

일반논문 (Regular Paper)

방송공학회논문지 제28권 제5호, 2023년 9월 (JBE Vol.28, No.5, September 2023)

<https://doi.org/10.5909/JBE.2023.28.5.623>

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

테스트 단계 도메인 적응 모델의 성능 향상을 위한 기법 분석

이 호 준^{a)}, 이 승 환^{b)}, 정 인 영^{b)}, 홍 성 은^{c)‡}

Investigating Performance-Boosting Strategies for Test-Time Adaptation

Hojoon Lee^{a)}, Seunghwan Lee^{b)}, Inyoung Jeong^{b)}, and Sungeun Hong^{c)‡}

요 약

테스트 단계 도메인 적응은 소스 데이터로 학습된 모델을 테스트 상황에서 소스 데이터에 대한 접근 없이 타겟 도메인에 적응시키는 것을 목표로 한다. 본 논문은 기존 테스트 단계 도메인 적응 분야에 적용 시 일관되게 성능 향상을 보이는 방법론을 소개하고 분석한다. 첫번째 기법은 각 도메인마다 클래스 별 중심값을 구하고 같은 클래스끼리 가까워지도록 학습시킨다. 두번째는 손실함수를 구하는 과정에서 Cross-Entropy 대신 Symmetric Cross-Entropy를 사용하여 모델이 노이즈 레이블에 강건하도록 학습시킨다. 세번째는 모델의 분별력을 증가시키고 동시에 모델이 특정 클래스로 과하게 예측하는 문제를 해결하기 위해 Information-Maximization 손실함수를 적용한다. 기존 테스트 단계 도메인 적응 모델 중 가장 좋은 성능을 보인 모델에 해당 기법들을 일관된 성능 향상을 확인할 수 있었으며, 해당 기법은 특정한 모델에 국한되지 않기 때문에 다양한 모델에 확장 적용 가능한 장점을 지닌다.

Abstract

Test-time adaptation aims to adapt a model trained on source data to the target domain without access to the source data during the test phase. This paper introduces and analyzes a methodology that consistently improves performance when applied to the existing test-time adaptation methods. The first technique involves calculating the centroid for each domain and training the model to bring instances of the same class closer together. The second one trains the model to be robust to noisy labels by using Symmetric Cross-Entropy instead of Cross-Entropy during the loss function calculation. The last technique applies the Information-Maximization loss function to increase the discriminative power of the model while addressing the issue of the model excessively predicting certain classes. These techniques were applied to the state-of-the-art model, confirming consistent performance improvement. Furthermore, these techniques can be applied to various models as they are not limited to a specific model.

Keyword : Test-Time Adaptation, Class Centroid, Symmetric Cross-Entropy, Information Maximization

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

b) 인하대학교 정보통신공학과(Department of Information and Communication Engineering, Inha University)

c) 성균관대학교 실감미디어공학과(Department of Immersive Media Engineering, Sungkyunkwan University)

‡ Corresponding Author : 홍성은(Sungeun Hong)

E-mail: : csehong@skku.edu

Tel: +82-2-740-1809

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

※ This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2021R1F1A1054569, No. 2022R1A4A1033549).

· Manuscript July 5, 2023; Revised August 29, 2023; Accepted August 29, 2023.

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

도메인 적응은 학습한 모델을 한 도메인에서 다른 도메인으로 전이하는 것을 의미한다. 트랜스퍼 러닝은 한 작업에서 학습한 모델을 다른 작업으로 전이하여 활용하는 반면, 도메인 적응은 동일한 작업을 수행하지만 데이터 분포가 다른 상황에서 발생한다. 도메인 적응의 목표는 도메인 간의 차이를 최소화하고 모델의 성능을 유지하는 것이다. 도메인 적응의 예시로 서로 다른 도메인에서 온 특정 대상을 분류하는 이미지의 스타일이나 특징이 조금씩 다를 수 있다. 이런 상황에서 두 도메인 사이의 차이를 최소화하여 모델이 어떤 도메인의 데이터가 있어도 더 일반화되게 만드는 작업을 의미한다. 이는 모델이 다른 도메인의 데이터를 볼 때도 성능을 유지하거나 향상시킬 수 있도록 돕는다.

도메인 적응(Domain-Adaptation)^[1]은 학습 과정에서 소스 도메인 데이터에 접근이 필수적이기 때문에 데이터 프라이버시 측면에서 민감할 수 있다. 테스트 단계 적응 분야는 소스 도메인 데이터에 접근하지 않고 타겟 도메인에 대해 적응을 하는 것을 목표로 한다. 그러므로 테스트 단계 적응은 실제 상황의 개인 정보가 들어있는 데이터를 인식하는 기기들에 있어 매우 중요한 분야이다. 현재 테스트 단

계 도메인 적응 분야에서 가장 좋은 성능을 보이는 CoTTA^[3] 모델은 테스트 단계 도메인 적응에 Continual Learning의 실험 세팅을 적용한 Continual Test-Time Adaptation을 위한 기법이며, 그림 1은 각 도메인 적응 시나리오 간 차이를 보여준다.

