

픽셀, 채널간 불필요한 상호연관 정보를 제거하는 초해상화 딥러닝 기법

*강정흠 **배성호

경희대학교 컴퓨터 공학과

*chhkang123@khu.ac.kr **shbae@khu.ac.kr

Development of Technique in Super Resolution domain that eliminates unnecessary Correlation information between Pixels & Channels.

*Jung-Heum Kang **Sung-Ho Bae

Kyung Hee University

요약

초해상화 딥러닝 기법은 학습 시 수렴하기까지 최소 수백 번의 에폭을 필요로 하며 오랜 시간이 걸린다. 최근, 영상 인식용 딥러닝 모델에서는 학습 수렴 속도를 향상시키기 위해 픽셀, 채널간 불필요한 상호연관 정보를 제거하는 Deconvolution 기술이 제안되었다. 본 논문에서는 최초로 Deconvolution 기술을 초해상화 딥러닝 방법에 적용하여 학습 수렴 속도 증가를 시도했다. 영상 인식 딥러닝 기법과 다르게 초해상화 딥러닝 기법은 이미지 특성 추출 부분과 이미지 복원 부분의 정보를 보존하는 것이 중요하기 때문에, EDSR을 Baseline 모델로 사용하여 양쪽 끝의 레이어는 기존의 Convolution 연산을 그대로 유지하고, 중간 레이어의 ResBlock 내의 Convolution 연산만 Deconvolution 연산으로 바뀌어서 구성하였다. 초해상화 벤치마크 데이터셋을 사용한 실험 결과, 수렴속도가 빨라지지 않는 결과를 도출했다. 본 논문에서는 Deconvolution 기술이 Baseline 모델의 성능을 개선하지 못하는 이유를 초해상화 분야에서 기본적으로 적용되는 Residual Learning 기법 때문으로 분석했다.

1. 서론

이미지 초해상화(Super-Resolution)기술은 단일 저해상도 이미지를 입력으로 받아, 그에 상응하는 고해상도 이미지를 복원하는 것을 목적으로 한다. 최근 Deep Neural Network(DNN)은 층 깊이를 깊게 만듦에 따라 복원 성능이 향상되었지만, FLOPs(floating poing operations)와 Parameter 수가 늘어났다. 이에 따라, 학습 시간이 오래 걸리기에 메모리를 효율적으로 사용하고 학습의 수렴 속도를 빠르게 하는 딥러닝에 대한 연구를 필요로 한다. 메모리 효율성에 관한 연구는 스마트폰/ 드론 같은 제한된 리소스 환경을 사용하는 모바일 기기들에 적용되므로, 경량화에 대한 연구가 활발하게 진행되고 있다.[4,7,9,11]. 학습의 수렴 속도를 빠르게 하는 기술에는 수학적, 통계학적 접근을 통해 고안된 여러 최적화 기술[12,17]이나 학습 과정 중에 네트워크 가중치, 파라미터의 변화에 의해 야기되는 네트워크 활성화 값들의 분포 변화를 완화시켜주는 정규화 기술[2,10,18]이 있다. 본 논문에서 적용해 보려는 논문인 Network Deconvolution 논문[2]은 정규화 연구의 일환으로 동물의 뇌에서 시각적 상호연관 정보를 제거해주는 시각뉴런에서 영감을 받아 데이터의 픽셀간, 채널간 상호연관 정보를 제거해주는 Deconvolution 연산을 고안해 영상 인식 딥러닝 분야에 적용했고, 학습 수렴 속도를 빠르게 만들고 더불어 정확도를 높게 만드는 데 성공했다. 본 논문은 데이터 및 채널 정규화기법인 Network Deconvolution 기술을 초해상화 분야에 적용해서 여러 모델들과 데이터셋을 사용해 실험해본 논문이다.

