
COSE474-2024F: Final Project Proposal

“Does a Good Beginning Really Guarantee a Good Ending?”

Yurim Lee / 2020390472 / Computer Science

1. Introduction

CLIP과 같은 대규모 사전 학습 모델의 등장으로, 모델을 효과적으로 다운스트림 작업에 적용하는 것은 중요한 과제가 되었다. 그러나 전체 모델을 파인튜닝하는 것은 많은 자원과 데이터를 필요로 한다. 이에 대한 대안으로, CoOp과 CoCoOp과 같은 텍스트 기반 프롬프트 학습 방법 및 MaPLe와 DAPT 등 텍스트와 이미지를 함께 프롬프팅하는 방법이 등장했다. 이러한 접근은 비전과 언어 프롬프트를 효과적으로 결합하여 CLIP의 성능을 극대화하고자 한다. 프롬프트 학습은 파라미터 효율성이 좋지만, 텍스트와 이미지 프롬프트의 align이 어렵고, 특히 few-shot 학습 환경에서 성능이 불안정하다. DAPT(Distribution-Aware Prompt Tuning)([Cho et al., 2023](#))는 Intra 및 Inter Dispersion을 적용하여 프롬프트 정렬을 개선했지만 극단적인 few-shot 상황에서는 여전히 한계를 갖는다. 이에 본 연구는 Learnable vector로 이루어진 prompt의 초기화 방식 변경과 프롬프트 길이 조정 등을 통해 극소수의 샘플만 있는 few-shot 상황에서도 안정적인 성능을 보이는 방법론을 탐구하고자 한다.

2. Problem definition & challenges

프롬프트 학습은 파라미터 효율적이지만 극단적인 few-shot 환경에서 성능이 다소 불안정하다는 한계를 갖는다. RPO(Read-only Prompt Optimization)([Lee et al., 2023](#))는 special token initialization을 통한 초기화 방식이 모델 성능을 향상하는데 효과적일 수 있다는 것을 밝혔다. 따라서 본 연구에서는 RPO의 특수 토큰 기반 초기화 방법을 활용하여 DAPT의 성능을 향상시키고자 한다. 또한, DAPT의 텍스트 및 이미지 프롬프트의 길이 변화와 클래스 토큰 위치 조정 등의 고찰을 통해 few-shot 학습과 같은 데이터가 부족한 상황에서도 모델의 성능을 개선하는 것을 목표로 한다.

3. Related Works

DAPT는 비전-언어 모델의 성능을 향상시키기 위해 프롬프트 튜닝 방식을 개선한 기법이다. 텍스트와 이미지 임베딩 간의 분포 정렬을 최적화하여 모델 성능을 높였으며, Inter-dispersion loss와 Intra-dispersion loss를 기존 CLIP 손실 함수에 결합해 새로운 손실 함수로 정의하였다. 레이어 추가 없이 프롬프트만으로 성능을

개선하고 다양한 데이터셋에서 일반화 성능을 향상시켰지만, 매우 적은 few-shot 상황에서는 여전히 성능이 불안정하다. RPO는 프롬프트 튜닝 과정에서 모델의 내부 표현에 미치는 영향을 최소화하며 성능을 높이기 위해 제안된 방법이다. Masked Attention을 사용해 사전 학습된 모델의 내부 표현 변화를 방지하고, Special Token-based Initialization를 통해 프롬프트를 초기화하여 few-shot 학습에서 뛰어난 성능을 발휘하였다.

4. State-of-the-art methods and baselines

Baseline은 DAPT의 1, 2, 16-shot으로 설정하였다. DAPT 16-shot 실험 결과, EuroSAT 데이터셋에서 CoOp에 비해 6.53%, StanfordCars 데이터셋에서는 VPT보다 12.43% 더 높은 정확도를 달성했다. 1-shot 실험에서는 EuroSAT에서 DAPT가 38.54%, StanfordCars에서 67.39%의 정확도를 보였다. EuroSAT에서는 CoOp, VPT, zero-shot CLIP보다 낮은 성능을 보였지만, StanfordCars에서는 동일하거나 더 높은 성능을 기록했다. 따라서 두 데이터셋을 활용해 새로운 개선 방법론을 평가하는 것은 DAPT가 기존에 강점을 보였던 분야뿐만 아니라, 부족했던 점을 해결할 수 있는지 검증하는 중요한 지표가 될 수 있다.

5. Schedule & Roles

1~3주차: DAPT 구조와 RPO ST-Initialization 방법론 및 코드 탐구, 다양한 프롬프트 개선 방안 탐구

4~7주차: DAPT 프롬프트 변경, 학습, 실험, 문제 보완 및 해결, 개선

8주차: 보고서 작성

References

- Cho, E., Kim, J., and Kim, H. J. Distribution-aware prompt tuning for vision-language models. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 22004–22013, 2023.
- Lee, D., Song, S., Suh, J., Choi, J., Lee, S., and Kim, H. J. Read-only prompt optimization for vision-language few-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1401–1411, 2023.