

Balanced Attention Mechanism을 활용한 CG/VR 영상의 초해상화

김소원¹, 박한훈^{1*}

¹부경대학교 전자공학과

CG/VR Image Super-Resolution Using Balanced Attention Mechanism

Sowon Kim¹, Hanhoon Park^{1*}

¹Department of Electronic Engineering, Pukyong National University

요 약 어텐션(Attention) 메커니즘은 딥러닝 기술을 활용한 다양한 컴퓨터 비전 시스템에서 활용되고 있으며, 초해상화(Super-resolution)를 위한 딥러닝 모델에도 어텐션 메커니즘을 적용하고 있다. 하지만 어텐션 메커니즘이 적용된 대부분의 초해상화 기법들은 Real 영상의 초해상화에만 초점을 맞추어서 연구되어, 어텐션 메커니즘을 적용한 초해상화가 CG나 VR 영상 초해상화에도 유효한지는 알기 어렵다. 본 논문에서는 최근에 제안된 어텐션 메커니즘 모듈인 BAM(Balanced Attention Mechanism) 모듈을 12개의 초해상화 딥러닝 모델에 적용한 후, CG나 VR 영상에서도 성능 향상 효과를 보이는지 확인하는 실험을 진행하였다. 실험 결과, BAM 모듈은 제한적으로 CG나 VR 영상의 초해상화 성능 향상에 기여하였으며, 데이터 특징과 크기, 그리고 네트워크 종류에 따라 성능 향상도가 달라진다는 것을 확인할 수 있었다.

• 주제어 : 어텐션 메커니즘, 초해상화, CG/VR, BAM, 딥러닝

Abstract Attention mechanisms have been used in deep learning-based computer vision systems, including single image super-resolution (SISR) networks. However, existing SISR networks with attention mechanism focused on real image super-resolution, so it is hard to know whether they are available for CG or VR images. In this paper, we attempt to apply a recent attention module, called balanced attention mechanism (BAM) module, to 12 state-of-the-art SISR networks, and then check whether the BAM module can achieve performance improvement in CG or VR image super-resolution. In our experiments, it has been confirmed that the performance improvement in CG or VR image super-resolution is limited and depends on data characteristics, size, and network type.

• Key Words : Attention mechanism, Super-resolution, CG/VR, BAM, Deep learning

Received 30 November 2021, Revised 26 December 2021, Accepted 28 December 2021

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

I. 서론

초해상화(Super-resolution)란 영상처리 및 컴퓨터 비전의 한 분야로, 저해상도(Low-resolution) 영상을 고해상도(High-resolution) 영상으로 변환하는 기술을 의미한다. 초해상화는 유일한 해가 존재하지 않는 문제(ill-posed problem)이기 때문에 다양한 초해상화 기법들이 제안되어 왔는데, 일반적으로 단일 시점 영상만 이용하는 SISR(Single Image Super-resolution)과 시점이 다른 여러 영상을 함께 이용하는 MISR(Multiple Image Super-resolution)로 나뉜다. 본 논문에서는 SISR 방법을 이용한 초해상화 연구에 초점을 둔다.

딥러닝 연구가 활성화되기 전에도 다양한 SISR 방법들이 존재했으나, 딥러닝 기술을 도입함으로써 SISR 방법의 성능은 크게 개선될 수 있었다. 딥러닝 기술을 이용한 SISR 방법의 시초인 SRCNN[1]의 성공적인 출현 이후, 딥러닝 기술의 발전과 함께 다양한 형태의 딥러닝 기술을 접목한 SISR 방법들이 제안되고 있다. 최근에는 어텐션(attention) 메커니즘을 적용한 초해상화 연구가 활발히 진행되고 있으며, 큰 오버헤드 없이 기존 딥러닝 기반 SISR 방법의 성능을 크게 향상시킬 수 있다[2-5].

BAM(Balanced Attention Mechanism) 모듈은 어텐션 메커니즘을 활용한 최신 모듈로써, 일반적인 딥러닝 모듈의 성능 향상에 초점을 두고 기존에 개발된 어텐션 모듈[6]을 초해상화에 적용하기 위해 변형, 개선된 것이다[7]. 그러나, BAM 모듈에서는 Real 영상의 초해상화만 초점을 두었기 때문에, Real 영상과는 다른 생성 원리나 과정을 가지는 CG나 VR 영상 초해상화에 해당 모듈이 유용한지 알 수 없다. 따라서, 본 논문에서는 BAM 모듈이 CG나 VR 영상의 초해상화에서도 효과가 있는지를 확인하기 위해 12개의 최신 SISR 딥러닝 모델에 BAM 모듈을 적용하여, BAM 모듈 적용 여부에 따른 CG나 VR 영상 초해상화의 성능 변화를 분석한다.

II. 본론

이 절에서는 본 연구에서 어텐션 모듈을 적용하기 위한 기준(baseline) 모델로 활용된 12개의 최신 SISR 딥러닝 모델과 어텐션 메커니즘, 그리고 BAM 모듈에 대해 간략히 설명한다.

2.1 기준 SISR 딥러닝 모델

본 연구에서 사용된 최신 SISR 딥러닝 모델은 어텐션 모듈을 갖고 있지 않은 모델 8개, 어텐션 모듈을 갖고 있는 모델 4개로 총 12개다.

