

## 대조 학습 기반 초해상도 모델 경량화 기법

문현철, 권용훈, 정진우, \*김성제  
한국전자기술연구원

{hcmoon23, yhkwon, jw.jeong, sungjei.kim}@keti.re.kr

### Compression of Super-Resolution model Using Contrastive Learning

HyeonCheol Moon, Yong-Hoon Kwon, JinWoo Jeong, \*SungJei Kim  
Korea Electronics Technology Institute

#### 요 약

최근 딥러닝의 발전에 따라 단일 이미지 초해상도 분야에 좋은 성과를 보여주고 있다. 그러나 보다 더 높은 성능을 획득하기 위해 네트워크의 깊이 및 파라미터의 수가 크게 증가하였고, 모바일 및 엣지 디바이스에 원활하게 적용되기 위하여 딥러닝 모델 경량화의 필요성이 대두되고 있다. 이에 본 논문에서는 초해상도 모델 중 하나인 EDSR(Enhanced Deep Residual Network)에 대조 학습 기반 지식 전이를 적용한 경량화 기법을 제안한다. 실험 결과 제안한 지식 전이 기법이 기존의 다른 지식 증류 기법보다 향상된 성능을 보임을 확인하였다.

#### 1. 서론

최근 합성곱 신경망(Convolution Neural Network)은 다양한 분야에서 활용되고 있으며, 특히 저화질의 영상을 고화질로 변환하는 초해상도 분야에 뛰어난 성능을 보이고 있다. 그러나, 보다 더 높은 성능을 획득하기 위해 네트워크의 깊이 및 파라미터의 수가 크게 증가하였고, 모바일 및 엣지 디바이스 같은 저전력 기기에 적용되기에는 한계가 존재한다. 이에 성능은 최대한 유지하면서 모델을 최대한 작게 만드는 경량화 기법들이 연구되고 있다.

초해상도 분야에서의 모델 경량화 연구는 크게 연산에 효율적인 신경망 구조를 설계하는 기법 [1]과 파라미터 수를 줄이는 경량화 기법들이 연구되고 있다 [2]. 그 중에 대표적인 경량화 기법은 가지치기(Pruning), 지식 증류(Knowledge Distillation) 등이 있다. 이 중에 지식 증류 기법은 보다 큰 모델인 교사 네트워크(Teacher Network)로부터 얻어진 지식을 작은 모델인 학생 네트워크(Student Network)로 전이하여 작은 모델의 성능을 향상시키는 기법이다. 초해상도 분야에서는 큰 모델과 작은 모델의 중간 특징 맵 혹은 네트워크의 결과를 L1 손실로 가깝게 하여 성능을 높여왔다 [3-4].

본 논문에서는 초해상도 모델 중 하나인 EDSR(Enhanced Deep Residual Network) 모델 [5]에 대조 학습 기반 지식 전이를 활용한 경량화 기법을 제안한다. 제안 기법은 우선 배치 내에서 생성된 교사와 학생 네트워크 결과물들을 긍정적 샘플(Positive samples)과 부정적 샘플(Negative samples) 둘로 구분한다. 그리고 긍정적 샘플의 유사성을 최대화, 부정적 샘플 간에는 최소화하여 교사 네트워크와 학생 네트워크 간의 상호의존정보(mutual information)를 최대화하게 된다.

#### 2. 제안 기법

대조 학습은 자가-지도 학습(self-supervised learning) 기법 중 하나로 특징에 대한 일반화를 최대화하고, 보다 더 적은 데이터로도 학습 효율을 향상시키는 기법이다 [6]. 대조 학습은 우선 긍정적 샘플과 부정적 샘플을 구분하여 각각에 유사성과 비유사성을 학습한다. 즉, 긍정적 샘플 간에 유사성을 최대화하는 것과 동시에 부정적 샘플 간에는 최소화하게 된다.

기존의 초해상도 네트워크의 학습법은 네트워크의 결과영상과 원본 영상(Ground-truth) 간의 L1 손실로 주로 학습하였고, 기존의 지식 증류 기법은 교사와 학생 네트워크의 중간 특징 맵 혹은 결과영상 간의 L1 손실로 학습하였다. 이는 교사 네트워크로부터 제대로 된 지식의 전달이 힘들었다.

