

## 상호증류를 통한 SRGAN 판별자의 성능 개선

이여진<sup>1</sup>, 박한훈<sup>1\*</sup>

<sup>1</sup>부경대학교 전자정보통신공학부

## Performance Improvement of SRGAN's Discriminator via Mutual Distillation

Yeojin Lee<sup>1</sup>, Hanhoon Park<sup>1\*</sup>

<sup>1</sup>Division of Electronics and Communications Engineering, Pukyong National University

**요약** 상호증류는 교사 네트워크 도움 없이 다수의 네트워크 사이에 지식을 전달함으로써 협력적으로 학습하도록 유도하는 지식증류 방법이다. 본 논문은 상호증류가 초해상화 네트워크에도 적용 가능한지 확인하는 것을 목표로 한다. 이를 위해 상호증류를 SRGAN의 판별자에 적용하는 실험을 수행하고, 상호증류가 SRGAN의 성능 향상에 미치는 영향을 분석한다. 실험 결과, 상호증류를 통해 판별자의 지식을 공유한 SRGAN은 정량적, 정성적 화질이 개선된 초해상화 영상을 생성하였다.

• **주제어** : 초해상화, 딥러닝, 지식증류, 상호증류, SRGAN, 판별자

**Abstract** Mutual distillation is a knowledge distillation method that guides a cohort of neural networks to learn cooperatively by transferring knowledge between them, without the help of a teacher network. This paper aims to confirm whether mutual distillation is also applicable to super-resolution networks. To this regard, we conduct experiments to apply mutual distillation to the discriminators of SRGANs and analyze the effect of mutual distillation on improving SRGAN's performance. As a result of the experiment, it was confirmed that SRGANs whose discriminators shared their knowledge through mutual distillation can produce super-resolution images enhanced in both quantitative and qualitative qualities.

• **Key Words** : Super-resolution, Deep learning, Knowledge distillation, Mutual distillation, SRGAN, Discriminator

---

Received 06 September 2022, Revised 27 September 2022, Accepted 29 September 2022

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

## I. 서론

초해상화(Super-resolution)란 저해상도 영상으로부터 고해상도 영상을 생성하는 기술로서, 의료, 보안 등 다양한 컴퓨터 비전 응용 분야에서 활용되고 있다. 최근 CNN(Convolutional Neural Network) 기반 딥러닝 기술의 발전과 함께 CNN을 이용한 초해상화 연구가 활발히 진행되고 있다. SRGAN은 대표적인 CNN 기반 초해상화 모델로, 지각 품질(perceptual quality)이 우수한 초해상화 영상을 생성할 수 있다[1].

CNN은 많은 메모리와 연산량을 요구하기 때문에 실시간 처리가 필요한 분야에 적용할 수 없고, 모바일 및 IoT 기기에 적용하기는 어렵다. 이런 한계를 극복하기 위해, 기존의 학습된 모델의 성능을 최대한 유지하며 크기를 줄이는 모델 경량화 연구가 활발히 진행되고 있다. 지식증류(knowledge distillation)은 네트워크 구조나 형태의 변형 없이 지식을 전달하는 모델 경량화 기술로서, 최근에는 모델의 성능을 최적화하는 방법으로 널리 활용되고 있다[2].

본 논문에서는 모델 경량화 대신 모델 성능 최적화에 초점을 두고, 지식증류 기법 중 상호증류(mutual distillation)를 SRGAN의 판별자에 적용하는 실험을 진행한다[3]. 실험을 통해, 상호증류를 통해 분류 성능이 향상된 판별자가 SRGAN의 초해상화 영상 생성 능력에 미치는 영향을 분석한다.

## II. 관련연구

### 2.1 지식증류

지식증류는 기 학습된 깊은 CNN 모델의 지식을 전달하여 가벼운 CNN 모델의 학습 능력을 향상시키는 기술을 말한다. 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 작고 얇은 학생 네트워크로 전달하는 방식이다[2]. 그러나, 교사 네트워크의 마지막 층의 지식만 전달하기 때문에 전달되는 지식에 한계가 있다. 이런 한계를 개선하기 위해 FitNets는 교사 네트워크의 중간층의 지식을 함께

전달하여 학생 네트워크가 학습할 수 있도록 하였다[4]. 이후, 교사 네트워크의 주의집중 지도(attention map)를 모방하거나, 교사 네트워크 없이 상호증류를 통해 학생 네트워크들이 서로 지식을 전달하면서 학습하거나, 다른 네트워크를 필요로 하지 않고 자가증류(self-distillation)를 통해 학습하는 등 다양한 지식증류 기법들이 제안되었다[3, 5, 6]. 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 기법이 제안되었다[7].

