_i, x_j)}{\sum_{\hat{j} \in \delta_i} f(x_i, x_{\hat{j}})} g(x_j). \quad (3)$$

여기서, $x_i, x_j, x_{\hat{j}}$ 는 각각 i, j, \hat{j} 위치에 대한 pixel-wise 특징이다. $f()$ 는 상호 유사성, 즉 두 패치 사이의 유사성을, $g()$ 는 특징 변화를 나타내는 함수이다. δ_i 는 0이 아닌 α_i 의 값을 나타내며 동시에 어텐션 해야 하는 위치들을 의미한다. 그리고 이러한 위치들을 그룹화한 것이 어텐션 버킷(attention bucket)으로, 어텐션 버킷은 LSH(Locality Sensitive Hashing)을 통해 그룹화되며, 계산 복잡성을 줄이는 효과를 가진다.

III. 실험 방법

3.1 어텐션 모듈 적용

기존 초해상화 신경망 모델 중에 IMDN, CARN, OISR-LF-s를 사용하여 어텐션 모듈 적용 실험을 진행했다. 그림 2에서 보는 것처럼 각 신경망에 다양한 구조를 가지는 어텐션 모듈을 새로 추가하거나 기존 어텐션 모듈 대신 사용하는 형태로 적용하였다. 실험에 사용된 어텐션 모듈의 구조 및 위치는 기존 연구를 참조하였다[10]. CARN 모델의 경우, 어텐션 모듈은 CARN 블록 뒤에 삽입하였으며, OISR-LF-s의 경우, 각 어텐션 모듈은 LF-block 맨 끝에 삽입하였다. IMDN은 원래 CCA라는 어텐션을 갖고 있으므로 CCA 자리를 다른 어텐션 모듈로 대체하였다. NLSA는 각 어텐션 모듈의 앞에 적용하였으며, 실험 결과표에서 NLSA를 적용한 경우 각 모델명에 NLSA를 삽입하여 표기하였다.

실험에 사용된 기본 어텐션 모듈은 CA, CA_max, SA, SA_avg, BAM, BAM_ADD, CBAM로 총 7개이며, 각 어텐션에 대한 설명은 다음과 같다.

- CA: 채널 어텐션(channel attention)으로, 전역 평균 풀링(global average pooling)을 사용함
- CA_max: 채널 어텐션 수행 후, 최대 풀링(max pooling)을 함
- SA: 공간 어텐션(spatial attention)으로, 최대 풀링을 사용함
- SA_avg: 공간 어텐션 수행 후 전역 평균 풀링을 함
- BAM: 채널 어텐션과 공간 어텐션을 병렬 연결함
- BAM_ADD: CA_max와 SA_avg를 병렬 연결함
- CBAM: 채널 어텐션과 공간 어텐션을 직렬 연결함

3.2 학습 및 테스트 데이터셋

실험에서 학습 데이터셋으로 DIV2K를 사용하였으며 [11], DIV2K 내 학습 데이터셋과 테스트 데이터셋을 합쳐 900장의 영상을 하나의 데이터셋으로 만들었다. 평가(evaluation) 데이터셋은 Set5를 사용하였다[12].

테스트 데이터셋은 총 4개로 Set5, Set14, BSDS100, Manga109를 사용하였다[13-15]. Set5, Set14, BSDS100, Manga109는 초해상화를 평가할 때 쓰이는 일반적인 테스트셋이다. Manga109는 512×512 크기로 잘라서 사용하였다.

3.3 실험 환경

최적화를 위해 Adam optimizer를 사용하였으며, 초기 학습률은 0.0001로 설정하고, L1 손실(loss)을 사용하였다. 배치 크기는 8이며, 64 × 64 크기의 패치를 사용하여 학습하였다. 학습 데이터 증가(augmentation)를 위해 학습 데이터 각 영상의 모퉁이에서 4장, 중심에서 1장씩 잘라내어 영상 수를 5배로 늘렸다. 학습 시 epoch는 100으로 설정하였다. 그 외 파라미터값들은 기존 연구에서 사용된 파라미터와 동일하게 설정하였다[10].

실험은 i7-8700, 3.20GHz CPU와 32GB RAM, Nvidia RTX 2080 Ti GPU를 가진 PC에서 수행되었다. 각 초해상화 모델 및 어텐션 모듈 구현은 PyTorch로 구현된 Python 코드를 활용하였다[16]. PyTorch, Python의 버전은 각각 1.7.1, 3.6.13이었다.

