

일반논문 (Regular Paper)

방송공학회논문지 제26권 제5호, 2021년 9월 (JBE Vol.26, No.5, September 2021)

<https://doi.org/10.5909/JBE.2021.26.5.608>

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

# 다중 스케일 얼굴 영역 디셔너리의 적대적 증류를 이용한 얼굴 초해상화

조병호<sup>a)</sup>, 박인규<sup>a)</sup>, 홍성은<sup>a)†</sup>

## Face Super-Resolution using Adversarial Distillation of Multi-Scale Facial Region Dictionary

Byungho Jo<sup>a)</sup>, In Kyu Park<sup>a)</sup>, and Sungeun Hong<sup>a)†</sup>

### 요약

최근 딥러닝 기반의 얼굴 초해상화 연구는 일반적인 영상에 대한 초해상화 연구와 달리 인간의 얼굴이 가지는 구조적 혹은 의미론적인 특성을 반영한 안면 랜드마크 정보, 주요 영역 디셔너리와 같은 사전 및 참조 정보를 사용하여 우수한 초해상화 결과를 보였다. 그러나 얼굴에 특화된 사전 정보를 사용할 시 추가적인 처리 소요 시간과 메모리를 요구하는 단점이 존재한다. 본 논문은 앞서 언급한 한계점을 극복하고자 지식 증류 기법을 활용한 효율적인 초해상화 모델을 제안한다. 주요 얼굴 영역 기반의 디셔너리 정보를 사용하는 선생 모델에 지식 증류 기법을 적용하여 추론 시 랜드마크 정보와 부가적인 디셔너리 사용이 필요 없는 학생 모델을 구축하였다. 제안하는 학생 모델은 특징맵 기반의 적대적 지식 증류를 통해 얼굴 주요 영역 디셔너리를 가지고 있는 선생 모델로부터 학습을 진행하였다. 본 논문은 제안하는 학생 모델의 실험 결과를 통해 정량 및 정성적으로 우수함을 보이며 선생 모델의 연산량에 비해 90% 이상 절감되는 효율성을 증명한다.

### Abstract

Recent deep learning-based face super-resolution (FSR) works showed significant performances by utilizing facial prior knowledge such as facial landmark and dictionary that reflects structural or semantic characteristics of the human face. However, most of these methods require additional processing time and memory. To solve this issue, this paper propose an efficient FSR models using knowledge distillation techniques. The intermediate features of teacher network which contains dictionary information based on major face regions are transferred to the student through adversarial multi-scale features distillation. Experimental results show that the proposed model is superior to other SR methods, and its effectiveness compare to teacher model.

Keyword : Image super-resolution, Face super-resolution, Knowledge distillation, Adversarial learning, Deep learning

a) 인하대학교 전기컴퓨터공학과(Inha University, Department of Electrical & Computer Engineering)

† Corresponding Author : 홍성은(Sungeun Hong)

E-mail: csehong@inha.ac.kr

Tel: +82-032-860-7427

ORCID: <https://orcid.org/0000-0003-1774-9168>

※ 이 논문의 연구 결과 중 일부는 “2021년 한국방송-미디어공학회 하계학술대회”에서 발표한 바 있음.

※ 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(2020-0-01389, 인공지능융합연구센터지원(인하대학교))과 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019R1A2C1006706).

※ This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2020-0-01389, Artificial Intelligence Convergence Research Center (Inha University)) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2019R1A2C1006706).

· Manuscript received August 2, 2021; Revised September 24, 2021; Accepted September 24, 2021.

Copyright © 2021 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

얼굴 영상 초해상화는 저해상도 얼굴 영상을 고해상도 얼굴 영상으로 초해상화를 수행하는 연구 분야이다. 최근 얼굴 영상을 사용하는 응용 분야가 발전함에 따라 얼굴 영상 초해상화 분야 역시 주목받고 있으며 관련되어 활발한 연구가 진행되고 있다. 최근 제안된 얼굴 영상 초해상화 기법은 단순히 학습 기반의 모델 구조만을 가지는 것이 아니라, 안면 영역이 가지는 랜드마크<sup>[1]</sup>, 3D 얼굴 모형<sup>[2]</sup>, 히트맵<sup>[3]</sup>과 같은 안면의 구조적 정보와 성별, 인종 등과 같은 의미론적인 사전 정보<sup>[4]</sup>를 사용하여 기존 단일 영상 초해상화 기법 대비 우수한 성능을 보이고 있다. 얼굴 영상의 사전 정보를 가지는 초해상화 기법과 더불어 참조 얼굴 영상 기반의 초해상화 연구<sup>[5]</sup>도 제안되었다. 참조 영상 기반 기법은 저해상도 입력 영상에서 구조적 혹은 의미론적인 정보를 사용하는 대신 입력 영상과 동일한 인물의 고화질의 영상을 참고삼아 복원을 수행하는 기법이다. 그러나 저해상도 영상과 동일한 인물의 고화질의 영상을 구하는 것은 어려움이 따르는 단점이 존재한다. 참조 영상의 한계점을 극복하기 위하여 DFDNet<sup>[7]</sup>은 참조 영상 대신 여러 고화질 얼굴 영상을 통해 구축된 디셔너리를 활용하는 복원 기법

을 제안하였다. DFDNet의 정량적, 정성적 우수성에도 불구하고, 이 기법은 입력 영상으로부터 랜드마크 정보와 사전에 구축된 디셔너리를 요구하는 단점이 존재한다.

요약하면 얼굴 초해상도의 성능 향상을 위해서 다양한 사전 정보 또는 참조 영상, 디셔너리 정보를 활용하는 연구가 제안되었다. 저해상도 입력 영상은 안면의 정보가 많이 손실되어 이로부터 구조적 정보인 랜드마크 혹은 성별 및 인종과 같은 의미론적인 속성 추정에는 한계가 존재한다. 추가로, 저해상도 얼굴 영상으로부터 해당 얼굴의 신원을 인지하고 동일 인물의 고품질 참조 영상을 구축하는 것 또한 어려움이 존재한다. 또한 디셔너리와 같은 사전에 구축된 얼굴의 정보를 사용하여 초해상화 수행 시 디셔너리가 메모리를 부가적으로 차지하며 연산량이 늘어나 추가적인 추론 시간이 소요되는 한계점이 존재한다. 그림 1을 통해 ground truth (GT) 영상의 랜드마크 추정이 정확하게 된 반면에 저해상도 영상에서는 부정확한 결과가 측정됨을 알 수 있다. 그림 1의 결과를 통해 랜드마크 추정 결과를 통해서 GT 영상에 비해 정보가 손실된 저해상도 영상으로부터 정확한 구조적 혹은 의미론적인 속성을 추정하는 것은 한계점이 존재하며 부정확한 초해상화 결과를 초래할 수 있음을 알 수 있다.