### 2.2 CNN 기반 초해상화 모델

초해상화는 저해상도 영상을 고해상도로 변환하는 기술이다. 기존 초해상화 기법들은 비교적 단순한 선형 매핑을 기반으로 하기 때문에 복잡하고 비선형적인 초해상화 모델을 구현하기 어려웠다. 이러한 문제를 해결하기 위해 최근 딥러닝 기술을 활용한 초해상화 알고리즘 기법들이 활발히 연구되고 있으며 다양한 구조나 형태의 CNN 기반 모델들이 제안되고 있다. SRCNN 모델은 초해상화 분야에 처음으로 딥러닝 기술을 도입한 사례로, 딥러닝 기술을 활용함으로써 고화질의 초해상화 영상 생성이 가능하게 되었다[8].

그러나, 초해상화 영상의 지각 품질을 개선하기 위해 GAN(Generative Adversarial Network)을 사용한 초해상화 모델들이 주목받기 시작했으며, SRGAN은 초해상화를 위한 초기 GAN 기반 초해상화 모델이다[1, 9]. SRGAN은 생성자(generator)와 판별자(discriminator)로 구성되는데, 생성자는 초해상화 영상과 ground-truth 영상의 VGG 특징 지도가 유사한 초해상화 영상을 생성하도록 학습되고, 판별자는 생성자가 생성한 영상과 실제(real) 영상 사이를 잘 구별하도록 학습된다[10]. 생성자는 학습이 진행될수록 다른 CNN 기반 초해상화 모델에 비해 정량적인 품질은 떨어지지만, 지각 품질이 우수한 초해상화 영상을 생성할 수 있다. 이후, SRGAN의 네트워크 구조를 개선하거나, 영상의 지각 품질을 측정하는 새로운 방법을 적용함으로써 생성된 초해상화 영상의 지각 품질을 보다 향상시키는 방법들이 제안되고 있다[11-12].

### III. 실험 방법

본 논문에서는 코호트 학습을 통해 지식을 공유하는 상호증류 기법을 SRGAN의 판별자에 적용하는 실험을 진행하고 분석하는 것을 목표로 한다. 실험의 편의를 위해 두 개의 SRGAN 모델만을 사용하였다.

상호증류를 적용한 SRGAN의 손실 함수는 다음과 같다. 우선, 생성자 학습을 위한 전체 손실  $L_G$ 은 기존 SRGAN과 같다. ground-truth인 고해상도 영상과 초해상화 영상 사이의 VGG 손실을 계산한 콘텐츠 손실(content loss)  $L_{con}$ 과 판별자가 초해상화 영상을 실제 영상이 아닌 생성자에 의해 생성된 영상임을 판별할 확률을 나타내는 적대적 손실(adversarial loss)  $L_{adv}$ 의 합으로 구성되며, 적대적 손실에 10-3을 곱하여 비율을 조절한다.

$$L_G = L_{con} + 10^{-3}L_{adv}. \quad (1)$$

다음으로, 판별자 학습을 위한 전체 손실  $L_D$ 은 판별자가 생성자에 의해 생성된 초해상화 영상을 실제 영상이라고 판별할 확률과 실제 영상을 생성자가 생성한 초해상화 영상이라고 판별할 확률의 합인  $L_{dis}$ 와 상호증류 손실인  $L_{md}$ 의 합으로 구성된다.

$$L_D = L_{dis} + L_{md}. \quad (2)$$

상호증류 손실  $L_{md}$ 은 판별자가 초해상화 영상을 실제 영상이 아니라고 판별할 확률과 동료 SRGAN의 판별자가 자신의 생성자가 생성한 초해상화 영상을 실제 영상이 아니라고 판별할 확률 사이의 쿨백-라이블러 발산(Kullback-Leibler divergence)인  $\nabla_{SR}$ 와 판별자가 실제 영상을 실제 영상이라고 판별할 확률과 동료 SRGAN의 판별자가 실제 영상을 실제 영상이라고 판별할 확률 사이의 쿨백-라이블러 발산인  $\nabla_{HR}$ 의 합으로 구성된다.

$$L_{md} = \nabla_{SR} + \nabla_{HR}. \quad (3)$$

상호증류 손실 계산을 위해 두 확률 사이의 직접적인 쿨백-라이블러 발산을 계산할 수도 있으나, 기존 연구에서와 같이 확률분포를 연화(softening)한 후 계산할 수도 있다[2]. 두 방법의 성능 차이는 실험 결과에서 주어진다.