2. 관련 연구

2.1. 초해상화 연구

SRCNN[3]은 처음으로 딥러닝 CNN 모델을 초해상화 연구에 적용해서 전통적인 머신러닝 기반 초해상화 기술의 성능을 능가한 모델이다. SRCNN은 Convolution 레이어 3개를 쌓아서 만든 모델로, 3개 이상의 레이어를 쌓아도 성능이 크게 개선되지 않았다고 결론지었다. 하지만 이후 VDSR[5]이 Skip Connection을 이용해서 Residual 이미지를 전달해 저해상도 원본 이미지와 고해상도 복원 이미지의 차이 부분을 집중적으로 학습하는 Residual Learning 기법을 사용하여 세계 최고 복원성능(State of The Art, SOTA)을 달성했다. Residual Learning 기법을 사용하면 복원 성능이 높아질 뿐만 아니라, Exploding/Vanishing Gradient문제를 해결해서 높은 Learning Rate를 사용해 빠른 수렴을 할 수 있게 만들기 때문에, 이후 초해상화 분야에서 SOTA를 달성한 EDSR[1] 과 RCAN[6] 등의 최근 연구들은 모두 Residual learning 기법을 사용한다. 초해상화 분야의 연구들은 i) 최종성능 향상 ii) 학습 개선을 통한 계산량 절감, 수렴시간 단축 iii) 모델사이즈, 파라미터 절감을 통한 추론시간(계산속도) 단축을 목표로 한다. 본 연구에서는 Deconvolution 기법을 적용해 i) 최종성능 향상과 ii) 수렴을 빠르게 하는 것을 목표로 한다.

2.2. Network Deconvolution

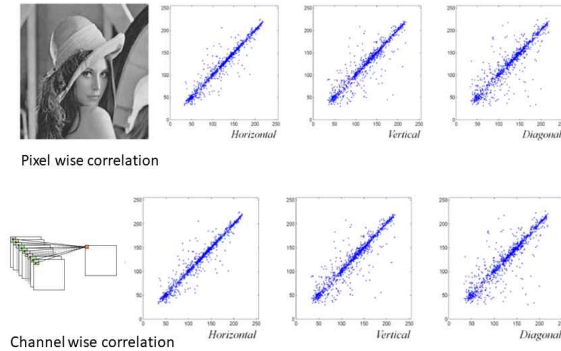


그림 1. 이미지 픽셀간, 네트워크 채널간 상호연관성

Convolution 연산은 Convolution 필터가 움직이며 이미지의 지역 정보를 추출하고 가중치를 학습하는 Convolutional Neural Network (CNN) 모델에서 사용되는 핵심 연산이다. 그러나 이 Convolution 연산은 학습하는 방식 때문에 픽셀 간 상호연관성이 강한 실생활 이미지에 대해서 학습한 정보를 중복 학습하는 비효율적인 특징을 지닌다. 이런 Convolution 연산의 단점을 Deconvolution 연산을 적용해서 해결할 수 있다.

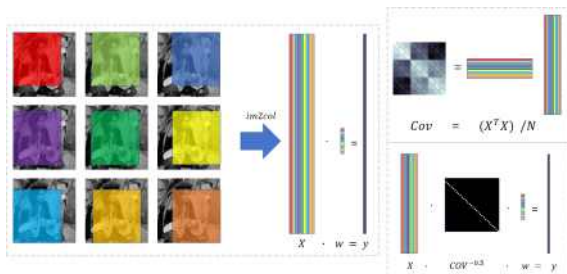


그림 2. 픽셀, 채널간 Decorrelation 연산

Deconvolution 연산은 Convolution 연산에서 픽셀간, 채널간의 연관 정보들을 제거해서 주요정보만을 학습하게 데이터를 정규화 시켜준다. Whitening이라고도 불리는 이 기술은, 대표적인 정규화 관련 논문인 Batch Normalization 논문[10]에서도 언급이 된다. 해당 논문에서는 Whitening 기법에 대해 학습 과정 중에 내부 파라미터의 변화로 인해 생기는 internal covariance shift 문제를 해결하지만, covariance matrix 계산과 inverse matrix 계산이 필요해서 계산량이 많으며 일부 파라미터들의 영향이 무시되기에 학습 성능이 저하되는 요소라고 주장했다. 그러나 Network Deconvolution[2]은 수학적 접근을 통해 간소화된 수식을 제시하고, GPU에 최적화된 코딩을 통해 Batch Normalization 기술과 학습 속도가 비슷하면서 더 뛰어난 성능을 보여주는 Correlation 방식을 고안했다. Network Deconvolution에서 Decorrelation을 적용하는 방법은 다음과 같다.

