

로짓 분포 변환 기반 지식 증류를 통한 얼굴 파싱 기법

배지호*¹, 이수원*²

Facial Parsing Techniques with Knowledge Distillation based on Logit Distribution Transformation

Jiho Bae*¹, Suwon Lee*²

요약

얼굴 조작과 같은 응용 분야에서 얼굴 파싱(Face parsing)은 핵심적인 역할을 하며, 특히 에지(Edge) 디바이스나 실시간 추론 환경에서는 모델 경량화가 필수적이다. 그러나 로짓(Logit) 기반 지식 증류는 분류 문제에서는 피쳐맵 기반 방식과 함께 활발히 연구됐지만, 이를 얼굴 파싱 작업에 적용할 경우 공통 온도 설정 문제와 픽셀 간 정보 전달의 한계와 같은 과제가 존재한다. 본 연구에서는 얼굴 파싱 작업에 로짓 기반 지식 증류를 도입하며, 기존 방법의 한계를 분석하고 이를 개선한 로짓 분포 변환 기법을 제안한다. 이 기법은 공통 온도 문제를 완화하고, 학생 모델이 교사 모델로부터 픽셀 간 정보를 효과적으로 전달받을 수 있도록 설계되었다. CelebA-Mask-HQ 데이터셋을 활용한 실험 결과, 본 연구에서 제안한 기법이 경량화된 모델에서 평균적으로 약 0.5의 성능을 향상함을 확인하였다.

Abstract

Face parsing plays a crucial role in applications such as face manipulation. In edge devices or real-time inference environments, model lightweighting is essential. While logit-based knowledge distillation has been actively studied in classification tasks alongside feature map-based methods, its application to face parsing poses challenges such as the common temperature setting issue and the limitation in transferring pixel-wise information. In this study, we introduce logit-based knowledge distillation to face parsing tasks, analyze the limitations of existing methods, and propose an improved logit distribution transformation technique. This technique alleviates the common temperature issue and is designed to enable the student model to effectively inherit pixel-wise information from the teacher model. Experimental results using the CelebA-Mask-HQ dataset demonstrate that the proposed method significantly enhances the performance of lightweight models.

Keywords

face parsing, knowledge distillation, logit standardization, lightweight

* 경상국립대학교 컴퓨터공학과(*² 교신저자)
- ORCID¹: <https://orcid.org/0009-0004-5942-4972>
- ORCID²: <https://orcid.org/0000-0003-2603-1385>

• Received: Nov. 06, 2024, Revised: Nov. 20, 2024, Accepted: Nov. 23, 2024
• Corresponding Author: Suwon Lee
Dept. of Computer Science and Engineering, Gyeongsang National University, 501, Jinju-daero, Jinju-si, Gyeongsangnam-do
Tel.: 82+55-772-1394, Email: leesuwon@gnu.ac.kr

1. 서 론

얼굴 파싱(Face parsing)은 얼굴 이미지에서 눈, 코, 입, 귀, 헤어라인 등 세부 부위를 픽셀 단위로 분할하는 기술로, 각 부위를 정밀하게 인식하고 구분하는 작업이다[1]. 이 기술은 얼굴 조작, 증강 현실, 보안 시스템, 의료 영상 분석 등 다양한 응용 분야에서 필수적인 역할을 수행하며, 특히 얼굴의 각 부위를 자연스럽게 조작하거나 변형해야 하는 분야에서 중요성이 부각된다[2].

최근 몇 년간 딥러닝 기술이 급격히 발전하면서 얼굴 파싱 분야에서도 딥러닝 기반 모델이 적극적으로 도입되고 있으며, 그 성능 또한 괄목할 만한 수준으로 향상되고 있다[3]. 특히 최신 딥러닝 모델들은 얼굴의 다양한 세부 부위를 더욱 정밀하게 구분할 수 있어 이전보다 높은 수준의 정확도와 자연스러운 결과물을 제공한다. 그러나 이러한 고성능 모델들은 대개 대규모 연산 자원을 요구하기 때문에, 에지 디바이스(Edge device)나 실시간 응용 환경에서 적용하기 어려운 문제가 여전히 존재한다. 따라서 이러한 제한된 환경에서 얼굴 파싱을 수행하기 위해서는 모델의 경량화가 필수적이며, 연산 자원을 효율적으로 활용하면서도 높은 정확도를 유지할 수 있는 경량 모델의 필요성이 크게 대두되고 있다[4].

이와 같은 맥락에서 경량화된 모델의 성능을 보완하기 위한 방법으로 지식 증류(Knowledge distillation)는 최근 효과적인 대안으로 주목받고 있다[5]. 지식 증류는 고성능 교사 모델(Teacher model)로부터 경량화된 학생 모델(Student model)에 지식을 전달하여 학생 모델이 보다 우수한 성능을 낼 수 있도록 학습시키는 기법이다. 이를 통해 경량 모델이 고성능 모델에 근접한 성능을 발휘할 수 있어 다양한 실시간 응용 분야에서 주목받고 있다. 그러나 현재까지의 연구에서 지식 증류 기법이 얼굴 파싱 작업에 직접적으로 적용된 사례는 드물며, 이는 해당 분야에서의 발전 가능성이 충분히 남아 있음을 시사한다.