2.1.1 어텐션 모듈이 없는 SISR 딥러닝 모델: EDSR, CARN, CARN-M, MDSR, s-LWSR, OISR-LF, OISR-LF-s, AWSRN

EDSR[8]은 NTIRE 2017 SISR Challenge에서 우승한 모델로 SRResNet[9]을 변형하여 ResNet[10]의 필요 없는 모듈을 삭제하고 깊이를 더 깊이 쌓은 모델이다. CARN은 계단식 메커니즘을 활용하여 로컬 및 글로벌 수준에서 여러 계층의 특징 지도(feature map)를 통합하기 위해 설계된 모델이고, CARN-M은 채귀 네트워크 아키텍처를 변형하여 CARN보다 가볍게 만든 모델이다[11]. MDSR은 특징 지도를 단일 스케일 구조가 아닌 다양한 스케일로 재구성하는 다중 스케일 구조의 모델이다[12]. s-LWSR[13]은 모바일 환경에서 실시간 초해상화를 위해 만들어진 모델로 U-Net[14]에서 아이디어를 얻어와 SISR에 처음으로 인코더-디코더의 구조를 가져와 적용하였다. OISR-LF와 OISR-LF-s는 상미분방정식을 활용해 새로운 residual 모듈을 만들어 적용한 모델이다[15]. 마지막으로, AWSRN은 효율적인 residual 학습을 위해 새로운 로컬 결합 블록을 만들어 설계한 모델이다[16].

2.1.2 어텐션 모듈이 있는 SISR 딥러닝 모델: RCAN, IMDN, PAN, DRLN

어텐션 메커니즘이 컴퓨터 비전 분야에서 주목받기 시작하면서 어텐션 메커니즘을 활용한 초해상화 딥러닝 모델에 대한 연구가 활발히 진행되었으며, 본 연구에서 사용된 어텐션 모듈이 있는 4개의 초해상화 딥러닝 모델에는 각각 다른 어텐션 메커니즘이 활용되었다.

RCAN은 채널 간의 중요도를 고려해 가중치를 주는 채널 어텐션이 사용된 SISR 딥러닝 모델이다[2]. IMDN은 contrast-aware 채널 어텐션을 사용한 모델이며[3], PAN은 채널 어텐션과 공간 어텐션은 스티한 픽셀 어텐션을 사용한 모델이다[4]. 마지막으로, DRLN은 Laplacian 어텐션을 새롭게 제안하고 활용한 모델이다[5].

2.2 어텐션 메커니즘

인간의 시각 시스템에서 어텐션이란 의미 있는 영역에 시선을 집중하여 시각 정보를 처리하는 것을 말한다. 어텐션 메커니즘이란 이러한 어텐션을 활용하여 입력 중에서 의미 있는 부분에 집중하여 학습하는 방식을 말한다[17]. 어텐션 메커니즘을 활용하여 학습하게 되면 입력의 모든 부분을 동등하게 활용하는 것이 아니라 중요한 정보가 있다고 판단되는 부분을 집중하여 학습하기 때문에 심층 신경망에서 더욱 향상된 결과를 얻을 수 있다. 최근에는 초해상화에서도 활용하고 있으며, 어텐션 메커니즘을 활용해 좋은 성능을 보여주었던 대표적인 초해상화 딥러닝 모델에는 본 연구에서 사용되는 RCAN, IMDN, PAN, DRLN을 들 수 있다.

2.3 BAM 모듈

본 연구에서 사용된 어텐션 모듈인 BAM 모듈[7]은 초해상화를 위해 기존의 어텐션 모듈[6]에서 아이디어를 가져와 설계된 모듈이다. BAM 모듈은 ACAM(Avgpool Channel Attention Module)과 MSAM(Maxpool Spatial Attention Module)을 병렬로 연결하여 기존 직렬로 연결했을 때보다 더 좋은 성능을 보여준다. ACAM은 공간 영역에 대해 global average pooling을 수행하여 각 채널의 중요도를 계산하고, MSAM은 채널 방향으로 max pooling 수행하여 공간적인 중요도를 계산하여, 채널 및 공간 영역에서 중요한 정보가 어디에 있는지를 주목한다.

본 연구에서는 BAM 모듈을 2.1절에서 설명한 어텐션 모듈이 없는 SISR 딥러닝 모델에서는 기존 모델의 기본 블록(basic block)의 뒤 또는 업스케일링 계층 앞에 부착하고, 어텐션 모듈이 있는 SISR 딥러닝 모델에서는 각각의 어텐션 모듈을 BAM 모듈로 대체하여, BAM 모듈의 유무나 기존 어텐션 모듈과의 성능 차이를 분석한다.

III. 실험 방법

BAM 모듈을 활용했을 때의 성능을 확인하기 위해 12개의 최신 SISR 딥러닝 모델을 사용하여 실험을 진행하였다. 8개의 어텐션 모듈이 사용되지 않은 모델에

는 BAM 모듈을 추가하기 전후의 초해상화 결과를 비교하고, 어텐션 모듈이 사용된 4개의 딥러닝 모델에는 어텐션 모듈을 BAM 모듈로 교체하기 전후의 초해상화 결과를 비교하였다.

