

논문 2020-15-06

블록 계층별 재학습을 이용한 다중 힌트정보 기반 지식전이 학습

(Multiple Hint Information-based Knowledge Transfer with Block-wise Retraining)

배 지 훈*
(Ji-Hoon Bae)

Abstract: In this paper, we propose a stage-wise knowledge transfer method that uses block-wise retraining to transfer the useful knowledge of a pre-trained residual network (ResNet) in a teacher-student framework (TSF). First, multiple hint information transfer and block-wise supervised retraining of the information was alternatively performed between teacher and student ResNet models. Next, Softened output information-based knowledge transfer was additionally considered in the TSF. The results experimentally showed that the proposed method using multiple hint-based bottom-up knowledge transfer coupled with incremental block-wise retraining provided the improved student ResNet with higher accuracy than existing KD and hint-based knowledge transfer methods considered in this study.

Keywords : Multiple hint information, Block-wise retraining, Knowledge transfer, Deep learning, Residual network

1. 서 론

현재, 4차 산업혁명 시대를 맞이하여 전 세계적으로 지능 정보화 및 융합을 추구하는 새로운 변화의 시기를 맞이하고 있다. 4차 산업혁명 시대의 주요 기술 중의 하나인 인공지능 기술은 빅데이터 기술과 더불어 앞으로 글로벌 시장 질서를 주도해 나갈 핵심 기술로 주목받고 있다. 특히, 인공지능 기술의 한 분야인 딥러닝 (Deep Learning)은 인간의 뇌를 모방한 신경망 (Neural Network) 학습 분야로서, 영상, 비디오 등의 컴퓨터 비전 분야에서 인간의 능력을 추월하는 탁월한 능력을 보여주고 있으며, 현재 음성인식 및 자연어 처리 등 다양한 분야로의 적용 범위가 확대되고 있는 추세이다. 하지만, 최근 딥러닝 성공의 주역인 합성곱 신경망 (convolutional neural networks: CNN) 구조 [1,

2]를 가지는 대부분의 최신 딥러닝 모델들은 일반적으로 수천 만개 이상의 수많은 학습 변수들로 이루어진 층수가 매우 깊고 네트워크 구조가 넓고 복잡한 구조를 가지고 있다 [3-7]. 특히, 합성곱 신경망 기반의 잔차 신경망 (residual networks: ResNet)은 지름 연결선 (shortcut connection)을 이용한 잔차 블록 (residual block) 구조를 제안함으로써, 30층 혹은 100층 이상의 심층 네트워크 구조를 가지고 있다 [6]. 비록, 이러한 딥러닝 모델들이 잘 최적화 및 학습되어 제공되더라도, 사물인터넷 (Internet of Things: IoT) 및 모바일, 임베디드 환경 [8, 9]과 같은 컴퓨팅 자원이 제한적인 분야에서는 이러한 복잡한 딥러닝 모델들을 직접 경량화 서비스에 적용하기가 쉽지 않을 수 있다.

이러한 문제점을 극복하기 위하여 사전 학습 완료된 모델을 선생 모델로 정의하여 유용한 정보를 추출하고, 상기 추출된 선생 정보를 이전받아 학습을 수행하는 또 다른 모델을 학생 모델로 정의한 선생-학생 프레임워크에 대한 연구가 진행되어 왔다 [10, 11]. 이러한, 선생-학생 프레임워크 구조에서는 선생 모델로부터 유용한 지식을 이전받아 학습을 수행하기 때문에 적용하고자 하는 분야에 적

*Corresponding Author (jihbae@cu.ac.kr)

Received: Feb. 29, 2020, Revised: Mar. 23, 2020,

Accepted: Mar. 28, 2020.

J.-H. Bae: Daegu Catholic University
(Assistant Prof.)

합하게 적응적으로 학생 모델의 층수를 조절함으로써 모델 경량화를 달성할 수 있다. [10]에서는 사전 학습 완료된 선생 모델로부터 지식을 추출하는 knowledge distillation (KD)이란 용어가 처음으로 등장하였다. 상기 기법에서는 선생 모델의 마지막 출력층에서의 완화된 클래스 확률값 정보를 KD로 간주하고 이를 학생 모델로 이전하여 학습을 수행하였다. [11]에서는 [10]의 KD 정보뿐만 아니라 선생 모델의 중간 은닉층 (hidden layer)인 힌트층 (hint layer)에서의 출력인 특징맵 (feature maps)을 닦아가도록 학습을 수행하는 힌트-KD 지식전이 학습기법을 제안하였다. 상기에서 소개한 연구 결과에 따르면, 선생 지식없이 전통적인 역전파 (backpropagation) 기법으로 학습한 학생 모델보다 상기 KD 혹은 힌트정보 등을 이전받아 학습을 수행한 학생 모델의 성능이 더 우수함을 보여주었다.