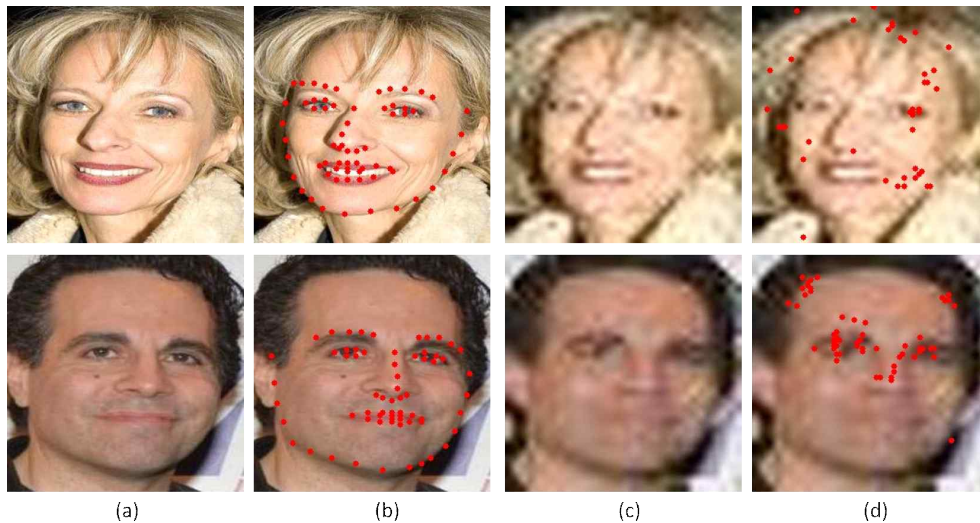


그림 1. 고해상도 영상 (GT)에서의 랜드마크 추정 결과와 저해상도 영상에서의 랜드마크 추정 결과 비교. (a) GT 영상, (b) GT 랜드마크 추정 결과, (c) 저해상도 영상, (d) 저해상도 영상 랜드마크 추정 결과

Fig. 1. Comparison of landmark estimation results between high-resolution images (GT) and low-resolution (LR) images. (a) GT image, (b) landmark estimation results from the GT image, (c) LR image, (d) landmark estimation results from the LR image

본 논문은 얼굴의 사전 정보를 사용하는 기법의 장점을 활용하면서 저해상도 영상으로부터의 구조적 및 의미론적 정보 추정 및 부가적인 정보 사용의 한계점을 극복하기 위해 지식 증류 기법을 적용한 효율적인 얼굴 초해상화 모델을 제안한다. 이를 위해 최근 우수한 성능을 보인 딥서너리 기반의 복원 기법에 지식 증류를 적용하여 복원 수행 시 입력 영상으로부터 랜드마크 정보 추정과 사전에 구축된 딥서너리가 필요 없는 모델을 제안한다. 제안하는 기법은 특징맵 기반의 지식 증류 기법 적용을 통해 훈련 시 학생 모델이 선생 모델의 특징맵으로부터 얼굴의 사전 정보를 학습하도록 설계되었다. 구체적으로, 본 연구는 최근 지식 증류 분야에서 우수한 성능을 보인 특징맵 기반의 적대적 온라인 지식 증류 기법<sup>[8]</sup>에서 착안하여 사전 훈련된 딥서너리 기반의 선생 모델로부터 다중 스케일 특징맵 기반의 적대적 오프라인 지식 증류 기법을 적용한 효율적인 학생 모델을 제안한다. 본 논문의 기여를 요약하면 다음과 같다.

- 얼굴 영상 초해상화에서 높은 성능을 보인 얼굴의 사전 정보를 사용하는 모델의 연산 측면에서의 단점을 극복하고자 다중 스케일 특징맵 기반의 적대적 오프라인 지식 증류 기법을 적용한 효율적인 학생 모델을 제안한다.
- 기존 초해상화 기법들과 정량적 비교 및 정성적 비교를 통해 제안하는 딥서너리 기반 기법의 우수성을 증명한다.
- 선생 모델과 학생 모델이 요구하는 구체적인 연산량 및 메모리 요구량 비교를 통해 최종적으로 제안하는 학생 모델의 효율성을 증명한다.

본 논문의 전체 구성은 다음과 같다. 2장에서는 얼굴 초해상화와 지식 증류에 대한 기존 연구를 소개하며, 3장에서는 본 논문에서 사용하는 선생과 학생 모델의 구조와 학생 모델을 훈련하기 위해 적용한 다중 스케일 특징맵 기반의 적대적 지식 증류 기법에 대해 설명한다. 4장에서는 제안하는 기법의 우수성을 검증하기 위해 성능을 정량적, 정성적으로 평가하고 타 기법과의 비교 평가를 수행한다. 추가로, 지식 증류 기법을 적용한 학생 모델의 효율성 및 유효성을 선생 모델과 지식 증류 기법을 적용하지 않은 학생 모델과

의 비교를 통해 나타낸다. 마지막으로, 5장에서는 본 논문의 결론을 맺는다.

## II. 관련 연구

### 1. 얼굴 초해상화

전통적으로 얼굴 초해상화 연구는 일반적인 영상에 대한 초해상화 기법과 달리 얼굴이 가지는 고유한 구조적 혹은 통계적 정보를 사용하여 사실적인 초해상화를 수행하도록 연구가 진행되었다. Wang et al.<sup>[9]</sup>은 PCA 기반의 저해상도 영상과 고해상도 영상간의 맵핑 함수 기법을 제안하였다. Kolouri et al.<sup>[10]</sup>은 학습을 통해 저해상도 영상에 대한 비선형 lagrangian 모델의 파라미터 추정 후에 얼굴 초해상화를 수행하였다.

SRCNN<sup>[11]</sup>에서 제안한 CNN 기반의 초해상화 연구를 시작으로, channel attention<sup>[12]</sup>, back-projection<sup>[13]</sup>, feedback mechanism<sup>[14]</sup>, GANs<sup>[15]</sup> (Generative Adversarial Networks)을 적용한 다양한 초해상화 연구 모델이 제안되었다. 얼굴 초해상화 연구에서는 CNN과 GANs 알고리즘 기반과 안면의 사전 정보를 사용하여 일반 영상 초해상화 기법과 달리 얼굴에 특화된 방법론이 제안되었다. FSRNet/FSRGAN<sup>[3]</sup>은 hourglass<sup>[16]</sup>기반의 prior estimation network를 제안하여 사전 정보를 획득 및 사용하였다. 이를 통해 저해상도 입력 영상으로부터 얼굴 영역의 랜드마크 히트 맵, 파싱 맵을 추정하여 사실적인 얼굴 영상 복원을 도왔다. Song et al.<sup>[17]</sup>는 눈, 코, 입과 같은 영역 별 컴포넌트 생성 모델을 두어 이를 사용해 얼굴 초해상화를 수행하였다. Ma et al.<sup>[18]</sup>는 두 개의 recurrent 모델을 두어 iterative collaboration을 수행하였다. 저해상도 얼굴 영상으로부터 랜드마크 추정과 복원을 수행하는 iterative collaboration을 통해 모델이 고품질의 얼굴 영상 초해상화를 수행하도록 돕는다.

그러나 알 수 없는 열화가 가해진 실제계의 저해상도 얼굴 영상에 대해서 사전 정보를 취득하는 것은 현실적으로 어려움이 따른다. HiFaceGAN<sup>[19]</sup>은 collaborative suppression and replenishment (CSR) 모듈을 제안하여 부가적인 얼굴에 대한 사전 정보가 필요 없이, 오로지 학습 기반의

모델만을 통해 얼굴 영상 복원을 수행하였다. CSR 모듈을 통해 모델은 입력 영상에서 추출한 특징맵 내부에서 열화와 중요한 특징을 구분하고 보강하도록 학습이 진행되어 실세계의 입력 영상에 대해서도 우수한 복원 성능을 보였다. 그러나 매우 낮은 저해상도의 영상에 대해 부자연스러운 복원 성능을 가지는 한계점이 존재한다. 또한 저해상도 영상에서의 사전 정보 획득의 한계점 극복을 참조 영상 기반의 사실적인 얼굴 영상 복원을 할 수 있는 기법이 제안되었다. GFRNet<sup>[5]</sup>은 warping 모듈을 제안하여 flow field를 예측하여 참조 영상으로부터 warping을 수행하였고 이를 바탕으로 저해상도 영상으로부터 복원을 수행하였다. GWAINet<sup>[6]</sup>은 안면 랜드마크 정보가 필요 없이 warping 수행과 더불어 훈련 과정에 Siamese<sup>[20]</sup> 모델을 사용한 identity loss를 도입하여 참조 영상 기반의 복원을 수행하였다.