3.1 데이터 세트와 평가 방법

본 연구에서는 SISR 딥러닝 모델을 CG나 VR 영상으로 학습할 때의 성능을 분석하기 위해 CG 영상으로 구성된 CG_train과 ReferenceDB[19]를 학습 데이터 세트로 활용하였다(그림 1 참조). CG_train은 인터넷으로부터 수집된 314장의 CG 영상을, ReferenceDB는 컴퓨터 게임의 스크린 샷인 1,832장의 CG 영상으로 구성된다.

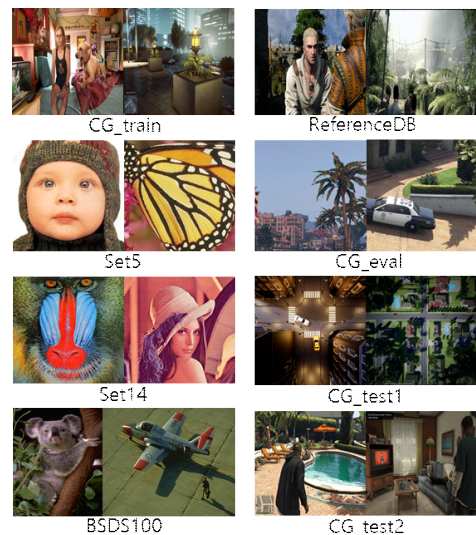


Fig. 1. Parts of images used in our experiments

평가(evaluation) 데이터 세트는 CG_eval로, 인터넷으로부터 수집된 8장의 CG 영상으로 구성된다. 실험 결과를 보여주는 표 1~4에서 평가 데이터로 사용된 데이터 셋은 볼드체로 표시하였다.

테스트 데이터 세트는 Real 영상을 포함하여 Set5[20], Set14[21], BSDS100[18], CG_eval, CG_test1[22], CG_test2[22]를 사용하였다. CG_test1과 CG_test2는 인터넷으로부터 수집된 각각 18장과 14장의 CG 영상을 가진다.

초해상화 결과를 평가하기 위한 성능 지표로는 영상 또는 동영상의 화질을 정량적으로 평가할 때 사용

Table 1. Results on 8 SISR models without attention module when training with CG images in CG_train

scale	Train dataset Test dataset Method	CG_train						
		Set5 PSNR / SSIM	Set14 PSNR / SSIM	BSDS100 PSNR / SSIM	CG_eval PSNR / SSIM	CG_test1 PSNR / SSIM	CG_test2 PSNR / SSIM	
x2	EDSR	36.67 / 0.95	32.65 / 0.91	33.25 / 0.92	38.47 / 0.97	37.15 / 0.95	39.83 / 0.97	
	EDSR-BAM	37.55 / 0.96 ↑	33.25 / 0.91 ↑	33.71 / 0.92 ↑	39.04 / 0.97 ↑	37.75 / 0.96 ↑	40.59 / 0.98 ↑	
	CARN	35.69 / 0.95	32.07 / 0.90	32.76 / 0.91	37.99 / 0.97	36.52 / 0.95	39.21 / 0.97	
	CARN-BAM	35.10 / 0.94	31.64 / 0.90	32.38 / 0.91	37.45 / 0.97	36.07 / 0.95	38.59 / 0.97	
	CARN-M	35.10 / 0.94	31.62 / 0.90	32.35 / 0.91	37.39 / 0.97	36.07 / 0.95	38.54 / 0.97	
	CARN-M-BAM	34.85 / 0.94	31.47 / 0.90	32.25 / 0.91	37.27 / 0.97	35.95 / 0.95	38.38 / 0.97	
	MSRN	36.72 / 0.96	32.67 / 0.91	33.30 / 0.92	38.53 / 0.97	37.19 / 0.95	39.92 / 0.97	
	MSRN-BAM	36.47 / 0.95	32.52 / 0.91	33.15 / 0.92	38.39 / 0.97	37.01 / 0.95	39.73 / 0.97	
	s-LWSR	36.01 / 0.95	32.29 / 0.90	32.95 / 0.91	38.17 / 0.97	36.73 / 0.95	39.44 / 0.97	
	s-LWSR-BAM	36.10 / 0.95 ↑	32.31 / 0.90 ↑	33.00 / 0.92 ↑	38.23 / 0.97 ↑	36.75 / 0.95 ↑	39.51 / 0.97 ↑	
	OISR-LF-s	36.26 / 0.95	32.41 / 0.91	33.02 / 0.92	38.25 / 0.97	36.83 / 0.95	39.52 / 0.97	
	OISR-LF-s-BAM	36.31 / 0.95 ↑	32.41 / 0.91	33.03 / 0.92 ↑	38.29 / 0.97 ↑	36.83 / 0.95	39.57 / 0.97 ↑	
	OISR-LF	37.00 / 0.96	32.84 / 0.91	33.40 / 0.92	38.61 / 0.97	37.38 / 0.96	40.03 / 0.97	
	OISR-LF-BAM	37.00 / 0.96	32.83 / 0.91 ↑	33.41 / 0.92 ↑	38.62 / 0.97 ↑	37.39 / 0.96 ↑	40.05 / 0.97 ↑	
	AWSRN	36.22 / 0.95	32.38 / 0.91	33.01 / 0.92	38.27 / 0.97	36.83 / 0.95	39.57 / 0.97	
	AWSRN-BAM	35.74 / 0.95	32.10 / 0.90	32.81 / 0.91	38.05 / 0.97	36.51 / 0.95	39.30 / 0.97	
x4	EDSR	30.50 / 0.86	27.71 / 0.76	27.83 / 0.74	31.64 / 0.87	31.78 / 0.86	33.02 / 0.89	
	EDSR-BAM	31.48 / 0.88 ↑	28.34 / 0.78 ↑	28.24 / 0.76 ↑	32.19 / 0.88 ↑	32.45 / 0.88 ↑	33.78 / 0.90 ↑	
	CARN	29.64 / 0.84	27.04 / 0.74	27.42 / 0.73	31.04 / 0.86	31.16 / 0.85	32.23 / 0.88	
	CARN-BAM	29.39 / 0.83	26.87 / 0.74	27.29 / 0.73	30.85 / 0.86	31.01 / 0.85	32.02 / 0.88	
	CARN-M	29.46 / 0.84	26.93 / 0.74	27.35 / 0.73	30.95 / 0.86	31.06 / 0.85	32.11 / 0.88	
	CARN-M-BAM	29.34 / 0.83	26.82 / 0.74	27.27 / 0.73	30.82 / 0.86	30.98 / 0.85	31.98 / 0.88	
	MSRN	30.52 / 0.86	27.70 / 0.76	27.83 / 0.74	31.65 / 0.87	31.76 / 0.86	33.04 / 0.89	
	MSRN-BAM	30.31 / 0.86	27.55 / 0.76	27.73 / 0.74	31.53 / 0.87	31.64 / 0.86	32.88 / 0.89	
	s-LWSR	30.05 / 0.85	27.38 / 0.75	27.63 / 0.74	31.36 / 0.87	31.48 / 0.86	32.71 / 0.89	
	s-LWSR-BAM	29.94 / 0.85	27.30 / 0.75	27.57 / 0.74	31.28 / 0.87	31.41 / 0.86	32.62 / 0.88	
	OISR-LF-s	30.17 / 0.85	27.46 / 0.75	27.67 / 0.74	31.44 / 0.87	31.56 / 0.86	32.75 / 0.89	
	OISR-LF-s-BAM	30.10 / 0.85	27.41 / 0.75	27.64 / 0.74	31.39 / 0.87	31.50 / 0.86	32.69 / 0.89	
	OISR-LF	30.80 / 0.87	27.91 / 0.76	27.95 / 0.75	31.79 / 0.88	31.91 / 0.87	33.22 / 0.89	
	OISR-LF-BAM	30.89 / 0.87 ↑	27.99 / 0.77 ↑	28.01 / 0.75 ↑	31.86 / 0.88 ↑	32.00 / 0.87 ↑	33.32 / 0.90 ↑	
	AWSRN	30.12 / 0.85	27.42 / 0.75	27.65 / 0.74	31.42 / 0.87	31.52 / 0.86	32.75 / 0.89	
	AWSRN-BAM	29.92 / 0.85	27.27 / 0.75	27.57 / 0.73	31.29 / 0.87	31.38 / 0.86	32.61 / 0.88	