본 연구에서는 기존 힌트정보 학습 기반의 지식전이 (knowledge transfer) 성능을 향상시키기 위하여 블록 계층별 재학습을 이용한 다중 힌트정보 기반 지식전이 학습 알고리즘을 제안하고자 한다. 먼저, 중간 은닉층에서의 단일 힌트정보를 이용하는 기존 힌트 학습방식 [11]과 달리 다중 힌트층에서 특징맵들을 추출하여 학생 모델이 선생 모델의 다중 힌트정보들을 닦아가도록 학습한다. 이를 위하여, 낮은 힌트층에서부터 높은 힌트층까지 순차적으로 힌트정보들을 학생모델로 전달한다. 이때, 각 힌트층에서 힌트정보 학습 완료 후, 학생 모델의 각 블록 계층별로 순차적으로 실제 레이블링 된 데이터를 이용하여 지도학습 방법으로 미세조정 학습을 수행한다. 마지막으로, 기존 지식전이 학습 방법과 동일하게 KD 정보를 이전하는 학습을 수행한다. 본 연구에서는 최신 딥러닝 모델로 널리 각광 받고 있는 ResNet을 적용한 선생-학생 프레임워크 상에서의 지식전이 성능을 분석하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문의 관련 연구들에 대하여 간략히 기술하고, 3장에서는 선생-학생 프레임워크 상에서 다중 힌트정보 기반의 블록 계층별 재학습에 대한 지식전이 방법을 제안한다. 4장에서는 공신력 있는 벤치마크 데이터셋들을 이용하여 제안 기법에 대한 성능 평가를 수행한 결과를 보여준다. 마지막으로, 5장에서는 본 연구 결과에 따른 결론을 도출한다.

II. 관련 연구

1. KD 학습 기반의 선생-학생 프레임워크

본 절에서는 선생-학생 프레임워크 기반의 기존 지식전이 기법으로 KD 학습 기법 [10]에 대하여 간략히 기술하고자 한다. 먼저, KD 학습 방법은 선생 모델의 최종 출력층에서의 소프트맥스 (softmax) 클래스 (class) 확률값을 완화상수 (softening factor)로 완화하여 선생 모델 내부에 축적된 지식을 추출한다. 다음으로, 상기 완화된 클래스 확률분포를 학생 모델로 전달하기 위하여 다음 식 (1)과 같은 비용 함수 J_{KD} 를 구성하여 KD 지식전이 학습을 수행한다 [10].

$$J_{KD}(\widehat{W}_s) = CE(Y_{true}, P_T) + \lambda CE(P_T, P_S) \quad (1)$$

여기서, Y_{true} 는 실제 레이블 (true label)을 가지는 참값 (ground truth)을 의미하고, λ 는 KD 가중치를, CE는 크로스 엔트로피 (cross-entropy) 손실함수를,

$$P_T = \text{softmax}(a_T/\tau) \quad \text{및}$$

$P_S = \text{softmax}(a_S/\tau)$ 를 각각 나타낸다. 완화된 소프트맥스 출력에서 a_T 는 선생 모델에서 소프트맥스 출력 전의 입력 벡터를, a_S 는 학생 모델에서 소프트맥스 출력 전의 입력 벡터를, τ 는 완화상수를 각각 나타낸다. [10]의 결과에 따르면 완화된 클래스 확률분포는 완화상수 적용 없이 클래스 확률분포 그대로를 학생 모델로 이전하는 경우보다 선생 모델이 학습한 지식을 좀 더 일반화하여 전달할 수 있음을 보여 주었다.

