

일반논문 (Regular Paper)

방송공학회논문지 제28권 제6호, 2023년 11월 (JBE Vol.28, No.6, November 2023)

<https://doi.org/10.5909/JBE.2023.28.6.709>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

# 대조 학습 기반 자가증류를 이용한 단일 영상 초해상화 모델 경량화 기법

문현철<sup>a)b)</sup>, 김재곤<sup>b)</sup>, 정진우<sup>a)</sup>, 김성제<sup>a)\*</sup>

## Contrastive Learning based Self-Distillation for Lightweight Single Image Super-Resolution Network

Hyeon-Cheol Moon<sup>a)b)</sup>, Jae-Gon Kim<sup>b)</sup>, Jinwoo Jeong<sup>a)</sup>, and Sungjei Kim<sup>a)\*</sup>

### 요약

CNN(Convolutional Neural Network) 기반의 단일 영상 초해상화 모델은 성능 향상을 위해 수많은 가중치 파라미터와 높은 계산 비용이 필요하기 때문에 모바일과 같은 저전력 기기에서 추론하기에는 제한적이다. 따라서, 다양한 딥러닝 모델 경량화 기법들이 연구되고 있으며, 그 중에 교사 모델에서 학생 모델로 지식을 전이하는 지식증류(Knowledge Distillation) 기법이 연구되고 있다. 그러나, 이러한 지식증류 기법은 적절한 교사 모델을 탐색하고 학습하는데 많은 비용이 발생한다. 본 논문에서는 교사 모델 없이 초해상화 모델을 효율적으로 학습하기 위한 대조 손실 기반 자가증류(self-distillation) 기법인 CLSD(Contrastive Learning based Self-Distillation)을 제안한다. 제안기법은 학생 모델 내 중간 특징 맵에 보조 작업 모델을 추가로 연결하고, 각 보조 모델의 출력 영상에 대해 대조 손실로 학습함으로써 기존 학생 모델의 전체 특징 맵을 작업 편향적(task-biased)으로 만들어 성능을 개선하였다. 실험결과 초해상화 모델에서 제안기법이 지식증류 및 기존 자가증류 모델 대비 성능이 향상됨을 확인하였다.

### Abstract

Convolutional neural network-based single image super-resolution (SISR) involves numerous parameters and high computational costs to ensure improved performance, limiting its applicability in resource-constrained devices such as mobile devices. Knowledge distillation (KD), which transfers knowledge from a teacher network to a student network, has been investigated as a method to make efficient and lightweight neural network models. However, KD requires the high cost to find appropriate teacher models. In this paper, we propose a contrastive learning based self-distillation (CLSD) for lightweight single image super-resolution model without teacher model. By constructing of an auxiliary task model within the intermediate feature map of the student model and the training of each auxiliary model using contrastive loss, the proposed method improved performance by directing the overall feature map of the existing student model toward task-specific biases. Experimental results show that the proposed method improves the task performance compared to KD and conventional self-distillation methods in the EDSR model.

Keyword : Contrastive learning, Deep supervision, Efficient super-resolution, Self-distillation, Single image super-resolution

Copyright © 2023 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

## 1. 서론

단일 영상 초해상화 기법(SISR: Single Image Super-Resolution)은 SRCNN(Super-Resolution using Convolutional Neural Network) 등장을 시작으로 CNN(Convolutional Neural Network)을 기반으로 하는 SISR 모델 기법이 기존의 영상처리 기반의 초해상화 기법보다 성능이 뛰어난 것으로 보고되었다<sup>[1]</sup>. 그러나 딥러닝 모델의 성능 향상을 위해 파라미터의 수와 추론 시 발생하는 계산 복잡도가 크게 증가하였다. 따라서 모바일 기기 같은 저전력 기기에 모델이 추론 및 응용되기에는 많은 제약이 따르며, 이러한 요구를 충족하기 위한 딥러닝 경량화 기법들이 연구되고 있다<sup>[2]</sup>. 딥러닝 경량화 기법은 일반적으로 모델 자체의 연산량 및 성능을 개선하는 네트워크 관점에서의 경량화 기법과 실제 추론 기기에서의 속도 및 성능을 개선 시키는 정수 기반 양자화(quantization) 같은 하드웨어 관점에서의 경량화 기법으로 나눌 수 있다. 그중에 네트워크 관점에서의 경량화 기법은 딥러닝 모델 구조를 효율적으로 설계하는 기법<sup>[3]</sup>, 가중치 혹은 채널 단위로 모델의 성능에 미치는 영향도를 측정하여 중요하지 않은 요소들을 없애는 가지치기(pruning)<sup>[4]</sup>, 주어진 가중치 행렬을 분해하는 행렬분해(matrix decomposition)<sup>[5]</sup>, 그리고 성능이 더 좋고 규모가 큰 교사 모델로부터 규모가 작은 학생 모델로 지식을 전이하여 성능을 개선하는 지식증류(KD: Knowledge Distillation)가 있다<sup>[6]</sup>. 이 중에서 KD 기법은 실제 응용 분

야에서 사용되는 딥러닝 모델의 구조를 수정하지 않고도 작업(task) 성능을 개선할 뿐만 아니라, 가지치기와 행렬분해 같은 다른 경량화 기법과 결합이 용이하다는 장점이 있다.

초해상화 분야에서 KD 기법은 주로 교사와 학생 모델의 중간 특징 맵(feature map)을 비슷하게 만드는 특징 증류(FD: Feature Distillation) 방식으로 연구되었다. 즉, FD는 교사와 학생 모델의 중간 특징 맵을 깊은 회귀자(DR: Deep Regressor) 혹은 별도의 공간적 유사(SA: Spatial Affinity) 행렬을 거리 기반 L1 손실로 유사하게 만들어 학생 모델의 성능을 개선했다<sup>[8-10]</sup>. 그러나 이러한 방식의 FD는 기존 SISR 모델 내 L1 손실이 가지는 이미지 흐림(image blurring) 및 패딩 손실 같은 문제를 완벽하게 해결하지 못했다. 이러한 문제를 해결하기 위해서 긍정적 샘플과 부정적 샘플들을 구분하여 각각에 유사성과 비유사성을 학습함으로써 긍정적 샘플 간의 상호의존정보(mutual information)를 최대화하는 대조 손실(contrastive loss) 기반의 FD 기법들이 연구되었다<sup>[11-12]</sup>.