Data matrix $X_{N \times F}$ (N 은 sample들의 개수, F 는 feature들의 수)에 대해 $Cov = \frac{1}{N}(X - \mu)^T(X - \mu)$ 을 계산해서 Covariance matrix 를 구한다. $D = Cov^{-0.5}$, $D^T(X - \mu)^T(X - \mu)D = I$ 식을 사용해 Channel, Pixel 상호 연관성을 제거해준다. ($D \approx (Cov + \epsilon \cdot I)^{-0.5}$) 최종식은 $y = (X - \mu) \cdot D \cdot w + b \approx X \cdot D \cdot w$ 이고, 다음 레

이어의 입력 x_{i+1} 는 $x_{i+1} = f_i \circ W_i \circ D_i \circ x_i$ 식으로 만들어진다. (\circ 는 행렬곱, f_i 는 활성화 함수, W_i 는 레이어의 가중치, D_i 는 입력에 대한 Deconvolution 연산, x_i 는 i 번째 레이어에서 나온 입력)

3. 제안 방법

3.1. 가정

Deconvolution 기술과 초해상화 연구에서 보편적으로 쓰이는 Residual Learning 기법은 학습할 필요가 없는 불필요한 정보를 학습대상에서 제거해줌으로써, 학습하는 정보량을 한정해서 집중적으로 학습할 수 있게 만든다는 점이 같다. 또한, ImageNet[19] 데이터 Classification에서 좋은 성능을 보인 모델들의 Batch Normalization을 없애고, Deconvolution 연산을 Convolution 레이어의 앞부분에 첨가해서 수렴속도가 빨라졌고, 최종수렴 정확도가 올라가는 성과를 보였다. 이러한 이유들로 Deconvolution을 초해상화 기술에 적용시키면 학습할 때 복원성능 및 수렴속도 측면에서 도움이 되고, Batch Normalization[10]의 단점을 극복하여 성능을 높이는 효과를 낼 수 있는 새로운 방법으로 적용될 수 있다고 가정했다.

3.2. 제안한 구조

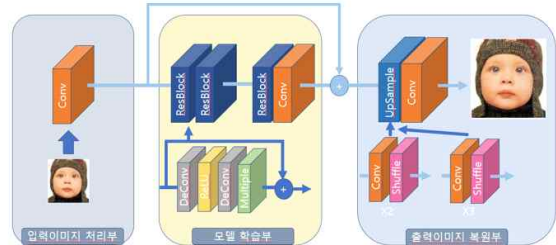


그림 3. Deconvolution 연산이 적용된 EDSR 네트워크 구조

실험에 사용한 네트워크 구조는 많은 초해상화 연구 논문들에서 결과 비교군으로 자주 쓰이는 초해상화 연구의 대표적인 모델인 EDSR 모델 내에서 Convolution 연산을 Deconvolution 연산으로 바꾼 모델이다. 초해상화 기술에서 입력이미지 처리부와 출력이미지 복원부는 중요한 정보들을 많이 다루는 부분이기 때문에 상호연관 정보를 제거하는 방법인 Deconvolution 연산을 모든 레이어에 적용하면 성능이 대폭 하락하는 문제를 발생시킨다. 따라서 초반부와 후반부는 기존 Convolution 연산을 유지하고 모델 학습부내의 ResBlock들의 내부에 있는 Convolution 연산들만 Deconvolution 연산으로 대체했다.

3.3. 실험 결과 및 분석

트레이닝 데이터는 표준화된 초해상화 데이터 셋중 하나인 DIV2K[13]를 사용했다. DIV2K를 이용해 모델 학습, 검증, 테스트를 진행하였다. 추가적으로 Set5[14], Set14[15], B100[16] 벤치마크 데이터 셋에 대해서 테스트를 진행했다.