2. 힌트-KD 학습 기반의 선생-학생 프레임워크

힌트정보 [11]는 사전 학습 완료된 선생 모델의 중간 은닉층인 힌트층 (hint layer)에서 출력된 특징맵을 의미한다. 힌트층에서 추출한 힌트정보를 학생 모델로 이전하기 위하여 학생 모델의 중간 은닉층인 가이드층 (guided layer)에서의 특징맵을 출력하고 다음의 식 (2)와 같이 두 특징맵들 사이의 유클리디안 (Euclidean) 거리가 최소가 되도록 비용함수 J_{HG} 를 설정하여 힌트정보 학습을 수행한다.

$$J_{HG}(\widehat{W}_G) = \frac{1}{2} \| H(X; W_H) - G(X; W_G) \|^2 \quad (2)$$

여기서, H 와 G 은 선생 모델의 입력층으로부터 힌트층까지의 가중치에 해당하는 W_H 로부터 활성화 함수 (activation function)들을 통과한 출력을, 학생 모델의 입력층으로부터 가이드층까지의 가중치에

해당하는 W_G 로부터 활성화함수들을 통과한 출력을 각각 나타낸다.

다음으로, 상기에서 구한 가중치에다가 그 나머지 가이드층에서부터 최종 출력층까지의 가중치들을 랜덤으로 초기화한 다음, II.1 절에서 기술한 식 (1)의 KD정보 학습을 단계적으로 수행한다. 기존 KD 학습 방식과 비교하여 KD 정보뿐만 아니라 은닉층에 숨겨진 힌트정보 또한 추가적으로 고려하기 때문에 지식전이 성능에 유리한 장점을 가지고 있다. 하지만, 선생 모델에 숨겨진 유용한 정보를 단일 힌트 정보만으로 표현하기에는 효과적인 지식전이 성능을 달성하는데 그 한계점을 가질 수 있다.

III. 제안 방법

1. 선생-학생 프레임워크 기반 다중 힌트정보

선생-학생 프레임워크 상에서 본 연구에서 제안하는 다중 힌트정보 추출 구조는 다음 그림 1과 같이 주어진다. 이때, 세 개의 잔차블럭들을 가지는 ResNet 딥러닝 모델을 고려한다 [6]. 단일 힌트정보만을 이용하는 기존 기법과 달리 본 연구에서는 낮은 힌트층에서부터 높은 힌트층까지 추출한 다중 힌트 정보들을 지식전이 프레임워크에 적용한다. 본 연구에서는 ResNet 모델에 대한 힌트정보 추출의 일관성을 유지하기 위하여 힌트정보 추출을 잔차블럭의 특징맵 필터개수가 변화하는 지점으로 선택한다. 예를 들어, 그림 1의 구조에서는 서로 다른 특징맵 필터개수를 가지는 세 개의 잔차블럭들로 구성되어 있기 때문에 세 개의 힌트정보들을 추출할 수 있다. 일반적으로 사전 학습 완료된 선생 모델의 합성곱 신경망 구조에서는 낮은 수준의 섬세한 특징에서부터 높은 수준의 추상적 특징까지 유용한 정보들을 포함하고 있기 때문에, 다중 힌트 정보들을 이용하게 되면 좀 더 효과적으로 은닉층에 숨겨진 선생 지식들을 활용할 수 있다. 따라서, 그림 1의 구조를 기반으로 추출된 다중 힌트정보들을 학생 모델로 이전하기 위한 방법은 다음의 III.2 절에서 상세히 기술한다.

2. 다중 힌트정보 학습과 블럭 계층별 재학습을 통한 미세조정

먼저, 식 (3)의 비용함수 J_{HG_1} 를 이용하여 유클리디안 거리가 최소가 되도록 그림 2와 같이 첫 번째 힌트층에 대한 힌트정보를 학생모델로 이전한다.

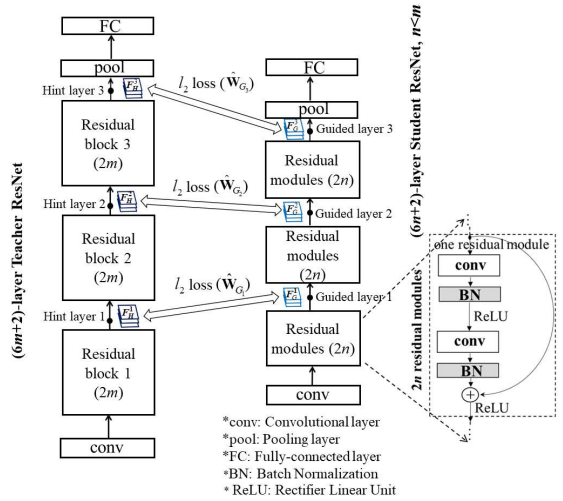


그림 1. 선생-학생 프레임워크 기반의 다중 힌트 정보 추출 구조

Fig. 1 Teacher-student framework structure for the extraction of multiple hint information

$$J_{HG_1}(W_{G_1}) = \frac{1}{2} \| H_1(X; W_{H_1}) - G_1(X; W_{G_1}) \|^2 \quad (3)$$