기존 참조 영상 기반의 초해상화 기법은 얼굴 영상 초해상도 분야에서 높은 성능을 보였음에도 불구하고 추가적인 고해상도 영상 획득과 연산을 요구하는 단점이 존재한다. DFDNet<sup>[7]</sup>은 참조 얼굴 기반의 방법에 대해 추가적인 영상 획득을 요구하는 한계점을 지적하면서 주요 얼굴 영역에 대한 디셔너리 기반의 복원 기법을 제안하였다. 사전에 고품질의 얼굴 데이터 셋에서 클러스터링을 통해 얼굴의 주요 영역인 눈, 코, 입에 대한 다양한 디셔너리를 구축하였다. 고화질의 영상에서 구축된 디셔너리를 통해 입력 영상의 랜드마크 정보와 디셔너리를 이용하여 모델의 특징맵에 주입함으로써 고품질의 얼굴 복원 결과를 보였다.

## 2. 지식 증류

지식 증류는 가볍고 효율적인 딥러닝 모델 구축을 위한 기법 중 하나로 Hinton et al.<sup>[21]</sup>에서 처음으로 제안되었다. 사전 학습된 깊은 모델을 선생 모델로 두고 비교적 가벼운 모델을 학생 모델로 두어 좋은 성능을 보이는 선생 모델로부터 학생 모델이 hard target labels뿐만 아니라 soft labels로부터 추가로 배우는 방법을 제안하였다. 이미지 인식 분야에 처음으로 적용되어 지식 증류를 통해 가벼운 학생 모델이 획기적인 성능 향상을 보였다. Zagoruyko and Komodakis<sup>[22]</sup>는 어텐션 기반의 지식 증류 기법을 적용하여 soft labels만 사용하는 것이 아니라 학생 모델의 특징맵을

선생 모델의 특징맵과 유사하도록 강제 하는 기법을 제안하여 더 큰 성능 향상을 보였다. Chung et al.<sup>[8]</sup>는 선생 모델과 학생 모델을 동시에 훈련시키는 온라인 학습 방식의 특징맵 기반 적대적 지식 증류 기법을 제안하였다. 선생과 학생 모델에 대응하는 구분자를 두어 사이클 학습 (cyclic learning) 기반으로 선생의 특징맵 분포를 따르도록 하여 가벼운 학생 모델의 우수한 성능 향상을 보였다.

이미지 인식 분야에서 시작된 지식 증류 기법을 바탕으로 최근 초해상화 분야에도 지식 증류 기법을 적용한 많은 연구<sup>[23,24,25,26,27]</sup>가 제안되었다. Gao et al.<sup>[25]</sup>는 선생 모델과 학생 모델의 특징맵간의 통계적 정보의 차이를 줄이는 특징맵 기반의 지식 증류 기법을 제안하였다. 사용된 통계적 정보는 최대 및 최솟값, 평균값이 사용되었다. He et al.<sup>[22]</sup>는 선생 모델과 학생 모델 간의 특징맵으로부터 Affinity matrix를 형성하고 차이를 줄이도록 지식 증류 기법을 수행하였다. Lee et al.<sup>[27]</sup>은 encoder-decoder 구조를 경량화하기 위한 기법을 제안하였다. 선생 모델은 입력 고해상도 이미지를 인코딩하고 다시 입력 영상으로 복원하는 decoder 구조를 가진다. 학생 모델은 사전 훈련된 선생 모델의 decoder 구조를 그대로 물려받아 저해상도 영상으로부터 초해상화를 수행하고 이때 선생과 학생 모델의 특징맵의 통계적 정보를 줄이도록 지식 증류 기법을 적용하였다. 일반적인 영상 초해상화 뿐만 아니라 얼굴 초해상화 연구에도 지식 증류 기법이 적용되어 제안되었다. Kim et al.<sup>[11]</sup>는 얼굴 초해상화 모델 훈련 시 초해상화 수행한 결과와 ground truth (GT) 영상에서 추출된 랜드마크 히트맵의 차이를 줄이도록 수행하였다. 히트맵 추출에 우수한 성능을 보이는 FAN<sup>[28]</sup> 모델을 사용 하여 훈련을 수행 할 시 많은 연산량을 요구하는 단점을 지식 증류 기법을 통해 distilled FAN을 통해 훈련 시 경량화된 랜드마크 추출 모델 사용 기법을 제안하였다.

요약하면 일반 영상 초해상도를 넘어 얼굴 영상 초해상도에 지식 증류를 접목시켜 훈련 시에 적은 메모리 사용을 위한 연구가 제안되었지만, 지식 증류 기법을 통해 추론 시 초해상화 모델의 구조 자체를 경량화 하기 위한 연구 혹은 연산량을 효율적으로 변환하기 위한 연구는 주목 받지 못하였다. 본 논문에서는 얼굴 초해상화 분야에서 좋은 성능을 보인 디셔너리 기반의 모델의 연산량을 추론 시에 효율

적으로 변환하고자 지식 증류 기법을 적용하였다. 이를 위해 랜드마크 정보를 바탕으로 고품질의 딥셔너리 특징 정보를 담고 있는 선생 모델의 다중 스케일의 특징맵을 학생 모델이 모사하도록 훈련을 진행하였다. 이러한 훈련을 통해 학생 모델이 추론 시에 랜드마크 정보와 딥셔너리가 필요 없이 초해상화를 수행할 수 있도록 설계하였다.

### III. 제안하는 기법

본 장에서는 베이스라인 모델인 선생과 학생 모델의 네트워크 구조와 훈련에 적용된 특징맵 기반의 오프라인 적대적 지식 증류 기법에 대해 기술하였다. 사용된 목적 함수는 학생 모델이 주요 얼굴 영역 딥셔너리 정보를 포함하고 있는 선생 모델의 특징맵을 모방하도록 설계되었다. 본 논문에서는 최근 특징맵 기반의 지식 증류 연구에서 우수한

성능을 보인 적대적 지식 증류 기법을 적용하였다. 사전에 학습이 완료된 선생 모델의 특징맵은 구분자에게 실제 입력으로 간주되고, 새로 학습을 시작하는 학생 모델의 특징맵은 구분자에게 가짜 입력으로 간주된다. 학생 모델은 적대적 학습 (adversarial learning)을 통해 자신의 특징맵을 선생 모델의 특징맵과 같은 분포를 가지도록 학습을 수행해 구분자가 실제 입력으로 판별하도록 한다. 전체적인 학습 구조는 그림 2에 도시하였다. 그림 2의 실선은 얼굴 영상 초해상화 수행 단계를 의미하며 점선은 각 모델의 특징맵 지식 증류 학습 단계를 의미한다.