이를 해결하기 위해 본 논문에서는 교사와 학생 네트워크의 상호의존정보를 극대화하는 대조 학습 기법을 제안한다. 그림 1은 제안기법의 구조이다. 학습 배치내 교사와 학생 네트워크의 결과물 중 같은 인덱스는 긍정적 샘플로 간주하고, 나머지 샘플은 부정적 샘플로 간주한다. 제안 기법인 대조 학습 손실을 수식으로 표현하면 수식 (1)과 같다.

$$L_{CL} = \sum_{i=1}^N \frac{d(SR_i^S, SR_i^T)}{\sum_k^K d(SR_i^S, SR_k^{Neg})} \quad (1)$$

여기서 N은 배치 크기를 의미하며,  $SR^T$ ,  $SR^S$ ,  $SR^{Neg}$ 는 각각 교사 네트워크, 학생 네트워크, 부정적 샘플의 초해상도 결과 영상을 의미하며, d는 샘플간의 유사도 측정법(L1, Cosine Similarity)을 의미한다. 유사도의 측정법을 L1을 사용한 경우는  $L_{CL}$ 에서 분모는 긍정적 샘플, 분자는 부정적 샘플의 손실 값을 의미하며, 각각 최소화/최대화함으로써  $L_{CL}$ 을 최소화하는 방향으로 학습한다. 반면에 코사인 유사도인 경우는 값이 클수록 유사성이 커지기 때문에 수식 (1)에  $-\log$  값을 취하여 손실 값을 최소화하는 방향으로 학습한다.

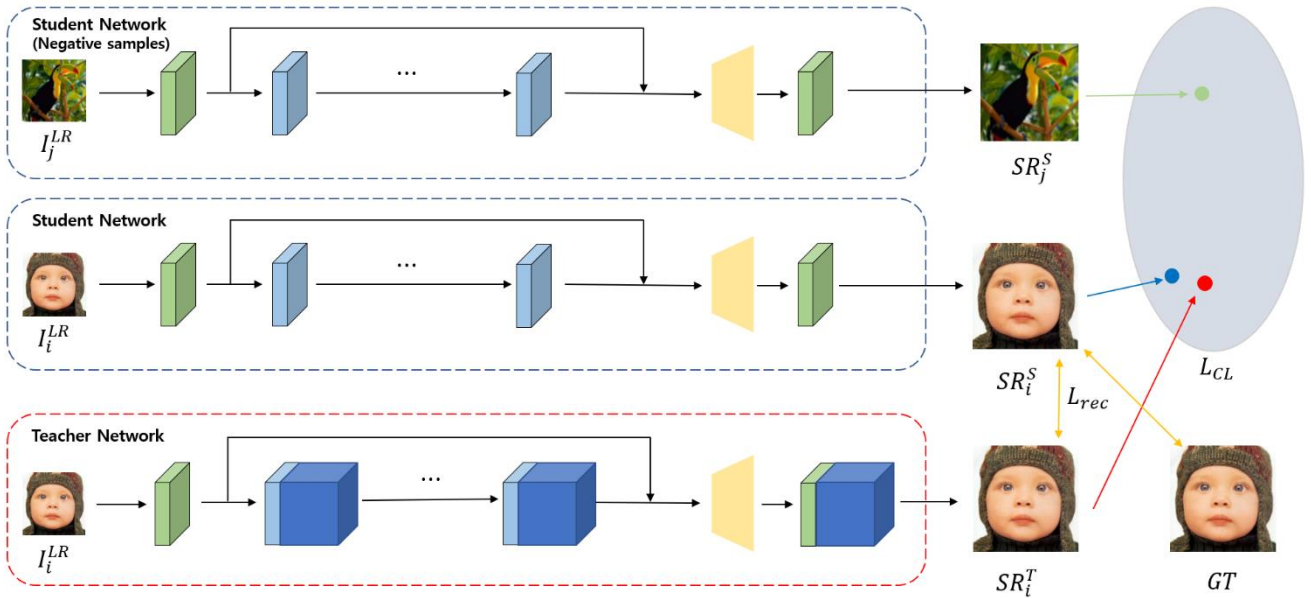


그림 1. EDSR 모델에서의 제안 기법 구조도

결국, 제안 기법의 최종 손실함수는 수식 (2)와 같다.

$$L_{total} = L_{rec} + \alpha L_{CL} \quad (1)$$

여기서  $L_{rec}$  는 기존 초해상도 학습 법인 원본 영상과의 L1 손실을 의미하며,  $\alpha$  는 두 손실 간의 균형에 필요한 하이퍼-파라미터(hyper-parameter)이다. 본 논문에서는  $\alpha$  값을 10 으로 설정하였다.