여기서, H_1 와 G_1 은 선생 모델의 입력층으로부터 첫 번째 힌트층까지의 가중치에 해당하는 W_{H_1} 로부터 활성화함수들을 통과한 출력을, 학생 모델의 입력층으로부터 첫 번째 가이드층까지의 가중치에 해당하는 W_{G_1} 로부터 활성화함수들을 통과한 출력을 각각 나타낸다.

식 (3)의 첫 번째 힌트정보 학습 종료 후, 그림 2와 같이 첫 번째 가이드층 위에 풀링층 (pooling layer) 및 전결합층 (fully-connected layer)들을 추가하여 실제 레이블 데이터에 대한 지도학습으로 미세조정을 수행한다.

다음으로, 그림 2의 첫 번째 블럭에 대한 미세조정 완료 후 식 (4)의 비용함수 J_{HG_2} 를 이용하여 유클리디안 거리가 최소가 되도록 그림 3과 같이 두 번째 힌트층에 대한 힌트정보를 학생모델로 이전한다.

$$J_{HG_2}(W_{G_2}) = \frac{1}{2} \| H_2(X; W_{H_2}) - G_2(X; W_{G_2}) \|^2 \quad (4)$$

여기서, H_2 와 G_2 는 선생 모델의 두 번째 힌트층까지의 가중치에 해당하는 W_{H_2} 로부터 활성화함수들을

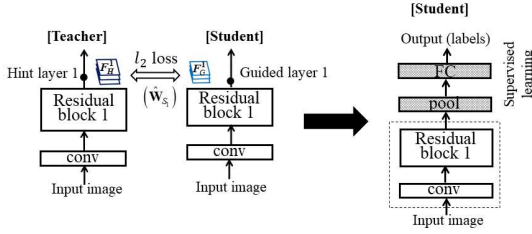


그림 2. 첫 번째 힌트정보 이전과 블록 재학습
Fig. 2 First hint information-based knowledge transfer and its block-wise retraining

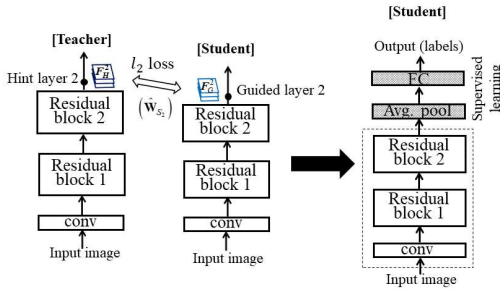


그림 3. 두 번째 힌트정보 이전과 블록 재학습
Fig. 3 Second hint information-based knowledge transfer and its block-wise retraining

통과한 출력을, 학생 모델의 두 번째 가이드층까지의 가중치에 해당하는 W_{G_2} 로부터 활성화함수들을 통과한 출력을 각각 나타낸다.

식 (4)의 두 번째 힌트정보 학습 종료 후, 그림 3과 같이 두 번째 가이드층 위에 풀링층 및 전결합층들을 추가하여 실제 레이블 데이터에 대한 지도 학습으로 미세조정을 수행한다.

마지막으로, 앞에서 수행한 학습 절차와 동일한 방식으로 식 (5)의 비용함수 J_{HG_3} 를 이용하여 그림 4와 같이 마지막 세 번째 힌트정보 학습 및 블록 재학습을 통한 미세조정을 수행한다. 이때, 각 블록에 대한 미세조정 시 선생 모델 없이 학생 모델만으로 실제 레이블 데이터를 이용하여 지도학습을 수행한다.

$$J_{HG_3}(\widehat{W}_{G_3}) = \frac{1}{2} \| H_3(X; W_{H_3}) - G_3(X; W_{G_3}) \|^2 \quad (5)$$

여기서, H_3 와 G_3 은 선생 모델의 세 번째 힌트층까지의 가중치에 해당하는 W_{H_3} 로부터 활성화함수들을 통과한 출력을, 학생 모델의 세 번째 가이드층까지