#### 1. 선생 모델

본 논문에서 사용한 선생 모델은 최근 얼굴 복원 분야에서 주요 얼굴 영역의 딥셔너리 사전 정보를 사용하여 우수한 성능을 보인 DFDNet<sup>[7]</sup> 모델을 베이스라인으로 선정하

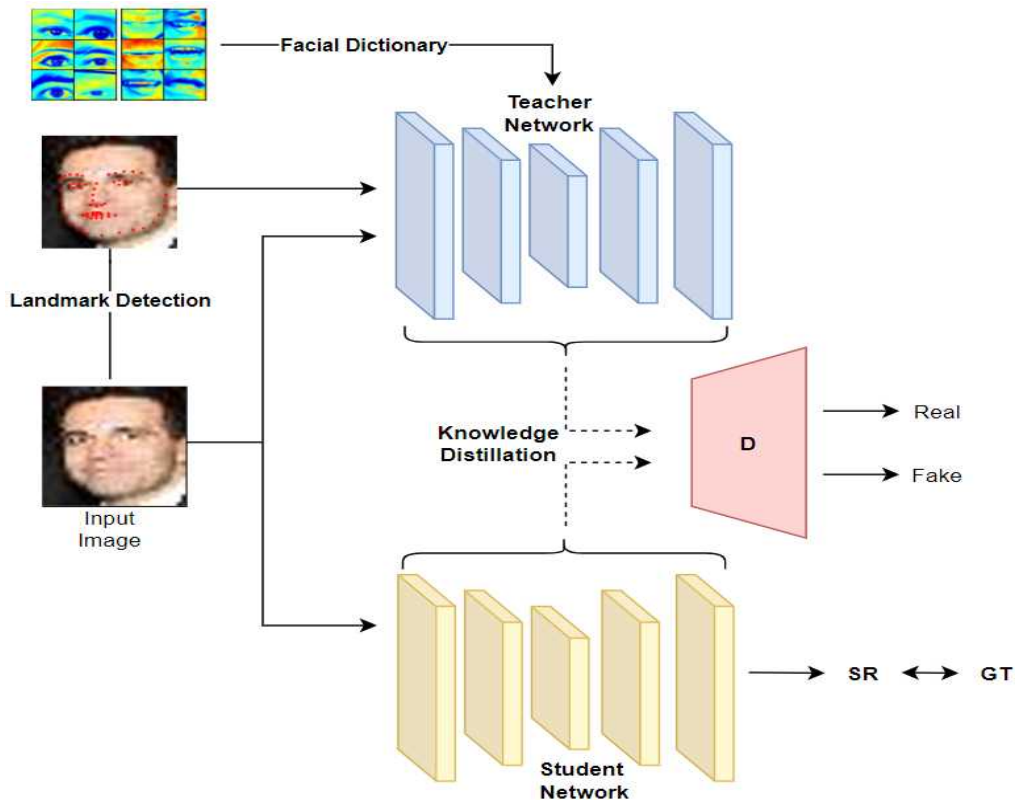


그림 2. 제안하는 학생 모델의 특징맵을 활용한 적대적 지식 증류 학습 방법  
Fig. 2. Overview of our proposed method

였다. 디셔너리는 고품질과 다양한 인종 정보를 가지고 있는 FFHQ<sup>[29]</sup> 데이터셋과 VggFace<sup>[30]</sup> 데이터셋에 사전 훈련된 VGG19<sup>[31]</sup> 네트워크의 다중 스케일의 특징맵으로부터 구축이 되었다. 주요 얼굴 영역의 디셔너리 ( $Dic_c$ ) 취득을 다음의 수식으로 나타낼 수 있다.

$$Dic_c = F_{Dic}(I_{HR}|L; \theta_{Vgg}), \quad (1)$$

고화질의 FFHQ 영상 ( $I_{HR}$ )로부터 추정된 랜드마크 정보 ( $L$ )를 사용하여 사전에 훈련된 VGG19 네트워크( $\theta_{Vgg}$ )의 각 layer에서 추출한 특징맵 ( $F_{Dic}$ )에서의 왼쪽/오른쪽 눈, 코, 입 영역 ( $c$ )을 잘라내어 고정된 크기로 재조정을 실시한다. 각 영상으로부터 샘플링된 영역에 K-means 알고리즘을 적용하여 영역별 클러스터를 생성하여 최종 디셔너리가 구축된다.

선생 모델의 구조는 U자 형태의 encoder-decoder 네트워크 구조를 가진다. 선생 모델은 입력 영상인 저해상도 얼굴 영상으로부터 encoder에서 추출한 특징맵에 구축한 디셔너리 정보를 전이 (dictionaries feature transfer, DFT)하도록 설계되었다. VGG19 구조를 가지는 encoder의 각 3개의 layer에서 추출된 특징맵에서 추정된 랜드마크 정보를 바탕으로 RoIAlign<sup>[32]</sup>을 수행하여 눈, 코, 입 위치를 추정한다. 구축한 디셔너리의 분포와 특징맵으로부터 추정된 영역의 분포를 맞추기 위하여 AdaIn<sup>[33]</sup>을 통해 정규화를 수행한다. Encoder의 특징맵에서 추정한 영역과 정규화된 디셔너리 집합 중 가장 가까운 원소를 찾기 위해 영역별 confidence score (CS) 추정이 수행되며 가장 높은 값을 가지는 디셔너리 원소를 encoder의 특징맵 ( $F_{enc}$ )에 주입하여 새로운 특징맵 ( $\hat{F}$ )을 형성한다.  $\hat{F}$ 를 사전 정보 (prior conditions)으로 간주하여 spatial feature transform (SFT)<sup>[34]</sup> 모듈을 통해 업샘플링이 수행하며 초해상화 영상이 생성된다. SFT 모듈을 통해 업샘플링 되는 수식은 다음과 같다.

$$SFT = \hat{F} \odot F_{enc} + \hat{F} \quad (2)$$

## 2. 학생 모델

본 연구에서 제안하는 학생 모델은 선생 모델과 동일한

encoder-decoder 구조를 가지지만 디셔너리 사용을 요구하는 RoIAlign, AdaIn, CS 모듈이 제거되었으며 SFT 모듈만 존재한다. 제안하는 기법은 학생 모델의 훈련을 SFT 모듈을 거친 선생과 학생 모델의 특징맵의 차이를 줄이도록 훈련 방향을 설계하였다. 설계한 학습 목표를 위해 선생과 학생 모델의 각 계층 (layer)에서 생성되는 특징맵이 가지는 분포의 차이를 줄이도록 다중 스케일의 구분자를 두어 특징맵 기반의 적대적 지식 증류 기법을 적용하였다. 다중 스케일의 구분자는 선생 모델의 특징맵을 실제 (real) 입력으로, 학생 모델의 특징맵을 가짜 (fake) 입력으로 판별하도록 학습이 진행된다. 학생 모델은 생성되는 특징맵을 구분자가 실제 입력으로 판별하도록 적대적 학습이 진행된다. 제안하는 학습 방법을 통해 학생 모델은 디셔너리 정보를 가지고 있는 선생 모델의 특징맵의 분포를 학습하게 된다. 이를 통해 추론 시에는 저해상도 입력 영상으로부터 랜드마크 정보 추정과 사전에 구축된 디셔너리가 필요 없이 초해상화를 수행할 수 있게 된다. 선생 모델에 비해 적은 모듈로 이루어져 있고 입력 영상으로부터 랜드마크 추정과 사전에 구축된 디셔너리가 필요 없는 학생 모델은 더 적은 메모리 사용량과 연산량을 가진다. 결론적으로 본 연구는 디셔너리 기반의 우수한 성능을 보이는 선생 모델의 단점인 입력 영상으로부터의 랜드마크 추정 및 디셔너리 사용으로 인한 부가적인 메모리 사용에 따른 연산량 증가를 다중 스케일의 특징맵 기반 적대적 지식 증류 기법을 통해 극복한 학생 모델을 제안한다.