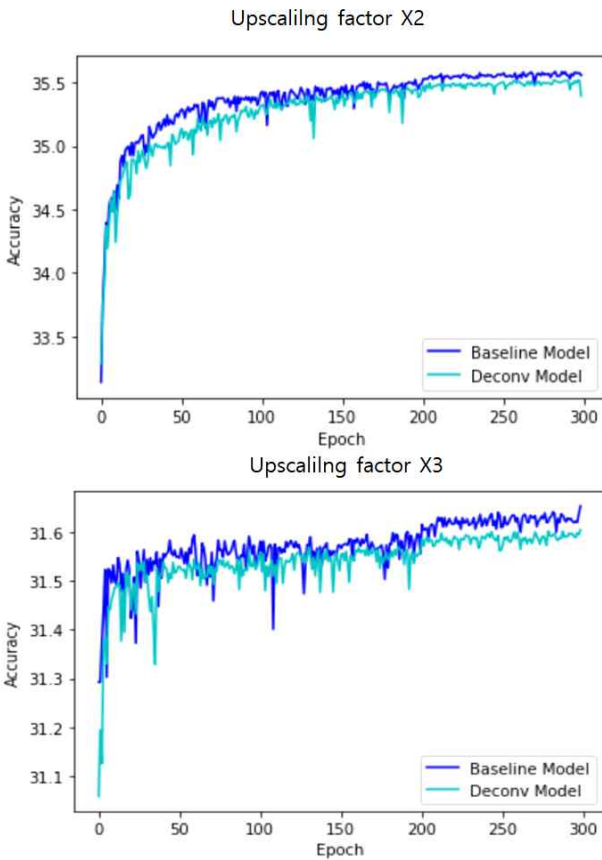


그림 4. EDSR baseline / Deconv 적용 EDSR 검증 성능 비교

Dataset	Scale	EDSR[1] (Baseline)	Ours
Set5[14]	×2	38.034	37.994
	×3	34.509	34.465
Set14[15]	×2	33.598	33.556
	×3	30.403	30.348
B100[16]	×2	32.213	32.174
	×3	29.141	29.134
DIV2K[13]	×2	34.670	34.589
	×3	31.001	30.957

표 1. 벤치마크 데이터셋 성능 PSNR(dB) 비교

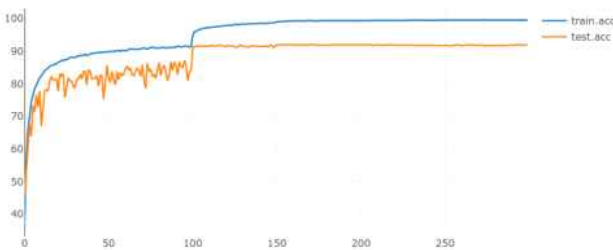


그림 5. 영상인식 딥러닝 분야 수렴그래프 예시

테스트 결과는 가정과는 다르게 최종 수렴속도가 빠르지 않았다. 원인은 baseline 모델인 EDSR은 Residual Learning을 사용해서 원본이

미지를 제외한 차이를 학습 및 복원하는 방식이기 때문으로 분석했다. 영상인식 딥러닝 분야에서의 Deconvolution의 장점은 Internal covariance shift문제를 해결하고 exploding/vanishing gradient문제를 해결해서 높은 learning rate를 사용해 빠르고 정확한 수렴을 하는 방식이었다. 그러나 Residual Learning이 적용된 초해상화 모델은 exploding/vanishing gradient문제가 해결되었기 때문에 이미 학습 수렴속도가 빠르고 Residual Learning을 사용한 모델에 Deconvolution을 적용하더라도 수렴속도가 더 빨라지지는 않는 것으로 추정했다.