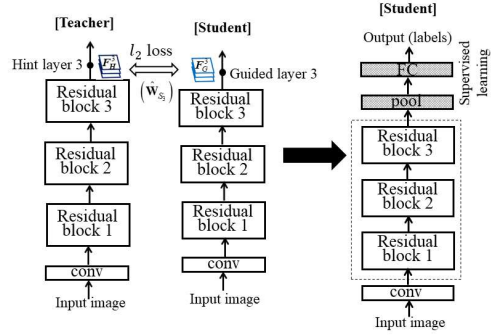


그림 4. 세 번째 힌트정보 이전과 블록 재학습
Fig. 4 Third hint information-based knowledge transfer and its block-wise retraining



그림 5. CIFAR-10 영상 데이터셋 [12]
Fig. 5 CIFAR-10 image dataset [12]

의 가중치에 해당하는 W_{G_3} 로부터 활성화함수들을 통과한 출력을 각각 나타낸다.

IV. 실험 결과

1. CIFAR-10 데이터셋을 이용한 학습 결과

본 연구에서 제안한 선생-학생 프레임워크 기반 지식전이 기법의 성능을 검증하기 위하여 다음 그림 5와 같이 국제적으로 널리 활용되고 있는 Canadian Institute For Advanced Research (CIFAR)-10 벤치마크 데이터셋을 이용한다 [12].

CIFAR-10 데이터셋은 총 10개 클래스로 구분되어 있으며, 60,000장의 32×32 훈련 (training) 컬러 영상들과 10,000장의 32×32 시험 (test) 컬러 영상들로 구성되어 있다. CIFAR-10 데이터셋을 이용한 학습 실험조건으로써, 사전 학습 완료된 91.6%의 인식 정확도 성능을 가지는 {16,32,64} 특징맵 필터개수를 가지는 26층 ResNet 선생 모델과 14층 ResNet 학생 모델로 구성된 선생-학생 프레임워크를 고려하고, 256 배치 단위로 학습을 수행한다.

먼저, KD 지식전이 학습 기법에 대하여, 초기 학습율 (learning rate)은 0.1로 시작하여 32,000 훈련횟수에서 0.01로 감소시킨 다음, 48,000 훈련 횟수에서 0.001로 다시 감소시킨 다음, 64,000 훈련 횟수에서 학습을 종료시킨다. 이때, 식 (1)의 KD 가중치는 학습 초기 시 $\lambda=4$ 로 시작하여 학습 종료 시 $\lambda=1$ 로 점진적으로 감소시킨다.

다음으로, 힌트-KD 지식전이 학습 기법에 대하여, 단계 1의 힌트정보 학습 시 초기 학습율을 0.0001로 시작하여 25,000 훈련횟수에서 1/10 비율로 감소시킨 다음, 35,000 훈련횟수에서 학습을 종료시킨다. 단계 2의 KD 학습 시에서는 상기 KD 지식전이 학습 방법과 동일하게 학습을 감소 스케줄링을 적용하여 총 64,000 훈련횟수 동안 학습시킨다.

마지막으로, 제안 방법에 대하여, 단계 1의 다중 힌트정보 학습 시 0.0001 학습율로 첫 번째 힌트정보에 대하여 3,000 훈련횟수를, 두 번째 힌트정보에 대하여 7,000 훈련횟수를, 세 번째 힌트정보에 대하여 15,000 훈련횟수를 순차적으로 수행하여 총 25,000 훈련횟수 동안 학습을 수행한 후 종료하였다. 이때, 각 힌트정보 학습 후 블록 계층별 재학습은 모두 동일하게 4,000 훈련횟수를 적용하였다. 다음으로, 단계 2의 KD 학습 시에서는 상기 기존 KD 지식전이 학습 기법과 동일하게 학습을 감소 스케줄링을 적용하여 총 64,000 훈련횟수 동안 학습을 수행하였다.