## 3. 목적 함수

본 연구에서는 사전 학습된 선생 모델을 사용하는 오프라인 지식 증류 기반의 훈련을 진행하였다. 제안하는 학생 모델과 구분자에 대한 목적 함수  $L_S L_D$ 는 다음과 같다.

$$L_S = \lambda_1 L_{recon} + \lambda_2 L_{vgg} + L_{kd} \quad (3)$$

수식 (1)에 사용된 목적 함수  $L_{recon}, L_{vgg}, L_{kd}$ 는 아래와 같이 정의된다.

$$L_{recon} = \|I_{SR} - I_{GT}\|_2 \quad (4)$$

$L_{recon}$ 는 픽셀 기반의 거리 함수 mean squared error (MSE)를 나타낸다. 학생 모델의 초해상화 결과 ( $I_{SR}$ ) 과 정답 영상 ( $I_{GT}$ ) 간의 픽셀 거리 차이를 줄이도록 수행하였다.

$$L_{vgg} = \sum_i^5 \|f_i(I_{SR}) - f_i(I_{GT})\|_1 \quad (5)$$

$L_{vgg}$ 은 사실적인 초해상화를 수행하기 위해 적용한 perceptual loss<sup>[35]</sup>에서 제안한 사전 훈련된 VGG-19에서 추출된 특징맵간의 거리 함수를 나타낸다. 본 훈련 과정에 사용된 perceptual loss는 ImageNet<sup>[36]</sup> 데이터 셋에 사전 훈련된 VGG-19 모델의 1에서 5까지의 계층이 사용되었으며,  $f_i(I)$ 는 입력 영상으로부터 각 계층에서 추출된 특징맵을 나타낸다.  $\lambda_1, \lambda_2$ 는 스케일링 파라미터로 DFDNet<sup>[7]</sup> 연구에서 사용한 동일한 값으로 각각 100, 20으로 설정되어 훈련에 사용되었다.

$$L_{kl} = \sum_{i=0}^3 \left( E[(D_i(F_{S_i}) - 1)^2] + E[(F_{S_i} - F_{T_i})^2] \right) \quad (6)$$

$L_{kl}$ 는 다중 스케일의 특징맵 기반의 적대적 지식 증류 목적 함수를 나타낸다.  $D_i, F_{S_i}, F_{T_i}$ 는 각각 스케일 별 구분자 모델과 학생, 선생 모델의 특징맵을 나타낸다. 학생 모델은 적대적 학습을 통해 선생 모델의 특징맵 분포에 근사하도록 학습이 진행된다. 추가로 학생 모델의 특징맵과 선생 모델의 특징맵의 L2 거리 함수를 적용하여 학습을 돕는다. 구분자는  $F_{T_i}$ 를 실제 데이터로 간주하고  $F_{S_i}$ 를 가짜 데이터로 간주하여 적대적 학습을 수행한다. 구분자의 목적 함수  $L_D$ 는 다음과 같다.

$$L_D = \sum_{i=0}^3 \left( E[(D_i(F_{T_i}) - 1)^2] + E[(D_i(F_{S_i}))^2] \right) \quad (7)$$

각 스케일 별 구분자와 학생 모델은 안정적인 적대적 학습이 가능하도록 제안된 LSGAN<sup>[37]</sup>기반의 적대적 함수를 사용하여 훈련을 진행한다. 최적화 진행 시, Adam<sup>[38]</sup> 최적화기가 사용되었으며 학습률은  $2 \times 10^{-4}$ , 하이퍼파라미터  $\beta_1, \beta_2$ 는 각각 0.9, 0.999로 설정하였다. 본 연구에서 사용하는 목적 함수를 통해 학생 모델은 정답 영상으로부터 학습

하는 것뿐만 아니라 특징맵 기반의 적대적 지식 증류 기법을 통해 선생 모델의 특징맵이 가지고 있는 디저너리 분포를 학습하도록 하였다.

## IV. 실험 결과

이번 장에서는 제안하는 기법의 유효성을 검증하기 위하여 진행한 실험에 대해 기술하였다. 실험 환경, 사용된 데이터 셋, 성능 평가 결과를 기술하였다. 추가로 제안하는 기법의 우수성을 보이기 위해 다양한 초해상화 기존 연구 기법과 정성적 및 정량적 평가 비교를 진행하였다. 비교에 사용된 기존 연구 기법은 general single image super-resolution (SISR) 모델인 RCAN<sup>[12]</sup>, DBPN<sup>[13]</sup>이 사용되었으며 얼굴 영역 초해상화 (FSR) 모델인 HiFaceGAN<sup>[19]</sup>이 사용되었다. 사용된 정량 평가 측정 기법은 전통적인 픽셀 기반의 평가 기법인 PSNR, SSIM, MS-SSIM와 perceptual 기반의 LPIPS<sup>[39]</sup>, FID<sup>[40]</sup>평가 기법이 사용되었다. 제안하는 지식 증류 기법의 효율성을 보이기 위하여 선생 모델 (teacher), 단일 스케일 특징맵 기반의 학생 모델 (student (SSKD)), 다중 스케일 특징맵 기반의 학생 모델 (student (MSKD)), 지식 증류 기법과 디저너리를 사용하지 않은 모델 (student w/o KD) 과의 결과 비교도 진행하였다. 마지막으로 선생 모델과 학생 모델의 추론 시 소요되는 시간과 사용되는 GPU 메모리 용량을 비교하여 학생 모델의 연산량의 효율성을 제시하였다.

### 1. 실험 환경

본 논문에서는 VGGFace2<sup>[30]</sup> 데이터셋을 훈련에 사용하였다. 제공되는 800,000장의 훈련 영상 중 10,000장의 영상을 임의로 추출하였다. 데이터 전처리 과정으로 MTCNN<sup>[41]</sup> 얼굴 탐지 모델을 사용하여 얼굴 영역만을 저장하였다. 추가로, 훈련 시 고해상도의 정답 영상을 사용하기 위해, 얼굴 영역 탐지 결과 해상도가  $256 \times 256$  보다 낮은 영상은 제외하였다. 제공되는 테스트 영상 중에서 임의로 2,000장을 추출하여 테스트를 진행하였다. 또한 제안하는 모델의 일반화 성능을 검증하기 위하여 훈련 과정에서 사용되지 않은

CelebA<sup>[42]</sup> 데이터 셋을 사용하여 추가적인 테스트를 진행하였다. 해당 데이터 셋에 앞서 언급한 동일한 전처리 과정을 적용하였으며 임의로 2,000장을 추출하여 테스트를 진행하였다.