실험을 위한 학습 데이터 셋으로 DIV2K, 검증 데이터 셋으로 Set5, Set14, BSD100, Urban100을 사용한다 [13-17] (그림 1 참조). 학습을 위한 저해상도 영상은 고해상도 영상을 바이큐빅 보간법(bicubic interpolation)을 사용하여 다운샘플링(down-sampling)하여 생성하였다. 실험에 사용된 SRGAN 모델은 32개의 resblock과 64개의 filter로 구성된다. 모델 구현은 TensorFlow(버전 2.6.0)를 사용하였으며, 모델 학습은 i7-8700 3.20GHz CPU와 16GB RAM, Nvidia RTX 3060 GPU를 가진 PC에서 진행되었다. Adam optimizer( $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ )를 사용하였으며, step 수는 50,000, 학습률(learning rate)은  $10^{-4}$ , 배치 크기(batch size)는 16으로 설정되었다. 초해상화 영상 결과의 화질 평가를 위해 정량적 화질 측정 지표인 PSNR(Peak Signal-to-Noise Ratio)과 정성적 화질 측정 지표인 PI(Perceptual Index)를 측정하였다. PI는 사람이 평가한 점수와 유사한 상관관계를 가지며, 무참조(no-reference) 영상 품질 점수인 Ma와 NIQE(Naturalness Image Quality Evaluator)로부터 계산된다[18, 19]. PI는 낮을수록 더 나은 성능을 나타낸다.

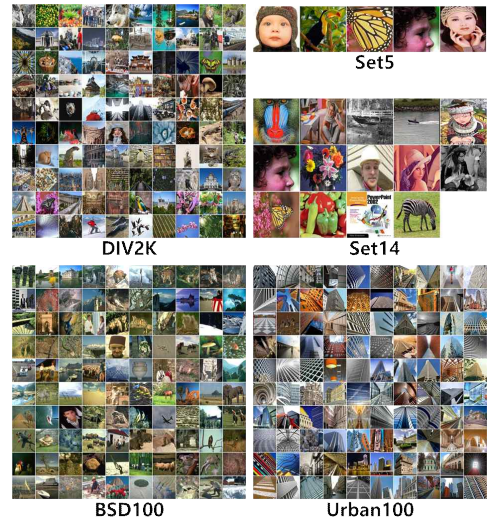


Fig. 1. Datasets used in our experiments.

Table 1. Mean PSNR and PI results of SRGAN with and without mutual distillation when scale factor = 2.

	PSNR			PI		
	Independent	ML	ML*	Independent	ML	ML*
Set5	32.191	32.789	32.465	2.766	2.716	2.629
Set14	29.649	30.365	30.139	2.763	2.637	2.638
BSD100	28.352	28.826	28.498	2.421	2.362	2.329
Urban100	27.640	28.123	27.802	3.828	3.832	3.772

Table 2. Mean PSNR and PI results of SRGAN with and without mutual distillation when scale factor = 4.

	PSNR			PI		
	Independent	ML	ML*	Independent	ML	ML*
Set5	26.960	27.469	27.119	3.124	3.194	3.111
Set14	25.129	25.740	25.421	2.811	2.935	2.853
BSD100	24.172	24.362	24.075	2.388	2.446	2.463
Urban100	23.407	23.675	23.429	3.338	3.400	3.383