Table 2. Results on 4 SISR models with attention module when training with CG images in CG_train

scale	Train dataset Test dataset Method	CG_train						
		Set5 PSNR / SSIM	Set14 PSNR / SSIM	BSDS100 PSNR / SSIM	CG_eval PSNR / SSIM	CG_test1 PSNR / SSIM	CG_test2 PSNR / SSIM	
x2	RCAN	37.13 / 0.96	32.91 / 0.91	33.48 / 0.92	38.74 / 0.97	37.49 / 0.96	40.19 / 0.97	
	RCAN-BAM	37.24 / 0.96 ↑	32.99 / 0.91 ↑	33.53 / 0.92	38.80 / 0.97 ↑	37.61 / 0.96 ↑	40.28 / 0.98 ↑	
	IMDN	36.43 / 0.95	32.48 / 0.91	33.12 / 0.92	38.37 / 0.97	36.96 / 0.95	39.69 / 0.97	
	IMDN-BAM	36.19 / 0.95	32.36 / 0.91	32.99 / 0.92	38.26 / 0.97	36.82 / 0.95	39.54 / 0.97	
	PAN	36.91 / 0.96	32.78 / 0.91	33.33 / 0.92	38.49 / 0.97	37.24 / 0.96	39.91 / 0.97	
	PAN-BAM	36.80 / 0.96	32.72 / 0.91	33.27 / 0.92	38.45 / 0.97	37.18 / 0.96	39.83 / 0.97	
	DRLN	37.41 / 0.96	33.14 / 0.91	33.66 / 0.92	39.00 / 0.97	37.76 / 0.96	-	
	DRLN-BAM	37.40 / 0.96	33.16 / 0.91 ↑	33.65 / 0.92	38.99 / 0.97	37.74 / 0.96	-	
x4	RCAN	30.78 / 0.87	27.92 / 0.76	27.96 / 0.75	31.81 / 0.88	31.98 / 0.87	33.28 / 0.90	
	RCAN-BAM	30.76 / 0.87	27.89 / 0.76	27.94 / 0.75	31.79 / 0.88	31.93 / 0.87	33.23 / 0.89	
	IMDN	29.63 / 0.84	27.10 / 0.74	27.44 / 0.73	31.09 / 0.86	31.21 / 0.85	32.35 / 0.88	
	IMDN-BAM	29.56 / 0.84	27.00 / 0.74	27.39 / 0.73	31.00 / 0.86	31.15 / 0.85	32.25 / 0.88	
	PAN	30.72 / 0.87	27.90 / 0.76	27.94 / 0.75	31.72 / 0.87	31.80 / 0.87	33.11 / 0.89	
	PAN-BAM	30.81 / 0.87 ↑	27.92 / 0.76 ↑	27.96 / 0.75 ↑	31.74 / 0.87 ↑	31.85 / 0.87 ↑	33.15 / 0.89 ↑	
	DRLN	31.17 / 0.88	28.16 / 0.77	28.17 / 0.75	32.03 / 0.88	32.21 / 0.87	33.56 / 0.90	
	DRLN-BAM	30.99 / 0.87	28.05 / 0.77	28.05 / 0.75	31.94 / 0.88	32.08 / 0.87	33.45 / 0.90	

