

동적 잔차 연결을 활용한 고해상도 복원 네트워크

박가람, *조남익

서울대학교

saturnian77@ispl.snu.ac.kr, *nicho@snu.ac.kr

Super-resolution Network Using Dynamic Residual Connection

Karam Park, Nam Ik Cho

Seoul National University

요 약

본 논문에서는 입력에 따라 합성곱 레이어 간의 잔차 연결을 변화시키는 동적 잔차 연결을 활용해 고해상도 복원 (Super-resolution) 작업을 위한 경량 네트워크를 설계하는 방법을 제안한다. 먼저, 동적 잔차 연결을 입력에 따라 변화시킬 수 있도록 경량화된 (Lightweight) 모듈을 설계하는 방법을 제안한다. 또한 이렇게 설계한 모듈로부터 얻은 잔차 연결에 대한 정보를 토대로 네트워크를 설계하는 방법을 제시한다. 제안된 방법을 통해 설계된 고해상도 복원 작업을 위한 네트워크는 적은 파라미터로도 입력에 따라 적응적으로 네트워크의 구조를 변화시킬 수 있어 효율적으로 다양한 입력 영상을 처리할 수 있다.

1. 서론

고해상도 복원 작업은 영상 처리 분야 중 한 갈래로 저해상도에서 고해상도 영상으로 복원을 수행하는 작업을 의미한다. 이 과정에서 다양한 해가 존재할 수 있어 해결이 어려운 불량 조건 (Ill-posed) 문제로 널리 알려져 있다. 최근 합성곱 신경망이 어려운 문제를 해결하는데 뛰어난 성능을 가지고 있음을 보여주면서 [1], 고해상도 복원 작업을 포함한 다양한 영상 처리 분야에서 합성곱 네트워크를 적용한 방식을 시도하고 있으며 결과 또한 성공적이었다.

SRCNN [2] 이 최초로 합성곱 신경망을 고해상도 복원 문제에 적용해 기존의 방식들보다 좋은 성능을 가진다는 것을 보여주었다. 이후 네트워크 설계에서 잔차 연결 (Residual connection) [3] 을 이용하면 네트워크의 깊이를 늘려 성능을 향상시킬 수 있다는 것이 널리 알려지면서, 점점 많은 합성곱 레이어를 갖는 네트워크들 [4-6] 이 연구되었다.

이런 깊이 방향으로의 확장은 네트워크에 좋은 성능을 부여했으나, 이러한 방식에도 큰 제약이 존재했다. 바로 네트워크의 깊이가 깊어지면서, 네트워크가 동작하는데 필요한 자원이 늘어났다는 점이다. 깊은 네트워크들은 파라미터가 많아 메모리 사용량이나 연산량 측면에서 높은 사양을 요구하게 된다. 게다가, 네트워크의 연산량이 많아 영상의 해상도 복원 작업에 시간이 오래 걸리는 것 또한 문제다. 이러한 제약들은 모바일 등의 연산 자원이 제한적인 환경에서 부적절하며, 현실적이지 못하다.

본 논문에서는 이러한 문제점들을 해결하기 위해 고해상도 복원 작업을 수행하는 경량화된 네트워크를 설계하는 방법을 제안한다. 제안한 방법은 네트워크 내 각 합성곱 레이어 간 연결을 입력에 따라 적응적으로 조절하는 경량화된 모듈을 설계하는 방법을 제시하고, 해당 방법으로 네트워크를 설계하는 방법을 제안한다. 제안하는 방법은 입력에 따라 동적으로 잔차 연결을 바꾸는 방식 [7] 으로 네트워크의 표현력을 높임으로써 경량화된 네트워크 구조로 인해 발생하는 성능의 저하를 상쇄해

네트워크의 성능과 효율을 조율한다.

본 논문의 구성은 다음과 같다. 2 절에서는 본 논문에서 제안하는 기법을 설명하고, 3 절에서는 제안한 기법의 성능을 실험을 통해서 확인한다. 마지막으로 4 절에서는 본 논문에 대한 결론을 맺는다.