지식 증류에서는 대표적으로 '로짓(Logit)' 기반 방식과 피쳐맵 기반의 방식이 있다. 로짓은 분류 문제에서 소프트맥스(Softmax) 연산 직전 모델의 최종 출력값으로, 교사 모델과 학생 모델 간의 학습 과정

에서 중요한 역할을 한다. 그러나 기존의 분류 문제에서 로짓 기반 지식 증류에서는 교사와 학생 모델이 동일한 온도(Temperature)를 사용하는 방식이 일반적이었으며, 최근 연구에 따르면 이러한 방식이 학습 과정에서 여러 문제를 일으킬 수 있음이 밝혀졌다[6]. 이를 해결하기 위해 일부 연구에서는 로짓 표준화(Standardization)와 같은 기법을 도입하였으나, 세그멘테이션(Segmentation)에서 중요한 픽셀 간 정보 전달이 제대로 되지 않는 문제[7][8]를 우리의 실험 결과로 확인할 수 있다.

이러한 한계를 극복하기 위해 본 연구에서는 로짓 분포 변환(Logit distribution transformation) 기반의 새로운 지식 증류 기법을 제안한다. 제안된 기법은 교사 모델과 학생 모델 간의 적절한 로짓 분포 조정을 통해 각 픽셀 간의 분포를 효과적으로 전달함으로써 경량 모델의 정확도를 극대화할 수 있도록 설계되었다. 우리는 CelebA-Mask-HQ 데이터셋[2]을 활용한 실험을 통해 제안된 방법이 기존의 로짓 표준화 기법에 비하여 경량 모델의 성능을 유의미하게 향상 시켰음을 입증하였다. 얼굴 파싱의 상위 분야인 dense prediction 지식 증류에서는 피쳐 맵 기반의 방식과 로짓 기반의 지식 증류 방식을 융합하여 사용하는데[7][8], 얼굴 파싱에서 개선된 로짓 기반의 방식은 향후 피쳐 맵과 로짓의 융합된 지식 증류에서 성능향상을 기대할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 얼굴 파싱과 지식 증류에 관한 기존 연구와 관련된 이론적 배경을 소개하고, 다양한 연구 결과와 방법론을 비교하여 본 연구의 필요성과 위치를 명확히 한다. 이를 통해 얼굴 파싱에서 지식 증류 기법의 잠재적 활용성과 현재의 한계점을 파악한다.

3장에서는 본 연구에서 제안하는 로짓 분포 변환 기법의 개념을 설명한다. 여기에서는 기존의 로짓 기반 지식 증류와의 차별점을 강조하고, 제안된 기법이 어떻게 교사 모델과 학생 모델 간의 적절한 로짓 분포 조정을 통해 성능을 극대화하는지 구체적인 절차와 알고리즘을 통해 논의한다.

4장에서는 제안된 방법의 실험적 유효성을 검증하기 위해 기존 학습 방법과 지식 증류 기반 학습 방법, 그리고 제안된 학습 방법을 비교한 실험 결과를 제시한다.

CelebA-Mask-HQ 데이터셋을 기반으로 다양한 지표에서 성능을 평가하며, 제안된 기법이 기존 방법에 비해 어떤 면에서 더 효과적인지 구체적으로 분석한다. 마지막으로, 5장에서는 본 연구의 주요 결론을 요약한다.

II. 관련연구

2.1 얼굴파싱을 위한 경량화

얼굴 파싱에서 대부분의 연구는 아키텍처(Architecture) 경량화를 중심으로 진행된다. 실시간 의미론적 분할을 위해 제안된 EHANet[4]은 경량화된 백본을 개발하고 하이브리드 주의 메커니즘을 통해 속도와 정확도 사이의 균형을 유지한다. BADANet[9]은 경량화된 네트워크와 확장된 주의 메커니즘을 활용하여 경계영역의 정확한 정보 학습을 통해 모델의 정확도와 효율성을 높인다. FP-LIIF[3]은 구조적으로 일관된 얼굴 특징을 활용해 효율적인 함수 기반 모델을 설계한다. S. Liu et al.[10]은 CNN과 RNN을 결합하여 계산 효율성을 높이면서도 높은 정확도를 유지한다.

2.2 지식 증류

지식 증류는 대규모 네트워크(교사 모델)의 학습 지식을 소형 네트워크(학생 모델)로 전달하여 효율적인 모델 압축을 이루는 접근법이다. 특히, 로짓(Logit)과 피쳐 맵(Feature map) 기반 지식 증류는 각각 교사와 학생 간의 상이한 정보 전달 방식에 중점을 두며, 주요 연구에서 다양한 기법들이 개발되었다.