각 실험에서의 PSNR과 SSIM의 수치를 구하여 초해상화 결과의 정량적인 화질을 비교하였다.

Table 1. Results of CARN with different attention modules

	Set5	Set14	BSDS100	Manga109
CARN	37.535 / 0.959	33.247 / 0.914	33.653 / 0.922	37.115 / 0.974
CARN_NLSA	37.589 / 0.959 ↑	33.302 / 0.914 ↑	33.710 / 0.923 ↑	37.277 / 0.974 ↑
CARN_CA	37.498 / 0.959	33.217 / 0.914	33.640 / 0.922	37.019 / 0.973
CARN_NLSA_CA	37.511 / 0.959 ↑	33.231 / 0.914 ↑	33.658 / 0.922 ↑	37.046 / 0.973 ↑
CARN_CA_max	37.449 / 0.959	33.171 / 0.914	33.596 / 0.922	36.873 / 0.973
CARN_NLSA_CA_max	37.541 / 0.959 ↑	33.206 / 0.914 ↑	33.656 / 0.922 ↑	37.029 / 0.973 ↑
CARN_SA	37.495 / 0.959	33.239 / 0.914	33.647 / 0.922	37.103 / 0.973
CARN_NLSA_SA	37.512 / 0.959 ↑	33.200 / 0.914	33.635 / 0.922	36.925 / 0.973
CARN_SA_avg	37.449 / 0.958	33.190 / 0.914	33.600 / 0.922	36.928 / 0.973
CARN_NLSA_SA_avg	37.506 / 0.959 ↑	33.218 / 0.914 ↑	33.651 / 0.922 ↑	37.070 / 0.973 ↑
CARN_BAM	37.417 / 0.958	33.137 / 0.913	33.563 / 0.921	36.832 / 0.972
CARN_NLSA_BAM	37.465 / 0.959 ↑	33.186 / 0.913 ↑	33.596 / 0.921 ↑	36.947 / 0.973 ↑
CARN_BAM_ADD	37.409 / 0.958	33.147 / 0.913	33.583 / 0.921	36.820 / 0.972
CARN_NLSA_BAM_ADD	37.401 / 0.958	33.134 / 0.913	33.577 / 0.921	36.739 / 0.972
CARN_CBAM	37.438 / 0.959	33.147 / 0.913	33.571 / 0.921	36.783 / 0.972
CARN_NLSA_CBAM	37.436 / 0.958	33.187 / 0.913 ↑	33.611 / 0.922 ↑	36.863 / 0.972 ↑

Table 2. Results of IMDN with different attention modules

	Set5	Set14	BSDS100	Manga109
IMDN	37.586 / 0.959	33.264 / 0.914	33.694 / 0.923	37.191 / 0.974
IMDN_NLSA	37.614 / 0.959 ↑	33.305 / 0.915 ↑	33.741 / 0.923 ↑	37.302 / 0.974 ↑
IMDN_CA	37.565 / 0.959	33.249 / 0.914	33.671 / 0.922	37.159 / 0.974
IMDN_NLSA_CA	37.597 / 0.959 ↑	33.285 / 0.915 ↑	33.743 / 0.923 ↑	37.286 / 0.974 ↑
IMDN_CA_max	37.575 / 0.959	33.233 / 0.914	33.644 / 0.922	37.030 / 0.973
IMDN_NLSA_CA_max	37.571 / 0.959	33.257 / 0.914 ↑	33.692 / 0.923 ↑	37.089 / 0.973 ↑
IMDN_SA	37.568 / 0.959	33.272 / 0.915	33.707 / 0.923	37.162 / 0.974
IMDN_NLSA_SA	37.590 / 0.959 ↑	33.305 / 0.915 ↑	33.737 / 0.923 ↑	37.209 / 0.974 ↑
IMDN_SA_avg	37.618 / 0.959	33.286 / 0.915	33.716 / 0.923	37.239 / 0.974
IMDN_NLSA_SA_avg	37.580 / 0.959	33.291 / 0.914 ↑	33.730 / 0.923 ↑	37.243 / 0.974 ↑
IMDN_BAM	37.591 / 0.959	33.275 / 0.914	33.687 / 0.923	37.103 / 0.974
IMDN_NLSA_BAM	37.587 / 0.959	33.289 / 0.915 ↑	33.720 / 0.923 ↑	37.229 / 0.974 ↑
IMDN_BAM_ADD	37.537 / 0.959	33.127 / 0.913	33.597 / 0.922	36.875 / 0.973
IMDN_NLSA_BAM_ADD	37.583 / 0.959 ↑	33.275 / 0.915 ↑	33.696 / 0.923 ↑	37.058 / 0.973 ↑
IMDN_CBAM	37.590 / 0.959	33.207 / 0.914	33.628 / 0.922	36.937 / 0.973
IMDN_NLSA_CBAM	37.558 / 0.959	33.224 / 0.914 ↑	33.675 / 0.922 ↑	37.062 / 0.973 ↑