2. 제안된 방법

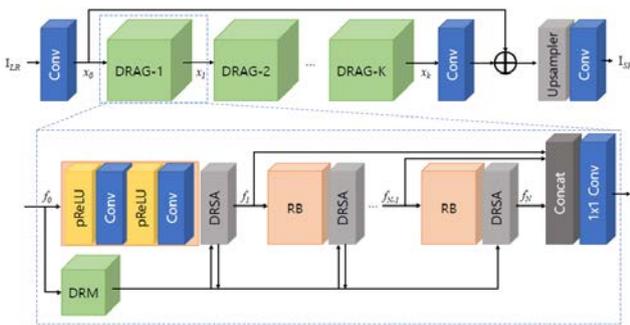


그림 1. 제안된 네트워크의 구조

그림 1 과 같이, 제안된 네트워크는 구조적으로 얇은 특징값 (Shallow feature)을 추출하는 합성곱 레이어, 의미있는 특징값을 추출하는 연속된 Dynamic residual attention group (DRAG) 구조 부분, 마지막으로 원하는 해상도로 복원하는 업샘플러 (Upsampler) 부분의 세 부분으로 나눌 수 있다. 가장 앞에 위치한 합성곱 레이어를 이용해 입력 I_{LR} 로부터 얇은 특징값 x_0 를 얻는다. x_0 는 여러 개의 DRAG를 거쳐 의미있는 특징값을 추출하게 되고, 마지막으로 업샘플러 모듈을 거치면 목적하는 해상도로 복원된 영상 I_{SR} 을 얻게 된다.

DRAG 구조는 여섯 개의 합성곱 레이어가 존재하며, 각 합성곱 레이어 앞에는 pReLU 레이어가 존재한다. 두 개의 합성곱 레이어들이 짝을 지어 하나의 Residual block (RB)를 형성하며, 각각의 RB로부터 출력되는 특징값들은 채널 방향으로 연결된 뒤 1×1 합성곱 연산을 거침으로써 DRAG의 출력이 결정된다. 하나의 RB에서 다음에 위치한 RB로 특징값을 넘겨줄 때엔, Dynamic residual module (DRM)이 관여하여 특징값의 계산을 입력에 따라 적응적으로 조절한다.

DRM의 구조는 그림 2에서 보이는 것과 같다. 입력으로 들어온 특징값 f_0 는 두 개의 1×1 합성곱 레이어를 거친 뒤, 전역평균풀링 (Global average pooling)을 거쳐 이전의 RB로부터 온 특징값들과 곱해지는 계수인 Dynamic residual attention (DRA)를 연산한다.

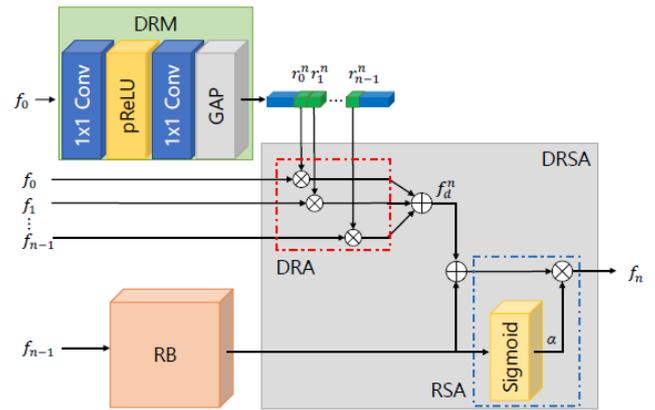


그림 2. 동적 잔차 연결을 계산하는 모듈의 구조

이와 같은 연산을 통해 이전의 RB들로부터 온 특징값들이 어떻게 조합할 지가 f_d^n 와 같이 결정되면, 해당 RB의 출력값을 이용해 어텐션 Residual self-attention (RSA)를 적용한다. RSA는 해당 RB의 출력값에 시그모이드를 적용함으로써 얻어지며, 이 과정에서 복잡한 연산을 요구하지 않기에 연산량 측면에서 효율적이다. 이렇게 얻어진 어텐션 값은 영상의 경계나 질감 부분에 집중하는 경향이 있어, 좀 더 의미있는 정보에 특징값이 집중되도록 돕는다.

