

가상 환경 기반 미세 조정 모델의 의미론적 분할 성능 분석 및 고찰

류경빈, 오정민, 강지수, 김준영

성신여자대학교

20211327@sungshin.ac.kr, 20211366@sungshin.ac.kr, 20221322@sungshin.ac.kr,
jkim@sungshin.ac.kr

Performance Analysis of Semantic Segmentation in Fine-tuned Models Based on Virtual Environment

Ryoo Gyeong Bin, Oh Jeong Min, Kang Ji Soo, Joon Young Kim

Sungshin Women's University

요 약

본 논문에서는 가상환경 데이터를 이용하여 미세 조정된 의미론적 분할(semantic segmentation) 모델의 성능을 분석한다. 이를 위해 NVIDIA Omniverse 플랫폼의 Isaac Sim 기반으로 구축된 가상환경에서 데이터를 수집하고 의미론적 분할 모델의 미세 조정을 진행하며, 이를 기반으로 가상환경 및 실제 환경에서의 의미론적 분할 성능에 대한 검증 및 분석을 진행한다.

I. 서 론

최근 가상 환경을 이용한 실제 환경에 대한 시뮬레이션 실험이 활발하게 진행되고 있다. [1] 특히 모빌리티 서비스에 대한 가상환경 기반 실험 또한 진행되는 추세인데, 가상환경 내에서 AI 기반 로봇 시뮬레이션이 가능한 NVIDIA Omniverse 플랫폼의 Isaac Sim을 이용한 실험이 이에 해당한다. 모빌리티 서비스 상에서는 안정적인 서비스 제공을 위하여 외부 객체가 존재하는 영역을 정확히 파악하는 데에 있어 카메라 데이터의 각 픽셀을 레이블로 분류하는 의미론적 분할의 중요성이 대두된다.

본 논문에서는 가상 환경 데이터를 학습한 모델의 의미론적 분할 성능을 분석하고자 한다. 이를 위해 NVIDIA Omniverse 플랫폼의 Isaac Sim 가상환경에서 추출한 데이터를 기반으로 의미론적 분할 모델의 미세 조정 후 가상환경 및 실제 환경에서의 성능 분석을 진행한다.

II. 사용 모델

본 논문에서는 의미론적 분할을 위한 아키텍처인 Fully-Convolutional Network (FCN) 모델을 기반으로 한 FCN-ResNet 50 모델을 사용한다. ResNet-50 아키텍처에서 특징이 추출되어 스킵 연결을 통해 FCN으로 전달된다. 훈련 과정에서는 사전 훈련된 모델의 가중치를 제외한 분류 색션의 가중치만 조정하며, 모델을 훈련 모드로 설정하여 이미지 및 레이블을 통과시켜 손실을 계산하고 역전파를 수행한다.

본 논문에서는 크로스 엔트로피 손실(Cross Entropy Loss) 함수를 손실 함수로 사용하였는데, 이는 분류 문제에서 실제 클래스와 예측 클래스 사이의 확률 분포 차이를 계산하는 방법으로 다음과 같이 정의된다. [2]

$$l_n = -w_{y_n} \log\left(\frac{\exp(x_{n,y_n})}{\sum_{c=1}^C \exp(x_{n,c})}\right) \quad (1) \quad \ell(x, y) = \frac{1}{N} \sum_{n=1}^N l_n \quad (2)$$

여기서 x 는 입력 이미지(RGB)이며, 각 이미지는 특정 클래스에 속하는 픽

셀의 확률을 예측하기 위해 모델에 의해 처리된다. y 는 대응하는 세그멘테이션 마스크(실제 클래스의 픽셀 집합)로, 각 픽셀의 실제 클래스를 나타내는 타깃값이다, w 는 가중치 값, C 는 클래스 개수, N 은 미니 배치의 차원을 나타내며 d_1, \dots, d_K 는 K 차원 케이스의 각 차원을 지칭한다. 참고로 (1)에서 $\ell(x, y)$ 의 경우 $\ell(x, y) = L = l_1, \dots, l_N$ 을 가정한다.

해당 손실 함수는 모델이 출력한 확률과 실제 레이블 간의 일치를 측정하며 잘못 분류된 클래스에 대해 강한 그래디언트를 부여함으로써 각 클래스에 대한 예측 확률을 높인다.