Table 3. Results of OISR-LF-s with different attention modules

	Set5	Set14	BSDS100	Manga109
OISR_LF_s	37.622 / 0.959	33.299 / 0.915	33.738 / 0.923	37.364 / 0.974
OISR_LF_s_NLSA	37.623 / 0.959 ↑	33.341 / 0.915 ↑	33.757 / 0.923 ↑	37.310 / 0.974
OISR_LF_s_CA	37.649 / 0.959	33.343 / 0.915	33.749 / 0.923	37.364 / 0.974
OISR_LF_s_NLSA_CA	37.667 / 0.959 ↑	33.345 / 0.915 ↑	33.747 / 0.923	37.320 / 0.974
OISR_LF_s_CA_max	37.673 / 0.959	33.350 / 0.915	33.739 / 0.923	37.345 / 0.974
OISR_LF_s_NLSA_CA_max	37.681 / 0.960 ↑	33.368 / 0.915 ↑	33.760 / 0.923 ↑	37.379 / 0.974 ↑
OISR_LF_s_SA	37.669 / 0.959	33.360 / 0.915	33.763 / 0.923	37.358 / 0.974
OISR_LF_s_NLSA_SA	37.674 / 0.960 ↑	33.361 / 0.915 ↑	33.762 / 0.923	37.414 / 0.974 ↑
OISR_LF_s_SA_avg	37.683 / 0.960	33.368 / 0.915	33.773 / 0.923	37.409 / 0.974
OISR_LF_s_NLSA_SA_avg	37.645 / 0.959	33.362 / 0.915	33.758 / 0.923	37.378 / 0.974
OISR_LF_s_BAM	37.685 / 0.960	33.350 / 0.915	33.753 / 0.923	37.379 / 0.974
OISR_LF_s_NLSA_BAM	37.679 / 0.959	33.348 / 0.915	33.757 / 0.923 ↑	37.336 / 0.974
OISR_LF_s_BAM_ADD	37.627 / 0.959	33.342 / 0.915	33.724 / 0.923	37.280 / 0.974
OISR_LF_s_NLSA_BAM_ADD	37.690 / 0.960 ↑	33.389 / 0.916 ↑	33.775 / 0.923 ↑	37.468 / 0.975 ↑
OISR_LF_s_CBAM	37.636 / 0.959	33.345 / 0.915	33.740 / 0.923	37.355 / 0.974
OISR_LF_s_NLSA_CBAM	37.637 / 0.959 ↑	33.353 / 0.915 ↑	33.745 / 0.923 ↑	37.356 / 0.974 ↑

IV. 실험 결과 및 분석

4.1 어텐션 모듈 적용

표 1은 CARN 모델에 7개의 어텐션(CA, CA_max, SA, SA_avg, BAM, BAM_ADD, CBAM)을 적용한 실험 결과를 보여준다. CARN의 경우, 어텐션 모듈의 구조나 형태에 상관없이 어텐션 모듈 적용으로 인한 성능 향상을 확인할 수 없었다. 반면, 표 2와 3은 IMDN과 OISR-LF-s 모델에 어텐션 모듈을 적용한 실험 결과로, 일부 어텐션 모듈을 적용함으로써 성능이 향상됨을 확인할 수 있었다. IMDN의 경우 SA_avg 모듈을 적용함으로써 모든 데이터셋에 대해 성능이 향상되었으며 성능 향상 정도도 가장 컸다. 각 데이터셋에 대해 최소 0.022dB, 최대 0.048dB 향상되었다. OISR-LF-s의 경우 Manga109 데이터셋을 제외한 모든 데이터셋에서 어텐션 모듈을 적용함으로써 성능이 향상되었으며, CA와 SA_avg 모듈은 Manga109를 포함한 모든 데이터셋에서 성능이 향상되었다. SA_avg 모듈의 성능이 가장 좋았으며, 각 데이터셋에 대해 최소 0.035, 최대 0.069dB 향상되었다.