로짓 기반 지식 증류는 교사 모델의 예측 확률 분포(로짓)를 학생 모델이 학습하는 방식이다. G. Hinton et al.[5]의 초기 연구에서는 소프트맥스 함수의 온도 매개변수를 도입하여, 교사와 학생 모델 간 확률 분포를 유연하게 조정하는 방식을 제안했다. 이후 여러 연구가 로짓 기반 증류의 성능 개선을 시도했다. CTKD[11]는 샘플마다 가변적인 온도를 예측하여 난이도에 따라 증류를 조정하는 방식을 제안하였으며, MLKD[12]는 다중 로짓 레벨에서 학생 모델이 교사 모델의 정보를 학습하도록 설계

되었다. 또한, DKD[13]는 온도 매개변수의 중요성을 강조하며, 학생이 교사의 로짓 분포에서 더 많은 정보를 학습할 수 있도록 했다. S. Sun et al.[6]은 분류 문제에서 로짓 기반 지식증류가 교사 모델과 학생 모델이 동일한 온도 매개변수를 사용하는 것에 문제를 제기하고 이를 보완한 로짓 표준화를 제안하였다.

피쳐 맵 기반 증류는 학생 모델이 교사 모델의 내부 표현, 즉 중간 계층의 피쳐 맵을 학습하도록 유도하는 방법이다. FitNet[14]은 초기 피쳐 맵 증류 방법으로, 교사 모델의 중간 피쳐 맵을 학생 모델이 학습하여 고차원 데이터에 대해 효과적인 특성 학습을 수행하도록 했다. 이후 등장한 다양한 연구들은 피쳐 맵 기반 증류의 성능을 확장하는 방식을 제안했다. AT[15]는 피쳐 맵 간의 유사성을 높이기 위해 어텐션 기법을 사용하였고, CRD[16]는 대조 학습을 통해 피쳐 맵의 유사성을 증진했다. ReviewKD[17]는 피쳐 맵을 교사 모델의 리뷰 단계에서 학습함으로써 더 정교한 피쳐 매칭을 달성했다.

최근 dense prediction 분야에서도 활발한 연구가 진행되고 있다. [7]은 픽셀 별 지식 증류 손실함수를 활용하여 픽셀 주변의 지식 증류의 중요성을 보여주었고, [8]은 채널별로 지식 증류 손실함수를 활용하였다. 얼굴 파싱에서도 [18]은 기존 경량화된 모델의 성능 개선을 목적으로한 일반적인 지식 증류 방식과 달리 자기 증류와 셀프 어텐션을 통해서 최첨단 수준으로 성능을 개선하였다.

III. 제안 방법

3.1 로짓 기반 지식 증류 방법

그림 1은 로짓기반 지식 증류 방법을 보여준다. 로짓은 분류 문제에서 소프트맥스(Softmax) 이전에 모델이 출력하는 최종 결과값을 의미한다. 이 로짓 값은 지식 증류 과정에서 중요한 역할을 하며, 특히 교사 모델의 로짓 값을 학생 모델이 학습하게 함으로써 학생 모델이 교사 모델의 고유한 지식을 더욱 효과적으로 습득할 수 있도록 한다.

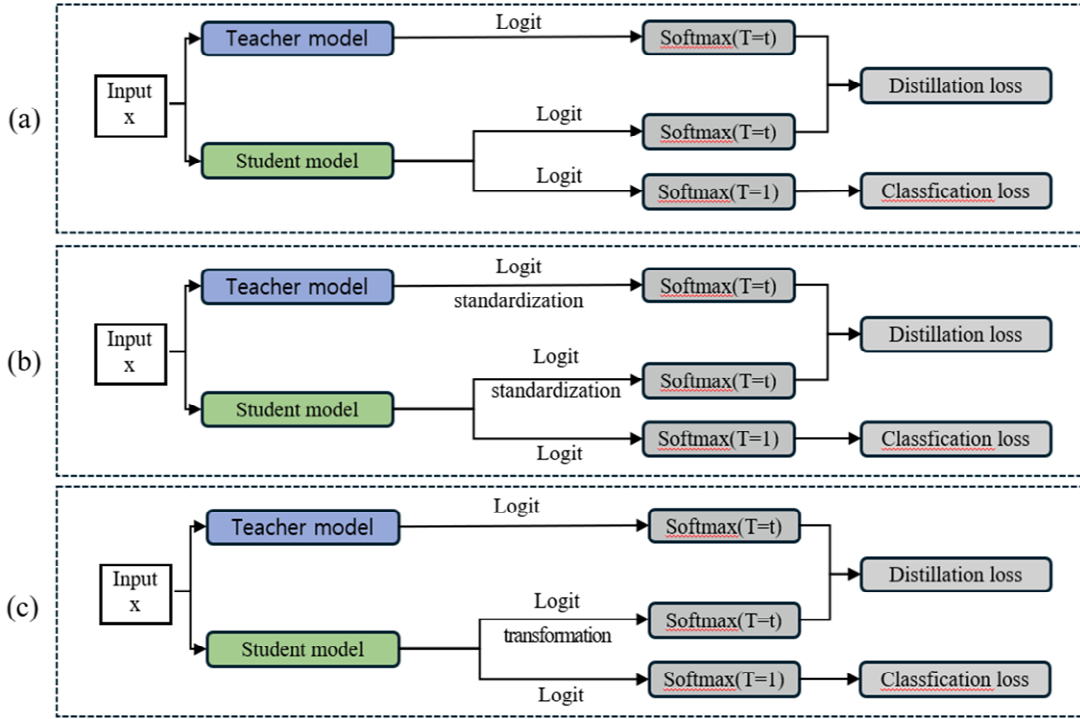


그림 1. 로짓 기반 지식 증류 방법 비교 (a) 초기 제안된 기본적인 지식 증류 방법, (b) 로짓 표준화를 적용한 지식 증류 방법, (c) 제안한 로짓 분포 변환 방법