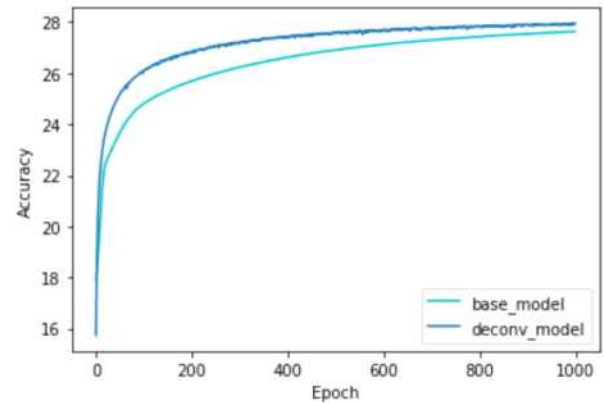


그림 6. SRCNN[3] / Deconv 적용 SRCNN 검증 성능 비교

그림 6은 Residual Learning을 사용하지 않는 간단한 3개의 Convolution 레이어로만 구성된 SRCNN 모델을 baseline으로 Deconvolution 연산을 적용해본 결과이다. 그래프를 보면, Deconvolution 연산을 적용한 모델이 수렴이 더 빠르고, 성능이 baseline에 비해서 더 높은 것을 알 수 있다. 이 실험결과를 통해, 그림 4의 결과는 Residual Learning 기법이 적용된 모델이기 때문에, 그림 6과 같은 빠른 수렴 결과가 도출되지 않았다고 할 수 있다. 이 실험을 통해 Deconvolution 기술은 영상 분류용 딥러닝 모델에서는 효과적으로 수렴 속도를 증가시키지만, 출력 타입이 고차원 실수 벡터형태를 가지는 초해상화 모델에서는 Residual Learning 기법이 적용되지 않는 경우에 한해서 수렴 속도를 향상시키는 것을 확인했다.

4. 결론

본 논문에서는 Network Deconvolution 기술을 초해상화 분야에 최초로 적용해보고, 채널 및 픽셀 단위의 Decorrelation이 수렴 속도에 주는 영향을 분석하는 실험을 했다. 초해상화 분야에서 Baseline으로 보편적으로 쓰이는 EDSR 모델에 대한 실험 결과는 가정과는 달리 수렴속도를 향상시키지 못했다. Residual Learning 기법을 적용하지 않는 초해상화 분야의 SRCNN 모델을 이용한 대조 실험을 통해 원인이 EDSR에 적용되어 있는 Residual Learning 기법 때문이라고 분석했다. Deconvolution 기술은 Residual Learning 기술을 사용하지 않는 초해상화 모델에 한해서 수렴 속도 향상 효과가 있다는 것을 알 수 있었다.

Acknowledgement

본 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2018R1C1B3008159) 본 논문은 과학기술정보통신부 및 정보통신산업진흥원의 ‘고성능 컴퓨팅 지원’ 사업으로부터 지원받아 수행하였음.

5. 참고 문헌

- [1] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *CVPR Workshops*, 2017. 5
- [2] Ye, Chengxi, et al. Network deconvolution, In *ICLR Conference*, 2020
- [3] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, pages 1646 - 1654, 2016. 4
- [4] S. Han, H. Mao, and W. J. Dally. Deep compression: Compressing deep neural network with pruning, trained quantization and huffman coding. *CoRR*, abs/1510.00149, 2, 2015.
- [5] Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee. "Accurate image super-resolution using very deep convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [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] Muqet, A.; Bin Iqbal, M.T.; Bae, S.-H. HRAN: Hybrid Residual Attention Network for Single Image Super-Resolution. *IEEE Access* 2019, 7, 137020 - 137029
- [8] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *CVPR*, 2016.
- [9] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv:1609.04802*, 2016.
- [10] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015.
- [11] Steven K. Esser, Jeffrey L. McKinstry, Deepika Bablani, Rathinakumar Appuswamy, and Dharmendra S. Modha. Learned step size quantization. *CoRR*, abs/1902.08153, 2019.
- [12] Kingma, Diederik P and Ba, Jimmy Lei. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [13] Agustsson, Eirikur, and Radu Timofte. "Ntire 2017 challenge on single image super-resolution: Dataset and study." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017.
- [14] M. Bevilacqua, A. Roumy, C. Guillemot and M. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding", Proc. Brit. Mach. Vis. Conf. (BMVC), pp. 1-10, 2012.
- [15] 33. R. Zeyde, M. Elad and M. Protter, "On single image scale-up using sparse-representations", Proc. Int. Conf. Curves Surf., pp. 711-730, 2010.
- [16] 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. 8th *IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 416-423, Jul. 2001.
- [17] S. Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1600.04747*, 2016.
- [18] Y. Wu and K. He. Group normalization. In *European Conference on Computer Vision (ECCV)*, 2018.
- [19] Deng, J. et al., 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. pp. 248 - 255.