기존 테스트 단계 적응에서의 기법^[2,3,4]들은 보통 정제된 소스 데이터로 학습시킨 모델을 테스트 상황에 투입하여 정제되지 않은 타겟 데이터를 입력 받아 소스 도메인과 타겟 도메인의 분포가 가까워지도록 초점을 맞춰왔다. 그러나 도메인의 분포만 적응시키는 방식은 각 데이터 샘플의 고유한 의미 있는 정보를 무시하는 경향을 보여 전혀 다른 클래스로 대응되는 문제가 발생하고 이 문제는 성능 하락으로 이어진다^[5,16].

본 논문은 기존 테스트 단계 적응에 있던 한계점을 해결하기 위해 도메인 분포뿐만 아니라 각 샘플 별 의미 있는 정보들도 활용하여 기존 보다 잘 적응할 수 있는 방법론을 제안한다. 첫번째 방법론은 각 도메인의 클래스마다 클래스 중심 값을 구하고 같은 클래스의 중심들이 서로 대응하게끔 L2 손실함수를 추가하는 방법이다. 두번째 방법은 기존 모델에서 최종 손실함수로 쓰이던 Cross-Entropy 손실함수 대신 Symmetric Cross-Entropy^[6] 손실함수를 적용한

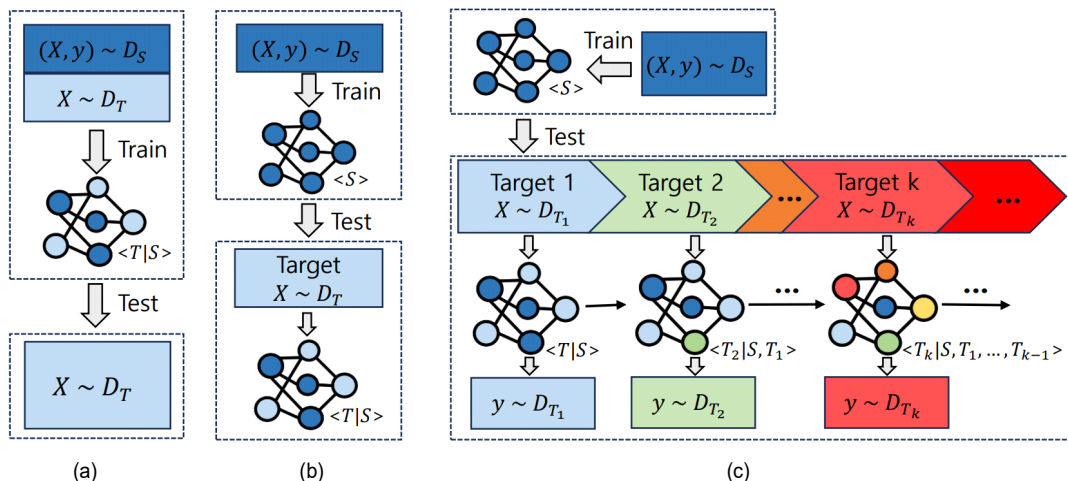


그림 1. 도메인 적응 시나리오 비교. (a) Unsupervised Domain Adaptation은 학습, 테스트 단계에서 소스, 타겟 데이터를 모두 이용할 수 있다. (b) Test Time-Adaptation은 학습 단계에만 소스 데이터로 모델을 학습시키고 테스트 단계에선 하나의 타겟 도메인만 적응한다. (c) Continual Test-Time Adaptation은 학습 단계는 Test Time-Adaptation과 같지만 테스트 단계에서 연속으로 오는 여러 타겟 도메인에 대해 적응하는 차이가 있다.

Fig. 1. Comparison of domain adaptation scenario

다. Symmetric Cross-Entropy 손실함수는 노이즈 레이블에 강건한 Reverse Cross-Entropy^[6] 항을 추가하여 모델이 어려운 경우에도 잘 예측하도록 할 뿐만 아니라 모델의 강건성도 향상시킨다. 마지막으로 모델이 타겟 데이터에 대한 결정력을 올려 줌과 동시에 모델의 예측이 전반적으로 특정 클래스에 쏠리지 않고 다양하게 예측할 수 있도록 도와주는 Information-Maximization^[7] 손실함수를 추가하여 CoTTA 모델 보다 높은 성능을 보인다.

본 논문의 주요 기여점을 요약하면 다음과 같다.

- 전반적인 도메인의 분포만 적응시키던 기존 테스트 단계 도메인 적응 방식에 클래스별 중심값과 Reverse Cross-Entropy 항 추가, 정규화를 통해 각 샘플들의 의미 있는 정보들을 활용하는 다양한 방법론들을 소개한다.
- 제안하는 방법으로 SOTA method에 비해 CIFAR10C에서 0.6%, CIFAR100C에서 0.8%, 그리고 ImageNet-C에서 6.2%로 모든 데이터셋에서 일관되게 향상된 성능을 보였다.
- 제안하는 방법들은 기존 네트워크 구조를 바꿀 필요가 없어 비슷한 네트워크 구조를 가지는 모델이라면 적용 가능하다.

II. 관련 연구

1. Unsupervised Domain Adaptation (UDA)

도메인 적응은 소스 데이터에서 훈련된 모델의 지식을 다른 분포를 가진 타겟 도메인으로 전이하는 문제를 연구한다. 초기 연구는 훈련 중에 타겟 도메인의 레이블에 대한 접근이 필요하다는 점에서 제한적이었다. 반면 비지도 도메인 적응은 타겟 도메인 레이블에 의존하지 않고 소스 도메인과의 분포를 일치시키는 것을 목표로 한다. 기존 비지도 도메인 적응 연구들은 불일치 손실^[8,9], 적대적 학습^[1,10,11], 대조 학습^[2,4], 자가 훈련^[5,13]을 통해서 다른 도메인 간 분포를 정렬했다. 그러나 실제 환경에서는 데이터 프라이버시 등과 같은 문제로 인해 소스 데이터를 사용할 수 없어야 한다. 테스트 단계 도메인 적응은 소스 데이터에 대한 접근 없이 모델을

테스트 단계동안 타겟 데이터에 조정한다.