Fig. 1. Comparing logit-based knowledge distillation methods (a) Initially proposed basic knowledge distillation method, (b) Knowledge distillation method applying logit standardization, (c) Proposed logit distribution transformation method

Hinton[5]의 지식 증류 방법에서는 로짓을 부드럽게(Smooth) 변환하기 위해 온도(Temperature) 매개변수를 활용하며, 이를 통해 학생 모델이 다양한 클래스 간의 확률 분포를 학습할 수 있게 한다. 온도 값을 높일수록 예측 확률 분포가 평탄해져 교사 모델이 학습한 각 클래스 간의 관계가 더욱 명확히 반영되며, 학생 모델은 정답 외에도 교사 모델의 확률 분포에 대한 정보를 풍부하게 학습할 수 있다.

$$q_i = \frac{\exp(z_i^t/\tau)}{\sum_i \exp(z_i^t/\tau)} \quad (1)$$

$$p_i = \frac{\exp(z_i^s/\tau)}{\sum_i \exp(z_i^s/\tau)} \quad (2)$$

식 (1)과 (2)는 이러한 지식 증류 과정을 나타내며, 그림 1 (a)처럼 각각 교사 모델의 로짓과 학생 모델의 로짓을 통해 소프트맥스와 온도 매개변수로 변환된 분포가 Kullback-Leibler 발산 손실(KL

divergence loss)을 기반으로 학생 모델이 교사 모델의 분포를 모방하도록 돕는다.

3.2 공유 온도 문제와 로짓 표준화

최근 연구에 따르면, 교사 모델과 학생 모델이 동일한 온도를 사용하는 방식에 대해 문제점이 제기되었다[6]. 이는 모델 간의 크기와 용량 차이로 인해 학생 모델이 교사 모델처럼 넓은 확률 분포 범위를 생성하기 어려울 수 있기 때문이다[19]. 특히 동일한 온도를 공유하는 경우, 학생 모델의 출력 표준편차가 제한받아 분포가 충분히 넓게 표현되지 못하는 문제가 발생할 수 있다. 이러한 한계를 극복하기 위해 최근 연구에서는 로짓 표준화(Logit standardization) 기법을 통해 교사 모델과 학생 모델 각각의 표준편차에 맞춘 온도를 설정하는 방식을 제안하며, 이를 통해 지식 증류 과정의 효율성을 높이고자 한다. 로짓 표준화를 통해 학생 모델이 교사 모델의 분포 특성을 보다 정확히 학습할 수 있으며, 지식 증류의 효과도 극대화될 수 있다.

3.3 얼굴 파싱에서의 로짓 표준화

$$\bar{X} = \frac{1}{K} \sum_{k=1}^K X^{(k)} \quad (3)$$

$$\sigma(X) = \sqrt{\frac{1}{K} \sum_{k=1}^K (X^{(k)} - \bar{X})^2} \quad (4)$$

$$Z(X; \tau) = (X - \bar{X}) / \sigma(X) / \tau \quad (5)$$

얼굴 파싱 작업에서 교사 모델과 학생 모델의 온도를 공유함으로써 발생하는 문제를 해결하기 위해, 본 연구에서는 로짓 표준화를 적용하여 그 효과를 검토하였다. 로짓 표준화는 그림 1 (b)와 같이 로짓의 평균과 표준편차를 사용해 교사와 학생 모델 각각의 출력 분포를 조정함으로써 온도 공유 문제를 보완한다. 식 (3)과 (4)는 로짓의 평균과 표준편차를, 식 (5)는 이를 통한 표준화 과정을 나타낸다. 여기서 식에서 X 는 모델의 최종 출력인 로짓을, τ 는 온도를 나타내며, 이를 소프트맥스 함수에 입력하여 지식 증류를 수행한다.

분류 문제에서 처음으로 제안된 로짓 표준화를 얼굴 파싱에 적용하기 위해서는 픽셀 별 모든 로짓에 대해서 평균이 0, 표준편차가 1이 되도록 표준화해 주어야 한다. 서로 다른 픽셀의 로짓은 서로 다른 평균, 표준편차 분포를 가지고 있는데, 이를 동일하게 동일한 평균, 표준편차로 변환하면, 픽셀 간의 정보가 손실되는 문제가 발생한다. 이러한 방식의 문제는 아래 실험을 통해서도 알 수 있다.