다음의 표 1은 상기에서 기술한 지식전이 기법들에 대하여 CIFAR-10 데이터셋을 이용한 인식 정확도들을 보여준다. 표 1의 결과에서와 같이 세 가지 기법 모두 선생 지식이 없는 14층 학생 모델만 있는 경우보다 더 좋은 정확도 성능을 보여주는 것을 관찰할 수 있다. 특히, KD 지식전이 학습기법 [10]과 힌트-KD 지식전이 학습기법 [11]과 비교 시, 후자 방법이 힌트정보와 KD정보의 두 가지 지식들을 고려하여 학생 모델로 전달하기 때문에 전자 방법보다 더 우수한 성능을 보여주는 것을 관찰할 수 있다. 하지만, 제안된 기법은 하나의 힌트정보만을 이용하는 기존 힌트-KD 지식전이 기법과 달리 다중 힌트 정보들을 고려하고 블록별 재학습을 통한 미세조정을 수행하기 때문에 보다 더 성능이 우수한 학생 모델을 최적화 할 수 있음을 관찰할 수 있다. 또한, 제안된 방법을 이용한 학생 모델의 네트워크를 구성하는 학습 파라미터 크기 및 층수가 52.57% (370.69k→175.81k)와 46.15% (26층→14층)으로 각각 감소하더라도, 26층 선생 모델과 거의 유사한 정확도 성능을 제공할 수 있는 반

표 1. CIFAR-10 데이터셋에 대한 인식 정확도[%]
Table 1. Accuracy [%] on CIFAR-10

Method	1 st	2 nd	3 rd	Avg.
*14-layer ResNet	90.36	90.23	90.30	90.29
KD [10]	90.58	90.69	90.4	90.55
Hint-KD [11]	90.46	90.85	90.79	90.7
Proposed method	91.59	91.61	91.57	91.59

* without teacher knowledge



그림 6. SVHN 영상 데이터셋 [13]
Fig. 6 SVHN image dataset [13]

면 (91.59% \cong 91.6%), 다른 기법들은 그렇지 못한 것을 관찰할 수 있다.

2. SVHN 데이터셋을 이용한 학습 결과

Street View House Number (SVHN) [13] 또한 국제적으로 널리 활용되고 있는 벤치마크 데이터 셋으로써, 그림 6과 같이 총 10개 클래스로 구분되어 있으며, 73,257장의 32×32 훈련 컬러 영상들과 26,032장의 32×32 시험 컬러 영상들로 구성되어 있다. SVHN 데이터셋을 이용한 학습 실험조건으로써, 사전 학습 완료된 96.32%의 인식 정확도 성능을 가지는 {64,128,256} 특징맵 필터개수를 가지는 26층 ResNet 선생 모델과 14층 ResNet 학생 모델로 구성된 선생-학생 프레임워크를 고려하였고, 100 배치 단위로 학습을 수행하였다. 세 가지 지식전이 학습 기법 모두 단계 2의 KD 학습을 36,000 훈련횟수를 적용한 것을 제외하고는 IV.1 절에서 수행하였던 CIFAR-10의 실험조건과 마찬가지로 동일한 학습을 감소 스케줄링 및 학습 조건들을 적용하였다.

그림 7의 결과에서와 같이 SVHN 데이터셋에 대해서도 다중 힌트정보와 각 블록별 재학습을 통

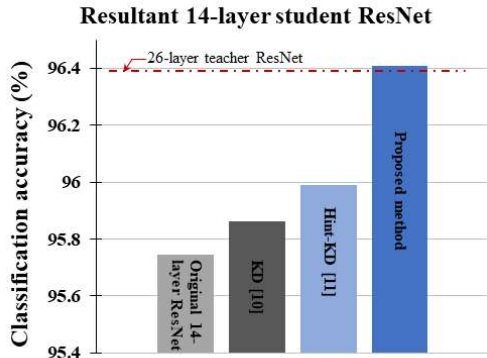


그림 7. SVHN 데이터셋에 대한 인식 정확도[%]
Fig. 7 Accuracy [%] on SVHN

한 미세조정을 수행함으로써 기존 두 지식전이 방법들보다 우수한 성능을 제공함을 관찰할 수 있다. 기존 기법 대비 본 연구에서 제안한 지식전이 기법을 적용한 학생모델 또한 네트워크 구조를 구성하는 학습 파라미터 복잡도가 5.88M에서 2.78M로 52.72% 경량화 되더라도 층수가 더 깊은 26층 선생모델보다 조금 더 높은 성능을 보여줄 수 있음을 관찰할 수 있다. 따라서, 표 1과 그림 7의 결과들을 토대로 제안한 기법이 본 연구에서 고려한 다른 기존 기법들보다 선생-학생 프레임워크 상에서 모델 경량화를 위한 지식전이에 더 유리함을 관찰할 수 있었다.

