

자가증류를 이용한 초해상화 네트워크 경량화 연구

*이여진 **박한훈

부경대학교

**hanhoon.park@pknu.ac.kr

A Study of Lightning Super-Resolution Networks Using Self-Distillation

*Lee, Yeojin **Park, Hanhoon

Pukyong National University

요약

최근 CNN(Convolutional Neural Network)은 초해상화(super-resolution)를 포함한 다양한 컴퓨터 비전 분야에서 우수한 성능을 보이며 널리 사용되고 있다. 그러나 CNN은 계산 집약적이고 많은 메모리가 요구되어 한정적인 하드웨어 자원인 모바일이나 IoT(Internet of Things) 기기에 적용하기 어렵다는 문제가 있다. 이런 한계를 해결하기 위해, 기 학습된 깊은 CNN 모델의 성능을 최대한 유지하며 네트워크의 깊이나 크기를 줄이는 경량화 연구가 활발히 진행되고 있다. 본 논문은 네트워크 경량화 기술인 지식증류(knowledge distillation) 중 자가증류(self-distillation)를 초해상화 CNN 모델에 적용하여 성능을 평가, 분석한다. 실험 결과, 정량적 평가지표를 통하여 자가증류를 통해서도 성능이 우수한 경량화된 초해상화 모델을 얻을 수 있음을 확인하였다.

1. 서론

인공신경망의 일종인 CNN(Convolutional Neural Network)은 컴퓨터 비전 분야, 자연어처리 등 다양한 분야에서 우수한 성능을 보이고 있으며, 영상 처리 및 컴퓨터 비전 기술 중의 하나인 초해상화(super-resolution) 분야에서도 CNN을 이용한 연구가 활발히 진행되고 있다. 하지만 많은 메모리와 연산량을 필요로 하는 CNN을 한정적인 하드웨어 자원인 모바일 및 IoT 기기에 적용하기는 어렵다. 이런 한계를 해결하기 위해, 기존의 학습된 모델의 성능을 최대한 유지하며 모델의 크기를 줄이는 인공신경망 경량화 연구가 활발히 진행되고 있다.

본 논문에서는 신경망 경량화를 위해 지식증류(knowledge distillation)[1]의 최신 기법 중 하나인 자가증류(self-distillation)[2]를 CNN 기반 초해상화 모델인 SRResNet[3]에 적용한다. 실험을 통해, 자가증류 기법의 초해상화 모델 경량화에 대한 성능을 평가한다.

2. 관련 연구

2.1 지식증류

지식증류는 기 학습된 깊은 CNN 모델의 지식을 전달하여 가벼운 CNN 모델의 학습 능력을 향상시키는 기술을 말한다. 전달되는 지식의 종류나 전달 방법 등에 따라 다양한 지식증류 기법들이 제안되고 있다. 처음 소개된 지식증류 기법은 크고 깊은 교사 네트워크의 소프트맥스(softmax) 층에서 출력된 확률분포 정보를 완화하여 작고 얇은 학생 네트워크로 전달하는 방식이다[1]. 그러나, 교사 네트워크의 마지막 층의 지식만 전달하기 때문에 전달되는 지식에 한계가 있다. 이런 한계를 개선하기 위해 FitNets[4]는 교사 네트워크의 중간 층의 지식을 함께 전달하여 학생 네트워크가 학습할 수 있도록 하였다. 실험을 통해 이전의 방

식[1]보다 학생 네트워크의 성능이 향상됨을 보였고, 지식증류를 통한 네트워크의 파라미터의 수, 연산 횟수, 속도 증가율, 압축률 등에 대한 정량적인 평가 결과를 제시하였다. 이후, 교사 네트워크 간의 주의집중 지도(attention map)를 모방하거나[5], 교사 네트워크 없이 학생 네트워크들의 코호트(cohort) 학습하거나[6], 다른 네트워크를 필요로 하지 않고 자가증류를 통해 학습하는[2] 등 다양한 지식증류 기법들이 제안되었다. 그러나, 이러한 대부분의 지식증류 기법들은 영상 분류 분야에서 실험이 진행되고 성능 또한 영상 분류에서만 입증되었다. 초해상화 분야에서 지식증류를 사용한 경량화 연구는 거의 없으며, 최근 특징 지도(feature map)의 유사도를 모방하도록 학습하는 기법[7]이 제안되었다.

2.2 자가증류

자가증류는 지식증류를 위한 최신 기법으로, 별도의 교사 네트워크 없이 스스로 증류를 진행하는 학습 기법이다[2]. 자가증류를 위해 네트워크는 여러 얇은(shallow) 구간으로 나뉘어 각각 학생 네트워크로써 증류 받게 되고, 이때 전체 네트워크의 결과를 교사 네트워크로 취급한다. 학생 네트워크 학습을 위해 자가증류에서 사용된 손실(loss)에는 세 종류가 있다.

