

SqueezeNet 기반의 지식 증류 기법을 활용한 초해상화 기법

서유림, 강석주

서강대학교 전자공학과

utjdbfla@sogang.ac.kr, sjkang@sogang.ac.kr

SqueezeNet based Single Image Super Resolution
using Knowledge Distillation

Yu lim Seo Suk-Ju Kang

Sogang University

요 약

근래의 초해상화 (super-resolution, SR) 연구는 네트워크를 깊고, 넓게 만들어 성능을 높이는데 주를 이뤘다. 그러나 동시에 높은 연산량과 메모리 소비량이 증가하는 문제가 발생하기 때문에 이를 실제로 하드웨어로 구현하기에는 어려운 문제가 존재한다. 그렇기에 우리는 네트워크 최적화를 통해 성능 감소를 최소화하면서 파라미터 수를 줄이는 네트워크 SqueezeSR을 설계하였다. 또한 지식 증류 (Knowledge Distillation, KD)를 이용해 추가적인 파라미터 수 증가 없이 성능을 높일 수 있는 학습 방법을 제안한다. 또한 KD 시 teacher network의 성능이 보다 student network에 잘 전달되도록 feature map 간의 비교를 통해 학습 효율을 높일 수 있었다. 결과적으로 우리는 KD 기법을 통해 추가적인 파라미터 수 증가 없이 성능을 높여 다른 SR네트워크 대비 더 빠르고 성능 감소를 최소화한 네트워크를 제안한다.

1. 서론

초해상화 (Super-Resolution, SR)은 저해상도 영상으로부터 고해상도 영상 복원하는 컴퓨터 비전의 한 분야이다. 2014 년 SRCNN[1]은 최초로 딥러닝을 사용하여 SR 을 수행하였다. 이에 영향을 받아 VDSR[2] 과 같은 보다 성능이 높은 네트워크들이 이후 소개되었다. 그러나 이러한 네트워크들은 레이어수가 많고 복잡하기 때문에 높은 연산량 및 많은 파라미터를 요구하는 문제가 존재한다. 위의 문제를 해결하기 위해 우선 우리는 SqueezeNet[3]에서 영감을 얻어 성능감소를 최소화하고 낮은 파라미터 수를 가지는 student network 인 SqueezeSR 를 설계하였고 지식 증류 (Knowledge Distillation, KD) [4]을 통해 학습방법을 변경하여 보다 높은 성능을 가지도록 네트워크를 학습해보았다. 이때 성능을 높이기 위해 동시 학습에 사용한 teacher network 로는 SR 네트워크 중 State-of-the art (Sota) 네트워크 중 하나인 EDSR 을 기반으로 설계한 EDSR-

T 를 사용하였다.

2. 제안하는 방법

2.1 네트워크 구조

그림 1 에서 확인할 수 있듯이 teacher network 인 EDSR-T 는 우선 feature extraction 을 위해 Conv layer 하나를 구성하였고 이후 Residual Block 3 개를 연결하였다. 각각의 Residual Block 은 총 9 개의 Residual Module 로 구성되어 있다. 각각의 Residual Module 은 Conv layer 2 개, ReLU 1 개 그리고 0.1 scaling 을 위한 Multiple layer 한 개로 구성 되어있다. 각각의 Residual Module 은 모두 skip connection 으로 연결되어 있고 Residual Block 또한 skip connection 으로 연결되어 있다. Student network 인 SqueezeSR 의 구조는 teacher network 와 같이 feature extraction 을 위해 처음에 Conv layer 하나를 구성 하였고 이후 Squeeze Block 3 개를 연결하였다. Squeeze Block

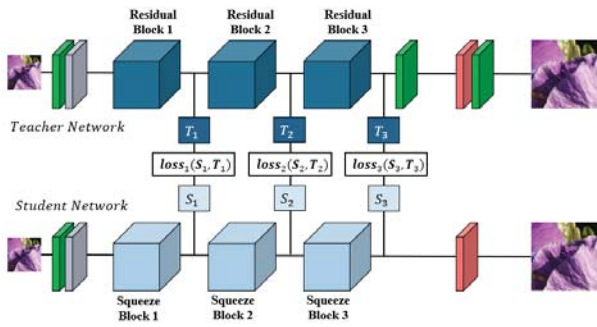


그림 1. Knowledge Distillation 을 위한 구조

하나는 총 1x1 filter 2 개와 3x3 filter 하나로 구성되어 있다. 각각의 Squeeze Block 은 모두 skip connection 으로 연결되어 있다. 기존의 SqueezeNet 과 다른 점은 downsampling layer 를 구성하지 않은 점이다.