### 3. 실험결과

본 논문의 실험을 위해 교사 및 학생 네트워크 모두 EDSR 구조로 설정하였다. 또한 공정한 비교를 위해 학습 데이터는 기존 기법과 제안 기법 모두 DIV2K [7]의 800 개의 이미지를 사용하였으며, 정량적 성능 비교를 위해 Benchmark 데이터셋인 Set 5, Set1, B100, Urban100 을 테스트 데이터 셋으로 사용하였다. 표 1 은 EDSR 내 학생과 교사 네트워크 모델 구성을 나타낸 것이다. 학생 네트워크는 교사 네트워크 대비 resblock 수와 채널수가 각각 2, 4 배 적으며, 전체 파라미터 수는 약 27.1배 차이난다.

표 2 는 제안된 기법의 실험 결과를 나타내었다. 실험 결과에서 알 수 있듯이, 제안 기법 모두 기존 지식 증류 기법 대비 전체 데이터 셋에서 PSNR 성능 향상을 보였다. 특히 texture 가 많은 Urban100 데이터 셋에서는 지식 증류를 적용하지 않는 학생 네트워크와 기존 기법 대비 각각 약 0.3, 0.15 dB 향상되었다. 또한, 제안 기법에서 유사도 측정법을 교사인 유사도와 L1 을 각각 실험하여 비교하였는데, 이 경우 전체적으로 L1 을 유사도 측정법을 사용한 경우가 성능이 더 좋았다.

표 1. Teacher 와 Student 모델 구성 (EDSR)

	# resblock	#channel	#parameters
Teacher (T)	32	256	40.7M
Student (S)	16	64	1.5M

표 2. 제안된 기법의 실험 결과 (PSNR, Scale 2)

	Set5	Set14	B100	Urban100
Student (no distill)	37.919	33.439	32.102	31.728
FAKD [3]	37.976	33.523	32.156	31.906
LSFD [4]	37.991	33.529	32.158	31.896
Ours (CS)	37.992	33.548	32.154	31.936
Ours (L1)	37.983	33.558	32.159	32.020

### 4. 결론

본 논문에서는 초해상도 모델 경량화를 위한 대조 학습 기반 지식 증류 기법을 제안하였다. 실험 결과 제안기법이 기존 기법대비 모든 데이터셋에서 성능 향상이 있음을 확인하였고, 특히 Urban100 데이터셋에서 0.15dB 향상되었다.

### 감사의 글

본 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-00802, 속성을 유지하는 지능적 미디어 화면비 변환 기술개발, 100%)

### 참고 문헌

[1] K. Hung, Z. Zhang, and J. Jiang, "Real-time Image Super-Resolution Using Recursive Depthwise Separable Convolution Network," IEEE Access, vol. 7, pp. 99804-99816, Jul. 2019..  
 [2] M. Lin, et al, "Filter ketch for Network Pruning," IEEE Transactions on Neural Networks and Learning Systems, Jun. 2021.

- [3] Z. He, et al., “Fakd: Feature-Affinity Based Knowledge Distillation for Efficient Image Super-Resolution,” IEEE International Conference on Image Processing (ICIP), 2020, pp.518-522.
- [4] S. Park, N. Kwak, “Local-Selective-Feature Distillation for Single Image Super-Resolution,” arXiv. Nov. 2021.
- [5] B. Lim, K. Lee, “ Enhanced Deep Residual Networks for Single Image Super-Resolution,” Proceedings of the IEEE conference on CVPR workshops, pp. 136-144, 2017.
- [6] Y. Tian, D. Krishnan, and P. Isola, “Contrastive Representation,” Proceedings of ICLR, 2020.
- [7] Agustsson, et al, “ NITRE 2017 challenge on SISR: Dataset and study,” IEEE Conferecne on CVPR Workshops, Vol. 3, 2017.