2. Continual Learning

Continual Learning은 이전 도메인의 지식을 잊지 않으면서 현재 도메인의 지식을 통합한다. 기존 연구들은 소스 데이터에 대한 재학습을 진행하는 방식^[14]과 매개변수의 중요도에 따라 가중치를 업데이트하는 정규화 방식^[15,16]으로 분류할 수 있다. 지속적인 학습은 모델을 다른 도메인에서 잘 작동하도록 훈련시킨다는 점에서 제안 기법과 유사하다. 하지만 테스트 단계 도메인 적응은 새로운 클래스를 학습하는 지속적인 학습과 달리 동일한 클래스를 다른 도메인에서 직면했을 때 치명적인 망각 문제를 해결하는 것에 초점을 맞춘다.

3. Test-Time Adaptation (TTA)

테스트 단계 도메인 적응(TTA)은 소스 데이터에 접근하지 않고 레이블이 지정되지 않은 타겟 데이터만 사용해 미리 훈련된 소스 모델을 타겟 데이터에 적응시킨다. 예를 들어 SHOT^[5]은 모델의 매개변수를 업데이트하기 위해 상호 정보 최대화를 사용하여 테스트 데이터에 적응한다. TENT^[2]는 엔트로피 손실을 최소화하여 배치 정규화 매개변수를 조정하는 데 중점을 둔다. 앞서 소개한 연구들은 단일 도메인 TTA 설정에서만 적응 성능을 개선하는 데 중점을 둔다. 나아가서 지속적인 도메인 이동을 고려하고 장기적인 적응을 수행하는 테스트 단계 도메인 적응은 적응 과정에서 소스 지식을 잊어버리는 치명적인 망각과 잘못 보정된 예측으로 인한 오류 누적이 필연적으로 발생한다. CoTTA^[3]는 이러한 문제를 확률적 복원과 가중 및 증강 평균 예측을 통해 해결한다. 본 논문은 지속적인 테스트 단계 도메인 적응을 위한 도메인 정렬, 클래스 정렬뿐만 아니라 각 샘플들의 정렬까지 수행할 수 있는 기법들을 소개한다.

III. 제안 기법

테스트 단계 적응의 목표는 소스 데이터로 학습된 모델이 레이블이 없는 타겟 데이터가 입력으로 들어오면 그에

맞는 레이블을 실시간으로 잘 예측하도록 하는 것이다. 제안하는 기법들은 기존 테스트 단계 도메인적응에 쓰이던 기법들이 각 샘플들의 의미 있는 정보를 무시하는 경향을 완화시켜준다.

본 논문에서 제안하는 기법을 설명하기 앞서 실험에 베이스라인으로 활용한 CoTTA^[3] 모델의 파이프라인을 소개한다. CoTTA는 테스트 단계 도메인 적응 분야에서 처음으로 Continual Learning 실험 세팅을 접목시켜 실제상황에 더 가까워지게 실험을 세팅하였다. CoTTA 네트워크 구조는 선생 네트워크와 학생 네트워크로 이루어져 있으며, 네트워크의 초기 세팅은 소스 데이터로 미리 학습된 네트워크의 가중치를 선생과 학생 네트워크에 그대로 복사하여 초기화하였다. 입력 데이터가 선생, 학생 네트워크에 들어가는 과정은 아래의 수식으로 표현할 수 있다.

$$\tilde{y}'_t{}^T = \frac{1}{N} \sum_{i=0}^{N-1} f_{\theta'}(aug_i(x_t^T)), \quad (1)$$

$$y'_t{}^T = \begin{cases} \hat{y}'_t{}^T & \text{if } f_{\theta_0}(x_t^T) \geq p_{th} \\ \tilde{y}'_t{}^T & \text{otherwise} \end{cases} \quad (2)$$

테스트 시 입력으로 타겟 데이터(x_t^T)는 학생 네트워크(f_{θ_s})에 데이터 증강(aug) 없이 들어간다. 반면, 선생 네트워크(f_{θ_t})에는 수식 2에 따라 테스트 입력에 대한 초기 네트워크(f_{θ_0})의 예측값이 설정한 임계값(p_{th})을 넘는다면 데이터 증강을 하지 않고 넘지 않는다면 데이터 증강을 적용하여 입력으로 넣는다. 각 네트워크의 입력에 대한 출력값($\tilde{y}'_t{}^T$)은 Cross-Entropy 손실함수를 통해 서로의 예측에 관한 차이를 줄이도록 학습된다. 이후 손실함수는 학생 네트워크에 전달되어 다음번의 학습 때 더 잘 예측하도록 학생 네트워크 가중치가 갱신된다. 갱신된 학생 네트워크의 가중치 θ 는 EMA 기법을 통해 선생 네트워크의 가중치 θ' 를 갱신시키며, EMA 수식은 아래와 같다.

$$\theta'_{t+1} = a\theta'_t + (1-a)\theta_{t+1} \quad (3)$$

네트워크의 최종 예측에는 갱신된 선생 네트워크의 출력값($y'_t{}^T$)을 사용한다. 그림 2는 CoTTA 네트워크에 제안하는 기법들을 적용한 파이프라인이다.

