

자가 지식 증류 기법을 적용한 객체 검출 기법의 성능 분석

*김동준, 이승현, 송병철

*인하대학교 전기컴퓨터공학과

[*hymki12@naver.com](mailto:hymki12@naver.com), lsh910703@gmail.com, bcsong@inha.ac.kr

Performance analysis of Object detection using Self-Knowledge distillation method

*Dong-Jun Kim, Seunghyun Lee, Byung-Cheol Song

Inha University

요 약

경량화 기법 중 하나인 Knowledge distillation 은 최근 object detection task 에 적용되고 있다. Knowledge distillation 은 3 가지 범주로 나뉘는데 그들 중에서 Self-Knowledge distillation 은 기존의 Knowledge distillation 에서의 pre-trained teacher 에 대한 의존성 문제를 완화시켜준다. Self-Knowledge distillation 또한 object detection task 에 적용되어 training cost 를 줄이고 고전적인 teacher-based methods 보다 좋은 성능을 성취했다.

1. 서론

Deep neural networks (DNN) 모델은 최근 object detection 분야에서 좋은 성능을 보여주고 있다. 하지만 DNN 모델이 좋은 성능을 내기 위해서는 더욱 깊고 복잡한 구조의 모델을 사용해야 하기 때문에 연산량과 메모리 소비량이 점점 증가하고 있다. 이 문제를 해결하기 위해 최근 DNN 모델을 경량화 하는 방법인 Knowledge distillation, Pruning, Quantization 과 같은 lightweight model 을 object detection 에 적용하는 연구가 진행되고 있다.[1]

본 논문에서는 Knowledge distillation 에서 Self-Knowledge distillation 을 object detection 에 적용한 사례를 소개하고 성능에 대한 비교 분석을 하고자 한다. Knowledge distillation 은 large teacher network 를 학습시킨 후 small student network 에게 지식을 전달하여 teacher network 의 높은 성능은 최대한 유지하고 training cost 를 줄여 student network 를

효율적으로 학습시키는 방식이다. 그 중 Self-Knowledge distillation 은 teacher network 가 student network 자기 자신이 되어 single model 에서 single training process 를 통해 효율적으로 지식을 증류하는 방식이다.

2. 본론

Knowledge distillation 은 teacher network 와 student network 의 지식 전달 방식에 따라 그림 1 과 같이 구분된다 [2]. 일반적인 경우인 Offline distillation 은 pre-trained teacher network 의 knowledge 를 추출하여 student network 를 학습하는 방법이다. Online distillation 의 경우는 teacher network 와 student network 가 동시에 학습을 진행하는 방식이다. 마지막으로 Self-knowledge distillation 은 자기 자신이 곧 teacher network 역할까지 맡아 single model 로 학습을 진행하는 방식이다. pre-trained teacher network 를

이용하지 않기 때문에 학습에 필요한 training cost 가 비교적 적은 편이며 teacher network 를 따로 사용하지 않는 만큼 성능은 낮은 경향이 있다 [3]. 지금까지 object detection task 에서 Pre-trained teacher network 를 갖는 Offline distillation 기법이 많이 적용되어 왔다. 하지만 Self-distillation 을 적용한 연구는 비교적 최근에 진행되고 있어 사례가 많지 않다. 따라서 Self-distillation 을 적용한 사례를 확인해보고 이에 대한 향후 연구 방향을 확인해 볼 필요가 있다.

