

Zero-Shot 이미지 분류를 위한 Transferable Contrastive Network 에 관한 연구

이정철, 이동규*

*경북대학교 인공지능학과

{jcleee2716, dglee}@knu.ac.kr

A Study on the of Transferable Contrastive Network for Zero-Shot Image Classification

Jeong-Cheol Lee, Dong-Gyu Lee*

*Kyungpook National University.

요 약

본 논문은 zero-shot 이미지 분류를 위한 transferable contrastive network (TCN)에 관한 연구이다. TCN은 이미지와 attribute의 정보를 추출하고 결합하는 information fusion 부분과, relation network를 통해 클래스 확률을 구하고 클래스 유사도를 레이블로 사용하여 모델을 학습하는 contrastive learning 부분으로 구성된다. 이러한 구조는 end-to-end 학습이 가능하게 하며, 학습 데이터셋에 포함되지 않은 카테고리를 예측하는데 있어서 강한 성능을 보인다. 제안하는 방법은 zero-shot 이미지 분류 벤치마크 데이터셋인 AwA2 데이터셋에 대하여 우수한 성능을 보임으로써 zero-shot 이미지 분류 문제에 효과적임을 보여주었다.

I. 서론

이미지 분류는 컴퓨터 비전 분야의 기초적인 문제들 중 하나로, 최근 수 년간 ImageNet [1]과 같은 수많은 레이블링된 데이터를 사용하여 빠른 속도로 발전되었다 [2]. 그러나 특정 도메인에서는 이러한 데이터를 수집하는 것은 비용과 시간이 많이 소모되는 일이며, 지도 학습의 경우 고정된 카테고리의 데이터만을 학습하기 때문에 유연하지 않다. 이와는 대조적으로 사람은 기존에 알고 있던 시각 정보와 추가적인 attribute만 보고도 기존에 알지 못했던 카테고리를 예측할 수 있다. 그래서 이와 유사하게 학습 데이터셋에 포함된 카테고리의 이미지 (seen class)로 모델을 학습시킨 후, 학습 데이터셋에 포함되지 않은 카테고리의 이미지 (unseen class)를 분류하는 zero-shot 이미지 분류가 연구되었다 [3].

Zero-shot 이미지 분류에 관한 연구는 크게 임베딩 기반 방법론 [4, 7]과 생성 모델 기반 방법론 [5]으로 구분된다. 임베딩 기반 방법론은 이미지의 특징 또는 attribute의 특징을 추출하고 정해진 차원의 임베딩 공간으로 매핑하여 문제를 해결하는 방법이다. 이와는 다르게 생성 모델 기반 방법론은 생성 모델을 seen class의 attribute로부터 이미지 특징을 생성하도록 학습하고, unseen class의 attribute로부터 이미지 특징을 생성하여 zero-shot 문제를 일반적인 문제로 변환하여 해결하는 방법이다.

본 논문에서는 임베딩 기반 방법론 중 transferable contrastive network (TCN) [4]에 기반한 zero-shot 이미지 분류에 관한 연구를 보여준다. TCN은 end-to-

end 학습이 가능하며, 클래스 유사도를 레이블로 사용함으로써 unseen class를 예측하는데 있어서 강한 성능을 보인다.

II. 본론

2.1 Transferable Contrastive Network 구조

본 연구에서는 zero-shot 이미지 분류 문제 해결을 위해 인공신경망 구조로 TCN [4]을 사용한다. TCN은 zero-shot 이미지 분류를 목적으로 제안된 임베딩 기반 방법론으로, information fusion 부분과 contrastive learning 부분으로 구성된다. Information fusion 부분은 convolutional neural network (CNN)에 이미지를 통과시켜 추출한 이미지 특징과 multi-layer perceptron (MLP)에 attribute vector를 통과시켜 추출한 semantic vector를 element-wise product를 적용하여 이미지와 attribute 정보를 결합한다. Contrastive learning 부분은 앞에서 결합한 정보를 relation network [6]를 통과시켜 클래스별 확률을 구한다. 이때, unseen class의 레이블로 클래스 유사도를 사용하여 seen class와 unseen class 간 대조적인 학습이 가능하도록 한다.