3. 실험 결과

해당 실험에서 제안된 네트워크를 학습시키기 위해 800장의 DIV2K 데이터셋을 사용하였다. 48×48 크기의 RGB 패치를 학습에 사용하였으며, 임의의 방향에 대한 회전과 반전을 통해 데이터를 증가시켰다.

제안된 방법을 평가하기 위해, 테스트 영상의 Y 채널 값에 대한 최대 신호 대 잡음비 (Peak Signal to Noise Ratio)를 사용하였다. 평가를 위해 Set5 [8], Set14 [9], B100 [10], 그리고 Urban100 [11] 데이터셋을 사용하였다.

학습에서 배치의 크기는 16으로 설정하였고, ADAM optimizer의 파라미터를 $\beta_1=0.9$, $\beta_2=0.999$, $\epsilon=10^{-8}$ 로 설정하여 학습에 사용하였다. 학습률 (learning rate)은 2×10^{-4} 로 설정하였으며, 매 2×10^5 의 반복에 따라 학습률을 0.85배씩 낮추며 학습을 진행하였다. 학습을 위한 손실 함수는 L_1 함수를 사용하였다.

배율	모델	Param	Set5	Set14	B100	Urban100
×2	bicubic	-	33.66	30.24	29.56	26.88
	FSRCNN [12]	0.01M	37.00	32.63	31.53	29.88
	DRRN [13]	0.30M	37.74	33.23	32.05	31.23
	A2FNet [14]	0.32M	37.91	33.45	31.79	31.79
	AWSRN [15]	0.40M	37.75	33.31	32.00	31.39
	FALSr [16]	0.41M	37.66	33.26	31.96	31.24
	IDN [17]	0.55M	37.83	33.30	32.08	31.27
	VDSR [4]	0.66M	37.53	33.03	31.90	30.76
	MemNet [18]	0.68M	37.78	33.28	32.08	31.31
	LapSRN [19]	0.81M	37.52	33.08	31.80	30.41
	Ours	0.37M	37.99	33.57	32.16	32.10

표 1. 제안한 방법과 기존의 경량화 모델과의 2 배율에 대한 비교

표 1 은 제안된 방법과 기존의 State-of-the-art (SOTA) 방식 중 경량화 네트워크라 부를 수 있는 작은 모델들과 그 성능을 비교한 것이다. 표에서 볼 수 있듯이, 제안한 방법은 매우 적은 파라미터를 가지고 기존의 네트워크들보다 우수한 성능을 보임을 확인할 수 있다.

Residual	DRAG			Param	PSNR
	DRA	RSA	Concat		
0				357K	31.54
0		0	0	375K	31.73
	0	0		358K	31.71
	0		0	377K	31.70
	0	0	0	377K	31.82

표 2. 제안한 DRAG 구조에 대한 실험

표 2 는 제안된 방법 중 DRAG 의 구조에 해당하는 DRA 와 RSA 그리고 채널 방향으로의 연결 (Concat)을 제거했을 때 나타나는 효과를 실험을 통해 정리한 것이다. 표 2 에서 DRAG 의 어느 한 요소가 제거되거나 기존의 잔차 연결 (Residual)로 대체되었을 때 성능이 저하되는 것을 확인할 수 있는데, 이는 DRAG 의 각 요소가 상호보완적으로 성능향상에 기여하며, 기존의 잔차 연결을 사용한 구조와 비교했을 때 성능면에서 더 뛰어난 구조임을 의미한다.