그러나, 이러한 KD 기법은 교사 모델이 반드시 필요하기 때문에 최적의 교사 모델의 구조를 탐색하기 위해 상당한 노력과 학습 비용이 발생한다. 따라서, 이러한 문제점을 해결하기 위해서 교사 모델을 활용하지 않고 내부 구조를 효율적으로 활용하여 학습 비용을 줄이면서 성능을 향상시키는 자가증류(SD: Self-Distillation) 기법들이 연구되었다<sup>[13-14]</sup>. 기존 SD 기법은 학생 모델 내 중간 특징 맵에 보조 작업 모델(auxiliary task network)을 구성하고, 각 보조 모델에 작업 손실을 최적화함으로써 기존 학생 모델의 특징 맵을 작업 편향적(task-biased)으로 만들어 성능을 개선하였다<sup>[13]</sup>. 그러나, 이러한 SD 기법들은 보조 모델의 손실 함수를 거리 기반 손실만을 사용하였기 때문에 마찬가지로 기존 L1 손실이 가지는 문제점을 해결하지 못했다.

본 논문에서는 대조 손실 기반의 SD 기법을 이용한 단일 영상 초해상화 모델 경량화 기법인 CLSD(Contrastive Learning based Self-Distillation)을 제안한다. 제안기법은 기존 기법 대비 다음과 같은 장점이 있다. 우선, 기존 KD 기법 대비 교사 모델을 따로 탐색하지 않기 때문에 학습 비용을 줄일 수 있다. 또한, 기존 학생 모델과 보조 작업 모델 간의 상호의존정보를 최대화 할 수 있는 대조 손실

a) 한국전자기술연구원(Korea Electronics Technology Institute, KETI)

b) 한국항공대학교 항공전자정보공학과(Korea Aerospace University, Department of Electronics and Information Engineering)

‡ Corresponding Author : 김성제(Sungjei Kim)

E-mail: sungjei.kim@keti.re.kr

Tel: +82-31-739-7477

ORCID: <https://orcid.org/0009-0008-6466-649X>

※이 논문의 연구 결과 중 일부는 한국방송·미디어공학회 2023년 하계 학술대회에서 발표한 바 있음.

※ 본 연구 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2021-0-00802, 속성을 유지하는 지능적 미디어 화면비 변환 기술 개발, 100%)

※ This work was supported by an Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean government [Ministry of Science and ICT (MSIT)] (Development of Intelligent Media Aspect Ratio Conversion Technology That Maintains Properties) under Grant 2021-0-00802.

· Manuscript September 1, 2023 ; Revised October 28, 2023; Accepted October 28, 2023.

<sup>[19]</sup>을 적용함으로써 보조 작업 모델의 지식 정보를 기존 거리 기반 L1 손실 보다 더 효율적으로 전달하여 기존 기법 대비 모델의 성능을 개선하였다. 제안기법의 성능 검증을 위해서 초해상화 CNN 모델인 EDSR(Enhanced Deep Residual Network) 학생 모 <sup>[15]</sup>에 각각의 KD 및 SD를 적용하여 PSNR(Peak Signal-to-Noise Ratio) 성능과 학습 비용을 비교했다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 KD 및 SD 기법에 대하여 설명한다. 3장에서는 제안기법을 설명하고 4장에서는 상세 실험 내용과 결과에 대해 분석하며, 마지막으로 5장에서 결론을 맺는다.

## II. 관련 연구

### 1. 지식증류

KD 기법은 보다 성능이 좋으면서 복잡도가 더 높은 교사 모델의 지식을 더 간단한 구조를 가지고 있는 학생 모델로 전이함으로써 학생 모델의 작업 성능을 높이는 기법이다. 분류(image classification)에서의 KD는 주로 교사 모델의 soft label과 학생 모델의 label 사이의 거리를 최소화하는 방향으로 시작하였다<sup>[16]</sup>. 교사 모델의 더 많은 지식을 흡수하기 위해 중간 특징 맵 간의 사이를 최소화하는 FD 기법들을 결합된 형태로 연구되었다<sup>[11]</sup>.

일반적인 FD는 교사와 학생의 중간 특징 맵 간의 채널 수가 상이하기 때문에 우선 채널 수를 맞추는 방향으로 연구되었다. 먼저, FitNet은 간단한 1x1 합성곱 계층으로 구성된 DR을 이용하여 교사와 학생 중간 특징 맵의 채널 수를 맞추고 난 뒤 L1 손실로 이 둘 간의 거리를 최소화하였다<sup>[8]</sup>. FAKD(Feature Affinity-based Knowledge Distillation)에서는 특징 내 상관 관계를 나타낼 수 있는 SA 행렬을 사용하여 FD를 수행하였다<sup>[9]</sup>. 더 나아가 LSFD(Local-Selective Feature Distillation)에서는 기존 DR을 사용하여 L1 손실로 FD하는 방식에서 교사와 학생 모델의 결과 이미지 간의 차이 맵(difference map)을 활용하여 특정 위치에 선택적으로 강조하여 정제된 특징을 증류하는 방식으로 기존 FD 기법을 개선하였다<sup>[10]</sup>. 이외에도 PISR(Privilege Information for

efficient Super Resolution)에서는 autoencoder 구조를 활용하여 입력 이미지의 privileged 정보를 증류하였다<sup>[17]</sup>. 그러나, 이러한 L1 손실 기반의 FD는 기존 L1 손실이 가지는 패턴 손실 및 이미지 흐림 현상, 그리고 magnitude가 큰 영역에 대해서만 민감하게 고려하는 점들을 완벽하게 해결하지 못했다. 또한, 일반적으로 교사와 학생의 계층의 깊이가 다르기 때문에 각각의 모델에서 적절히 매칭해야 할 특징 맵을 탐색하는데 비용이 더 발생한다.