III. 실험 및 결과

3.1 데이터 수집

본 논문에서는 '배경', '칸', '상자' 등 warehouse with forklifts 환경에 기본적으로 라벨링된 객체를 포함한 22가지 클래스에 대한 데이터를 수집하였다. 데이터 수집은 NVIDIA Isaac Sim 환경에 Turtlebot3 Burger를 배치하여 진행하였다. 이 과정에서 NVIDIA Isaac Sim Prop Assets의 box 및 can 객체들을 배치하였다. 데이터 기록을 진행한 과정은 그림1과 같다.

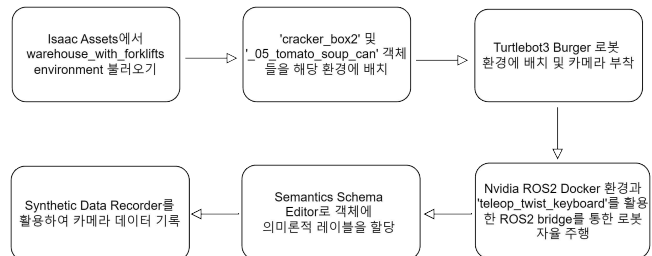


그림1. FCN ResNet-50 아키텍처내 데이터 기록 과정

Isaac ROS Docker 개발 환경에서 ROS2 명령어를 사용하여 가상 환경 내 Turtlebot의 작동이 수행되었으며, 다양한 환경에서 정밀한 위치 조정과 효과적인 데이터 수집이 가능하게 하였다. 의미론적 레이블의 경우 Omniverse 플

랫폼의 확장 도구인 Semantics Schema Editor 도구를 활용하여 할당하였다. Synthetic Data Recorder를 이용하여 이미지 데이터, 의미론적 분할을 위한 색상-클래스 매핑 데이터 및 의미론적 분할 마스크 데이터를 기록 및 저장하였다.

3.2 실험

가상환경 내 실험에서는 Realsense D405 카메라를 사용하였다. 모델은 PyTorch 프레임워크를 이용하여 구현되었으며, 학습 절차는 모든 백본 매개변수를 동결하고 주 분류기와 필요에 따라 보조 분류기를 조정하여 사용자 정의 클래스 수에 맞추었다. 하이퍼파라미터는 표 1과 같으며 모델의 학습은 NVIDIA A100 GPU를 사용하여 진행하였다.

| Hyperparameter | Value |
|------------------------------|--------------------|
| Learning Rate | 0.01 |
| Number of Classes | 22 |
| Optimizer | Adam |
| Loss Function | Cross Entropy Loss |
| Model Architecture | FCN ResNet50 |
| Training Device | GPU/CPU |
| Backbone Parameters Freezing | Yes |
| Train Auxiliary Classifier | Yes |

표1. FCN ResNet-50 아키텍처내 모델 하이퍼파라미터

FCN ResNet-50 모델을 사용하여 22개 클래스에 대한 의미론적 분할 작업을 수행하였으며, 사전 훈련된 가중치로 초기화된 후 대상 데이터에 맞게 미세 조정되었다. 모델은 torchvision에서 제공되는 fcn_resnet50을 사용하였으며, 손실 함수로 Cross Entorpy Loss를 사용하였다.

FCN ResNet-50 모델은 '배경', '칸', '상자', 그리고 소화가, 지게차, 천장 등 warehouse with forklifts 환경에 기본적으로 라벨링된 객체를 포함한 스물 두 가지 클래스에 대해 의미론적 분할을 수행하였다. 훈련 과정에서는 CUDA를 지원하는 GPU 상에서 Jaccard Index, F1 Score, Recall, Precision 지표를 통해 모델의 성능을 평가하였다.

본 연구에서 사용된 모델은 이미지의 각 픽셀에 대한 클래스 레이블을 예측함으로써 이미지의 전체적인 의미론적 분할을 수행하였다. 모델의 의미론적 분할 성능을 표 2에서 확인할 수 있다.

| Metrics | Values |
|-----------------|--------|
| Training Loss | 0.09 |
| Validation Loss | 0.09 |
| IOU | 0.63 |
| F1 | 0.96 |
| Precision | 0.96 |
| Recall | 0.96 |

표2. FCN ResNet-50 아키텍처 기반 훈련 결과

그림 2은 NVIDIA Issac Sim의 warehouse with forklift 환경 데이터의 can 및 box를 예측한 결과이며, 가상환경 데이터에 대하여 미세 조정을 거친 모델의 의미론적 분할 예측 성능을 확인할 수 있다.