3.4 로짓 표준화 적용 및 시각화

ResNet101+FCN을 교사 모델로, MobileNetV4+DeepLabV3를 학생 모델로 설정하여 세 가지 방식으로 실험을 수행하였다: 첫째, 기본 학습 방법, 둘째, 전통적 지식 증류 방법(vanilla -KD), 셋째, 로짓 표준화가 포함된 지식 증류 방법이다. 표 1의 실험 결과는 지식 증류를 적용하지 않은 경우보다 vanilla-KD 방법이 높은 성능을 보였음을 나타낸다. 그러나, 로짓 표준화를 적용했을 때 성능이 오히려 감소하는 현상이 확인되었다. 또한 그림 2와 같이 다양한 모델에서 픽셀별 로짓의 평균과 표준편차를 시각화하였다.

표 1. 기존 방법의 기존 증류 성능 (F1-score)

Table 1. Performance of the existing methods for knowledge distillation (F1-score)

Not applied	vanilla-KD	KD-STD
83.97	85.37	83.51

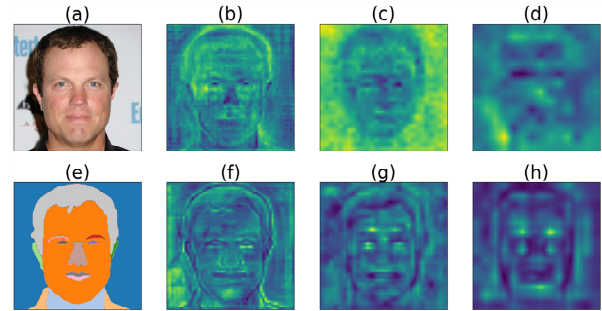


그림 2. 여러 모델에서 픽셀별 로짓의 평균과 표준편차 (a) 입력 이미지, (b) Resnet101 + FCN 모델 로짓 평균, (c) Mobilenet3 + Deeplabv3 모델 로짓 평균, (d) Resnet18 + FCN 모델 로짓 평균, (e) 출력 레이블, (f) Resnet101 + FCN 모델 로짓 표준편차, (g) Mobilenet3 + Deeplabv3 모델 로짓 표준편차, (h) Resnet18 + FCN 모델 로짓 표준편차

Fig. 2. Mean and standard deviation of per-pixel logit across multiple models (a) Input image, (b) Logit mean from the ResNet101 + FCN model, (c) Logit mean from the MobileNetV3 + DeepLabV3 model, (d) Logit mean from the ResNet18 + FCN model, (e) Output labels, (f) Logit standard deviation from the ResNet101 + FCN model, (g) Logit standard deviation from the MobileNetV3 + DeepLabV3 model, (h) Logit standard deviation from the ResNet18 + FCN model

그림 2의 (b)와 (f)에 나타난 바와 같이, ResNet101+FCN과 같은 고성능 모델은 출력 레이블과 유사한 뚜렷한 분포를 보이며, 이는 픽셀별 로짓의 평균과 표준편차가 명확하게 드러남을 보여준다. 반면, 성능이 낮은 ResNet18 + FCN 모델은 분포가 상대적으로 흐릿하게 나타나 로짓의 명확성이 떨어진다. 이는 교사 모델이 픽셀별 로짓 분포뿐만 아니라, 픽셀 간 로짓 분포에서 중요한 정보를 포함하고 있음을 시사한다. 이러한 픽셀 간 정보는 기존 세그먼트 문제에서 시사하는 바와 동일하다 [7][8]. 그러나 교사 모델과 학생 모델의 로짓을 동일하게 표준화할 경우, 교사 모델의 픽셀 간 로짓 분포 정보가 학생 모델로 효과적으로 전달되지 않는 문제가 발생한다.

3.5 로짓 분포 변환

그림 1은 전통적인 지식 증류 방법과 로짓 표준화 지식 증류 방법, 제안된 로짓 분포 변환 지식 증류 방법을 비교한 결과를 보여준다. 그림 1 (a)는 기본적인 지식 증류 방법, 그림 1 (b)는 3.3장에서 설명한 로짓 표준화가 적용된 지식 증류 방법을 나타낸다. 세 모델의 주요한 차이점은 logit를 처리하는 부분이 전통적인 지식 증류 방식은 로짓에 추가적인 처리를 하지 않고, 로짓 표준화 방식은 교사 모델과 학생모델의 로짓을 표준화해 준다. 제안한 방식은 학생 모델의 로짓만 분포 변환을 진행한다. 로짓 표준화 방식은 교사 모델과 학생 모델 모두에 표준화를 적용하여 모든 픽셀별 로짓이 평균 0, 표준편차 1의 분포를 따르도록 한다. 그러나 이러한 접근법은 교사 모델의 픽셀 간 로짓 분포를 그대로 학생 모델에 전달하기 어렵고, 중요한 분포 정보를 손실하는 문제가 발생한다.

$$T(X_s, \tau) = ((X_s - \bar{X}_s) / \sigma(X_s) \times \sigma(X_t) + \bar{X}_t) / \tau \quad (6)$$

이를 해결하기 위해 본 연구에서는 교사 모델의 픽셀 간 로짓 분포를 유지하고, 그림 1 (C) 와 같이 학생 모델에만 로짓 분포 변환을 적용하여 최적의 온도 값을 조정하는 방식을 제안한다. 여기서 X_s , X_t 는 각각 학생 모델과 교사 모델의 로짓을 나타내며, 식 (6)에서는 학생 모델의 로짓을 표준화한 후 이를 교사 모델의 로짓 분포로 변환하는 방식을 보여준다. 이를 통해 학생 모델이 교사 모델의 중요한 분포 정보를 더욱 효과적으로 학습할 수 있다.