### 2. 대조 학습

대조 학습은 자가-지도 학습(self-supervised learning) 기법 중 하나로 주어진 데이터에 대해 긍정적 샘플과 부정적 샘플을 구분하여 각각에 유사성과 비유사성을 학습함으로써 데이터의 특성을 잘 타내는 특징 맵을 만드는 것을 유도한다<sup>[19]</sup>. 즉, 대조 학습은 긍정적 샘플과 부정적 샘플 간의 Kullback-Leibler(KL) 발산을 최대화함으로써 긍정적 샘플 간의 상호의존정보를 최대화하는 동시에 두 샘플의 분포가 명확하게 구별한다. 이를 통해 자가 지도 학습 시나리오에서의 대조 학습은 특징들을 잘 군집화함으로써 큰 성능 향상을 유도했다.

기존의 L1 손실 기반의 FD는 학생 모델 관점에서 교사 모델의 지식을 상대적으로 적절히 전달하지 못하였다. 따라서, 이러한 문제를 해결하기 위해 대조 손실 기반의 KD 기법 등이 연구됐다<sup>[20,26]</sup>. 대조 손실 기반의 KD 기법은 교사 및 학생 모델의 특징 및 결과 영상 영역에서 긍정 및 부정적 샘플을 구성한 뒤, 이를 대조 학습으로 증류하여 교사와 학생 모델 간의 상호의존정보를 최대화하여 성능 향상되었다. 즉, L1 손실 기반의 FD는 오직 교사 모델의 지식에 대한 상한선(upper bound)만을 제공하지만, 대조 손실 기반의 FD는 부정적 샘플이 주는 하한선(lower bound)를 같이 제공함으로써 보다 더 많은 교사 모델의 지식을 전달하고 흡수함으로써 성능을 향상시켰다.

### 3. 자가증류

기존의 지식증류 기법은 우선 교사 모델이 필요하기 때문에 교사 모델의 학습 및 탐색에 상당히 많은 비용이 발생

한다. 더불어 학생 모델의 관점에서 성능이 좋은 교사 모델이라도 지식을 완전히 흡수하지 못한다는 문제점이 발생할 수 있고, 이는 증류 손실 관점에서 학습에 대한 안정성이 떨어진다<sup>[14]</sup>. 이러한 문제를 해결하기 위해 별도의 교사 네트워크 없이 학생 모델이 스스로 지식을 전이하는 SD 기법이 연구되었다<sup>[18]</sup>.

SD 기법은 일반적으로 학생 모델 내 보조 작업 모델을 추가로 구성하여 학습하는 DS(Deep Supervision) 기법 등으로 연구되어지고 있다<sup>[14]</sup>. 기존의 CNN은 일반적으로 연쇄 구조(cascading architecture)를 가지고 있기 때문에 모델의 상단 부분이 작업 무관한 특징(task-irrelevant feature)을 가지고 있다. 이를 해결하기 위해 DS 기법은 중간 특징 맵 구간에 얕은(shallow) 보조 작업 모델을 추가로 구성하고 이를 학습하게 된다. 이 때, 상대적으로 학생 모델의 상단 부분이 보조 작업 모델 관점에서 상대적으로 하단에 위치하게 된다. 결국 DS 기법은 이러한 학습을 통해 학생 모델의 상단 부분의 특징 맵을 작업 편향적으로 유도하여 이미지 분류 분야에서 성능을 크게 향상시켰다. 따라서 기존 SD 기법은 보조 및 학생 모델의 협력을 통해 학생 모델의 전반적인 특징 맵을 작업 편향적으로 만들어 성능 향상을 유도했다.

### III. 제안기법

#### 1. 기존 SD 기법

기존 SD 기법 중에 하나인 DS 기법은 일반적으로 사용하는 손실 함수는 크게 3가지 방식으로 나눌 수 있다. 기존 학생 모델의 출력과 정답(ground-truth)을 비교하는 기존 작업 모델 손실함수, 보조 작업 모델의 출력과 정답을 비교하는 보조 작업 손실함수, 그리고 마지막으로 기존 작업 모델과 보조 작업 모델 간의 지식증류 손실로 구성된다<sup>[14]</sup>. 이 때, 자기증류 내 지식증류 손실은 모델의 출력 및 중간 특징 영역 모두에서 적용될 수 있다. 또한 보조 작업 모델은 실제 추론 시에는 작동이 되지 않기 때문에 기존 학생 모델 대비 계산 복잡도는 동일하다.

기존의 초해상화 모델 및 심층 감독 기반의 SD 기법의

손실 함수를 수식으로 표현하면 수식은 다음과 같다.

$$L_{SR} = \frac{1}{2N} \sum_{i=1}^N \|SR_i - HR_i\|_1 \quad (1)$$

$$L_{ASR} = \frac{1}{2N} \sum_{i=1}^N \frac{1}{T} \sum_{j=1}^T \|SR_i^j - HR_i\|_1 \quad (2)$$

$$L_{SD} = \frac{1}{2N} \sum_{i=1}^N \frac{1}{T} \sum_{j=1}^T \|SR_i^j - SR_i\|_1 \quad (3)$$

여기서  $T$ 는 보조 네트워크의 수를 의미한다.  $N$ ,  $SR_i$ ,  $SR_i^j$ 은 각각 배치 크기, 기존 모델의 결과 초해상화 영상, 그리고  $j$ 번째 보조 작업 모델의 결과 초해상화 영상을 의미한다. 본 논문에서는 기존 지식증류 기법과 공정한 비교를 위해 보조 작업 네트워크의 수를 3으로 설정하였다. 즉, 보조 작업 모델은 거리 기반 초해상화 손실 함수  $L_{ASR}$ 와 학생 모델 결과 영상으로부터 증류 손실  $L_{SD}$ 를 결합한 형태로 학습을 수행한다. 심층 기법의 최종 손실 함수는 다음과 같다.

$$L_{DS} = L_{SR} + L_{ASR} + \alpha L_{SD} \quad (4)$$

여기서  $\alpha$ 는 두 손실 간의 균형에 필요한 하이퍼-파라미터(hyper-parameter)이며 본 논문에서는 10으로 값을 설정하였다.

