커리큘럼 학습을 통한 자기 교사 학습의 성능 향상

윤의현, 이재구*
국민대학교

*jaekoo@kookmin.ac.kr

Enhancing Performance through Curriculum Learning in Self-Supervised Learning

Euihyun Yoon, Jaekoo Lee*
College of Computer Science, Kookmin University.

요 약

최근 컴퓨터 비전 과업은 많은 발전을 이뤘다. 하지만 기존의 컴퓨터 비전 과업은 정답 값이 존재하는 데이터 집합으로 학습을 하기 때문에 정답 값에 의존하는 특징이 존재한다. 이를 개선하고자 정답 값이 없는 데이터로 이미지의 일반적인 특징을 추출하는 자기 교사 학습 방법이 발전되고 있다. 다만 자기 교사 학습은 이미지의 일반적인 특징을 추출하기 위해 많은 반복 학습을 해야 한다. 본 논문은 다양한 관점의 이미지를 사용해 대조 학습하는 자기 교사 학습법에 커리큘럼 학습을 적용하여 개선하였다. 제안한 모델은 500 번 더 적은 반복 학습으로 기존의 자기 교사 학습 방법보다 0.43% 높은 성능을 달성 하였다.

Ⅰ. 서 론

컴퓨터 비전 과업은 교사 학습 방식으로 많은 발전을 이뤘다[1]. 하지만 교사 학습의 경우 정답 값 기반으로 학습을 진행한다. 정답 값이 존재하는 다양한 데이터를 수집하는 일은 비용이 많이 들기 때문에 정답 값 없이도 학습이 가능한 자기 교사 학습 방법이 발전되고 있다.

최근 긍정적인 쌍만 사용하는 자기 교사 학습 방법인 DINO[2]가 제안됐다. DINO 는 두 개의 동일한 네트워크를 설정하여 각각 교사, 학생 네트워크로 학습을 진행한다. 학습 방법은 교사 네트워크의 출력을 정답으로 설정하여 학생 네트워크를 학습하는 방식이다. 하지만 DINO 와 같은 자기 교사 학습 방법은 다양한 관점으로 변형 (Augmentation)된 이미지를 학습하기 위해 대량의 반복 학습을 요구한다.

기존 많은 컴퓨터 자원을 사용해야 하는 문제를 해결하고자 본 논문은 커리큘럼 학습[3]을 적용하였다. 커리큘럼 학습은 인간의 학습 방법과 유사한 학습을 하기위해 쉬운 데이터부터 어려운 데이터를 사용하는 방법으로 수렴속도 향상과 높은 성능을 낸다.

본 논문은 긍정적인 쌍만 사용한 자기 교사 학습 방법에 그림[1]과 같이 이미지의 영역을 점차적으로 줄이는 커리큘럼 학습 (Curriculum Learning)을 적용하였다. 제안한 모델은 500 번 더 적은 반복 학습으로 기존의 자기 교사 학습 방법보다 0.43% 높은 성능을 달성하였다. 또한, 100 번의 반복 학습 기준 기존의 자기 교사 학습 방법보다 2.71% 높은 성능을 달성 하였다.

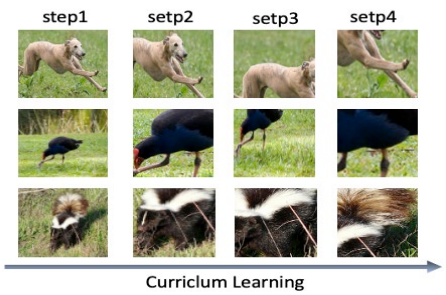

그림 2. 커리큘럼 학습 방법

Ⅱ. 본 론

본 논문은 대조 학습 방법 중 긍정적인 쌍만 사용하는 자기 교사 방법인 DINO 를 사용했다. DINO 는 하나의 이미지에 대해 각각 다른 형태의 변형을 진행한다. 변형은 이미지를 96x96 으로 줄인 지역적인 관점 (Local view)과 이미지를 224x224 로 줄인 전역적인 관점 (Global view)이 존재한다. 변형된 이미지는 각각 네트워크에 입력된다. 이때 네트워크는 교사 네트워크와 학생 네트워크로 나누어진다. 그림[2]와 같이 학생 네트워크는 크고 작은 이미지를 학습해서 교사 네트워크의 출력을 예측해야 한다. 모델의 목적 함수는