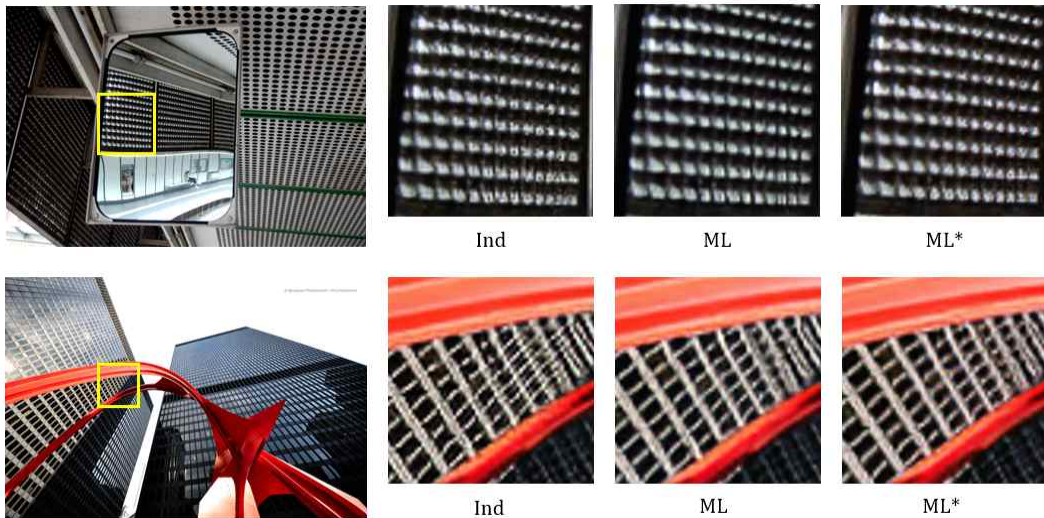


Fig. 2. Visual comparison between super-resolved images with and without mutual distillation (scale factor = 2).

#### IV. 실험 결과 및 분석

표 1은 상호증류 유무에 따른 SRGAN 모델의 2배 초해상화 결과 영상의 평균 PSNR과 PI값을 보여준다. Independent는 상호증류를 받지 않고 독립적으로 학습한 모델이고 ML은 판별자에 상호증류를 적용한 모델이다. ML\*은 판별자의 결과(확률분포)를 연화한 후 상호증류한 모델이다.

우선, 모든 데이터 셋에 대하여 상호증류한 모델이 독립적으로 학습한 모델보다 높은 PSNR 값을 가진다. 이는 상호증류한 판별자의 분류 성능 개선이 생성자로 하여금 정량적으로 더 나은 화질의 영상을 생성하도록

유도한다는 것을 나타낸다. 단, 정량적 화질 향상을 위해서는 확률분포를 연화하지 않고 상호증류하는 것이 더 우수한 결과를 보였다. 정성적 화질 평가 지표인 PI의 경우, 대부분의 데이터 셋에서 상호증류를 한 모델이 우위를 차지했으며, 특히 확률분포를 연화한 ML\*이 가장 높은 성능을 보였다. 이는 상호증류를 통해 SRGAN의 판별자의 분류 성능을 향상시킴으로써 초해상화 영상은 정성적으로도 향상된다는 것을 의미한다. 또한, 정성적인 화질 향상을 위해서는 확률분포를 연화하여 상호증류를 하는 것이 낫다는 것을 보여준다.

표 2는 상호증류 유무에 따른 SRGAN 모델의 4배 초해상화 결과 영상의 평균 PSNR과 PI값을 보여준다.

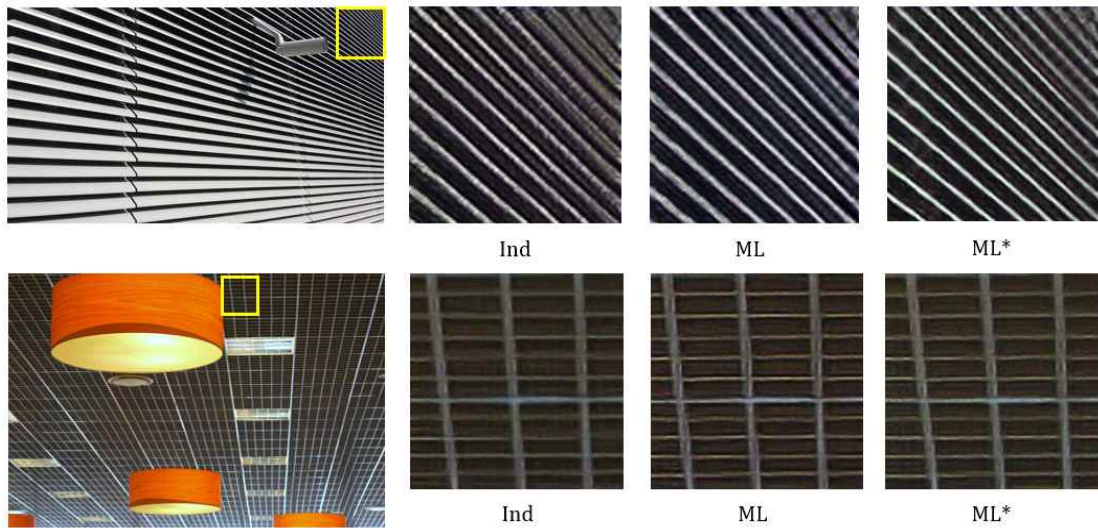


Fig. 3. Visual comparison between super-resolved images with and without mutual distillation (scale factor = 4).