#### 2. CLSD

그림 1은 EDSR 모델에서의 제안기법인 CLSD의 구조도를 나타내었다. 제안기법도 기존 심층 감독 기법과 마찬가지로 보조 작업 모델을 KD 관점에서의 학생 모델로 간주하고, 이를 제외한 학생 모델의 최종 결과 영상을 교사로 간주하여 지식을 전이한다. 본 논문에서는 기존 심층 감독 기법 대비 보조 작업 모델과 학생 모델 간의 상호의존정보를 극대화하기 위해 기존의 심층 기법에 대조 손실 기반의 증류 손실을 추가하는 방식을 제안한다.

대조 학습의 긍정 및 부정적 샘플을 구성하는 방법은 크게 학습 배치내 이미지를 활용하는 방법과 외부 이미지를 활용하는 방법으로 나뉘지만, 일반적으로 학습의 안정성을

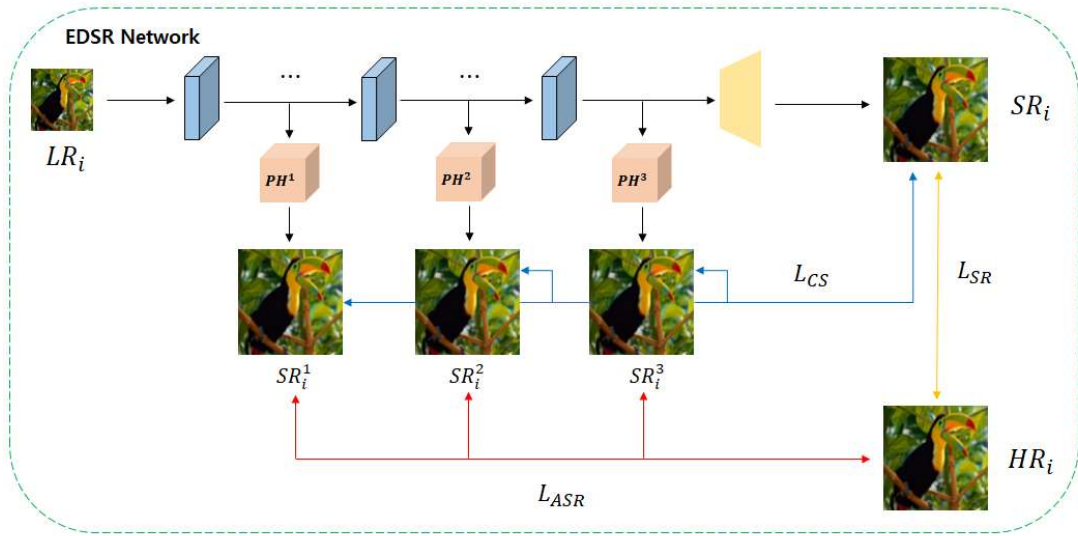


그림 1. 제안기법인 대조 학습 기반 자가증류 기법(CLSD)의 구조도  
 Fig. 1. Overall architecture of contrastive learning based self-distillation(CLSD)

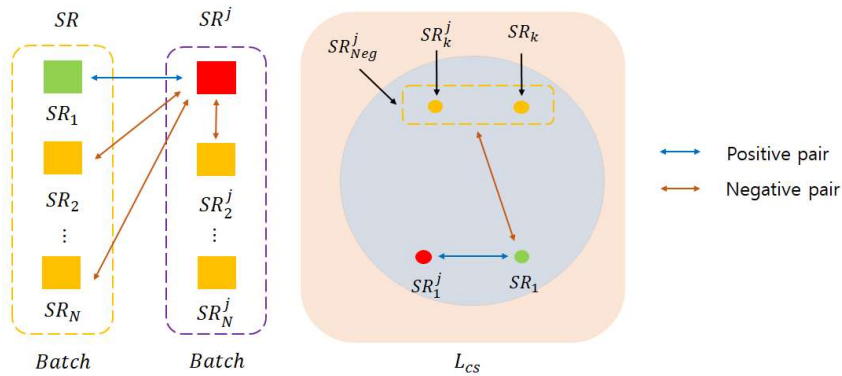


그림 2. CLSD의 대조 학습 예시  
 Fig. 2. An example of the contrastive learning in CLSD

위해 학습 배치내 인덱스(index)가 다른 이미지를 부정적 샘플로 간주한다. 초해상화 모델에서의 대조 학습 기반 지식증류는 대부분 출력 영상 영역에서 수행되어 기존 L1 손실 기반의 FD 방법보다 성능 향상을 이끌어 냈다<sup>[20,26]</sup>. 그림 2는 제안기법의 대조 학습의 예시를 나타내었다. 우선 제안기법은 대조 학습의 안정성을 위해 부정적 샘플의 구성을 외부 이미지를 사용하지 않고 주어진 학습 배치  $N$ 개의 입력 이미지를 사용하였다. 학습 배치 내 보조 작업 모델의 결과 초해상화 영상과 보조 작업 모델의 결과 초해상화 영상을 기준으로 배치 내 같은 인덱스를 가지는 영상을 긍정

적 샘플로 간주하고, 인덱스가 다른 나머지 샘플을 부정적 샘플로 간주한다. 또한, 더 많은 부정적 샘플의 구성을 위해 동일 보조 작업 모델 내 결과 영상 중 다른 인덱스를 가지는 영상도 추가하였다. 즉, 부정적 샘플은 각 모델에서 출력되는 결과 영상에서 인덱스가 다른 영상간의 쌍과 각 모델 내 초해상화 영상에서 인덱스가 다른 영상간의 쌍으로 구성될 수 있다. 제안기법의 대조 학습은 학습 배치  $N$ 개를 입력받았다고 가정할 때,  $N$ 개의 긍정적 샘플,  $2N(N-1)$ 개의 부정적 샘플을 구성할 수 있다. 제안된 CLSD를 수식으로 표현하면 다음과 같다.

$$L_{CS}^j = \frac{\sum_{i=1}^N d(SR_i, SR_i^j)}{\sum_k^K d(SR_i, SR_k^j)} \quad (5)$$

$$L_{CLSD} = L_{DS} + \beta \sum_{j=1}^T L_{CS}^j \quad (6)$$

여기서  $d$ 는 대조 학습을 위한 샘플 간의 유사도 측정을 할수 있는 함수를 의미한다. 일반적으로 거리 기반 L1 손실 혹은 코사인 유사도(cosine similarity)를 사용할 수 있으며, 본 논문에서는 L1 손실을 사용하였다.  $L_{CS}^j$ 는 j번째 보조 모델에서 계산되는 대조 손실 값을 의미하며, 최소화하기 위해 분모인 부정적 샘플값은 최대화되는 반면에 분자인 긍정적 샘플 간의 거리는 최소화가 된다. 또한,  $\beta$ 는 두 손실 간의 균형에 필요한 하이퍼-파라미터(hyper-parameter)이며 본 논문에서는 10으로 값을 설정하였다.