되는 PSNR(Peak Signal-to-Noise Ratio)과 SSIM (Structural Similarity Index Measure)을 사용하였다.

3.2 실험 조건

실험은 i7-8700 3.20GHz CPU와 32GB RAM, NVIDIA RTX 2080 Ti GPU를 가진 PC에서 수행되었다. 각 SISR 딥러닝 모델 및 BAM 모듈 구현은 PyTorch로 구현된 Python 코드[23]를 활용하였다. PyTorch, Python의 버전은 각각 1.7.1, 3.6.13이었다.

각 SISR 딥러닝 모델에 대해 2배와 4배에 대한 초해

상화 실험을 진행하였다. 해상도를 2배로 초해상화 할 때는 64×64 패치 사이즈로 학습하였으며, 4배로 초해상화 할 때는 32×32 패치 사이즈로 학습을 하였다. 또한, 학습 데이터 증가(augmentation)를 위해 학습 데이터 세트의 각 영상의 모퉁이에서 4장, 중심에서 1장씩 잘라내어 영상 수를 5배로 늘렸다. 학습 시 epoch는 100으로 설정하였다. 나머지 파라미터는 BAM 모듈 논문[7]에서 사용된 파라미터를 그대로 활용하였다. 다만, GPU 메모리 부족으로 인해 파라미터가 큰 네트워크에 대해서는 배치 사이즈를 적절하게 줄여 학습하였다.

Table 3. Results on 8 SISR models without attention module when training with CG images in ReferenceDB