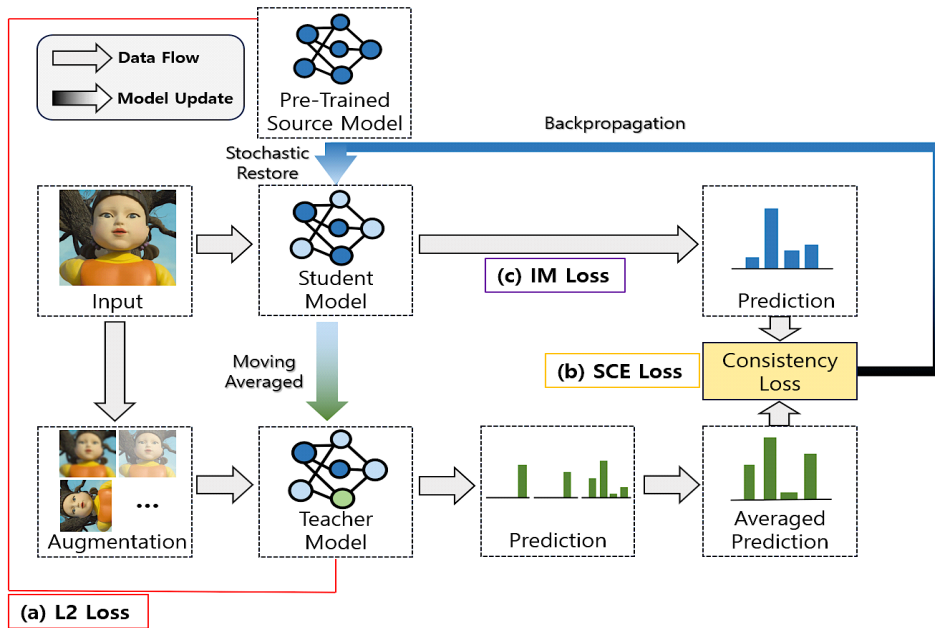


그림 2. 제안 기법을 적용한 파이프라인. (a) 도메인의 클래스 별 중심값을 적응하는 경우 (b) Consistency 손실함수로 SCE 손실함수 적용 (c) 학생 모델에 IM 손실함수를 추가 적용
Fig. 2. Proposed method pipeline

본 논문은 앞서 설명한 CoTTA 네트워크를 베이스라인으로 삼고 샘플 고유의 정보를 잘 추출할 수 있는 세가지 방법을 소개한다.

1. 클래스 별 중심값 적응

첫번째 방법은 같은 클래스의 중심값이 서로 가까워지도록 학습하여 클래스 대응을 잘 하도록 만드는 방법이다. 클래스 중심값을 적응하는 손실함수는 다음과 같이 표현된다.

$$L_{MC}(X_S, Y_S, X_T) = \sum_{k=1}^K \delta(C_S^k, C_T^k) \quad (4)$$

수식에서 X_S, Y_S 는 소스 데이터, X_T 는 타겟 데이터를 나타내며, C_S^k, C_T^k 는 각각 소스 도메인, 타겟 도메인에서 K 개의 클래스에 관한 클래스 중심값을, δ 는 L2 손실함수를 의미한다.

본 논문에서 적용한 방식과 기존 클래스 중심을 활용하는 방식의 차이점은 소스 도메인의 클래스 중심값을 구하는 과정에서 본 논문은 테스트 단계 도메인 적응이기 때문에 소스 데이터에 접근할 수 없다는 점이다. 이러한 문제를 해결하기 위해 본 논문은 소스 데이터에 접근하지 않고 이미 소스 데이터로 충분히 학습시킨 네트워크의 마지막 layer에는 각 클래스를 대표할 수 있는 특징값들이 [클래스 개수, 특징 개수]차원의 형태로 저장되어 있다는 점을 이용했다⁵⁾. 해당 layer를 클래스 별로 특징값에 대해 평균을 구하고 이 평균값을 각 클래스 별 중심값으로 활용하였다.

타겟 데이터에 대한 클래스 별 중심값은 테스트 시 들어오는 데이터에 따라 계속 이동하기 때문에 매 배치단위의 입력이 들어올 때 마다 pseudo labeling을 통해 클래스를 예측한다. 임의로 레이블이 정해진 타겟 데이터들은 같은 레이블로 정해진 데이터들끼리 평균을 적용하여 클래스 별 중심값을 획득한다.

위의 과정으로 얻어진 소스 도메인, 타겟 도메인의 클래스 별 중심값들은 같은 클래스의 중심값들끼리 L2 손실함수를 적용하여 서로의 거리를 줄여가도록 학습한다. 기존 CoTTA의 학습과정에 클래스 별 중심이 가까워지는 손실함수를 추가하여 전반적인 도메인뿐만 아니라 클래스끼리 대응하며 적응할 수 있다.