본 연구에서는  $32 \times 32$  해상도를 가지는 입력 얼굴 영상을  $256 \times 256$  해상도로 초해상화를 수행하도록 실험을 진행하였다. 훈련 시, 저해상도 입력 영상 구축에서 단순히 보간법만을 사용하여 다운 샘플을 수행하지 않고 임의로 열화를 가하였다. 본 훈련에서 진행한 열화는 다음 수식으로 표현된다.

$$I_{LR} = ((I_{HR} \otimes k) \downarrow_r + n_\sigma)_{JPEG_q} \quad (8)$$

$k$ 는 가우시안 블러, 모션 블러 중 임의로 선택되는 블러 커널을 의미한다. 블러 커널의 크기는  $\{3.0:5.0\}$  범위 내에서 임의로 설정된다.  $\downarrow_r, n_\sigma, JPEG_q$  순서대로 다운 샘플링 함수, 가우시안 노이즈, JPEG 압축 기호를 나타낸다. 수식 8에서의 각각의 열화 함수의 파라미터 값은 순서대로  $\{3.0:5.0\}$ ,  $\{0.8:0.4\}$ ,  $\{1.5:15.0\}$ ,  $\{0.4:0.8\}$  범위 내에서 임의로 설정되어 입력 얼굴 영상에 적용된다. 본 연구의 실험에서 수행한 모든 초해상화 모델의 훈련 과정에서는 24개의 미니 배치가 사용되었으며 공식 저자가 공개한 코드를

바탕으로 훈련이 진행되었다. 학생 모델의 훈련은 Intel i7-9700 3.6GHz CPU와 NVIDIA Quadro RTX 8000 GPU를 장착한 컴퓨터 환경에서 3일간 진행되었다.

## 2. 정량 평가 및 비교

본 연구에서 수행한 정량적 평가 및 비교를 표 1에 나타내었다. 측정 결과에 따르면, 대체적으로 SISR 모델은 통계 기반의 전통 평가 지표 (PSNR, SSIM, MS-SSIM)에서 좋은 성능을 보인다. 그러나 perceptual 평가 지표 (LPIPS, FID)에서는 FSR 모델에 비해 낮은 성능을 보인다. HiFaceGAN의 경우 perceptual 평가 지표에서 선생 모델과 동등한 혹은 조금 더 나은 성능을 보이지만, PSNR, SSIM 평가 지표에서 bicubic 보간법을 사용한 경우보다 낮은 성능을 보인다. 선생 모델은 대부분의 평가 기법에 있어 우수한 성능을 보임을 알 수 있다. SSIM, MS-SSIM의 평가 지표의 경우 RCAN 모델이 근소하게 더 높은 성능을 보이지만, perceptual 평가 기법인 LPIPS, FID에 있어 기존 일반 영상에 대한 초해상화 모델 대비 20% 이상 향상된 평가 결과를 보인다. 학생 모델의 평가 결과를 보았을 때 단일 스케일의 지식 증류 기법을 적용한 모델은 SISR 모델과 FSR 모델보다 더 높은 성능을 보인다. 제한하는 모델인 다중 스케일의

표 1. VGGFace2와 CelebA 데이터 셋에 대한 정량 평가 및 비교 결과  
 Table 1. Comparison of SR models in VGGFace2 and CelebA

Dataset	Method	Evaluation Metric				
		PSNR $\uparrow$	SSIM $\uparrow$	MS-SSIM $\uparrow$	LPIPS $\downarrow$	FID $\downarrow$
VGGFace2 <sup>[30]</sup>	Bicubic	25.74	0.731	0.852	0.5124	149.495
	RCAN <sup>[12]</sup>	26.59	0.779	0.891	0.3191	73.905
	DBPN <sup>[13]</sup>	26.29	0.768	0.881	0.2971	48.336
	HiFaceGAN <sup>[19]</sup>	21.01	0.698	0.831	0.2635	31.375
	Teacher	26.74	0.771	0.889	0.2413	31.431
	Student (SSKD)	26.62	0.769	0.886	0.2511	34.764
	Student (MSKD)	26.68	0.768	0.888	0.2411	32.347
	CelebA <sup>[40]</sup>	Bicubic	26.1	0.744	0.856	0.4854
RCAN <sup>[12]</sup>		27.17	0.797	0.899	0.2794	65.551
DBPN <sup>[13]</sup>		26.83	0.787	0.89	0.2594	41.799
HiFaceGAN <sup>[19]</sup>		21.37	0.714	0.842	0.2315	25.342
Teacher		27.33	0.79	0.898	0.207	26.713
Student (SSKD)		27.21	0.788	0.895	0.2152	29.429
Student (MSKD)		27.28	0.787	0.897	0.2073	27.17



지식 증류 기법을 적용하였을 경우 단일 스케일을 적용한 모델보다 향상된 성능을 보였으며 특히 perceptual 평가 기법에 있어서 5~10% 정도 향상을 보였다. 정량 평가 결과를 통해 제안하는 학생 모델이 선생 모델의 평가 결과와 근사함을 확인할 수 있다.

### 3. 정성 평가 및 비교

본 절에서는 제안하는 기법과 다양한 초해상화 모델의 정성적 비교 분석을 작성하였다. 각 모델 별 초해상화 결과를 그림 3에 나타내었다. 그림 3의 첫 번째 행부터 네 번째



그림 3. VGGFace2와 CelebA 데이터 셋에 대한 정성 비교 결과  
 Fig. 3. Qualitative comparison of SR models in VGGFace2 and CelebA

행의 영상은 VGGFace2 테스트 결과를 나타내며 다섯 번째 행부터 여덟 번째 행의 영상은 CelebA 테스트 결과를 나타낸다. SISR 모델인 RCAN과 DBPN 모델의 경우 bicubic 보간 기법에 비해 사실적인 복원 결과를 보이지만, 지나치게 매끄러운 초해상화 결과를 보인다. 하지만 중요한 얼굴 영역인 눈, 코, 입에 대해 비현실적인 형태를 가지는 복원 결과를 보임을 알 수 있다. FSR 모델인 HiFaceGAN은 SISR 모델보다 사실적인 얼굴 형태 및 주요 얼굴 영역을 복원한 결과를 보인다. 그러나 입력 영상의 색상을 유지하지 못하고 변색 된 초해상화 결과를 보인다. 디서너리 기반의 선생 모델 (teacher) 의 경우 가장 사실적인 초해상화 결과를 보인다. 제안하는 학생 모델 (student) 의 경우 선생 모델과 비슷하게 전체적인 얼굴 영역을 보임과 얼굴의 주요 영역인 눈, 코, 입에 대해서 우수한 복원 성능을 보임을 알 수 있다.

#### 4. 지식 증류 효과

본 절에서는 제안하는 지식 증류 기법을 적용한 학생 모델의 효율성을 보이기 위하여 지식 증류 기법을 적용하지 않은 학생 모델과의 정량적 및 정성적 비교를 나타내었으며 선생 모델과 학생 모델의 추론 시간과 GPU 메모리 사용량을 비교하여 나타내었다.

VGGFace2와 CelebA 테스트 데이터 셋을 사용하여 정량 평가 및 비교를 수행한 결과를 표 2에 나타내었다. 표 2의 결과에 따르면, 제안하는 지식 증류 기법을 적용한 학생 모델의 경우, 모든 평가 지표에서 성능 향상을 보이며 특히 FID 평가 지표에서 큰 성능 향상을 보인다. 또한 지식

증류 기법을 적용하지 않은 학생 모델을 bicubic 보간법의 평가 결과와 비교했을 때, PSNR, SSIM 계열의 평가 지표와 비슷한 측정 결과를 보이며 이를 통해 지식 증류 기법을 적용하지 않은 모델이 열등한 성능을 보임을 알 수 있다.