#### IV. 실험결과

##### 1. 실험환경

본 논문의 실험을 위한 학습 데이터 셋으로 DIV2K<sup>[21]</sup>을 사용하고 테스트 데이터 셋으로 Set5<sup>[22]</sup>, Set14<sup>[23]</sup>, BSD100<sup>[24]</sup>,

표 1. EDSR 모델 내 교사 및 학생 모델 구성

Table 1. Configuration of teacher and student in the EDSR network

Network	#Resblocks	#Channels	#Parameters
Teacher (T)	32	256	40.7M
Student (S)	16	64	1.5M

Urban100<sup>[25]</sup>을 사용하였다. 제안기법을 기존의 SD 기법 뿐만 아니라 KD 기법과도 비교하기 때문에 EDSR에서 교사 및 학생 모델의 구성을 표 1로 나타냈다. 제안기법을 포함한 SD 기법은 별도의 교사 모델 없이 EDSR 학생 모델과 보조 작업 모델만을 사용하였다. 또한, 보조 작업 모델인 PH(Projection Heads)의 구성을 기존 EDSR 네트워크에서 사용되는 ResBlocks을 각 보조 작업 모델로 5개씩 사용하였으며, 채널 수는 기존 학생 모델과 동일하게 64로 설정하였다. 학습을 위한 하이퍼-파라미터는 학습율(learning rate), 배치 크기를 각각  $2e-4$ , 16으로 설정하였고, 최적화 함수를 Adam으로 설정하였다. 또한, 제안된 CLSD의 손실 값의 포화(saturation)이 늦게 진행되어 epoch를 기존 지식증류 기법에서 사용하는 300보다 2배 많은 600으로 설정하였다. 따라서, 공정한 실험 비교를 위해 Scratch로 학습한 Student baseline, FAKD, LFD, 그리고 작업 손실 기반의 심층 감독 기법인 DS에도 같은 하이퍼-파라미터를 적용하였으며, DS와 제안된 CLSD의 PH의 구성은 동일하다.

표 2. EDSR 모델에서의 제안기법 실험결과(Scale 2)

Table 2. Experimental results of proposed methods on the EDSR Network(Scale 2)

Methods	Distillation type	Scale	Set5 (PSNR)	Set14 (PSNR)	B100 (PSNR)	Urban100 (PSNR)
Baseline	-	x2	37.919	33.439	32.102	31.728
FAKD <sup>[9]</sup>	KD		37.976	33.523	32.156	31.906
LFD <sup>[10]</sup>			<b>37.986</b>	33.528	<b>32.159</b>	31.935
DS	SD		37.957	<b>33.536</b>	32.150	<b>31.968</b>
CLSD (ours)			<b>37.977</b>	<b>33.602</b>	<b>32.174</b>	<b>32.028</b>

표 3. EDSR 모델에서의 제안기법 실험결과(Scale 3, PSNR)

Table 3. Experimental results of proposed methods on the EDSR Network(Scale 3)

Methods	Distillation type	Scale	Set5 (PSNR)	Set14 (PSNR)	B100 (PSNR)	Urban100 (PSNR)
Baseline	-	x3	34.272	30.266	29.044	27.959
FAKD <sup>[9]</sup>	KD		<b>34.356</b>	30.296	29.066	28.016
LFD <sup>[10]</sup>			34.333	<b>30.301</b>	<b>29.077</b>	28.029
DS	SD		34.330	30.299	29.068	<b>28.033</b>
CLSD (ours)			<b>34.336</b>	<b>30.307</b>	<b>29.078</b>	<b>28.078</b>

표 4. EDSR 모델에서의 제안기법 실험결과(Scale 4, PSNR)

Table 4. Experimental results of proposed methods on the EDSR Network(Scale 4)

Methods	Distillation type	Scale	Set5 (PSNR)	Set14 (PSNR)	B100 (PSNR)	Urban100 (PSNR)
Baseline	-	x4	32.102	28.526	27.538	25.905
FAKD <sup>[9]</sup>	KD		32.138	28.547	27.557	25.972
LFD <sup>[10]</sup>			32.107	28.524	27.552	25.962
DS	SD		32.112	28.545	27.549	25.970
CLSD (ours)			32.099	28.556	27.560	25.991

표 5. 각 손실 함수 요소별 효율성에 대한 실험결과(Scale 2)

Table 5. Experimental results for the efficiency of each loss component

Loss component				B100 (PSNR)	Urban100 (PSNR)
$L_{SR}$	$L_{ASR}$	$L_{SD}$	$L_{CS}$		
✓				32.102	31.728
✓	✓			32.145	31.916
✓	✓	✓		32.150	31.968
✓	✓		✓	32.174	32.028

## 2. 실험결과

표 2-4는 제안기법의 정량적 성능(PSNR)에 대한 실험결과를 나타냈고, 각 데이터 셋 내 best 및 second 성능은 각각 빨간색, 파란색으로 표시하였다. 또한, 표 1의 Student 모델에 각 기법을 적용하였기 때문에 표기된 기법의 모델 복잡도 및 연산량은 모두 동일하다. 실험결과에서 알 수 있듯이, 상대적으로 텍스처가 적은 Set5 데이터셋을 제외하고 제안

기법이 상대적으로 큰 교사 모델로부터 지식을 전이 받은 KD 및 DS 기법 대비 성능 향상이 되었음을 확인할 수 있다. 특히 텍스처 및 패턴이 많은 Urban100 데이터셋에서 기존 기법 대비 PSNR 기준 Scale 2/3/4에서 각각 0.09/0.05/0.03dB 향상됨을 확인하였다. 이는 제안기법이 교사 모델이 없이도 기존 KD보다 전체 테스트 데이터 셋에서 평균적으로 Scale 2/3/4에서 0.05/0.03/0.02 dB 향상됨을 확인하였다. 이는 학생 모델의 특징을 교사 모델의 특징을 비