2. Symmetric Cross-Entropy 손실함수

본 논문은 선생과 학생 네트워크가 서로 일관된 예측을 하는데 주로 사용되던 Cross-Entropy 손실함수 대신 Symmetric Cross-Entropy⁶⁾ 손실함수를 적용한 방법을 제안한다. Symmetric Cross-Entropy는 레이블에 노이즈가 끼 있는 경우에도 Cross-Entropy 보다 더욱 강건하게 예측할 수 있다는 장점이 있고, 아래의 수식과 같이 표현된다.

$$L_{SCE}(q,p) = -\sum_{c=1}^C q_c \log p_c - \sum_{c=1}^C p_c \log q_c \quad (5)$$

여기서 p, q 는 각각 학생 네트워크, 선생 네트워크의 예측값을 의미한다. 수식 5의 첫번째 항은 기존에 알고 있던 Cross-Entropy의 수식과 같고 두번째 항은 Cross-Entropy를 뒤집은 형태로 Reverse Cross-Entropy⁶⁾라 d부른다. Cross-Entropy는 모델이 좋은 방향으로 수렴하는데 큰 도움을 주지만 레이블 노이즈에는 취약하다는 점이 있다. 그러므로 레이블 노이즈에 강건한 Reverse Cross-Entropy 손실함수를 추가하여 최종적인 Symmetric Cross-Entropy 손실함수를 설계했다. Reverse Cross-Entropy를 사용하면 Cross-Entropy만 사용했을 때 보다 예측분포를 좀 더 넓은 관점으로 해석할 수 있기 때문에, 선생과 학생 네트워크에서 나온 의미 있는 정보들을 더 많이 활용할 수 있다.

3. Information Maximization 손실함수

Information Maximization⁷⁾ 손실함수인 L_{IM} 는 기존 도메인 적응 분야에서 타겟 데이터에 대한 모델의 결정력을 높이도록 학습되는 L_{ent} 항과 특정 클래스로 예측이 모이지 않도록 완화시켜주는 L_{div} 항으로 구성되어 있으며, 수식은 다음과 같다.

$$L_{IM} = L_{ent} + L_{div}$$

$$L_{ent}(f; X_t) = -E_{x_t \in X_t} \sum_{k=1}^K \delta_k(f_t(x_t)) \log \delta_k(f_t(x_t)), \quad (6)$$

$$L_{div}(f; X_t) = \sum_{k=1}^K \hat{p}_k \log \hat{p}_k$$

위 수식에서 $f_t(x)$ 는 타겟 데이터에 대한 모델의 출력값

을 의미하며, δ_k 는 softmax 함수를 의미한다. 즉, L_{ent} 는 Cross-Entropy의 수식과 같다고 이해할 수 있다. L_{div} 에서 \hat{p} 는 $E_{x_t \in X_t}[\delta(f_t^{(k)}(x_t))]$ 를 의미하며, 전체 타겟 도메인의 출력 특징맵에 대한 평균값이다.

기존에는 네트워크 구조가 하나의 네트워크로만 학습되지만 CoTTA 네트워크 구조는 학생, 선생으로 이루어진 두 개의 네트워크인 점이다. CoTTA 네트워크에서 모델의 결정력을 올려주는데 중요한 역할을 하는 것은 CoTTA의 Consistency 손실함수이기 때문에, L_{ent} 의 역할은 CoTTA 네트워크의 Consistency 손실함수로 대체하였고, L_{div} 는 학생 네트워크의 출력값으로 사용했다.

IV. 실험 결과

1. 데이터셋 소개 및 파라미터 세팅

실험에는 CIFAR10C, CIFAR100C, 그리고 ImageNet-C^[18] 데이터셋들이 사용됐다. 각 데이터셋들은 분류 네트워크의 강건성을 높이기 위해 기존의 CIFAR10, CIFAR100^[19] 그리고 ImageNet-1k^[20] 데이터셋에 15가지 종류의 노이즈마다 5단계의 노이즈 강도를 적용시킨 데이터로 구성된다.

2. 제안 기법을 적용한 결과

먼저 CIFAR10C 데이터셋에서 제안하는 기법의 효용성을 평가하였다. Source의 방식은 입력된 타겟 데이터에 대해 도메인 적응 없이 바로 레이블을 예측한다. 표 1을 보면, Source 방식을 적용한 성능은 43.5%의 오차율을 보이며 이는 도메인 적응과정이 필요함을 시사한다. 표 1의 * 기호는 재현된 성능을 의미하며, 해당 성능을 기준으로 삼고 제안하는 기법들을 적용하였다. 표 1의 L2 Loss는 클래스 별 중심값 적용, SCE Loss는 Symmetric Cross-Entropy 손실 함수를 IM Loss는 Information Maximization을 의미한다. 베이스라인 네트워크에 제안하는 기법들을 적용하면 일관되게 성능이 향상함을 알 수 있다.