scale	Train dataset		ReferenceDB					
	Test dataset	Method	Set5 PSNR / SSIM	Set14 PSNR / SSIM	BSDS100 PSNR / SSIM	CG_eval PSNR / SSIM	CG_test1 PSNR / SSIM	CG_test2 PSNR / SSIM
x2	EDSR		37.50 / 0.96	33.22 / 0.91	33.71 / 0.92	39.17 / 0.97	37.75 / 0.96	40.74 / 0.98
	EDSR-BAM		37.55 / 0.96 ↑	33.32 / 0.91 ↑	33.73 / 0.92 ↑	39.63 / 0.98 ↑	37.78 / 0.96 ↑	41.31 / 0.98 ↑
	CARN		37.26 / 0.96	33.04 / 0.91	33.56 / 0.92	38.80 / 0.97	37.53 / 0.96	40.32 / 0.98
	CARN-BAM		37.04 / 0.96	32.88 / 0.91	33.46 / 0.92	38.66 / 0.97	37.36 / 0.96	40.15 / 0.97
	CARN-M		36.97 / 0.96	32.85 / 0.91	33.41 / 0.92	38.61 / 0.97	37.38 / 0.96	40.09 / 0.97
	CARN-M-BAM		36.51 / 0.95	32.58 / 0.91	33.21 / 0.92	38.42 / 0.97	37.02 / 0.95	39.81 / 0.97
	MSRN		37.53 / 0.96	33.28 / 0.91	33.72 / 0.92	39.27 / 0.98	37.79 / 0.96	40.83 / 0.98
	MSRN-BAM		37.53 / 0.96	33.27 / 0.91	33.72 / 0.92	39.08 / 0.97	37.80 / 0.96 ↑	40.61 / 0.98
	s-LWSR		37.31 / 0.96	33.08 / 0.91	33.62 / 0.92	38.90 / 0.97	37.65 / 0.96	40.40 / 0.98
	s-LWSR-BAM		37.31 / 0.96 ↑	33.11 / 0.91 ↑	33.63 / 0.92 ↑	38.91 / 0.97 ↑	37.65 / 0.96	40.42 / 0.98 ↑
	OISR-LF-s		37.38 / 0.96	33.10 / 0.91	33.63 / 0.92	38.89 / 0.97	37.66 / 0.96	40.39 / 0.98
	OISR-LF-s-BAM		37.42 / 0.96 ↑	33.11 / 0.91 ↑	33.62 / 0.92	38.90 / 0.97 ↑	37.64 / 0.96	40.40 / 0.98 ↑
	OISR-LF		37.60 / 0.96	33.30 / 0.91	33.74 / 0.92	39.26 / 0.98	37.81 / 0.96	40.84 / 0.98
	OISR-LF-BAM		37.59 / 0.96	33.30 / 0.91	33.74 / 0.92 ↑	39.29 / 0.98 ↑	37.86 / 0.96 ↑	40.87 / 0.98 ↑
	AWSRN		37.45 / 0.96	33.18 / 0.91	33.68 / 0.92	39.00 / 0.97	37.79 / 0.96	40.51 / 0.98
	AWSRN-BAM		37.36 / 0.96	33.09 / 0.91	33.61 / 0.92	38.89 / 0.97	37.67 / 0.96	40.39 / 0.98
x4	EDSR		31.62 / 0.89	28.41 / 0.78	28.28 / 0.76	32.24 / 0.88	32.49 / 0.88	33.81 / 0.90
	EDSR-BAM		31.96 / 0.89 ↑	28.62 / 0.78 ↑	28.43 / 0.76 ↑	32.52 / 0.89 ↑	32.74 / 0.88 ↑	34.16 / 0.91 ↑
	CARN		31.20 / 0.88	28.16 / 0.77	28.12 / 0.75	31.97 / 0.88	32.13 / 0.87	33.47 / 0.90
	CARN-BAM		30.87 / 0.87	27.96 / 0.77	27.99 / 0.75	31.82 / 0.88	31.93 / 0.87	33.29 / 0.89
	CARN-M		30.86 / 0.87	27.96 / 0.77	27.99 / 0.75	31.80 / 0.88	31.93 / 0.87	33.24 / 0.89
	CARN-M-BAM		30.58 / 0.87	27.76 / 0.76	27.85 / 0.74	31.66 / 0.87	31.71 / 0.86	32.99 / 0.89
	MSRN		31.60 / 0.89	28.43 / 0.78	28.28 / 0.76	32.53 / 0.89	32.50 / 0.88	33.83 / 0.90
	MSRN-BAM		31.51 / 0.89	28.38 / 0.78	28.24 / 0.76	32.18 / 0.88	32.40 / 0.88	33.74 / 0.90
	s-LWSR		31.29 / 0.88	28.23 / 0.77	28.15 / 0.75	32.06 / 0.88	32.25 / 0.88	33.57 / 0.90
	s-LWSR-BAM		31.26 / 0.88	28.23 / 0.77 ↑	28.16 / 0.75 ↑	32.06 / 0.88 ↑	32.29 / 0.88 ↑	33.59 / 0.90 ↑
	OISR-LF-s		31.37 / 0.88	28.28 / 0.77	28.19 / 0.75	32.07 / 0.88	32.30 / 0.88	33.61 / 0.90
	OISR-LF-s-BAM		31.39 / 0.88 ↑	28.30 / 0.77 ↑	28.20 / 0.76 ↑	32.09 / 0.88 ↑	32.32 / 0.88 ↑	33.65 / 0.90 ↑
	OISR-LF		31.70 / 0.89	28.45 / 0.78	28.31 / 0.76	32.28 / 0.89	32.53 / 0.88	33.86 / 0.90
	OISR-LF-BAM		31.71 / 0.89 ↑	28.45 / 0.78 ↑	28.31 / 0.76	32.28 / 0.89 ↑	32.53 / 0.88	33.88 / 0.90 ↑
	AWSRN		31.45 / 0.88	28.35 / 0.78	28.23 / 0.76	32.18 / 0.88	32.36 / 0.88	33.72 / 0.90
	AWSRN-BAM		31.40 / 0.88	28.31 / 0.77	28.20 / 0.76	32.12 / 0.88	32.35 / 0.88	33.66 / 0.90

Table 4. Results on 4 SISR models with attention module when training with CG images in ReferenceDB

scale	Train dataset		ReferenceDB					
	Test dataset	Method	Set5 PSNR / SSIM	Set14 PSNR / SSIM	BSDS100 PSNR / SSIM	CG_eval PSNR / SSIM	CG_test1 PSNR / SSIM	CG_test2 PSNR / SSIM
x2	RCAN		37.62 / 0.96	33.36 / 0.91	33.73 / 0.92	39.69 / 0.98	37.88 / 0.96	41.36 / 0.98
	RCAN-BAM		37.57 / 0.96	33.35 / 0.91	33.73 / 0.92	39.74 / 0.98 ↑	37.86 / 0.96	41.41 / 0.98 ↑
	IMDN		37.48 / 0.96	33.19 / 0.91	33.68 / 0.92	39.03 / 0.97	37.82 / 0.96	40.56 / 0.98
	IMDN-BAM		37.50 / 0.96 ↑	33.20 / 0.91 ↑	33.69 / 0.92 ↑	39.01 / 0.97	37.80 / 0.96	40.56 / 0.98
	PAN		37.41 / 0.96	33.14 / 0.91	33.60 / 0.92	38.86 / 0.97	37.67 / 0.96	40.39 / 0.98
	PAN-BAM		37.40 / 0.96	33.13 / 0.91	33.60 / 0.92	38.88 / 0.97 ↑	37.68 / 0.96 ↑	40.39 / 0.98
	DRLN		37.57 / 0.96	33.41 / 0.92	33.71 / 0.92	39.63 / 0.98	37.94 / 0.96	-
	DRLN-BAM		37.61 / 0.96 ↑	33.37 / 0.92	33.75 / 0.92 ↑	39.58 / 0.98	37.89 / 0.96	-
x4	RCAN		31.97 / 0.89	28.63 / 0.78	28.44 / 0.76	32.53 / 0.89	32.77 / 0.88	34.14 / 0.91
	RCAN-BAM		31.93 / 0.89	28.58 / 0.78	28.42 / 0.76	32.48 / 0.89	32.70 / 0.88	34.10 / 0.91
	IMDN		31.17 / 0.88	28.14 / 0.77	28.10 / 0.75	31.99 / 0.88	32.20 / 0.87	33.49 / 0.90
	IMDN-BAM		31.09 / 0.88	28.10 / 0.77	28.07 / 0.75	31.94 / 0.88	32.15 / 0.87	33.44 / 0.90
	PAN		31.30 / 0.88	28.21 / 0.77	28.14 / 0.75	31.96 / 0.88	32.19 / 0.87	33.47 / 0.90
	PAN-BAM		31.31 / 0.88 ↑	28.20 / 0.77	28.14 / 0.75	31.96 / 0.88	32.16 / 0.87	33.45 / 0.90
	DRLN		31.82 / 0.89	28.56 / 0.78	28.39 / 0.76	32.42 / 0.89	32.67 / 0.88	34.05 / 0.90
	DRLN-BAM		31.80 / 0.89	28.51 / 0.78	28.36 / 0.76	32.36 / 0.89	32.61 / 0.88	33.97 / 0.90