정성 비교 결과를 그림 4를 통해 나타내었다. 그림 4의 결과에 따르면, 지식 증류 기법을 적용하지 않은 학생의 모델은 bicubic 보간법에 비해 비교적 얼굴 구조를 우수하게 복원하지만 얼굴의 핵심 영역인 눈, 입 부분에 있어 부자연스러운 복원 결과를 보인다. 반면에 지식 증류 기법을 적용



그림 4. 지식 증류 기법을 적용한 학생 모델과 적용하지 않은 학생 모델의 VGGFace2와 CelebA 데이터셋에 대한 정성 비교

Fig. 4. Qualitative Comparison between student and student without KD network in VGGFace2 and CelebA

표 2. 지식 증류 기법을 적용한 학생 모델과 적용하지 않은 학생 모델의 VGGFace2와 CelebA 데이터셋에 대한 정량 평가 및 비교  
 Table 2. Quantitative Comparison between student and student w/o KD network in VGGFace2 and CelebA

Dataset	Method	Evaluation Metric				
		PSNR ↑	SSIM ↑	MS-SSIM ↑	LPIPS ↓	FID ↓
VGGFace2	Bicubic	25.745	0.731	0.852	0.5124	149.495
	Student w/o KD	25.837	0.75	0.868	0.2746	43.406
	Student	26.682	0.768	0.888	0.2411	32.347
CelebA	Bicubic	26.1	0.744	0.856	0.4854	165.602
	Student w/o KD	26.306	0.766	0.875	0.2345	33.536
	Student	27.288	0.787	0.897	0.2073	27.17

한 학생 모델의 경우 사실적인 초해상화 결과를 보임을 알 수 있다. 결과적으로, 제안하는 학생 모델은 지식 증류 기법을 적용하지 않은 학생 모델보다 정량적 및 정성적으로 우수한 초해상화 성능을 가짐을 알 수 있다. 제안하는 학생 모델의 연산 효율성을 나타내기 위해 추론 시 소요 되는 시간과 메모리 사용량을 선생 모델과의 비교 결과를 표 3에 나타내었다.

추론 시간은 테스트에 사용한 2,000장을 사용하여 측정 하였으며 전체 영상에 소요된 시간의 평균값을 나타내었다. 측정 결과를 비교하였을 때, 학생 모델은 선생 모델보다 10 배 이상의 빠른 추론 시간을 보인다. 딥셔너리 셋을 메모리에 올릴 경우 비교적 작은 크기의 메모리가 요구 되지만, 많은 GPU 연산을 요구함에 따라 필요한 GPU 메모리는 선생 모델에 비해 절반 이하의 메모리 용량을 사용하는 것을 알 수 있다. 최종적으로, 특징맵 기반의 적대적 지식 증류 기법을 적용한 제안하는 학생 모델은 지식 증류 기법을 적용하지 않은 학생 모델보다 정량 및 정성적으로 우수한 성능을 보이고 있으며 테스트 진행 시 소요되는 추론 시간과 GPU 메모리 사용량을 선생 모델과의 비교를 통해 연산량 측면에서의 효율성을 확인할 수 있다.

표 3. 선생과 학생 모델 별 추론 시간 및 메모리 사용량 비교  
Table 3. The inference time and memory usage comparison between teacher and student network

	Inference Time (m/s)	Memory Usage (GB)	Dictionary Size (GB)
Teacher	0.4308	9.8	1.8
Student	0.0342	4.3	-

## V. 결 론

본 논문에서는 기존 얼굴 초해상화 연구에서 우수한 성능을 보인 주요 얼굴 영역의 딥셔너리 정보를 사용하는 선생 모델로부터 지식 증류 기법을 통해 효율적인 학생 모델을 제안하였다. 선생 모델이 가지는 단점인 랜드마크 정보 추정 및 딥셔너리 셋 요구를 특징맵 기반의 적대적 지식 증류 기법을 통해 극복하고 선생 모델의 성능에 근사하는 학생 모델을 제안하였다. 더 나아가 추론 시간 단축 및 메모리 사용량 감소를 달성할 수 있었다. 기존의 CNN 기반의

초해상화 모델과 정성적 및 정량적 평가 결과 비교를 통해 학생 모델의 우수함을 보이고, 나아가 지식 증류를 적용하지 않은 학생 모델과의 정성적 및 정량적 성능 비교를 통해 제안하는 기법의 우수성을 검증하였다. 향후 제안하는 기법의 실환경 적용을 위해서, 선생 모델보다 더 가벼운 모델 아키텍처로 구성된 학생 모델에 기반한 지식 증류 기법을 연구할 계획이다.

## 참 고 문 헌 (References)

- [1] D. Kim, M. Kim, G. Kwon, and D. S. Kim, "Progressive face super-resolution via attention to facial landmark," In Proc. of British Machine Vision Conference, pp. 192, 2017
- [2] X. Hu et al., "Face super-resolution guided by 3D facial priors," In Proc. of European Conference on Computer Vision, pp. 763-780, 2020, doi:10.1007/978-3-030-58548-8\_44.
- [3] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang. "Fsrnet: End-to-end learning face super-resolution with facial priors," In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2492-2501, 2018, doi:10.1109/CVPR.2018.00264.
- [4] X. Yu, B. Fernando, R. Hartley and F. Porikli, "Super-resolving very low-resolution face images with supplementary attributes," In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 908-917, 2018, doi:10.1109/CVPR.2018.00101.
- [5] X. Li, M. Liu, Y. Ye, W. Zuo, L. Lin, and R. Yang. "Learning warped guidance for blind face restoration," In Proc. of European Conference on Computer Vision, pp. 278-296, 2018, doi:10.1007/978-3-030-01261-8\_17.
- [6] B. Dogan, S. Gu, and R. Timofte. "Exemplar guided face image super-resolution without facial landmarks," In Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1814-1823, 2019, doi:10.1109/CVPRW.2019.00232.
- [7] X. Li, C. Chen, S. Zhou, X. Lin, W. Zuo, and L. Zhang. "Blind face restoration via deep multi-scale component dictionaries," In Proc. of European Conference on Computer Vision, pp. 399-415, 2020, doi:10.1007/978-3-030-58545-7\_23.
- [8] I. Chung, S. Park, J. Kim, and N. Kwak, "Feature-map-level online adversarial knowledge distillation," In Proc. of International Conference on Machine Learning, pp. 2006-2015, 2020.
- [9] W. Xiaogang, and T. Xiaoou. "Hallucinating face by eigentransformation," IEEE Trans. on Systems, Man, and Cybernetics, Part C, 35(3):425-434, 2005.
- [10] S. Kolouri and G. K. Rohde. "Transport-based single frame super resolution of very low resolution face images," In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 4876-4884, 2015, doi:10.1109/CVPR.2015.7299121.
- [11] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," In Proc. of European Conference on