클래스 별 중심값 적용을 적용한 경우 베이스라인의 성능보다 0.4% 향상되었다. 향상된 이유로 기존의 방식들은 도메인 분포 적응만 고려하여 클래스가 잘못 대응되는 문제를 클래스 별 중심값을 적용시킴으로써 완화했기 때문이다. Symmetric Cross-Entropy 손실함수를 적용한 경우에도 기존에 비해 0.4% 향상된 성능을 보였는데, 이는 레이블 노이즈에 취약한 Cross-Entropy 손실함수 대신 Reverse Cross-Entropy 손실함수를 적용하여 노이즈에도 강건한 예측을 할 수 있도록 했기 때문이다. Information

표 1. CIFAR10C 데이터셋으로 기존 TTA 방식과 제안하는 기법을 적용한 Error Rate(%)
Table 1. Error Rates(%) for applying conventional TTA and proposed technique on the CIFAR10C dataset

Method	$t \rightarrow$															Mean ↓			
	L2 Loss	SCE Loss	IM Loss	Gaussian	Shot	Impulse	defocus	glass	motion	zoom	snow	frost	fog	brightness	contrast		elastic_trans	pixelate	jpeg
Source				72.3	65.7	72.9	46.9	54.3	34.8	42.0	25.1	41.3	26.0	9.3	46.7	26.6	58.5	30.3	43.5
BN Adapt				28.1	26.1	36.3	12.8	35.3	14.2	12.1	17.3	17.4	15.3	8.4	12.6	23.8	19.7	27.3	20.4
TENT				24.8	20.6	28.6	14.4	31.1	16.5	14.1	19.1	18.6	18.6	12.2	20.3	25.7	20.8	24.9	20.7
CoTTA*				24.6	22.0	26.7	12.0	27.6	12.5	10.6	15.1	14.6	12.9	7.7	11.4	18.7	13.7	18.1	16.6
Ours	✓			24.3	21.6	25.2	11.8	27.2	12.2	10.4	15.1	14.2	12.7	8.0	11.1	18.3	13.8	17.7	16.2
Ours		✓		24.3	21.5	25.4	11.7	26.8	12.2	10.4	15.0	14.2	12.7	7.9	11.2	18.2	13.4	17.5	16.2
Ours			✓	24.1	21.4	25.8	11.6	27.2	12.2	10.5	14.7	13.9	12.5	7.8	10.7	17.9	13.1	17.1	16.0
Ours	✓	✓		24.3	21.5	25.4	11.8	26.7	12.3	10.6	14.9	14.1	12.5	7.9	10.8	18.3	13.5	17.4	16.1
Ours		✓	✓	24.1	21.1	25.7	11.8	27.22	12.3	10.5	14.8	14.2	12.5	7.8	10.7	18.0	13.4	17.3	16.1
Ours	✓	✓	✓	24.1	21.6	25.4	11.9	26.7	12.3	10.8	15.1	14.3	13.0	8.4	11.2	18.2	13.7	17.6	16.3
Ours	✓	✓	✓	24.3	21.5	25.4	11.9	26.6	12.3	10.9	15.2	14.2	12.7	8.2	11.5	18.4	13.8	17.6	16.3

Maximization을 적용한 방식은 클래스의 다양성을 고려하여 예측이 잘못된 클래스로 모이지 않도록 학습하는 차이점 때문에 더 좋은 성능을 냈다.

제안하는 기법들의 기여도를 확인하기 위해 ablation study를 실시했다. 실험결과 Symmetric Cross-Entropy와 조합한 경우는 단일로 사용했을 때 보다 성능 향상이 있었지만, L2와 IM 손실함수를 둘 다 적용한 경우는 단일로 사용했을 때 보다 성능이 감소했다. Symmetric Cross-Entropy는 데이터 분포를 이동시키기보다 레이블 노이즈에 강건하게 만드는 효과를 가지기 때문에, 데이터 분포를 이동시키는 L2나 IM 손실함수를 같이 적용한 경우 상승효과가 발생한다. 성능이 좋아지는 이유로 Symmetric Cross-Entropy의 Reverse Cross-Entropy term은 높은 확신을 가진 샘플에 대해 큰 영향을 발휘한다. 그러므로 데이터

의 분포를 이동시켜 높은 확신을 가지게 하는 L2나 IM 손실함수를 같이 적용한 경우 상승효과가 발생한다. 하지만 L2와 IM 손실함수를 모두 적용한 경우에는 기존보다 성능이 감소한다. 두 기법을 같이 사용하게 되면 네트워크가 너무 높은 확신을 갖는데 이 경우 네트워크의 예측이 잘못된다면 이후 네트워크를 업데이트하는 과정에서 오류 누적이 더 크게 발생할 수 있다.

제안 기법의 효용성을 더 증명하기 위해, CIFAR10C 보다 복잡한 데이터셋인 CIFAR100C 데이터셋에서 실험을 진행한다. CIFAR100C의 실험 결과인 표 2를 보면 CIFAR10C와 비슷한 경향성을 보이며 제안하는 기법을 적용한 경우 일관되게 성능이 증가함을 확인할 수 있다. 특이한 점은 SCE와 IM 손실함수를 같이 사용한 경우 31.8%로 기존보다 0.8% 향상된 최고의 성능을 보였는데, 이는

표 2. CIFAR100C 데이터셋으로 Baseline과 제안하는 기법을 적용한 Error Rate(%)
 Table 2. Error Rates(%) for applying conventional TTA and proposed technique on the CIFAR100C dataset

