SemiVL(Semi-Supervised Semantic segmentation with Vision-Language Guidance) 모델의 임베디드 시스템 포팅(Porting)에 관한 연구

박승민[1], 김두상[2]
[1]동서울대학교 컴퓨터소프트웨어학과 학부생
[2]동서울대학교 컴퓨터소프트웨어학과 교수
7195sm@naver.com, dskim@du.ac.kr

Research on embedded system porting of SemiVL(Semi-Supervised Semantic segmentation with Vision-Language Guidance) model

Seung-Min Park[1]
Du-Sang Kim[1]
[1]Dept. of Computer software, Dong seoul University

요 약

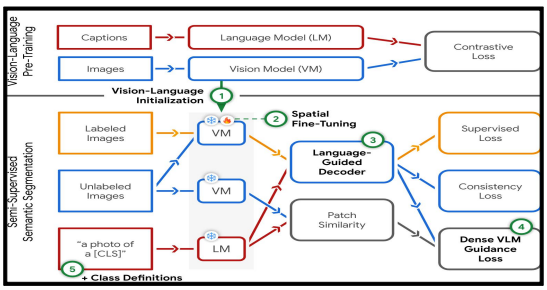
SemiVL(Semi-Supervised Semantic Segmentation with Vision-Language Guidance) 모델은 자원이 제한된 환경에서도 높은 이미지 분할 성능을 발휘하는 준지도 학습 기반의 시맨틱 세그멘테이션 모델이다. 본 논문은 PyTorch 프레임워크에서 TorchScript 프레임워크로 변환된 SemiVL 모델을 임베디드 시스템 환경(Google Pixel 2)에 적용하여 온디바이스 AI를 구현한 연구이다. 목표는 데스크톱 GPU 환경과 유사한 추론 성능을 달성하는 것이었다. 성능 평가는 Pascal VOC 데이터셋을 사용하였으며, mIoU(mean Intersection over Union)와 추론 시간을 주요 지표로 측정하였다. 실험 결과, TorchScript로 변환된 SemiVL 모델은 데스크톱 PC에서 77.5%의 mIoU와 6438.99ms의 추론 시간을 기록하였고, Google Pixel 2에서는 62.8%의 mIoU와 6658.45ms의 추론 시간을 달성하였다. 이 결과는 임베디드 시스템 환경에서 SemiVL 모델이 온디바이스 AI 솔루션으로 활용될 수 있음을 보여준다.

1. 서론

임베디드 시스템의 성능 향상으로 인해 인공지능 모델을 스마트폰과 같은 Portable device에 적용하는 연구가 인공지능 분야에서 중요한 주제로 떠오르고 있다. 특히, 컴퓨터 비전 모델을 임베디드 시스템에 적용함으로써 실시간 데이터 처리와 분석이 가능해져 자율주행, IoT, 의료 영상 분석과 같은 다양한 분야에서 활용할 수 있는 가능성을 열어주고 있다. 본 논문에서는 SemiVL모델을 기존의 Pytorch 프레임워크(Framework)에서 TorchScript 프레임워크로 변환하여, 특정 GPU 제조사인 엔비디아(Nvidia) GPU가 없는 환경에서도 GPU가 있는 환경과 유사한 이미지 분할 성능을 유지할 수 있게 연구를 수행하였다.

2. SemiVL

SemiVL[1] 모델은 "Semi-Supervised Semantic Segmentation with Vision-Language Guidance" 모델로, 객체 분할 작업 중에서도 Semantic Segmentation 작업을 수행하며, 제한된 양의 라벨링된 데이터와 많은 양의 라벨링되지 않은 데이터를 활용해 학습하는 준지도 학습(Semi-supervised learning) 문제를 해결한다. 기존의 준 지도 학습 기반 컴퓨터 비전 모델들은 제한된 양의 라벨링 데이터로 인해 유사한 시각적 특징을 가진 클래스 간의 구분이 어려운 경우가 많았다. SemiVL 모델은 사전 학습된 CLIP(Contrastive Language-Image Pre-Training)[2] 모델을 활용하여 이러한 문제를 해결하였다. 또한, 라벨링된 데이터셋에 과적합되지 않도록 하기 위해 학습 중 라벨링되지 않은 이미지를 가중치가 고정된 CLIP 인코더에 입력해 예측된 마스크와 비교하는 손실 함수를 계산한다. 이와 같은 특징 덕분에 SemiVL 모델은 제한된 라벨 데이터로도 높은 성능을 발휘할 수 있다.



(그림 1) SemiVL 모델의 구조도

3. PyTorch에서 TorchScript로 변환

PyTorch는 딥러닝 연구와 개발에서 널리 사용되는 오픈 소스 프레임워크로, 직관적인 인터페이스와 동적 계산 그래프(define-by-run) 특성을 통해 복잡한 신경망 모델의 설계와 훈련을 쉽게 할 수 있게 해준다. 그러나 이러한 동적 특성은 메모리 사용량이 높고 실행 속도가 느려, 임베디드 시스템과 같은 자원이 제한된 환경에서는 최적화된 배포가 어려움을 초래할 수 있다. TorchScript는 PyTorch 모델을 정적 그래프로 변환하여 다양한 플랫폼에서 최적화된 성능을 발휘할 수 있도록 지원하는 프레임워크다. 이를 통해 변환된 모델은 메모리와 연산 자원을 효율적으로 사용하여 임베디드 시스템에서도 높은 성능을 유지할 수 있다. 이러한 최적화는 모델 배포와 성능 향상에 필수적이며, 제한된 자원 환경에서 PyTorch 모델의 실용성을 크게 높여준다.