하지만 표 1~3의 결과로부터 어텐션 모듈을 적용하는 것이 초해상화 모델의 성능 향상에 효과적이라고 하기에는 어렵다. OISR-LF-s만 보았을 때는 어텐션 모듈을 적용하는 것이 효과적으로 보이지만 CARN의 성능은 오히려 떨어졌으며, IMDN의 경우 성능 개선 효과가 제한적이었다. 결과적으로 모델이나 데이터셋에 따라 어텐션 모듈의 성능 개선 효과 차이가 크다는 것을 알 수 있었다.

4.2 NLSA 적용

표 1~3의 결과로부터 CARN, IMDN, OISR-LF-s 모델에 7개의 어텐션 모듈과 함께 NLSA를 추가적으로 적용하였을 때의 성능 변화를 확인할 수 있다. NLSA를 추가하기 전에 비해 NLSA를 추가했을 때 성능이 향상되는 경우 그 결괏값에 “↑” 표시를 하여 강조하였다. 어텐션 모듈 및 데이터셋에 따라 차이는 있지만, 모든 초해상화 모델에서 NLSA를 추가함으로써 성능이 더욱 향상되는 양상을 보였다.

CARN은 SA와 BAM_ADD 모듈 적용 시 NLSA를 추가함으로써 성능이 오히려 저하되었으나, 나머지 어텐션 모듈을 적용할 경우 NLSA를 함께 사용함으로써 최소 0.013dB의 성능 향상 효과를 보였다. 어텐션 모듈

적용 없이 CARN에 NLSA만을 추가했을 때 (CARN_NLSA) 성능이 가장 우수했으며, 모든 데이터셋에 대해 어텐션 모듈을 적용하지 않은 기존 CARN 보다 초해상화 성능이 크게 향상되었다. 특히 Manga109 데이터셋의 경우 0.162dB 향상되었다.

IMDN의 경우, Set5 데이터셋에 대해서는 일부 어텐션 모듈 적용 시 NLSA를 추가함으로써 성능이 저하되었으나, 나머지 테스트 데이터셋에 대해서는 NLSA를 추가함으로써 0.004dB부터 0.183dB까지 항상 성능이 향상되었다. CARN의 경우와 마찬가지로 다른 어텐션 모듈을 사용하지 않고 IMDN에 NLSA만을 추가했을 때의 성능이 가장 우수했으며, 기존 IMDN에 비해 최대 0.111dB 향상되었다.

OISR-LF-s의 경우, 데이터셋에 따라 다르지만, SA_avg와 BAM 모듈을 적용한 경우를 제외한 대부분의 결과에서 NLSA를 추가함으로써 성능이 향상되었다. BAM_ADD 모듈에 추가했을 때 성능 향상 효과가 가장 컸으며 최대 0.188dB까지 향상되었다. 결과적으로 BAM_ADD 모듈과 NLSA를 함께 사용했을 때 초해상화 성능이 가장 우수했으며, 모든 데이터셋에 대해 기존 OISR-LF-s보다 최소 0.037dB, 최대 0.104dB 향상되었다.

실험 결과로부터 어텐션 모듈 앞에 NLSA를 추가하면 초해상화에 도움이 되는 더 많은 정보를 어텐션 모듈의 입력으로 전달해 줌으로써 초해상화 성능을 향상시킬 수 있다는 것을 확인하였다. 그러나, 초해상화 모델의 구조, 어텐션 모듈의 성향, 데이터셋 종류에 따라 성능 향상 효과가 없거나 크지 않은 경우도 있음으로 NLSA 추가로 인한 성능 개선 효과는 제한적이다.