2.2 지식 증류 기법

Teacher network 로부터 보다 나은 정보를 얻기 위해서는 feature map 단위에서 정보를 얻는 것이 필요하다. 그림 1. 에서 보이듯이 teacher network 의 feature map T_1, T_2, T_3 는 각각 Residual Block 을 통과한 출력 feature map 이고 student network 의 feature map S_1, S_2, S_3 는 Squeeze Block 을 통과한 출력 feature map 이다. 각각의 feature map 은 순서에 맞게 짝으로 구성되어 손실 함수의 입력으로 사용된다.

3. 실험 결과

이번 장에서는 정성적, 정량적 실험적 결과와 분석을 보여줄 것이다. 표 1 에서는 KD 을 적용하지 않은 SqueezeSR 과 KD 를 적용한 SqueezeSR-KD 그리고 teacher network 인 EDSR-T 와 정량적 결과 비교를 확인할 수 있다. SqueezeSR 은 다른 SR 방법 대비 매우 낮은 파라미터 수를 필요로 하지만 성능 감소 폭이 크지 않은 것을 확인할 수 있다. 또한 EDSR-T 의 지도 아래 학습시킨 SqueezeSR-KD 의 경우 추가적인 파라미터 수 증가 없이 성능이 올라간 것을 확인할 수 있다. 또한 이는 Bicubic 보간법과 유사한 수준의 수행시간을 가진다. 그림 2 에서는 다른 SR 기법들과 정성적 비교를 확인할 수 있다. 출력 이미지를 보면 확인할 수 있듯이 제안된 방법은 보다 선명한 영상을 출력하는 것을 알 수 있다.

4. 결론

우리는 네트워크 최적화 기법을 통해 다른 SR 네트워크에 비해 낮은 파라미터수를 가지고 연산 속도가 빠르면서 성능감소를 최

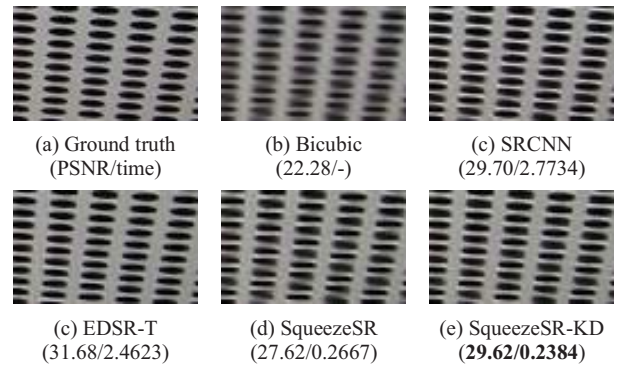


그림 2. Urban100 의 “img_004” 영상의 정성적 결과

표 1. 평가 데이터셋을 활용한 정량적 결과

Dataset	Scale	Bicubic	EDSR-T	Squeeze SR	Squeeze SR-KD
Set5	2	33.66	37.89	36.62	36.90
	3	30.39	34.17	32.33	32.70
	4	28.42	31.75	30.67	30.86
Set14	2	30.24	33.28	32.34	32.56
	3	27.55	29.97	28.97	29.24
	4	26.00	28.20	27.63	27.70
Urban 100	2	26.88	31.29	29.43	29.60
	3	24.46	27.43	25.71	26.10
	4	23.14	25.33	24.59	24.73

소화한 SqueezeSR 을 설계하였고 KD 기법을 활용해 해당 네트워크의 성능을 높이는 학습 기법을 제시하였다. 또한 정량적, 정성적 결과 모두에서 확인할 수 있듯이 성능 감소폭에 비해 해당 네트워크가 다른 네트워크에 비해 매우 많이 경량화된 것을 확인할 수 있다.

5. 참조

[1] Dong, Chao, et al. "Image super-resolution using deep convolutional networks." IEEE transactions on pattern analysis and machine intelligence 38.2 (2015): 295-307.
 [2] J. Kim, J.K. Lee, and K.M. Lee. "Accurate image super-resolution using very deep convolutional networks." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1637-1645, 2016.
 [3] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer. "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 1mb model size." arXiv preprint arXiv:1602.07360, 2016.
 [4] G. Hinton, O. Vinyals, and J. Dean. "Distilling the knowledge in a neural network." arXiv preprint arXiv:1503.02531, 2015.

6. 사사

본 연구는 2020 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원 (No. 2020M3H4A1A02084899), 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업의 연구 결과 (IITP-2020-2018-0-01421) 및 문화체육관광부 및 한국콘텐츠진흥원의 연구개발지원사업으로 수행되었음(과제번호: R2020040058)