Method	t →																	Mean ↓	
	L2 Loss	SCE Loss	IM Loss	Gaussian	Shot	Impulse	defocus	glass	motion	zoom	snow	frost	fog	brightness	contrast	elastic.trans	pixelate		j/peg
Source				73.0	68.0	39.4	29.3	54.1	30.8	28.8	39.5	45.8	50.3	29.5	55.1	37.2	74.7	41.2	46.4
BN Adapt				42.1	40.7	42.7	27.6	41.9	29.7	27.9	34.9	35.0	41.5	26.5	30.3	35.7	32.9	41.2	35.4
TENT				37.2	35.8	41.7	37.9	51.2	48.3	48.5	58.4	63.7	71.1	70.4	82.3	88.0	88.5	90.4	60.9
CoTTA*				40.4	37.9	39.8	27.0	37.9	28.3	26.2	33.2	32.0	40.3	24.8	27.1	32.4	27.8	33.4	32.6
Ours	✓			39.9	37.7	39.2	26.6	37.5	28.1	26.2	32.7	31.8	39.6	24.6	26.6	31.9	27.9	33.0	32.2
Ours		✓		40.1	37.7	39.1	26.8	36.9	27.9	25.9	32.8	31.5	39.7	24.5	26.4	31.7	27.6	33.0	32.1
Ours			✓	39.9	37.4	39.0	26.6	37.1	28.1	25.8	32.5	31.4	39.6	24.4	26.4	31.5	27.4	32.9	32.0
Ours	✓	✓		40.3	37.6	39.7	26.8	37.9	28.2	26.1	33.1	32.0	40.7	24.8	26.8	32.4	28.3	33.7	32.6
Ours		✓	✓	39.7	36.7	38.3	26.4	36.6	27.7	25.7	32.5	31.3	39.2	24.7	26.8	31.1	27.7	33.0	31.8
Ours	✓		✓	40.1	37.2	38.7	26.8	37.1	28.0	25.8	32.8	31.5	40.1	24.8	26.6	31.8	27.5	32.9	32.1
Ours	✓	✓	✓	40.4	37.0	38.4	26.4	37.1	27.8	25.6	32.3	31.3	39.3	24.7	26.7	31.1	27.5	32.7	31.9

표 3. ImageNet-C 데이터셋으로 Baseline과 제안하는 기법을 적용한 Error Rate(%)
 Table 3. Error Rates(%) for applying conventional TTA and proposed technique on the ImageNet-C dataset

Methods	Source	BN	TENT	CoTTA*	Ours	Ours	Ours	Ours	Ours	Ours	Ours
L2 Loss					✓			✓	✓		✓
SCE Loss						✓		✓		✓	✓
IM Loss							✓		✓	✓	✓
Mean ↓	82.4	72.1	66.5	68.5	66.7	63.2	64.4	63.0	64.1	62.3	62.5

CIFAR100C가 CIFAR10C에 비해 분류할 클래스 종류와 데이터 개수가 많아졌기 때문에 두 손실함수의 효과가 더욱 크게 작용했다.

제안 기법을 앞선 데이터셋들 보다 복잡한 ImageNet-C 데이터셋에 실험을 진행하였다. CIFAR10C, CIFAR100C는 노이즈 종류의 순서가 고정되었으나 ImageNet-C는 순서를 다양하게 바꾸어 10번을 실행했기 때문에 본 연구도 기존의 실험 세팅을 따랐다. 표 3은 10번의 실행에 대한 평균 Error Rate를 명시하였다. 표 3에 따르면, 제안하는 기법이 1000개의 클래스를 가지는 ImageNet-C 데이터셋에서도 일관되게 성능향상에 도움됨을 확인할 수 있다. 최고 성능은 SCE와 IM 손실함수를 같이 적용한 62.3%로 베이스라인인 68.5% 보다 6.2%의 향상을 이루었다. ImageNet-C에서 제안한 3개의 손실함수 중 성능 향상에 가장 높은 기여를 한 것은 SCE 손실함수이다. L2와 IM 손실함수는 규제하는 항이 없어 모델이 과적합하는 경향이 있었다. 하지만 SCE 손실함수는 기존 Cross-Entropy 항은 낮은 confidence의 샘플에 영향을 받으며, Reverse Cross-Entropy 항에서는 높은 confidence의 샘플에 영향을 받아 전반적인 손실함수가 균형적하도록 규제하는 역할을 했기 때문이다.

V. 결론

본 논문은 전반적인 도메인 분포만 적응시켰던 테스트 단계 도메인 적응 기법에 샘플의 의미있는 정보를 활용하여 모델이 더 나은 예측을 하기 위한 세가지 방법을 제시했다. 첫 번째 방법은 각 도메인의 클래스 별 중심값을 구하여 같은 클래스의 중심끼리 가까워지도록 하는 손실함수를 추가하였다. 두 번째 방법은 레이블 노이즈에 취약한 Cross-Entropy 손실함수 대신, 레이블 노이즈에 강건한 Symmetric Cross-Entropy 손실함수를 적용하였다. 마지막 방법은 도메인 적응분야에서 특정 클래스로 예측이 모이는 현상을 조절하기 위해 제안된 Information Maximization 손실함수를 테스트 도메인 적응에 맞게 적용하였다. 다양한 검증 실험을 통해 제안 기법들이 샘플의 중요한 정보들을 추출할 수 있고, 모델의 효율성 대비 좋은 성능을 얻을 수 있음을 확인할 수 있었다.