V. 결론

본 논문에서는 선생-학생 프레임워크 기반의 다중 힌트정보 추출과 블록별 재학습을 이용한 단계별 지식전이 기법을 제안하였다. 첫 번째 단계에서는 다중 힌트정보를 학생모델로 이전하기 위하여 낮은 힌트층에서부터 높은 힌트층까지 출력된 힌트 정보들을 순차적으로 이전하고, 각 힌트정보 학습 종료 후 지도학습 방법으로 블록별 미세조정을 수행하였다. 두 번째 단계에서는 선생 모델의 출력층에서의 완화된 클래스 확률분포를 학생 모델로 이전하는 학습을 수행하였다. 본 연구에서 제안한 지식전이 기법의 성능을 검증하기 위하여 국제적으로 공신력 있는 CIFAR-10 및 SVHN 데이터셋들에 대하여 인식 정확도 성능을 검증하였다. 본 연구의 실험결과에 따르면 기존 KD 및 힌트정보 학습 기법에 비하여 인식 정확도 성능이 더 우수함을 관찰할

수 있었다. 또한, 본 제안방법에서는 기존 기법 대비 다중 힌트정보들을 활용하여 지식전이를 수행하기 때문에 학생 모델이 경량화 되더라도 심층 선생 모델과 동등한 성능을 유지할 수 있음을 관찰할 수 있었다.

References

- [1] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based Learning Applied to Document Recognition," *Proceedings of IEEE*, Vol. 86, No. 11, pp. 1 - 46, 1998.
- [2] S. L. Phung, A. Bouzerdoum, "A Pyramidal Neural Network for Visual Pattern Recognition," *Journal of IEEE Transactions on Neural Networks*, Vol. 18, No. 2, pp. 329 - 343, 2007.
- [3] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Proceedings of Advances in Neural Information Processing Systems*, pp. 1106 - 1114, 2012.
- [4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going Deeper with Convolutions," *Proceedings of IEEE Conference on Computer Vision Pattern and Recognition*, pp. 1-9, 2015.
- [5] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-scale Image Recognition," *Proceedings of International Conference on Learning Represent*, pp. 1-14, 2015.
- [6] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-12, 2016.
- [7] G. Huang, Z. Liu, L. V. D. Maaten, K. Weinberger, "Densely Connected Convolutional Networks," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2261-2269, 2017.
- [8] D. Jeon, M.-S. Kim, S.-J. Ryu, D.-H. Lee, J.-K. Kim, "Fully Printed Chipless RFID Tags Using Dipole Array Structures with Enhanced

- Reading Ranges,” Journal of Korea Electromagnetic Engineering and Science, Vol. 17, No. 3, pp. 159 - 164, 2017 (in Korean).
- [9] Y.S.Yoon, H. Zo, M. Choi, D. Lee, and H.-W. Lee, “Exploring the Dynamic Knowledge Structure of Studies on the Internet of Things: Keyword Analysis,” Journal of ETRI Journal, Vol. 40, No. 6, pp. 745-758, 2018.
- [10] G. Hinton, O. Vinyals, J. Dean, “Distilling the Knowledge in a Neural Network,” Proceeding of Neural Information Processing Systems Workshop, pp. 1-19.
- [11] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, Y. Bengio, “Fitnets: Hints for Thin Deep Nets,” Proceedings of 5th International Conference on Learning Represent, pp. 1-13, 2015.
- [12] “CIFAR-10 and CIFAR-100 Datasets,” Available at: <https://www.cs.toronto.edu/~kriz/cifar.html>.
- [13] Y. Netzer, T. Wang, A. Coates, “Reading Digits in Natural Images with Unsupervised Feature Learning,” Proceedings of Neural Information Processing Systems Workshop on Deep Learning and Unsupervised Feature, pp. 1-9, 2011.

Ji-Hoon Bae (배지훈)



Ji-Hoon Bae received the B.S. degree in electronic engineering from Kyungpook National University, Daegu, Korea, in 2000, and the M.S. and Ph.D. degrees in Electrical Engineering from Pohang University of Science and Technology (POSTECH), Pohang, Gyeongsangbuk-do, Korea, in 2002 and 2016, respectively. From 2002 to 2019, Dr. Bae was with Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, as a principal researcher. He joined the faculty of the Department of AI • Big Data Engineering, Daegu Catholic University (DCU), Gyeongsan, Gyeongbuk, Korea in 2019. He is interested in deep learning and transfer learning, radar imaging, and optimized techniques.

Email: jihbae@cu.ac.kr