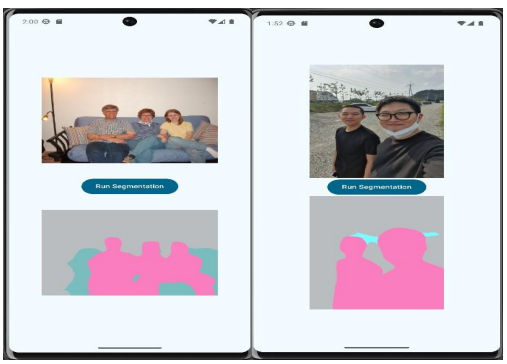
4. 실험환경 및 결과

<표 1>은 실험을 수행한 Google Pixel 2와 성능 비교 대상인 데스크톱 PC의 하드웨어, 소프트웨어 규격이다.

<표 1> Google Pixel 2의 하드웨어, 소프트웨어 규격

구분		항목	내용
Google Pixel 2	H/W	CPU	Snapdragon 835
		GPU	Adreno 540
		RAM	4 GB
	S/W	OS	Android 14
데스크톱 PC	H/W	CPU	i7-12700K
		GPU	RTX 3080
		RAM	16GB
	S/W	OS	Windows 11

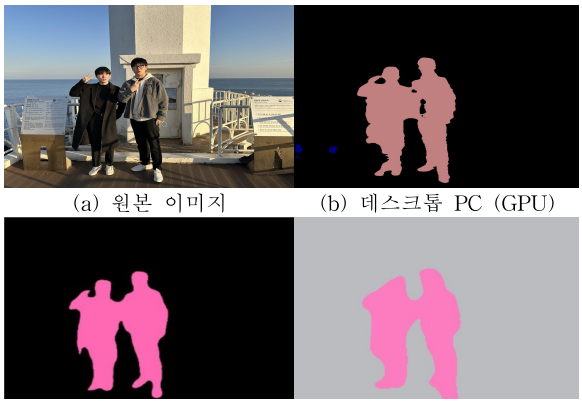
SemiVL을 TorchScript 프레임워크를 사용하여 모바일 환경 Google Pixel 2에 탑재하였고, Pascal VOC Dataset에 대한 모델의 성능을 mIoU(mean Intersection over Union)로 비교 평가하였다. 실험 결과, Google Pixel 2는 데스크톱 PC(GPU)와 비교해 21.2%(mIoU)의 성능 저하와 423.91(msec)의 추론 시간 증가를 보였다. 그림 2는 Google Pixel 2에서 구현한 어플리케이션의 결과화면이다.



(그림 2) Google Pixel 2에서 구현된 결과화면

<표 3> Pascal VOC Dataset에 대한 성능 비교

구분		mIoU(%)	추론시간 (msec)
PyTorch	데스크톱 PC (GPU)	84.0	6234.54
TorchScript	데스크톱 PC (CPU)	77.5	6438.99
	Google Pixel 2	62.8	6658.45



(a) 원본 이미지 (b) 데스크톱 PC (GPU)

(c) 데스크톱 PC(CPU) (d) Google Pixel 2

(그림 3) 각 환경의 Zero-shot 예측 수행 결과.

5. 결론

본 연구에서는 SemiVL 모델을 TorchScript 프레임워크로 변환한 후 모바일 환경에 적용해 실험을 수행하였다. 변환 과정에서 TorchScript의 트레이싱 방식을 사용하여 모델을 최적화하고 배포 가능하게 만들었다. 실험 결과, 변환된 모델은 임베디드 시스템에서도 원활하게 작동하며 효과적인 성능을 보였다. 구체적으로, Google Pixel 2는 데스크톱 PC(GPU)에 비해 21.2%의 mIoU 성능 저하와 423.91ms의 추론 시간 증가를 나타냈지만, 이 정도의 성능 저하는 임베디드 시스템에서 On-device AI로 활용하기에 충분히 수용 가능하다고 판단된다. 이를 통해 SemiVL 모델의 임베디드 환경에서의 활용 가능성을 입증하였으며, 앞으로도 다양한 최적화 기법을 통해 임베디드 시스템에서 인공지능 모델의 성능을 더욱 향상시킬 수 있을 것으로 기대된다. 본 연구 결과는 향후 SAM2와 같은 다른 segmentation 모델을 On-device AI로 활용하기 위한 방안 연구에도 기여할 수 있을 것이다.

참고문헌

[1] Lukas Hoyer, et.al "SemiVL: Semi-Supervised Semantic Segmentation with Vision-Language Guidance" , ECCV24, 2024

[2] Feng Liang, et.al “Open-vocabulary semantic segmentation with mask-adapted clip”. In CVPR, pages 7061 - 7070, 2023