제안된 방법은 [2]에서 언급된 로짓 표준화의 장점을 그대로 살리며, 평균 0, 유한 표준편차, 단조성, 경계성 등의 특성을 유지한다. 학생 모델의 분포를 평균 0, 표준편차 1로 표준화한 후 교사 모델의 분포로 변환하여 기존 연구에서와 동일하게 교사와 학생 모델 모두 평균이 0인 상태를 유지할 수 있다. 또한 교사 모델의 표준편차가 무한한 경우에도 학생 모델은 이를 반영하여 변환되므로 유한 표준편차의 장점을 확보할 수 있다. 뿐만 아니라, 이전 연구와 마찬가지로 학생 모델의 로짓 변환이 선

형 변환 함수이므로 동일한 순위를 유지하며, 교사 모델의 로짓 분포가 과도하게 커지더라도 학생 모델의 분포는 교사 모델의 표준편차를 곱하여 계산되므로 학생 모델의 분포가 과도하게 커지는 것을 방지할 수 있다. 이러한 특성을 통해 로짓 분포 변환은 경계성 속성을 유지하며, 교사 모델의 픽셀 간 로짓 분포 정보를 효과적으로 전달할 수 있는 장점을 가진다.

IV. 실험

4.1 데이터셋 및 평가지표

본 연구에서는 얼굴 파싱 문제에서 지식 증류의 효과를 검증하기 위해 CelebAMask-HQ 데이터셋[9]을 사용하였다. CelebAMask-HQ 데이터셋은 학습 세트, 검증 세트, 테스트 세트로 나뉘며, 각각 24,183개, 2,993개, 2,824개의 데이터로 구성되어 있다. 이 데이터셋은 다양한 성별과 인종을 포함한 얼굴 이미지를 다루며, 얼굴의 각 부위를 19개의 세부 클래스(눈, 코, 입, 헤어라인 등)로 분할한 마스크 레이블을 제공한다. 이러한 세부 클래스를 통해 얼굴의 각 부위에 대한 파싱 정확도를 측정할 수 있으며, 이는 연구에서 제안한 지식 증류 기법이 다양한 얼굴 속성에 대해 얼마나 효과적으로 작동하는지 평가하는 데 적합한 기준이 된다.

평가 지표로는 정확도(Accuracy)와 F1-score를 사용하여 성능을 측정하였으며, 이를 통해 모델이 얼굴의 각 부위를 얼마나 정밀하게 분할하고 예측하는지 평가하였다. 정확도는 각 클래스의 예측 정확도를 종합한 값으로 모델의 전반적인 성능을 나타내며, F1-score는 정밀도와 재현율의 조화를 반영하여 예측 성능의 균형을 평가하는 지표이다. 이를 통해 다양한 얼굴 부위에서 모델의 예측 정확도를 세부적으로 비교하고 분석할 수 있다.

4.2 구현 상세

본 연구에서는 ResNet101+FCN 모델을 교사 모델로 설정하여, 해당 모델의 지식을 학생 모델에 전달

하는 지식 증류 과정을 수행하였다. 학생 모델로는 MobileNet-v3+DeepLabV3, ResNet34+FCN, ResNet18+FCN 세 가지 경량화 모델을 선정하였으며, 다양한 수준의 복잡도와 성능을 갖춘 모델을 활용하여 제안한 지식 증류 기법이 각 모델에서 어떤 성능 향상을 이루는지 비교 분석하였다.

학습 과정에서 모든 모델에 대해 배치 크기 (Batch size)는 5, 학습률(Learning rate)은 0.01로 설정하였으며, 최적화 기법으로는 SGD(Stochastic Gradient Descent)를 사용하여 총 100 에폭(Epoch) 동안 학습을 진행하였다. 이러한 설정은 교사 모델과 학생 모델 간의 일관된 학습 환경을 제공하여, 각 지식 증류 방법이 성능 향상에 미치는 영향을 공정하게 비교할 수 있도록 한다.

4.3 실험결과

표 2와 표 3은 지식 증류를 적용하지 않은 기본 방법, 전통적인 지식 증류(vanilla-KD), 로짓 표준화 지식 증류(KD-STD), 그리고 본 연구에서 제안한 방법을 비교한 성능 결과를 보여준다. 실험 결과, 지식 증류를 적용하지 않은 경우, 교사 모델인 ResNet101+FCN이 F1-score, Accuracy 각각 87.1 95.4로 가장 높은 성능을 기록하였으며, 그 뒤를 이어 DeepLabv3+MobileNet, ResNet34+FCN, ResNet18+FCN 순으로 성능 차이를 보였다. 이는 각 모델의 구조적 복잡도가 성능에 직접적으로 영향을 미친다는 점을 확인할 수 있는 결과이다.