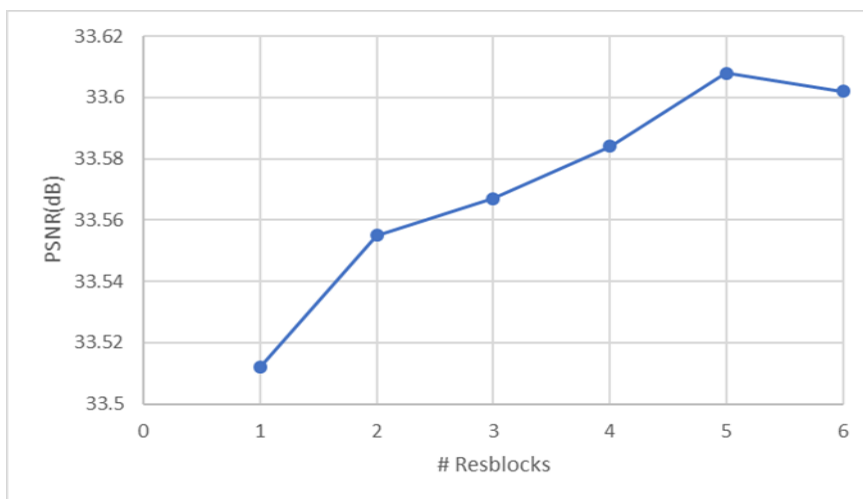


그림 3. 성능과 PH의 Resblocks 수 사이의 관계(Set14, Scale2)

Fig. 3. Relationship between performance and the number of Resblocks in the PH (Set14, Scale2)

숫해지게 유도하는 것보다 전반적인 특징 맵을 작업 편향적으로 만드는 것이 성능 향상에 더 영향을 준다는 것을 의미한다.

표 5는 각 손실 함수 요소 별 효율성에 대한 실험결과를

나타냈다. 실험결과에서 알 수 있듯이 각 손실 함수 요소들이 결합된 형태가 기존  $L_{SR}$ 만 적용할 때보다 성능 향상 효과가 있음을 확인하였고, 제안한 가 기존의 보다 PSNR 향상 효과가 약 0.02dB 됨을 확인하였다.