참고 문헌 (References)

- [1] Y. Ganin, and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," International conference on machine learning, PMLR, pp. 1180-1189, 2015.
- [2] D. Wang, E. Shelhamer, S. Liu, B. Olshausen, and T. Darrell, "Tent: Fully Test-Time Adaptation by Entropy Minimization," International Conference on Learning Representations, ICLR, 2021.
- [3] Q. Wang, O. Fink, L. Van Gool, and D. Dai, "Continual test-time domain adaptation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7201-7211, 2022. doi: <https://doi.org/10.1109/cvpr52688.2022.00706>
- [4] D. Chen, D. Wang, T. Darrell, and S. Ebrahimi, "Contrastive test-time adaptation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 295-305, 2022. doi: <https://doi.org/10.1109/cvpr52688.2022.00039>
- [5] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation," International Conference on Machine Learning, PMLR, pp. 6028-6039 2020.
- [6] Y. Wang, X. Ma, Z. Chen, Y. Luo, J. Yi, and J. Bailey, "Symmetric cross entropy for robust learning with noisy labels," Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 322-330, 2019. doi: <https://doi.org/10.1109/iccv.2019.00041>
- [7] A. Krause, P. Perona, and R. Gomes, "Discriminative clustering by regularized information maximization," Advances in neural information processing systems 23, 2010.
- [8] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," International Conference on Machine Learning, PMLR, pp. 97-105, 2015.
- [9] B. Sun, and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," Computer Vision - ECCV 2016 Workshops, Amsterdam, The Netherlands, Proceedings, Part III 14. Springer International Publishing, pp. 443-450, 2016. doi: https://doi.org/10.1007/978-3-319-49409-8_35
- [10] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7167-7176, 2017. doi: <https://doi.org/10.1109/cvpr.2017.316>
- [11] L. Chen, H. Chen, Z. Wei, X. Jin, X. Tan, Y. Jin, and E. Chen, "Reusing the task-specific classifier as a discriminator: Discriminator-free adversarial domain adaptation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7181-7190, 2022. doi: <https://doi.org/10.1109/cvpr52688.2022.00704>
- [12] G. Kang, L. Jiang, Y. Yang, and A. G. Hauptmann, "Contrastive adaptation network for unsupervised domain adaptation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4893-4902, 2019. doi: <https://doi.org/10.1109/cvpr.2019.00503>

- [13] H. Liu, J. Wang, and M. Long, "Cycle self-training for domain adaptation," *Advances in Neural Information Processing Systems* 34, 2021.
- [14] H. Shin, J. K. Lee, J. Kim, and J. Kim, "Continual learning with deep generative replay," *Advances in Neural Information Processing Systems* 30, 2017.
- [15] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the national academy of sciences* 114.13, pp. 3521-3526, 2017.
- [16] D. Lopez-Paz, and M. A. Ranzato, "Gradient episodic memory for continual learning," *Advances in Neural Information Processing Systems* 30, 2017.
- [17] S. Xie, Z. Zheng, L. Chen, and C. Chen, "Learning semantic representations for unsupervised domain adaptation," *International Conference on Machine Learning*, PMLR, pp. 5423-5432, 2018.
- [18] D. Hendrycks, and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," *International Conference on Learning Representations*, ICLR, 2019.
- [19] A. Krizhevsky, and G. Hinton, Learning multiple layers of features from tiny images, Technical report, Citeseer, <http://www.cs.utoronto.ca/~kriz/learning-features-2009-TR.pdf> (accessed April. 8, 2009).
- [20] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. F. Fei, "Imagenet: A large-scale hierarchical image database," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248-255, 2009. doi: <https://doi.org/10.1109/cvpr.2009.5206848>
- [21] F. Croce, M. Andriushchenko, V. Schwag, E. Debenedetti, N. Flammarion, M. Chiang, P. Mittal, and M. Hein, "Robustbench: a standardized adversarial robustness benchmark." In *NeurIPS Datasets and Benchmarks Track*, 2021.
- [22] S. Zagoruyko, and N. Komodakis, "Wide Residual Networks," *British Machine Vision Conference 2016*, British Machine Vision Association, 2016. doi: <https://doi.org/10.5244/c.30.87>

저 자 소 개

이 호 준



- 2016년 3월 ~ 2022년 2월 : 인하대학교 정보통신공학과 학사
- 2022년 3월 ~ 현재 : 인하대학교 전기컴퓨터공학과 석사
- ORCID : <https://orcid.org/0000-0001-8150-2472>
- 주관심분야 : Domain Adaptation, Continual Learning, Parameter Efficient Transfer Learning

이 승 환



- 2017년 3월 ~ 현재 : 인하대학교 정보통신공학과 학사
- ORCID : <https://orcid.org/0009-0006-5762-4107>
- 주관심분야 : Domain Adaptation, Multimodal Learning

저 자 소 개



정 인 영

- 2021년 3월 ~ 현재 : 인하대학교 정보통신공학과 학사
- ORCID : <https://orcid.org/0009-0006-3683-6514>
- 주관심분야 : Domain Adaptation, Continual Learning, Multimodal Learning



홍 성 은

- 2010년 2월 : 한양대학교 컴퓨터공학과 학사
- 2012년 8월 : 카이스트 전산학과 석사
- 2018년 2월 : 카이스트 전산학과 박사
- 2018년 1월 ~ 2020년 8월 : SK telecom (T-Brain, AI Center) 연구원
- 2020년 9월 ~ 2023년 8월 : 인하대학교 정보통신공학과 조교수
- 2023년 9월 ~ 현재 : 성균관대학교 실감미디어공학과 조교수
- ORCID : <https://orcid.org/0000-0003-1774-9168>
- 주관심분야 : Domain Adaptation, Multimodal Learning, Face Understanding, Video Object Segmentation