표 2. 지식 증류 성능 비교 (F1-score)
Table 2. Knowledge distillation performance comparison (F1-score)

Student model	Knowledge distillation method			
	Not applied	vanilla-KD	KD-STD	ours
Resnet101+FCN (teacher)	87.1	-	-	-
Deeplabv3_mobilenet	83.97	85.37	83.51	85.37
Resnet34 +FCN	73.78	74.56	66.94	75.42
Resnet18 +FCN	73.41	74.51	67.3	75.01

표 3. 지식 증류 성능 비교 (정확도)
Table 3. Knowledge distillation performance comparison (Accuracy)

Student model	Knowledge distillation method			
	Not applied	vanilla-KD	KD-STD	ours
Resnet101+FCN (teacher)	95.04	-	-	-
Deeplabv3_mobilenet	94.34	94.72	94.32	94.77
Resnet34 +FCN	91.89	92.72	91.48	92.76
Resnet18 +FCN	91.56	92.53	91.15	92.56

전통적인 지식 증류 방법(vanilla-KD)을 적용한 경우, 모든 모델과 평가 지표에서 지식 증류를 적용하지 않은 기본 방법보다 F1-score 각각 DeepLabv3+MobileNet 1.4 ResNet34+FCN 0.78, ResNet18+FCN 1.1 향상되었으며, Accuracy 에서도 비슷한 추세를 보인다. 이는 얼굴 파싱에서 로짓 기반 지식 증류의 효과를 입증한다. vanilla-KD 방식은 학생 모델이 교사 모델의 출력을 학습함으로써 성능을 끌어올릴 수 있음을 확인하게 한다. 그러나 로짓 표준화를 적용한 KD-STD 방식의 경우, F1-score에서 DeepLabv3+MobileNet, ResNet34+FCN, ResNet18+FCN 각각 83.51, 66.94, 67.3로 모든 지식 증류 모델보다 성능이 낮았으며, 심지어 지식 증류를 적용하지 않은 모델보다도 낮은 성능을 기록하였다. 이는 2.3장에서 언급된 대로, 교사 모델의 픽셀 간 정보가 학생 모델에 제대로 전달되지 않아 발생한 문제로 판단된다.

본 연구에서 제안한 방법은 vanilla-KD의 공유 온도 문제와 KD-STD 방식의 교사 모델 픽셀 간 정보 손실 문제를 보완하여, F1-score에서 DeepLabv3+MobileNet, ResNet34+FCN, ResNet18+FCN 각각 85.37, 75.42, 75.01, Accuracy에서 각각 94.77, 92.76, 92.56으로 vanilla-KD보다 우수한 성능을 보였다. 제안된 기법은 학생 모델이 교사 모델의 중요한 분포 정보를 더 효과적으로 학습할 수 있도록 하여 지식 증류의 성능을 극대화할 수 있었다. 이를 통해 얼굴 파싱 작업에서 지식 증류 기법이 모델의 경량화와 실시간 성능을 유지하면서도 높은 정확도를 제공할 수 있음을 확인할 수 있었다.

V. 결론 및 향후 계획

본 연구에서는 얼굴 파싱 분야에 지식 증류를 효과적으로 적용하기 위해 기존 지식 증류 방식의 단점인 공유 온도 문제를 보완하는 새로운 접근을 제안하였다. 이를 위해 먼저 로짓 표준화 방식을 얼굴 파싱 작업에 적용하고, 성능 저하의 원인을 세부적으로 분석하였다. 이후, 분석 결과를 바탕으로 기존 방식의 한계를 극복할 수 있는 로짓 분포 변환 기법을 제안하였으며, 이를 통해 학생 모델이 교사 모델의 중요한 픽셀 간 분포 정보를 보다 정확하게 학습할 수 있도록 하였다. CelebAMask-HQ 데이터셋을 활용한 실험을 통해 본 연구에서 제안한 방법이 기존 지식 증류 방식에 비해 성능을 유의미하게 향상함을 확인하였다.

향후 연구에서는 다양한 로짓 기반 지식 증류 기법과 본 연구에서 제안한 로짓 분포 변환 방법을 융합하여 제안된 방법의 추가적인 활용 가능성을 탐구하고, 더욱 폭넓은 응용 분야에서의 효용성을 입증할 수 있을 것으로 기대된다.