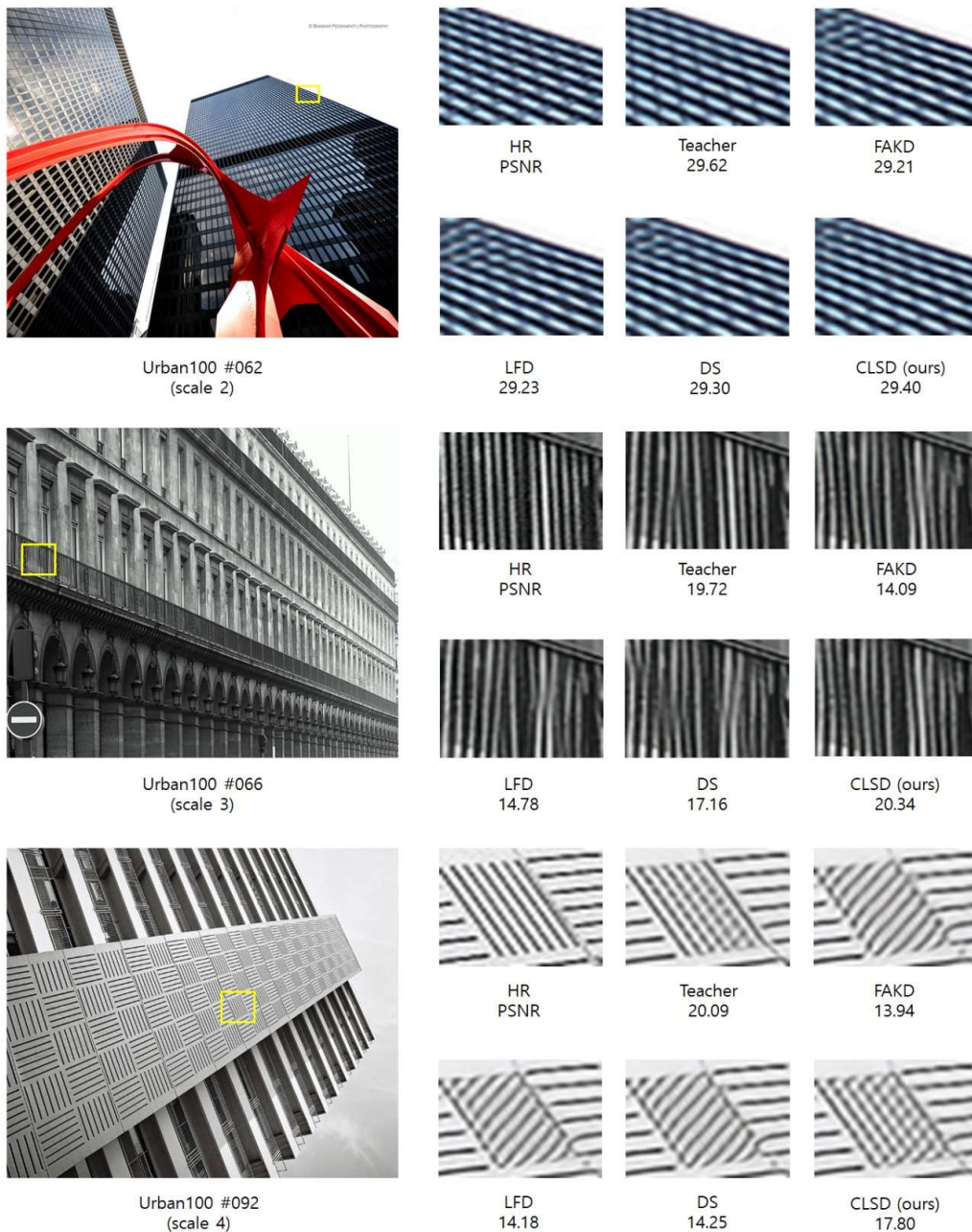


그림 4. 주관적 화질 비교  
Fig. 4. Comparison of Subjective Image Quality



그림 3은 보조 작업 모델인 PH의 Resblocks의 수에 따른 Set14 데이터셋에서 최종 학생 모델의 성능 관계도를 나타낸다. 일반적으로 PH가 복잡할 수록 성능이 증가하지만 블록수가 5개를 초과한 경우 약간의 성능 저하가 발생함을 확인하였다. 그림 4는 기존 기법과 CLSD의 Urban100 데이터 셋에 대해서 주관적 화질 비교를 나타냈다. 여기서 PSNR 값은 각 영상의 crop된 영역에 대해서만 계산하였다. 우선 62번 영상에서는 기존 지식증류 및 DS 기법은 왼쪽 위 영역의 대각 선 형태가 ‘ㄱ’자 형태로 되어 있는 반면에, CLSD 기법은 일자 형태로 복원이 제일 잘된 것을 알 수 있다. 또한, 66/92번 영상에서도 일자 형태의 선이 더 잘 복원됨을 알 수 있으며, 특히 92번 영상은 기존 기법들은 일자 형태의 선을 전혀 따라가지 못한 채 대각선 형태로 복원된 반면에 제안기법은 일부분 일자 형태로 복원됨을 볼 수 있다. 따라서, CLSD 기법이 교사 모델 없이도 기존 지식증류 기법 보다 교사 모델의 텍스처 관점에서 더 복원이 잘 됨을 알 수 있다.

표 6은 각 지식 및 자가증류 기법의 학습 비용(시간)에 대해 비교하였다. 학습 시간 측정을 위해 Intel CPU와 NVIDIA RTX Titan GPU 2대가 이용되었으며, Pytorch 1.9.0 프레임워크에 기반하여 구현되었다. 제안된 CLSD 기법은 단일 학생 모델 관점에서 기존 지식증류 기법 대비 1.46배, 교사 모델도 같이 고려할 때 2.16배 빠르게 학습이 됨을 알 수 있다. 따라서, 제안된 자가증류 기법이 기존 지식증류 기법보다 모델 성능 뿐만 아니라 학습 비용에도 이점이 있음을 확인할 수 있다.

표 6. 각 증류 기법의 학습 비용 비교  
 Table 6. Comparison of training costs for each distillation method

Methods	Training time (h)		Total training time (h)
	Teacher	Student	
FAKD	20.3	40.9	61.2
LFD	20.3	48.8	69.1
DS	-	25.7	25.7
CLSD	-	30.2	30.2

## V. 결론

본 논문에서는 더 효율적인 지식 전이를 위한 대조 손실 기반의 SD 기법을 이용한 단일 영상 초해상화 모델 경량화 기법을 제안하였다. 단일 영상 초해상화 모델 중 하나인 EDSR에 대하여 기존 지식증류 및 자가증류 기법과 제안기법의 성능을 비교하여 검증하였다. 실험결과 제안된 자가증류 기법이 기존 자가증류 기법 뿐만 아니라 교사 모델을 활용한 지식증류 기법과 비교해서도 각각 PSNR 0.09/0.05dB 성능 향상이 있음을 확인하였고, 주관적 화질 측면에서도 텍스처 복원의 효과가 더 뛰어난 것을 확인하였다.

따라서, 제안기법이 기존의 자가증류와 달리 교사 모델을 탐색하고 학습 비용 없이도 성능 향상을 얻었기 때문에, 모바일 같은 연산량이 제한된 저전력 기기에서 활용되는 초해상화 모델의 성능을 개선하는 연구에 도움이 될 것을 기대한다.

## 참고 문헌 (References)

- [1] C. Dong, C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 38, No. 2, pp. 295-307, February 2016. doi: <https://doi.org/10.1109/TPAMI.2015.2439281>
- [2] S. Han, H. Mao, and W. Dally, "Deep Compression: Compressing Deep Neural Network with Pruning, Trained Quantization and Huffman Coding," In Proceeding of 4th International Conference on Learning Representations (ICLR), San Juan, USA, 2016.
- [3] G. Gankhuyag, J. Huh, M. Kim, K. Yoon, H. Moon, S. Lee, J. Jeong, S. Kim, and Y. Choe, "Skip-Concatenated Image Super-Resolution Network for Mobile Devices," IEEE Access, Vol. 11, pp. 4972-4982, December 2022. doi: <https://doi.org/10.1109/ACCESS.2022.3232258>
- [4] Z. Hou and S. Kung, "Efficient Image Super Resolution Via Channel Discriminative Deep Neural Network Pruning," In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), Barcelona, Spain, pp.3647-3651, 2020. doi: <https://doi.org/10.1109/ICASSP40776.2020.9054019>
- [5] J. Liu, Z. Wu, L. Xiao, J. Sun, and H. Yan, "A Truncated Matrix Decomposition for Hyperspectral Image Super-Resolution," IEEE Transactions on Image Processing, Vol. 29, pp. 8028-8042, July 2020. doi: <https://doi.org/10.1109/TIP.2020.3009830>
- [6] Q. Huang, Y. Zhang, H. Hu, Y. Zhu, and Z. Zhao, "Binarizing Super-Resolution Networks by Pixel-Correlation Knowledge Distillation," In Proceedings of IEEE International Conference on