2.2 클래스 유사도

Unseen class를 분류하기 위해 클래스 유사도를 사용한다. 클래스 유사도는 희소 코딩을 사용하여 각 seen class별 unseen class에 대한 유사도를 계산하고 정규화한다. 클래스 유사도를 구하는 식은 다음과 같다.

$$s_k = \arg \min_{s_k} \|a_k - \sum_{j=K+1}^{K+L} a_j s_{kj}\|_2^2 + \beta \|s_k\|_2 \quad (1)$$

K 와 L 은 학습 클래스와 unseen 의 개수이며, a 는 클래스의 attribute 벡터를 의미한다. 이렇게 구한 클래스 유사도를 레이블로 사용하여 unseen class 를 분류할 수 있도록 한다. 이때 학습을 위해 adaptive margin loss [7]를 차용하여 정답 클래스의 확률에 따라 자동으로 적절한 margin 을 부여하는 방식을 적용한다.

III. 실험 및 분석

3.1 데이터셋

제안하는 방법을 평가하기 위해 animals with attributes2 (AwA2) [3] 데이터셋을 활용한다. AwA2 데이터셋은 85 가지 attributes 와 50 가지 카테고리의 동물 이미지가 포함된 데이터셋으로, 37322 장의 이미지로 구성되어 있다. 이 중 23527 장은 학습, 5882 장은 평가, 7913 장은 시험데이터로 사용하였다.

3.2 실험 결과

본 실험에서 성능적으로 Top-1 정확도를 사용하였으며, seen 과 unseen 각각 정확도를 측정하고 두 정확도의 조화평균을 사용하여 최종 성능을 측정하였다.

방법	S	U	H
TCN [4]	65.8	61.2	63.4
DVBE [7]	70.8	63.6	67.0
HSVA [8]	79.8	56.7	66.3
제안하는 방법	74.3	64.2	68.9

표 1 AwA2 데이터셋에 대한 이미지 분류 성능 비교

표 1 은 AwA2 데이터셋에 대한 임베딩 기반 방법들의 성능을 비교한 것이다. S, U, H 는 각각 seen, unseen, harmonic mean 을 의미한다. 제안하는 방법이 DVBE 보다 seen 과 unseen 에 각각 3.5%, 0.6% 높은 정확도를 보였으며, HSVA 보다 seen 은 5.5% 낮지만 unseen 이 7.5% 높아 H 에서 2.6% 높은 정확도를 보임으로써 제안하는 방법이 다른 방법들보다 zero-shot 이미지 분류에 효과적임을 보였다.

III. 결론

본 논문에서는 zero-shot 이미지 분류 방법인 transferable contrastive network (TCN)에 대한 연구를 보여준다. 이미지에서 추출한 특징과 attribute 에서 추출한 특징을 결합하고 relation network 를 통과시켜 구한 클래스 확률과 최소 코딩을 사용하여 unseen 에 대한 레이블을 생성하여 unseen 에 대한 학습을 진행하였다. 벤치마크 데이터셋에서 다른 임베딩 기반 방법들의 성능을 능가하여 제안하는 방법이 zero-shot 이미지 분류에 효과적임을 보여주었다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음 (IITP-2022-2020-0-01808*)

참 고 문 헌

- [1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge", International Journal of Computer Vision, 2015, Vol. 115, No. 3, pp. 211-252.
- [2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 26 - July 1, 2016, pp. 770-778.
- [3] Y. Xian, B. Schiele, and Z. Akata, "Zero-Shot Learning - The Good, the Band and the Ugly", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, pp. 4582-4591.
- [4] H. Jiang, R. Wang, S. Shan, and X. Chen., "Transferable contrastive network for generalized zero-shot learning.", Proceedings of the IEEE/CVF International Conference on Computer Vision, Oct. 22-29, 2019, pp. 9765-9774.
- [5] Y. Xian, T. Lorenz, B. Schiele, and Z. Akata, "Feature Generating Networks for Zero-Shot Learning", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-22, 2018, pp. 5542-5551.
- [6] F. Sung, Y. Yang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to Compare: Relation Network for Few-Shot Learning", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-22, 2018, pp. 1199-1208.
- [7] S. Min, H. Yao, H. Xie, C. Wang, Z. J. Zha, and Y. Zhang, "Domain-Aware Visual Bias Eliminating for Generalized Zero-Shot Learning", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 6-19, 2020, pp. 12664-12673.
- [8] S. Chen, G. S. Xie, Y. Liu, Q. Peng, B. Sun, H. Li, X. You, and L. Shao, "HSVA: Hierarchical Semantic-Visual Adaptation for Zero-Shot Learning", Advances in Neural Information Processing Systems, Dec. 6-14, 2021, pp. 16622-16634.