IV. 실험 결과

표 1-4는 각각 서로 다른 학습 데이터 세트를 사용하여 2배와 4배로 초해상화한 Real, CG 영상의 PSNR, SSIM 결과를 보여준다. 각 결과는 BAM 모듈을 적용했을 때와 적용하지 않았을 때의 결과를 가지며, BAM 모듈을 적용한 결과는 각 모델명 뒤에 “-BAM” 을 붙여 표시하였다. 일부 데이터 세트에 대해 DRLN과 DRLN-BAM은 GPU 메모리 오류로 인해 결과를 확인할 수 없어 생략했다.

4.1 업스케일 크기에 따른 차이

표 1, 2는 CG_train을 사용하여 학습된 각 딥러닝 모델의 BAM 모듈의 적용 유무에 따른 PSNR과 SSIM 결과를 보여준다. 어텐션이 없는 모델의 경우(표 1 참조) 2배 초해상화 결과에서는 EDSR, s-LWSR, OISR-LF-s에 BAM 모듈을 적용함으로써 성능이 향상되지만, 4배 초해상화 결과에서는 EDSR과 OISR-LF만 성능이 향상되었다. 어텐션이 있는 모델의 경우(표 2 참조) 2배 초해상화 결과에서는 RCAN의 성능이 향상되

었고, 4배 초해상화 결과에서는 PAN이 성능이 향상되어 업스케일 크기에 따라 BAM 모듈 적용으로 인해 성능이 향상되는 모델이 다른 것을 볼 수 있다.

표 3, 4는 ReferenceDB를 사용하여 학습된 각 딥러닝 모델의 BAM 모듈의 적용 유무에 따른 PSNR과 SSIM 결과를 보여준다. 어텐션이 없는 모델의 경우(표 3 참조) 2배 초해상화 결과에서는 EDSR, s-LWSR, OISR-LF-s에 BAM 모듈을 적용함으로써 성능이 향상되었지만, 4배 초해상화 결과에서는 EDSR과 OISR-LF-s, OISR-LF의 성능이 향상되었다. 어텐션이 있는 모델의 경우(표 4 참조) 2배나 4배 초해상화 결과 모두 BAM 모듈로 변경함으로써 성능 향상이 뚜렷한 모델이 없는(일부 데이터 세트에서만 성능 향상이 나타남) 것을 볼 수 있다.

결과적으로, 업스케일 크기에 따라 BAM 모듈을 적용했을 때의 성능 개선 효과는 달라졌으며, 업스케일 크기가 증가할수록 BAM 모듈의 성능 개선 효과는 떨어졌다. 이는 BAM 모듈을 사용할 때 업스케일 크기도 함께 고려해야 함을 보여주었다.

4.2 학습 데이터 크기에 따른 차이

표 1-4의 결과를 통해 학습 데이터의 크기에 따라 BAM 모듈의 성능 개선 효과가 다를 수 있다. 학습 데이터 크기가 작은 CG_train을 사용하여 학습하면, 2배 초해상화 결과에서는 EDSR, s-LWSR, OISR-LF-s, OISR-LF, RCAN, 4배 초해상화 결과에서는 EDSR, OISR-LF, PAN이 BAM 모듈 적용으로 성능이 향상되었지만, 학습 데이터 크기가 큰 Reference DB를 사용하여 학습하면, EDSR, s-LWSR, OISR-LF-s, OISR-LF, 4배 초해상화 결과에서는 EDSR, s-LWSR, OISR-LF-s, OISR-LF의 성능이 향상되었다. 결과적으로, 어텐션이 없는 모델의 경우 학습 데이터의 크기가 클수록 BAM 모듈의 성능 개선 효과가 향상되지만, 어텐션이 있는 모델의 경우 오히려 학습 데이터의 크기가 클 경우 BAM의 성능 개선 효과가 떨어졌다.