2배 초해상화에서와 마찬가지로, 상호증류한 모델이 높은 PSNR을 보이며, 확률분포를 연화하지 않는 것이 좋다는 것을 확인할 수 있다. 반면, PI의 경우, 상호증류를 받지 않은 모델이 더 좋은 결과를 보였다. 그러나 초해상화 영상을 시각적으로 비교해 보면 상호증류를 받은 모델의 결과가 더 좋은 경우 많아 상호증류가 정성적인 화질에 부정적인 영향을 미친다고 보기는 어렵다.

그림 2와 3은 각각 2배, 4배 초해상화에서 상호증류 유무에 따라 SRGAN 모델이 생성한 초해상화 영상을 시각적으로 비교하여 보여준다. 상호증류한 모델이 증류를 받지 않은 모델보다 왜곡이 적고 선명한(경계선이 뚜렷한) 영상을 생성하는 것을 알 수 있다.

## V. 결론

본 논문에서는 상호증류를 통한 판별자의 분류 능력 향상이 SRGAN의 초해상화 영상 생성에 미치는 영향을 분석하는 것을 목표로 하였다. 이를 위해 두 개의 SRGAN 모델의 판별자에 대해 상호증류를 적용하여 학습하는 실험을 진행하였으며, 판별자에 상호증류를 적용한 모델이 증류를 받지 않은 모델보다 정량적, 정성적(인지적)으로 더 우수한 초해상화 영상을 생성한다는 것을 확인하였다.

향후 상호증류를 다양한 구조의 초해상화 모델에 적용하기 위한 방안을 마련하고, 이를 통해 초해상화 분야에서의 상호증류 기법의 유용성을 확인하기 위한 추가 연구가 필요하다.

## ACKNOWLEDGMENTS

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

## REFERENCES

- [1] C. Ledig, et al., “Photo-realistic single image superresolution using a generative adversarial network,” Proc. of CVPR, pp. 105-114, 2017.
- [2] G. Hinton, et al., “Distilling the knowledge in a neural network,” Proc. of NIPS, 2014.
- [3] Y. Zhang, et al., “Deep mutual learning,” Proc. of CVPR, pp. 4320-4328, 2018.
- [4] A. Romero, et al., “FitNets: hints for thin deep nets,” Proc. of ICLR, 2015.
- [5] N. Komodakis and S. Zagoruyko, “Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer,” Proc. of ICLR, 2017.

[6] 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 ICCV, pp. 3713-3722, 2019.

[7] Z. He, et al., "Fakd: feature-affinity based knowledge distillation for efficient image super-resolution," Proc. of ICIP, pp. 518-522, 2020.

[8] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," Proc. of ECCV, pp. 184-199, 2014.

[9] I. J. Goodfellow, et al., "Generative adversarial networks," arXiv preprint arXiv:1406.2661, 2014.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint, arXiv:1409.1556, 2014.

[11] X. Wang, et al., "ESRGAN: enhanced super resolution generative adversarial networks," Proc. of ECCV, pp. 63-79, 2018.

[12] Y. Choi and H. Park, "Improving ESRGAN with an additional image quality loss," Multimedia Tools and Applications, 2022.

[13] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: dataset and study," Proc. of CVPRW, pp. 126-135, 2017.

[14] M. Bevilacqua, et al., "Low complexity single image super-resolution based on nonnegative neighbor embedding," Proc. BMVC, 2012.

[15] R. Zeyde, et al., "On single image scale-up using sparse-representations," Proc. of Int. Conf. on Curves and Surfaces, pp. 711-730, 2010.

[16] D. Martin, et al., "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.

[17] J.-B. Huang, et al., "Single image super-resolution from transformed self-exemplars," Proc. of CVPR, pp. 5197-5206, 2015.

[18] C. Ma, et al., "Learning a no-Reference quality metric for single-image super-resolution," CVIU, vol. 158, pp. 1-16, 2017.

[19] A. Mittal, et al., "Making a completely blind image quality analyzer," IEEE Signal Processing Letters, vol. 20, no. 3, pp. 209-212, 2013.

---

## 저자 소개

---

이 여 진 (Yeojin Lee)



2021년 3월~현재 : 부경대학교  
전자공학과 학부생  
관심분야 : 지식 증류, SLAM

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

관심분야 : 증강현실, 인간컴퓨터상호작용,  
컴퓨터비전/그래픽스, 딥러닝 응용