V. 결론 및 향후 연구

본 논문은 초해상화 성능 향상을 위해 어텐션 모듈과 NLSA를 함께 사용하는 것을 제안하였다. 비록 데이터셋 종류와 초해상화 모델의 구조에 따라 성능 향상이 제약적이었지만, 어텐션 모듈과 NLSA를 함께 사용하였을 때 대부분의 경우 어텐션 모듈을 사용하지 않거나 어텐션 모듈만을 사용하였을 때보다 성능이 향상된 것을 확인할 수 있었다. 그러나, 어텐션 모듈 적용에 의한 성능 개선 효과가 초해상화 모델에 따라 크게 달라지기 때문에 NLSA를 함께 사용하여 달성될 수 있는 성능 향상에는 한계가 있었다. 이를 보완하기 위해서는 초해상화 모델의 구조에 상관없이 높은 성능을

보여줄 수 있는 어텐션 모듈 개발을 위한 향후 연구가 필요하다.

ACKNOWLEDGMENTS

본 연구는 산업통상자원부와 한국산업기술진흥원의 “지역혁신클러스터육성사업(R&D, P0004797)” 으로 수행된 연구결과입니다.

REFERENCES

- [1] Z. Wang, J. Chen, and S. C. H. Hoi, “Deep learning for image super-resolution: a survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, pp. 3365-3387, 2021.
- [2] S. Kim and H. Park, “CG/VR image super-resolution using balanced attention mechanism,” *Journal of Korea Institute of Convergence Signal Processing*, vol. 22, no. 4, pp. 156-163, 2021.
- [3] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” *Proc. ECCV*, 2018.
- [4] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, “Second-order attention network for single image super-resolution,” *Proc. CVPR*, pp. 11065-11074, 2019.
- [5] N. Ahn, B. Kang, and K.-A. Sohn, “Fast, accurate, and lightweight super-resolution with cascading residual network,” *Proc. ECCV*, 2018.
- [6] Z. Hui, X. Wang, and X. Gao, “Fast and accurate single image super-resolution via information distillation network,” *Proc. CVPR*, pp. 723-731, 2018.
- [7] Z. Hui et al., “Lightweight image super-resolution with information multi-distillation network,” *Proc. ACM MM*, 2019.
- [8] X. He et al., “ODE-inspired network design for single image super-resolution,” *Proc. CVPR*, 2019.
- [9] Y. Mei, Y. Fan, and Y. Zhou, “Image super-resolution with non-local sparse attention,” *Proc. CVPR*, 2021.
- [10] F. Wang, H. Hu, and C. Shen, “BAM: a lightweight and efficient balanced attention mechanism for single image super resolution,” *arXiv preprint arXiv:2104.07566*, 2021.
- [11] E. Agustsson and R. Timofte, “NTIRE 2017 challenge on single image super-resolution: dataset and study,” *Proc. CVPRW*, 2017.
- [12] M. Bevilacqua, A. Roumy, C. Guillemot, and A. Morel, “Low complexity single image super-resolution based on nonnegative neighbor embedding,” *Proc. BMVC*, 2012.
- [13] R. Zeyde, M. Elad, and M. Protter, “On single image scale-up using sparse-representations,” *Proc. International Conference on Curves and Surfaces*, pp. 711-730, 2010.
- [14] 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,” *Proc. ICCV*, vol. 2, pp. 416-423, 2001.
- [15] R. Narita, K. Tsubota, T. Yamasaki, and K. Aizawa, “Sketch-based manga retrieval using deep features,” *Proc. ICDAR*, pp. 49-53, 2017.
- [16] <https://github.com/dandingbudanding/BAM>

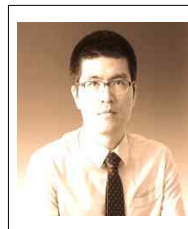
저자 소개

김 소 원 (Sowon Kim)



2019년 3월~현재 : 부경대학교
전자공학과 학부생
관심분야 : 딥러닝, 초해상화

박 한 훈 (Hanhoon Park)



2000년 2월 : 한양대학교
전자통신전파공학과(공학사)
2002년 2월 : 한양대학교
전자통신전파공학과(공학석사)
2007년 8월 : 한양대학교
전자통신전파공학과(공학박사)
2012년 3월~현재 : 부경대학교

전자공학과 교수

관심분야 : 증강현실, 인간컴퓨터상호작용,
컴퓨터비전/그래픽스, 딥러닝 응용