$$L = \min_{\theta_s} H(P_t(x), P_S(x)) \quad (1)$$

$$H(a,b) = -a \log b$$

식(1) 과 같이 교차 엔트로피 오차 (Cross Entropy Error)를 사용하여 교사 네트워크의 예측 값을 정답으로

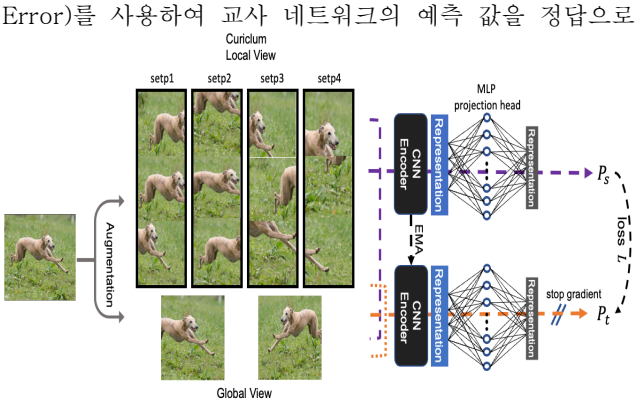

그림 1. 제안 모델 프레임워크

표 1. 커리큘럼 스텝

	1 단계	2 단계	3 단계	4 단계
영역 크기	30%~40%	15%~40%	10%~30%	5%~20%
반복 수	50	50	100	100

설정하여 학생 네트워크를 학습하도록 구성 돼있다.

본 논문은 기존 DINO 의 지역적인 변형을 개선하였다. 기존 지역적인 변형은 96x96 의 크기로 이미지를 작게 할 때 선택되는 영역의 크기를 전체 이미지의 5%에서 40% 사이 무작위로 선택하였다. 무작위 선택은 이미지를 다양한 관점에서 특징을 학습할 수 있게 한다. 하지만 무작위 선택은 아무런 객체가 존재하지 않는 배경 또한 선택된다. 학습 초기 객체가 존재하지 않는 배경이 선택되면 학습을 방해하여 성능 열화를 일으킨다. 따라서, 본 논문은 커리큘럼 학습 방법을 적용해 학습 초기 노이즈로 인한 성능 열화를 개선하였다.

커리큘럼 학습 방법은 인간 및 동물의 학습 법을 묘사한 학습 방법이다. 학습 데이터를 무작위로 사용하는 것이 아닌 쉬운 데이터부터 어려운 데이터 순으로 점진적으로 더욱 복잡한 데이터를 학습하여 빠른 수렴 속도와 높은 성능을 낸다[3]. 하지만 학습 난이도를 정하기 위해 데이터를 직접 나눠야 하며 모델 입장에서 쉬운 데이터와 어려운 데이터를 구분하기 힘든 문제점이 존재한다.

본 논문에서 제안한 모델은 그림[1] 과 같이 학습 초기 큰 영역의 이미지를 학습시키고 점점 작은 영역의 이미지를 순차적으로 학습하는 방식으로 커리큘럼 학습을 구성하였다. 학습 구성은 모델에 입력되는 정보량으로 구분할 수 있어 데이터의 난이도를 효과적으로 나눌 수 있었다. 모델의 커리큘럼 단계의 상세한 정보는 표[1]과 같이 설정하였다. 초기 30% ~ 40%, 15% ~ 40%의 크기로 자른 큰 이미지에 대해 각각 50 번 반복 학습하여 모델을 안정화하였다. 그 후 10% ~ 30%, 5% ~ 20%의 크기로 자른 작은 이미지에 대해 100 번의 반복 학습을 설정하여 어려운 문제를 더욱 많이 학습하게 하였다. 또한 선택영역의 크기에 관한 실험을 추가로 진행하였다.