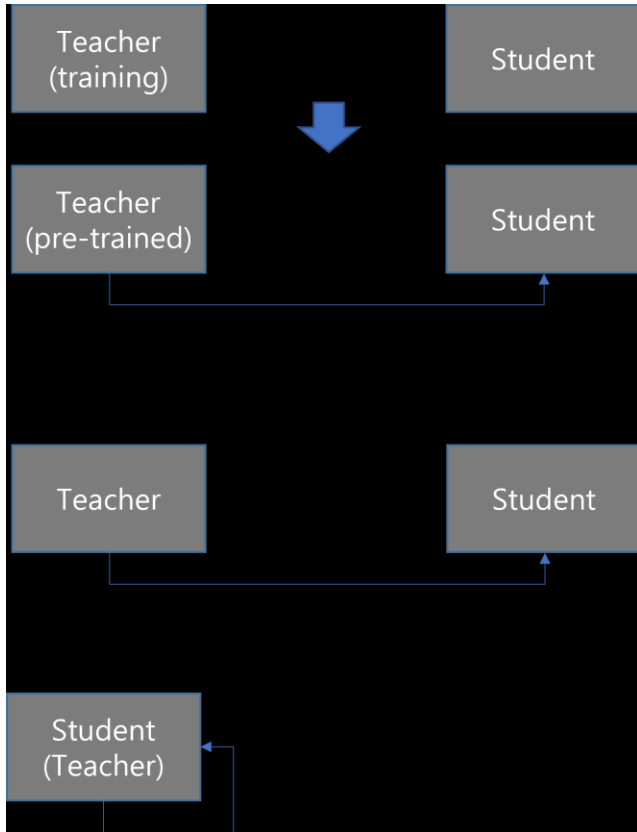


그림 1. Types of Knowledge distillation

본 논문에서 소개하는 사례들은 object detection 에 Self-knowledge distillation 을 적용하여 정량적인 성능을 개선하거나 teacher 가 없음에도 teacher-based method 와 비교하여 비슷한 수준의 성능을 보여준다. Blurred image 에 강인한 object detection 을 수행할 때(S.-J. Cho et al.,2022)[4] Original model 과 Knowledge distillation 을 이용한 model 보다 높은 성능을 보여주기도 한다. Fine-tuning 기법과 함께 object detection task 에 Self-knowledge distillation 을 적용하여 (Xu et al.2021)[5] 성능을 개선하기도 하고 새로운 framework 를 제안(Zhang et al 2022)[6]하여 고전적인 teacher-based method 보다 높은 성능을 보여주기도 한다. 또한 제안한 기법에서 Self-knowledge distillation 을 도입하여 기법의 문제점을 보완하고 성능을 개선하기도 한다.[7]

3. 실험 및 결과

표 1 과 표 2 는 앞에서 제시한 teacher-based method 와 Self-knowledge distillation 을 적용한 기법의 성능 차이를 보여준다. 표 1 의 경우 blurred image 에서 Knowledge distillation 이 적용되지 않은 모델과 pre-trained teacher 를 가진 Knowledge distillation 모델 보다 Self-Knowledge distillation 을 적용한 모델에서 더 높은 성능을 보여주고 있다.

Model		Original	KD	S-KD
mAP	VOC_blurry	66.5	67.4	69.0
	VOC_sharp	73.4	74.8	78.0

표 1. Comparison on blurred image

표 2 의 경우 고전적인 teacher-based method(FGFI)는 ResNet-101 을 teacher 로 설정하였고 각각 Student(Baseline, LGD)의 경우 ResNet-50 과 101 로 설정하여 성능을 비교하였다. 비록 최신의 높은 성능을 가진 teacher-based model 은 아니지만 Self-Knowledge distillation 을 이용하였음에도 정량적으로 더 나은 성능을 보인 것을 확인할 수 있다.

Model	Teacher	Student(R50)	Student(R101)
Baseline	N/A	38.8	40.6
FGFI	R101	39.8	40.7
LGD	N/A	40.3	42.1

표 2. Comparison with classical teacher-based method

표 3 의 경우 Weakly supervised object detection 기법인 SLV 에 Self-knowledge distillation(SKD)을 적용하여 성능을 개선하였다.

Model	mAP	mAP(applied SKD)
SLV(Thresholding)	55.0	56.6
SLV(Adaptive search)	55.7	57.2

표 3. Comparison with Self-knowledge distillation method

4. 결론 및 향후 연구 방향

Self-Knowledge distillation 을 이용하였을 경우 teacher-based method 로 학습하는 것보다 training cost 가 줄어들게 된다. 그리고 고성능의 최신 기법은 아니지만 teacher-based method 보다 정량적인 성능이 더 좋을 수 있었다. 하지만 classic teacher-based method 와의 비교이기 때문에 여전히 성능적으로 teacher-based method 에 비해 낮기 때문에 성능을 보완할 수 있는 추가적인 연구가 진행되어야 한다.

5. Acknowledgement

This work was supported by IITP grants funded by the Korea government (MSIT) (No. 2021-0-02068, AI Innovation Hub and RS-2022-00155915, Artificial Intelligence Convergence Research Center (Inha University)), and was supported by the NRF grant funded by the Korea government (MSIT) (No. 2022R1A2C2010095 and No. 2022R1A4A1033549).

6. 참고문헌

- [1] Han, Song, et al. "Learning both weights and connections for efficient neural network." *Advances in neural information processing systems* 28 (2015)
- [2] Gou, Jianping, et al. "Knowledge distillation: A survey." *International Journal of Computer Vision* 129.6 (2021): 1789-1819.
- [3] Xu, Ting-Bing, and Cheng-Lin Liu. "Data-distortion guided self-distillation for deep neural network." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. No. 01. 2019
- [4] S-J. Cho et al. "Blur-Robust Object Detection Using Feature-Level Deblurring via Self-Guided Knowledge Distillation" *IEEE Access* (2022)
- [5] Weipeng Xu et al. "Robust and Accurate Object Detection Via Self-Knowledge Distillation" *arXiv* (2021)
- [6] Peizhen Zhang, et al. "LGD: Label-guided Self-distillation for Object Detection" on *AAAI Conference* (2022)
- [7] Ze Chen, et al. "Spatial likelihood voting with Self-knowledge distillation for weakly supervised object detection" *arXiv*(2022)