4.3 BAM 모듈 적용 효과

BAM 모듈 적용 여부에 따른 각 모델의 성능 변화를 살펴보면 네트워크의 특징과 학습 데이터 크기에 따라 성능이 달라진다. BAM 모듈을 제안한 기존 연구

에서는 일반적으로 BAM 모듈을 활용하면 대부분의 초해상화 모델의 성능이 향상된다고 했지만, 본 논문의 실험에서는 EDSR만 실험 조건에 상관없이 일반적으로 성능이 향상되었고, 그 외 모델들은 BAM 모듈을 적용하더라도 성능이 향상되지 않거나, 일부 실험 조건에서만 제한적으로 성능이 향상되었다. 일부 모델의 경우 BAM 모듈을 적용하였을 때, 오히려 성능이 저하되었다. 이러한 결과는 Real 영상 초해상화와 CG 영상 초해상화에서 다르지 않았다. 결과적으로, BAM 모듈을 적용하는 것이 항상 SISR 모델의 성능을 개선에 긍정적인 영향을 주지 못하며, 모델의 특징이나 학습 데이터 세트의 크기 등을 함께 고려해야 한다는 것을 확인할 수 있었다.

V. 결론

본 논문은 초해상화 딥러닝 모델에 BAM 모듈을 적용하여, BAM 모듈에 의한 초해상화 성능 개선 효과를 실험적으로 분석하였다. 실험 결과, CG 영상 초해상화에도 BAM 모듈 적용으로 인한 성능 개선 효과가 있음을 확인했으나, 학습 데이터의 크기에 따라 초해상화 성능이 개선되는 모델도 있었지만, 그렇지 못한 모델들이 더 많았다. BAM 모듈을 제안한 논문[7]의 결과와는 다르게 BAM 모듈을 적용함으로써 항상 초해상화 성능이 향상되는 것은 아님을 확인할 수 있었다.

다만, 일부 모델은 BAM 모듈을 사용하였을 때 일반적으로 향상된 결과를 보여주는 것으로 보아 BAM 모듈을 적용하였을 때 성능이 향상될 수 있는 모델의 특징이 무엇인지 알아낼 필요가 있다. 또한 BAM 모듈을 개선하여 딥러닝 모델과 학습 데이터의 크기에 상관없이 성능 향상이 가능한 어텐션 모듈에 대한 추가 연구가 필요하다.

ACKNOWLEDGMENTS

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

REFERENCES

- [1] C. Dong, C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," PAMI, vol. 38, no. 2, pp. 295-307, 2015.
- [2] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image superresolution using very deep residual channel attention networks," ECCV, pp. 286-301, 2018.
- [3] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image superresolution with information multi-distillation network," ACM MM, pp. 2024-2032, 2019.
- [4] H. Zhao, X. Kong, J. He, Y. Qiao, and C. Dong, "Efficient image superresolution using pixel attention," arXiv preprint arXiv:2010.01073, 2020.
- [5] S. Anwar and N. Barnes, "Densely residual Laplacian super-resolution," PAMI, 2020.
- [6] J. Park, S. Woo, J.Y. Lee, and I.S. Kweon. "BAM: bottleneck attention module," arXiv preprint arXiv:1807.06514, 2018.
- [7] 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.
- [8] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," CVPRW, pp. 136-144, 2017.
- [9] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," arXiv preprint arXiv:1609.04802, 2016.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," CVPR, 2016.
- [11] N. Ahn, B. Kang, and K. A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," ECCV, pp. 252-268, 2018.
- [12] J. Qin, Y. Huang, and W. Wen, "Multi-scale feature fusion residual network for single image super-resolution," Neurocomputing, vol. 379, pp. 334-342, 2020.
- [13] B. Li, B. Wang, J. Liu, Z. Qi, and Y. Shi, "s-LWSR: super lightweight super-resolution network," IEEE Transactions on Image Processing, vol. 29, pp. 8368-8380, 2020.
- [14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networksfor biomedical image segmentation," International Conference on Medical Image Computing and Computer-assisted Intervention, 2015.
- [15] X. He, Z. Mo, P. Wang, Y. Liu, M. Yang, and J. Cheng, "ODE-inspired network design for single image super-resolution," CVPR, pp. 1732-1741, 2020.
- [16] C. Wang, Z. Li, and J. Shi, "Lightweight image super-resolution with adaptive weighted learning network," arXiv preprint arXiv:1904.02358, 2019.
- [17] A. Allport, "Visual attention," Foundations of Cognitive Science, The MIT Press, pp. 631-682, 1989.
- [18] 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," ICCV, vol. 2, pp. 416-423, 2001.
- [19] Reference database, <http://level-design.org/referencedb/>
- [20] M. Bevilacqua, A. Roumy, C. Guillemot, and A. Morel, "Low complexity single image super-resolution based on nonnegative neighbor embedding," BMVC, 2012.
- [21] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," International Conference on Curves and Surfaces, pp. 711-730, 2010.
- [22] S. Kim and H. Park, "Dependency on training data in CG image super-resolution," Summer Conference of Korea Institute of Convergence Signal Processing, pp. 57-59, 2021.
- [23] BAM code, <https://github.com/dandingbudanding/BAM>

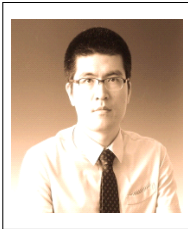
저자 소개

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