그림2. 가상환경 데이터셋을 활용한 모델 예측

실제 데이터에 대한 예측 성능 확인을 위해 Turtlebot3 Burger에 D405와 유사한 기능을 탑재한 Realsense D435 카메라를 부착하여 주행하여 실제 데이터를 수집하였다. can에 해당하는 원기둥 모형과 box에 해당하는 직육면체 모형을 촬영하였으며, 수집 데이터 경우 labelme를 이용하여 분할 마스크를 생성하였다. 실제 환경 데이터 대상 초기 예측 경우 낮은 성능을 기

록하였는데 이는 가상환경내 학습된 모델이 실제 환경에서 동일 성능을 발휘하지 못할 수 있음을 시사하며, 이를 극복할 방안을 모색할 필요가 있다.



그림 3. 훈련된 ResNet-50 활용한 초기 실 환경 데이터 예측 결과
본 연구에서 가상환경에서의 모델 학습이 실제 환경 데이터에 대한 예측 능력이 떨어지는 주된 이유를 분석한 결과, 가상환경이 제공하는 복잡한 시나리오가 오히려 모델이 간단한 형태를 인식하는 능력을 저하시킬 수 있음을 확인하였다. 이에 실제 데이터에 대한 미세조정을 진행하였다.

원본 FCN-ResNet50 모델, 가상환경 데이터로 미세 조정된 모델, 그리고 실제 데이터로 다시 미세조정된 모델의 성능을 표 3에서 비교하였다. 실제 데이터를 통한 2차 미세 조정을 진행하였을 경우 모델의 인식 정확도가 유의미하게 향상되었다는 사실을 확인할 수 있었다. 이러한 결과는 가상환경에서 학습된 데이터가 모델의 성능을 개선에 활용될 수 있음을 시사하며, 이는 가상환경 데이터의 유용성을 입증하는 근거로 볼 수 있다.

| Model | Validation Loss | IOU | F1 Score | Precision | Recall |
|----------|-----------------|-------|----------|-----------|--------|
| 원본 모델 | 0.33 | 0.32 | 0.96 | 0.96 | 0.96 |
| 1차 미세 조정 | 32.7 | 0.01 | 0.01 | 0.01 | 0.01 |
| 2차 미세 조정 | 0.007 | 0.974 | 0.998 | 0.998 | 0.998 |

표3. 미세조정 단계에 따른 FCN-ResNet50 모델의 예측 결과

IV. 결론

본 논문에서는 가상환경 데이터를 이용하여 미세 조정된 의미론적 분할 모델의 성능을 분석하였다. 가상환경 데이터에 대하여 미세 조정된 FCN-ResNet50 모델은 가상환경 데이터에 대한 예측은 적절히 수행하였으나 실제 환경 데이터에 대해서는 낮은 정확도를 나타내었다. 그러나 소량의 실제 환경 데이터에 대한 2차 미세튜닝을 수행한 결과 유의미한 성능 개선을 확인할 수 있었다. 특히, 예측된 마스크와 실제 마스크간의 높은 유사도를 보이며 그림 4와 같이 효과적인 의미론적 분할을 수행한다.

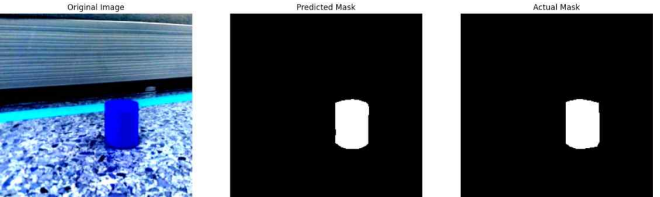


그림 4. 2차 미세 조정된 모델을 사용한 실제 환경 데이터 예측 결과
본 결과는 미세 조정된 모델이 초기에는 단순한 패턴 인식에 대한 문제점을 추가 미세조정을 통해 원본 모델보다 효과적으로 의미론적 분할이 가능함을 보여준다. 향후 실제 환경에 대한 추가적인 학습 진행 없이 유의미한 예측이 가능한 범용적인 의미론적 분할 모델에 대한 연구를 진행할 예정이다.

참 고 문 헌

[1] J. Collins, S. Chand, A. Vanderkop and D. Howard, "A Review of Physics Simulators for Robotic Applications," in IEEE Access, vol. 9, pp. 51416-51431, 2021

[2] Pytorch, "CrossEntropyLoss", 2023, Available From: <https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html>