References

- [1] S. Liu, J. Yang, C. Huang, and M.-H. Yang, "Multi-objective convolutional learning for face labeling", Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, pp. 3451-3459, Jun. 2015. <https://doi.org/10.1109/cvpr.2015.7298967>.
- [2] C.-H. Lee, Z. Liu, L. Wu, and P. Luo, "MaskGAN: Towards diverse and interactive facial image manipulation", Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 5549-5558, Jun. 2020. <https://doi.org/10.1109/cvpr42600.2020.00559>.
- [3] M. Sarkar, N. SR, M. Hemani, R. Jain, and B. Krishnamurthy, "Parameter efficient local implicit image function network for face segmentation", Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, pp. 6789-6798, Jun. 2023. <https://doi.org/10.1109/cvpr52729.2023.02009>.
- [4] Z. Xue, W. Mao, and W. Jiang, "EHANet: Efficient hybrid attention network towards real-time semantic segmentation", 2020 IEEE 6th International Conference on Computer and Communications (ICCC), Chengdu, China, pp. 1234-1241, Dec. 2020. <https://doi.org/10.1109/iccc51575.2020.9345050>.
- [5] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network", arXiv preprint arXiv:1503.02531, pp. 1-9, Mar. 2015. <https://doi.org/10.21437/interspeech.2023-190>.
- [6] S. Sun, W. Ren, J. Li, R. Wang, and X. Cao, "Logit standardization in knowledge distillation", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 4567-4576, Jun. 2024. <https://doi.org/10.1109/cvpr52733.2024.01489>.
- [7] Y. Liu, K. Chen, C. Liu, Z. Qin, Z. Luo, and J. Wang, "Structured knowledge distillation for semantic segmentation", Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, pp. 4556-4565, Jun. 2019. <https://doi.org/10.1109/cvpr.2019.00271>.
- [8] C. Shu, Y. Liu, J. Gao, Z. Yan, and C. Shen, "Channel-wise knowledge distillation for dense prediction", Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, pp. 5168-5177, Oct. 2021. <https://doi.org/10.1109/iccv48922.2021.00526>.
- [9] S. Raghavendra, S. K. Abhilash, V. M. Nookala, N. N. Srinidhi, and N. D. Adesh, "BADANet: Boundary Aware Dilated Attention Network for Face Parsing", IEEE Access, Vol. 11, pp. 12345-12355, Jun. 2023. <https://doi.org/10.1109/access.2023.3319561>.
- [10] S. Liu, J. Shi, J. Liang, and M.-H. Yang, "Face

- parsing via recurrent propagation", arXiv preprint arXiv:1708.01936, pp. 1-10, Aug. 2017. <https://doi.org/10.48550/arXiv.1708.01936>.
- [11] Z. Li, X. Li, L. Yang, B. Zhao, R. Song, L. Luo, J. Li, and J. Yang, "Curriculum temperature for knowledge distillation", Proc. of the AAAI Conference on Artificial Intelligence (AAAI), Washington DC, USA., Vol. 37, No. 2, pp. 1234-1241, Feb. 2023. <https://doi.org/10.1609/aaai.v37i2.25236>.
- [12] Y. Jin, J. Wang, and D. Lin, "Multi-level logit distillation", Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, pp. 1234-1245, Jun. 2023. <https://doi.org/10.1109/cvpr52729.2023.02325>.
- [13] B. Zhao, Q. Cui, R. Song, Y. Qiu, and J. Liang, "Decoupled knowledge distillation", Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, pp. 5549-5560, Jun. 2022. <https://doi.org/10.1109/cvpr52688.2022.01165>.
- [14] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "Fitnets: Hints for thin deep nets", arXiv preprint arXiv:1412.6550, pp. 1-13, Dec. 2014. <https://doi.org/10.48550/arXiv.1412.6550>.
- [15] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer", arXiv preprint arXiv:1612.03928, pp. 1-12, Dec. 2016. <https://doi.org/10.48550/arXiv.1612.03928>.
- [16] Y. Tian, D. Krishnan, and P. Isola, "Contrastive representation distillation", arXiv preprint arXiv:1910.10699, pp. 1-10, Oct. 2019. <https://doi.org/10.48550/arXiv.1910.10699>.
- [17] P. Chen, S. Liu, H. Zhao, and J. Jia, "Distilling knowledge via knowledge review", Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, pp. 1234-1243, Jun. 2021. <https://doi.org/10.1109/cvpr46437.2021.00497>.
- [18] S. Han and H. Yoon, "Advancing Face Parsing in Real-World: Synergizing Self-Attention and Self-Distillation", IEEE Access, Vol. 12, pp. 29812-29823, Feb. 2024. <https://doi.org/10.1109/access.2024.3368530>.
- [19] S. I. Mirzadeh, M. Farajtabar, A. Li, N. Levine, A. Matsukawa, and H. Ghasemzadeh, "Improved knowledge distillation via teacher assistant", Proc. of the AAAI Conference on Artificial Intelligence (AAAI), New York, USA, Vol. 34, No. 4, pp. 1234-1242, Apr. 2020. <https://doi.org/10.1609/aaai.v34i04.5963>.

저자소개

배 지 호 (Jiho Bae)



2019년 3월 ~ 현재 :
경상국립대학교 컴퓨터공학과
학부과정
관심분야 : 인공지능, 컴퓨터비전,
증강현실

이 수 원 (Suwon Lee)



2012년 7월 : 한국과학기술원
전산학과(공학석사)
2017년 7월 : 한국과학기술원
전산학과(공학박사)
2018년 3월 ~ 현재 :
경상국립대학교 컴퓨터공학과
부교수
관심분야 : 증강현실, 컴퓨터비전, 인간-컴퓨터 상호작용