Ⅲ. 실험 및 결과

표[2]를 보면 제안한 모델의 성능이 기존 DINO 보다 500 번 더 적은 반복 학습 수로 ImageNet[4] 데이터집합에서 0.41% 더욱 높은 성능을 달성 하였다. 또한 Flowers101[5] 데이터 집합에서 1.1% 더욱 높은 성능을 달성 하였다.

표[3]은 영역 크기에 따른 성능 표이다. 각각 반복 학습 수를 100 으로 고정 하여 실험을 진행하였다. Curi-DINO/0.3 은 2 단계를 20~40%로 설정하였으며 4 단계를 5%~30%로 설정하였다. Curi-DINO/0.2 는 표[1] 과 같이 설정하였다. 두 모델 모두 기존 DINO 대비 최소 2.5% 최대 2.7% 높은 성능을 달성 하였다. 또한 Curi-

표 3. 기본모델과 제안모델 성능 비교

	Epoch	Dataset	Accuracy
DINO	800	ImageNet	75.31
	800	Flowers	97.80
Curi-DINO (Ours)	300	ImageNet	**75.74**
	300	Flowers	**98.90**

표 2. 선택영역 비율에 따른 성능

	DINO	Curi-DINO /0.3	Curi-DINO /0.2
Top-1	71.26	73.76	**73.97**
Top-5	90.58	91.75	**91.86**

DINO/0.3 과 Curi-DINO/0.2 두 모델 간 0.21% 성능 차이를 보이는데 이는 커리큘럼 학습을 진행할 때 더 작은 이미지를 학습하여 어려운 문제를 해결하는 것이 더욱더 효과적임을 볼 수 있다.

기존 DINO 는 100 번의 적은 반복 수로 학습을 하면 초기에 아무런 객체가 포함되지 않는 이미지가 선택되어 성능 열화가 일어난다. 반면 제안된 모델은 큰 이미지부터 작은 이미지로 순차적으로 학습하기 때문에 100 번의 반복 학습으로 최대 2.7% 높은 성능을 달성하였다.

Ⅳ. 결 론

기존 DINO 는 이미지 영역 크기를 5%에서 40% 사이 무작위로 정하여 학습을 진행하였다. 기존 무작위 학습의 경우 아무런 객체가 포함되지 않는 노이즈가 선택되어 성능 열화를 일으킨다. 또한 이런 문제를 해결하기 위해 많은 반복 학습 수를 가지게 된다. 반면 본 논문에서 제안한 모델은 쉬운 이미지부터 어려운 이미지로 학습하기 때문에 더 적은 반복 학습으로도 더욱 높은 성능을 달성할 수 있었다.

본 논문이 제안하는 커리큘럼 학습은 이미지를 전역적인 관점과 지역적인 관점으로 변형하여 학습하는 다른 자기 교사 학습 방법에 적용할 수 있으며 더 다양하게 선택영역의 비율을 변경할 수 있다. 이를 토대로 다양한 실험을 진행해 다른 자기 교사 방법의 성능 향상 또한 기대하는 바이다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.RS-2023-00212484,복잡한 실제 주행환경에서 설명 가능한 움직임 예측).

참 고 문 헌

[1] Canziani, Alfredo, Adam Paszke, and Eugenio Culurciello. "An analysis of deep neural network models for practical applications." *arXiv preprint arXiv:1605.07678* (2016).

[2] Li, Chunyuan, et al. "Efficient self-supervised vision transformers for representation learning." *arXiv preprint arXiv:2106.09785* (2021).

[3] Bengio, Yoshua, et al. "Curriculum learning." *Proceedings of the 26th annual international conference on machine learning*. 2009.

[4] Nilsback, Maria-Elena, and Andrew Zisserman. "Automated flower classification over a large number of classes." *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*. IEEE, 2008.

[5] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer International Publishing, 2014.