- Label 손실: 각 학생 네트워크, 즉 얇은 구간의 결과들과 레이블(label)을 비교한다. 교차 엔트로피(cross-entropy)를 통해 계산되며, 이 손실을 최소화함으로써 학생 네트워크는 영상 분류를 위한 지식을 데이터 셋(dataset)에서 직접 추출하여 학습한다.
- Teacher 손실: 가장 깊은 네트워크, 즉 교사 네트워크의 결과(즉, 레이블의 확률값)를 각 학생 네트워크의 결과들과 비교한다. KL 발산(Kullback-Leibler divergence)를 통해 계산되며, 이 손실을 최소화함으로써 학생 네트워크들이 깊은 교사 네트워크의 지식(레

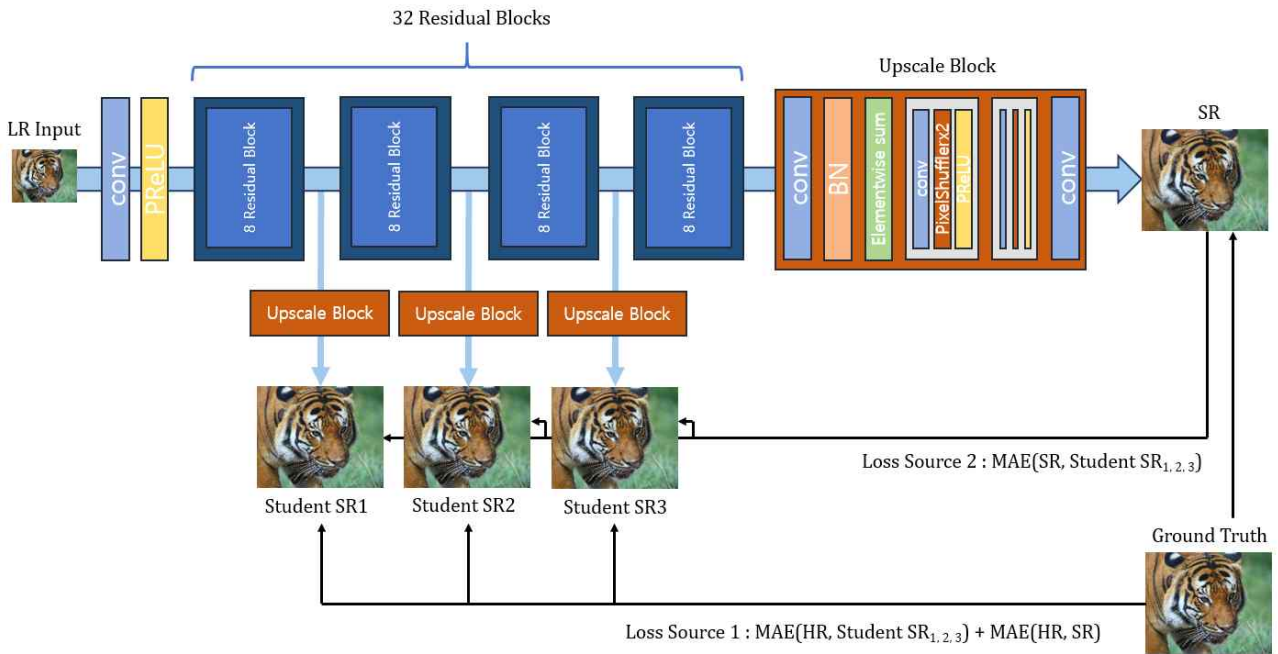


Fig 1. The architecture of SRResNet-32 to which self-distillation is applied.

이블 분류 능력)을 받는다.

- Feature 손실: 교사 네트워크와 학생 네트워크의 특징 지도를 비교한다. L2 norm을 통해 계산되며, 이 손실을 최소화함으로써 학생 네트워크의 특징 지도가 교사 네트워크의 특징 지도와 유사해 지도록 유도한다.

3. 실험 방법

본 연구에서는 하나의 신경망 모델을 중심모델과 여러 하위모델로 나누어 중심모델의 지식을 하위모델로 증류하는 자가증류 기법을 초해상화에 적용하는 실험을 진행하고 분석하는 것을 목표로 한다.

초해상화에 자가증류를 적용하기 위해서는, 영상 분류 모델에 자가증류 기법을 적용한 기존 연구와는 다르게 네트워크 구조 및 손실 계산을 위한 함수를 바꿔야 할 필요가 있다.

우선, 기존 초해상화 모델에 자가증류를 적용하기 위해 그림 1에서와 같이 중간 층에서 분기된 각 하위모델에 업스케일링 블록을 추가한다. 다음으로, 자가증류를 통해 하위 모델, 즉 학생 네트워크 학습을 위한 손실 함수는 픽셀 손실과 증류 손실로 구성된다. 픽셀 손실은 Ground-truth 영상과 학생 네트워크의 초해상화 영상 사이의 평균 절대 오차(mean absolute error)로 계산되며, 증류 손실은 중심모델의 초해상화 영상과 각 하위모델의 초해상화 영상 사이의 평균 절대 오차로 계산된다. 증류 손실을 최소화함으로써 중심모델과 각 하위모델은 서로 가까운 초해상화 영상을 생성하도록 학습된다. 여기서, 손실을 구하기 위해 평균 절대 오차를 사용한 것은 실험을 통해 평균 제곱 오차(mean square error)를 사용하는 것보다 정량적, 직각적으로 더 우수한 결과를 얻는 것을 확인하였기 때문이다. 두 가지 손실은 하이퍼 파라미터를 통해 가중치를 조절할 수 있으며 본 논문의 실험에서는 픽셀 손실과 증류 손실의 비율을 9:1로 설정하였다.