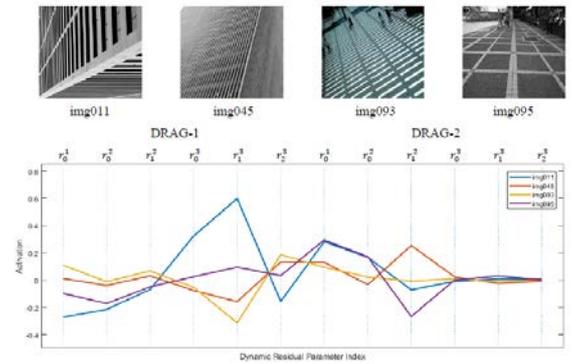


그림 3. 입력 영상에 따른 잔차 연결 방식의 변화

그림 3 은 제안된 방법이 입력 영상에 따라 잔차 연결이 어떻게 변화하는지를 보여주는 그래프이다. 그림 3 을 통해 입력 영상이 변화하면, 입력 영상의 특징을 반영해 잔차 연결 방식이 다양한 방식으로 조절함을 직접 확인할 수 있다.

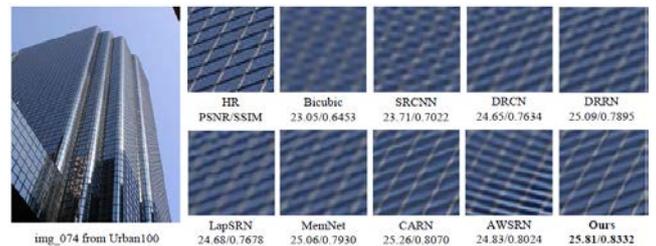


그림 4. 시각화된 네트워크의 출력 비교

그림 4 는 기존의 방식들과 네트워크의 출력을 시각화하여 직접 비교한 것이다. 그림을 통해 알 수 있듯 제안된 방법이 기존의 경량화 네트워크들과 비교했을 때 가장 선명하고 정확한 영상을 출력하는 것을 확인할 수 있다.

4. 결론

해당 논문에서는 입력의 특징값을 반영한 동적인 잔차 연결을 활용하여, 네트워크 내의 특징값들을 다양한 조합으로 사용함으로써 네트워크의 표현력을 높이는 방법을 제안하였다. 이 방법은 경량화된 네트워크가 갖는 성능의 저하를 경감시키는 방식으로 네트워크의 효율과 성능을 조율한다. 또한 제안하는 네트워크 설계방식은 다양한 분야의 잔차 연결을 활용하는 네트워크에 적용될 수 있다. 실험 결과들은 제안된 방식이 기존의 방식들과 비교해 우수함을 보였다.

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2021R1A2C2007220).

참고 문헌

- [1] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012): 1097-1105.
- [2] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295-307, 2015.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [4] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR Oral)*, June 2016.
- [5] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136-144.
- [6] Zhang, Yulun, et al. "Image super-resolution using very deep residual channel attention networks." *Proceedings of the European conference on computer vision (ECCV)*. 2018.
- [7] Park, Karam, Jae Woong Soh, and Nam Ik Cho. "Single Image Super-Resolution with Dynamic Residual Connection." *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021.
- [8] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low complexity single-image super-resolution based on nonnegative neighbor embedding," 2012.
- [9] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *International conference on curves and surfaces*. Springer, 2010, pp. 711-730.
- [10] 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 *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 2. IEEE, 2001, pp. 416-423.
- [11] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5197-5206.
- [12] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European conference on computer vision*. Springer, 2016, pp. 391-407.
- [13] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3147-3155.
- [14] X. Wang, Q. Wang, Y. Zhao, J. Yan, L. Fan, and L. Chen, "Lightweight single-image super-resolution network with attentive auxiliary feature learning," in *Proceedings of the Asian Conference on Computer Vision*, 2020.
- [15] C. Wang, Z. Li, and J. Shi, "Lightweight image super-resolution with adaptive weighted learning network," *arXiv preprint arXiv:1904.02358*, 2019.
- [16] X. Chu, B. Zhang, H. Ma, R. Xu, J. Li, and Q. Li, "Fast, accurate and lightweight super-resolution with neural architecture search," *arXiv preprint arXiv:1901.07261*, 2019.
- [17] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super resolution via information distillation network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 723-731.
- [18] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4539-4547.
- [19] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 11, pp. 2599-2613, 2018.