- computer Vision, pp. 184-199, 2014, doi:10.1007/978-3-319-10593-2\_13.
- [12] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. "Image super-resolution using very deep residual channel attention networks," In Proc. of European Conference on Computer Vision, pp. 294-310, 2018, doi:10.1007/978-3-030-01234-2\_18.
- [13] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1664-1673, 2018, doi: 10.1109/CVPR.2018.00179.
- [14] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 3867-3876, 2019, doi: 10.1109/CVPR.2019.00399.
- [15] I. J. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," In Proc. of Advances in Neural Information Processing Systems, 2014.
- [16] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," In Proc. of European Conference on Computer Vision, pp. 483-499, 2016, doi:10.1007/978-3-319-46484-8\_29.
- [17] Y. Song, J. Zhang, S. He, L. Bao, and Q. Yang. "Learning to hallucinate face images via component generation and enhancement," In Proc. of International Joint Conference on Artificial Intelligence, pp. 4537-4543, 2017, doi:10.5555/3171837.3171921.
- [18] C. Ma, Z. Jiang, Y. Rao, J. Lu, and J. Zhou. "Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation," In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 5568-5577, 2020, doi:10.1109/CVPR42600.2020.00561.
- [19] L. Yang, C. Liu, P. Wang, S. Wang, P. Ren, S. Ma, and W. Gao. "Hifacegan: Face renovation via collaborative suppression and replenishment," In Proc. of ACM International Conference on Multimedia, pp. 1551-1560, 2020, doi:10.1145/3394171.3413965.
- [20] G. Koch, R. Zemel, and R. Salakhutdinov. "Siamese neural networks for one-shot image recognition," In Proc. of International Conference on Machine Learning, 2015.
- [21] G. Hinton, O. Vinyals, and J. Dean. "Distilling the knowledge in a neural network," In Proc. of Neural Information Processing Systems, 2014
- [22] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," In Proc. of International Conference on Learning Representations, 2017
- [23] S. Ahn and S. J. Kang, "Deep learning-based real-time super-resolution architecture design," Journal of Broadcast Engineering, Vol. 26, No. 2, pp. 167-174, March, 2021.
- [24] O. S. Kwon "Real-time low-resolution face recognition algorithm for surveillance systems," Journal of Broadcast Engineering, Vol.25, No.1, p. 105-108, Jan, 2020.
- [25] Q. Gao, Y. Zhao, G. Li, and T. Tong, "Image super-resolution using knowledge distillation," In Proc. of Asian Conference on Computer Vision, pp. 527-541, 2018, doi:10.1007/978-3-030-20890-5\_34.
- [26] Z. He, T. Dai, J. Lu, Y. Jiang, and S. T. Xia. "FAKD: Feature-affinity based knowledge distillation for efficient image super-resolution," In Proc. of IEEE International Conference on Image Processing, pp. 518-522, 2020, doi:10.1109/ICIP40778.2020.9190917.
- [27] W. Lee, J. Lee, D. Kim, and B. Ham. "Learning with privileged information for efficient image super-resolution," In Proc. of European Conference on Computer Vision, pp. 465-482, 2020, doi:10.1007/978-3-030-58586-0\_28.
- [28] A. Bulat, and G. Tzimiropoulos. "How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks)," In Proc. of IEEE International Conference on Computer Vision n. pp. 1021-1030, 2017, doi:10.1109/ICCV.2017.116.
- [29] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 4396-4405, 2019, doi:10.1109/CVPR.2019.00453.
- [30] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. "Vggface2: A dataset for recognising faces across pose and age," In Proc. of IEEE International Conference on Automatic Face & Gesture Recognition, pp. 67-74, 2018
- [31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," In Proc. of International Conference on Learning Representations, 2015.
- [32] K. He, G. Gkioxari, P. Dollár, and R. Girshick. "Mask r-cnn," In Proc. of IEEE International Conference on Computer Vision, pp. 2980-2988, 2017, doi:10.1109/ICCV.2017.322.
- [33] X. Huang and S. Belongi. "Arbitrary style transfer in real-time with adaptive instance normalization," In Proc. of IEEE International Conference on Computer Vision, pp. 1510-1519, 2017, doi:10.1109/ICCV.2017.167.
- [34] X. Wang, K. Yu, C. Dong, and C. C. Loy. "Recovering realistic texture in image super-resolution by deep spatial feature transform," In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 606-615, 2018, doi:10.1109/CVPR.2018.00070.
- [35] J. Johnson, A. Alahi, F.F. Li. "Perceptual Losses for real-time style transfer and super-resolution," In Proc. of European Conference on Computer Vision, pp. 694-711, 2016, doi:10.1007/978-3-319-46475-6\_43.
- [36] J. Dong, W. Dong, R. Socher, L. J. Li, K. Li, and F. F. Li. "ImageNet: A large-scale hierarchical image database," In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255, 2009, doi:10.1109/CVPR.2009.5206848.
- [37] X. Mao, Q. Li, H. Xie, R.Y.K Lau, Z. Wang, and S. PaulSmalley. "Least squares generative adversarial networks," In Proc. of IEEE International Conference on Computer Vision, pp. 2813-2821, 2017, doi:10.1109/ICCV.2017.304.
- [38] D.P. Kingma and J. Ba. "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
- [39] R. Zhang, P. Isola, A. Efros, E. Shechtman, and O. Wang. "The unreasonable effectiveness of deep features as a perceptual metric," In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1664-1673, 2018, doi:10.1109/CVPR.2018.00068.
- [40] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. "Gans trained by a two time-scale update rule converge to a local nash

equilibrium,” In Proc. of Advances in Neural Information Processing Systems, pp. 6629-6640, 2017.

[41] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. “Joint face detection and alignment using multitask cascaded convolutional networks,” IEEE Signal

Processing Letters, 23(10):1499-1503, 2016.

[42] Z. Liu, P. Luo, X. Wang, and X. Tang. “Deep learning face attributes in the wild,” In Proc. of International Conference on Computer Vision, p. 3730-3738, 2015, doi:10.1109/ICCV.2015.425.

---

## 저 자 소 개



### 조 병 호

- 2020년 2월 : 인천대학교 컴퓨터공학과 학사
- 2020년 3월 ~ 현재 : 인하대학교 전기컴퓨터공학과 석사과정
- ORCID : <https://orcid.org/0000-0002-7652-0999>
- 주관심분야 : 컴퓨터비전, deep learning, face restoration



### 박 인 규

- 1995년 2월 : 서울대학교 제어계측공학과 학사
- 1997년 2월 : 서울대학교 제어계측공학과 석사
- 2001년 8월 : 서울대학교 전기컴퓨터공학부 박사
- 2001년 9월 ~ 2004년 2월 : 삼성종합기술원 멀티미디어랩 전문연구원
- 2007년 1월 ~ 2008년 2월 : Mitsubishi Electric Research Laboratories (MERL) 방문연구원
- 2014년 9월 ~ 2015년 8월 : MIT Media Lab 방문부교수
- 2018년 7월 ~ 2019년 6월 : University of California, San Diego (UCSD) 방문학자
- 2004년 3월 ~ 현재 : 인하대학교 정보통신공학과 교수
- ORCID : <http://orcid.org/0000-0003-4774-7841>
- 주관심분야 : 컴퓨터비전 및 그래픽스 (영상기반 3차원 형상 복원, 증강현실, computational photography), GPGPU



### 홍 성 은

- 2010년 2월 : 한양대학교 컴퓨터공학과 학사
- 2012년 8월 : 카이스트 컴퓨터공학과 석사
- 2018년 2월 : 카이스트 컴퓨터공학과 박사
- 2018년 1월 ~ 2020년 8월 : SK telecom T-Brain, AI Center 연구원
- 2020년 9월 ~ 현재 : 인하대학교 정보통신공학과 교수
- ORCID : <https://orcid.org/0000-0003-1774-9168>
- 주관심분야 : domain adaptation, face recognition, face manipulation, instance segmentation