4. 실험 결과

실험을 위한 학습 데이터 셋으로 DIV2K[8]을 사용하고 검증 데이터 셋으로 Set5[9], Set14[10], BSD100[11], Urban100[12]을 사용한다. 저해상도 영상은 바이큐빅 보간법을 통해 생성하였다. 학습을 위한 하이퍼 파라미터들은 다음과 같다. 학습률은 10^{-4} 로 설정하고 최적화 알고리즘은 Adam을 사용하며 배치(batch)의 크기는 4로 설정하였다. 각 실험은 200번의 학습을 진행한다.

실험에 사용한 모델은 SRResNet으로, 총 32개의 resblock으로 구성된 모델을 중심모델(SRResNet-32)로 하고, 처음 8개의 resblock을 하나의 하위모델(SRResNet-8)로 설정하였다(그림 1 참조).

표 1은 SRResNet-32에 자가증류를 적용하여 하위모델로 얻어진 SRResNet-8과 자가증류 없이 독립적으로 학습된 SRResNet-8 사이의 성능을 비교한 결과를 보여준다. BSD100 데이터 셋의 PSNR을 제외하고 자가증류를 적용한 모델이 그렇지 않은 모델에 비하여 높은 PSNR과 SSIM이 나온 것을 확인할 수 있다.

Table 1. PSNR/SSIM results of SRResNet-8 trained with or without self-distillation

	SRResNet-8	SRResNet-8 with self-distillation
Set5	27.343/0.779	28.758/0.879
Set14	25.410/0.679	26.093/0.747
BSD100	25.393/0.648	24.361/0.684
Urban100	22.667/0.636	24.313/0.656

표 2는 하위모델을 얻기 위한 중심모델의 크기를 달리 하였을 때의 결과를 보여준다. 중심모델의 크기가 클수록 좋은 성능을 보이는 것을 알 수 있으며, 중심모델의 크기가 클수록 양질의 지식을 하위모델로 전

달하기 때문인 것으로 판단된다.

Table 2. PSNR/SSIM results of SRResnet-8 trained with different sized teacher models

	SRResNet-32	SRResNet-50	SRResNet-100
Set5	28.758/0.879	28.803/0.881	29.225/0.892
Set14	26.093/0.747	26.097/0.748	26.340/0.754
BSD100	24.361/0.684	24.336/0.685	24.478/0.697
Urban100	24.313/0.656	23.652/0.656	24.091/0.656

5. 결론

본 논문에서는 네트워크 경량화 기술인 지식증류 중 자가증류를 초해상화에 적용하여 성능을 평가, 분석함으로써 자가증류를 통해서도 효과적으로 초해상화 모델을 경량화할 수 있음을 보였다. 자가증류를 적용하여 학습하여 경량화된 모델의 성능은 지식증류없이 학습된 같은 크기의 모델보다 성능이 우수한 것을 확인하였고, 경량화된 모델의 성능은 중심모델의 크기가 클수록 우수한 것을 알 수 있었다.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) Grant by the Korean Government through the MSIT under Grant 2021R1F1A1045749

REFERENCES

- [1] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," Proc. of NIPS 2014 Deep Learning and Representation Learning Workshop, 2014.
- [2] L. Zhang, J. Song, A. Gao, J. Chen, C. Bao, and K. Ma, "Be your own teacher: improve the performance of convolutional neural networks via self distillation," Proc. of International Conference on Computer Vision, pp. 3713-3722, 2019.
- [3] C. Ledig, et al., "Photo-realistic single image super-resolution using a generative adversarial network," Proc. of CVPR, pp. 4681-4690, 2017.
- [4] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "FitNets: hints for thin deep nets," Proc. of the International Conference on Learning Representations, 2015.
- [5] N. Komodakis and S. Zagoruyko, "Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer," Proc. of the International Conference on Learning Representations, 2017.
- [6] Y. Zhang, T. Xiang, T. M. Hospedales, and H. Lu, "Deep

- mutual learning," Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 4320-4328, 2018.
- [7] Z. He, T. Dai, J. Lu, Y. Jiang, and S. -T. Xia, "Fakd: feature-affinity based knowledge distillation for efficient image super-resolution," Proc. of IEEE International Conference on Image Processing, pp. 518-522, 2020.
- [8] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: dataset and study," Proc. CVPRW, 2017.
- [9] M. Bevilacqua, A. Roumy, C. Guillemot, and M. A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," Proc. of British Machine Vision Conference, pp. 135.1-135.10, 2012.
- [10] 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.
- [11] 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.
- [12] R. Narita, K. Tsubota, T. Yamasaki, and K. Aizawa, "Sketch-based manga retrieval using deep features," Proc. ICDAR, pp. 49-53, 2017.