- Image Processing (ICIP), Anchorage, USA, 2021.  
doi: <https://doi.org/10.1109/ICIP42928.2021.9506517>
- [7] Y. Zhang, H. Chen, X. Chen, Y. Deng, C. Xu, and Y. Wang, "Data-Free Knowledge Distillation for Image Super-Resolution," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2021.  
doi: <https://doi.org/10.1109/CVPR46437.2021.00776>
- [8] D. Neogi, N. Das, and S. Deb, "FitNet: A deep neural network driven architecture for real time posture rectification," In Proceedings of International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Zallaq, Bahrain, 2021.  
doi: <https://doi.org/10.1109/3ICT53449.2021.9582113>
- [9] Z. He, T. Dai, Y. Jiang, and S. Xia, "Fakd: Feature-Affinity Based Knowledge Distillation for Efficient Image Super-Resolution," In Proceedings of IEEE International Conference on Image Processing (ICIP), Abu Dhabi, UAE, 2020.  
doi: <https://doi.org/10.1109/ICIP40778.2020.9190917>
- [10] S. Park and N. Kwak, "Local-selective feature distillation for single image super-resolution," arXiv:2111.10988, 2021.
- [11] L. Chen, D. Wang, Z. Gan, J. Liu, R. Henao, and L. Carin, "Wasserstein Contrastive Representation Distillation," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2021.  
doi: <https://doi.org/10.1109/CVPR46437.2021.01603>
- [12] Y. Tian, D. Krishnan, and P. Isola, "Contrastive Representation Distillation," Proceeding of 4th International Conference on Learning Representations (ICLR), Addis Ababa, Ethiopia, 2020.
- [13] X. Wnag, J. Zhu, Z. Yan, Z. Zhang, Y. Chen, and H. Li, "LaST: Label-Free Self-Distillation Contrastive Learning With Transformer Architecture for Remote Sensing Image Scene Classification," IEEE Geoscience and Remote Sensing Letters, Vol. 19, pp.1-5, 2022.  
doi: <https://doi.org/10.1109/LGRS.2022.3185088>
- [14] L. Zhang, J. Song, A. Gao, J. Chen, C. Bao, and K. Man, "Be Your Own Teacher: Improve the Performance of Convolutional Neural Networks via Self Distillation," In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, pp.3713-3722, 2019.  
doi: <https://doi.org/10.48550/arXiv.1905.08094>
- [15] B. Lim, S. Son, H. Kim, S. Nah, and K. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Honolulu, Hawaii, pp.136-144, 2017.
- [16] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," Arxiv: 1503.02531, 2015.
- [17] W. Lee, J. Lee, D. Kim, and B. Ham, "Learning with Privileged Information for EFFICIENT Image Super-Resolution," In Proceedings of European Conference on Computer Vision (ECCV), Online, 2020.
- [18] L. Zhang, C. Bao, and K. Ma, "Self-Distillation: Towards Efficient and Compact Neural Networks," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 44. No. 8, pp. 4388-4403, August 2022.  
doi: <https://doi.org/10.1109/TPAMI.2021.3067100>
- [19] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," In Proceedings of the 37th International Conference on Machine Learning (PMLR), Online, pp.1597-1607, 2020.
- [20] H. Moon, J. Jeong, and S. Kim, "Feature-domain Adaptive Contrastive Distillation for Efficient Single Image Super-Resolution," arXiv: 2211.15951, 2023.
- [21] R. Timofe, E. Agustsson, L. Gool, M. Yang, and L. Zhang, "NTRIE 2017 challenge on single image super-resolution: Methods and results," In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) workshops, Honolulu, Hawaii, pp.114-125, 2017.
- [22] M. Bevilacqua, A. Roumy, C. Guillemot, and M. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," In Proceedings of the 23rd British Machine Vision Conference (BMVC), Surrey, UK, pp. 135.1-10, 2012.  
doi: <https://doi.org/10.5244/C.26.135>
- [23] [Availiabel at Online] [https://link.springer.com/chapter/10.1007/978-3-642-27413-8\\_47](https://link.springer.com/chapter/10.1007/978-3-642-27413-8_47)
- [24] 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," In Proceedings of the IEEE International Conference on Computer Vision, Vancouver, Canada, pp.416-423, 2001.  
doi: <https://doi.org/10.1109/ICCV.2001.937655>
- [25] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA, pp.5197-5206, 2015.
- [26] Y. Wang, S. Lin, Y. Qu, H. Wu, Z. Zhang, Y. Xie, and A. Yao, "Towards Compact Single Image Super-Resolution via Contrastive Self-distillation," In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence(ICJAI-21), Montreal, Canada, pp. 1122-1128, 2021.

---

저 자 소 개

---



**문 현 철**

- 2018년 2월 : 한국항공대학교 항공전자정보공학과 학사
- 2020년 8월 : 한국항공대학교 항공전자정보공학과 석사
- 2021년 9월 ~ 현재: 한국전자기술연구원 연구원
- 2023년 9월 ~ 현재: 한국항공대학교 항공전자정보공학과 박사과정
- ORCID : <https://orcid.org/0000-0002-1672-2345>
- 주관심분야 : 인공지능 기반 미디어 신호 처리, 인공지능 경량화



**김 재 곤**

- 1990년 2월 : 경북대학교 전자공학과 학사
- 1992년 2월 : KAIST 전기 및 전자공학과 석사
- 2005년 2월 : KAIST 전기 및 전자공학과 박사
- 1992년 3월 ~ 2007년 2월 : 한국전자통신연구원(ETRI) 선임연구원/팀장
- 2001년 9월 ~ 2002년 7월 : Columbia University 연구원
- 2015년 12월 ~ 2016년 1월 : UC San Diego, Visiting Scholar
- 2007년 9월 ~ 현재 : 한국항공대학교 항공전자정보공학부 교수
- ORCID : <http://orcid.org/0000-0003-3686-4786>
- 주관심분야 : 비디오 부호화 표준, 비디오 신호처리, Immersive Video, Deep Learning



**정 진 우**

- 2004년 2월 : 연세대학교 전기전자공학과 학사
- 2006년 8월 : 연세대학교 전기전자공학과 석사
- 2011년 8월 : 연세대학교 전기전자공학과 박사
- 2011년 9월 ~ 2015년 4월 : 삼성전자 VD사업부 책임연구원
- 2016년 1월 ~ 현재 : 한국전자기술연구원 책임연구원
- ORCID : <https://orcid.org/0000-0003-0528-8755>
- 주관심분야 : 인공지능 기반 미디어신호 처리, 수증 영상 처리



**김 성 제**

- 2004년 2월 : 연세대학교 전기전자공학과 학사
- 2006년 8월 : 연세대학교 전기전자공학과 석사
- 2011년 8월 : 연세대학교 전기전자공학과 박사
- 2011년 3월 ~ 2015년 8월 : 삼성전자 S.LSI 책임연구원
- 2015년 9월 ~ 현재 : 한국전자기술연구원 책임연구원
- ORCID : <https://orcid.org/0009-0008-6466-649X>
- 주관심분야 : 미디어 신호처리, 인공지능