트랜스포머 모델을 이용한 비협조 위성의 상대 Pose 추정

조상현[1]*, 조한철[1]
연세대학교 천문우주학과[1]

Pose estimation of a non-cooperative satellite using transformer model

Sanghyeon Cho[1*], Hancheol Cho[1]

Key Words : Vision-based navigation(비전 기반 항법), Satellite pose estimation(인공위성 포즈 추정), Computer vision(컴퓨터비전), Artificial intelligence(인공지능), Deep learning(딥러닝)

서 론

최근 활발히 연구되고 있는 궤도상 서비스 임무에서는 target 위성의 상대위치와 자세 정보가 필수적이다. 비협조 위성은 통신이 어려워 궤도상에서 이를 직접 추정해야 하는데, chaser 위성의 카메라로 얻은 이미지를 분석하는 것이 효과적이다.

이미지 기반 상태 추정에는 딥러닝이 널리 쓰인다. 하지만 실제 우주 기반 이미지는 확보가 어려워, 합성 이미지와 소량의 실제 이미지를 학습에 사용한다. 대표적인 dataset이 Stanford SLAB의 SPEED+이다. 이를 활용한 pose 추정 연구는 많았지만 기존 연구의 상당수는 CNN 기반 모델이었다.

이번 연구에서는 트랜스포머 기반의 Pyramid Vision Transformer[(2)]를 backbone으로 사용한 모델을 설계하였다. 해당 모델은 객체 탐지, 쿼터니언과 상대 위치 벡터 추정을 동시에 진행한다. SPEED+ 데이터에 해당 모델을 학습 및 평가하여 모델의 성능을 분석하였다.

본 론

모델 설계

설계한 모델은 객체 탐지, 쿼터니언과 상대 위치 벡터 추정을 목적으로 만들어졌다. 객체 탐지는 multi-scale feature map을 이용하면 성능이 증가한다. 따라서 feature extractor 역할을 하는 backbone으로서 Pyramid Vision Transformer(PVT)[(2)]를 선택했다. PVT는 트랜스포머 기반 모델으로서 4개의 multi-scale feature map을 출력한다. PVT는 레이어의 깊이에 따라 Tiny, Small, Medium, Large로 구분되는데 해당 연구에서 사용한 것은 PVT-small이다.

객체 탐지에는 Faster R-CNN[(3)]이라는 모델을 사용하였다. Faster R-CNN은 PVT에서 나온 feature map을 바로 입력으로 받지 않는다. 객체 탐지의 성능을 높이기 위해 Feature Pyramid Network(FPN)[(4)]를 사용하여 정보가 풍부한 고해상도의 feature map을 입력으로 받게 설계하였다.

쿼터니언과 상대 위치 벡터 추정은 2단계로 행해진다. 먼저 모델은 위성의 11개 keypoint에 대한 heatmap을 추정한다. 그 후, 추정된 heatmap이 PnP 알고리즘의 입력으로 들어가면 쿼터니언과 상대 위치 벡터가 출력된다. Heatmap 추정에 사용하는 신경망은 ViTPose[(5)] 모델의 classic decoder를 수정한 것을 사용하였다. 기존 classic decoder는 Deconv-BN-ReLU로 구성된 2개의 block으로 이루어져 있다. 하지만 이번 연구에서는 모델에 맞게 1개의 block으로만 신경망을 만들었다.

모델의 전체적인 구조는 Fig. 1.과 같다. Backbone에서 나온 feature map이 각 head로 들어가서 객체 탐지와 pose 추정에 사용된다.

Loss 설계

객체 탐지에 사용한 loss는 Faster R-CNN의 loss를 그대로 사용하였다. Faster R-CNN의 loss는 binary cross-entropy, cross-entropy, smooth L1로 이루어져있다. Binary cross-entropy는 배경, 물체 이진분류에 사용되며, cross-entropy는 객체의 종류 추정, 마지막으로 smooth L1은 box 회귀에 사용된다.

Heatmap 추정에 사용한 loss는 Mean Squared Error를 사용하였다.

Dataset

이번 연구에서 사용된 데이터는 Stanford SLAB에서 생성한 SPEED+[(1)] dataset이다. SPEED+는 OpenGL로 생성된 합성사진과 SLAB의 TRON 시설을 이용하여 만든 실제사진으로 이루어져있다. 학습에 사용되는 합성사진은 총 59960장이며 성능 평가에 사용되는 실제사진은 총 9531장이다.

모델은 합성사진으로 학습되지만 성능 평가는 실제사진으로 진행되기 때문에 domain gap 문제가 발생한다. 이 문제를 해결하기 위해 data augmentation을 진행하였다. 합성사진에 noise 또는 blur를 추가하거나 위성 일부를 가리는 기법 등을 사용하여 domain gap 문제를 완화시켰다.

추가적으로, chaser 위성이 target 위성의 사진을 궤도상에서 얻었다고 가정하여 실제사진 일부를 학습에 사용하여 모델 성능이 얼마나 향상되는지도 비교해보았다.

결 론

Fig. 2.~ 5.는 합성사진의 검증 데이터를 이용한 결과이다. Fig. 2.와 3.을 보면 객체 탐지와 heatmap 추정이 잘 되는 것을 확인할 수 있다. Fig. 4.와 5.는 추정된 heatmap을 이용하여 pose를 추정한 결과이다. 상대 위치 벡터에 비해 쿼터니언 추정이 비교적 어려

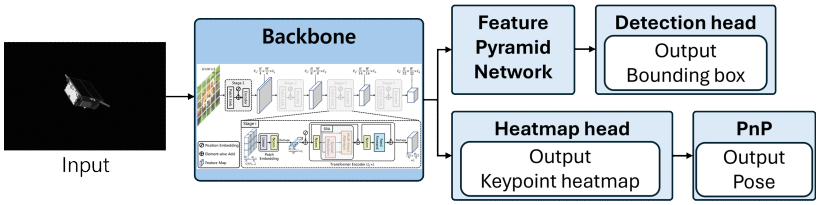

Fig. 1. Pose estimation model

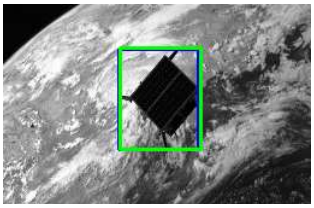

Fig. 2. 객체 탐지 결과

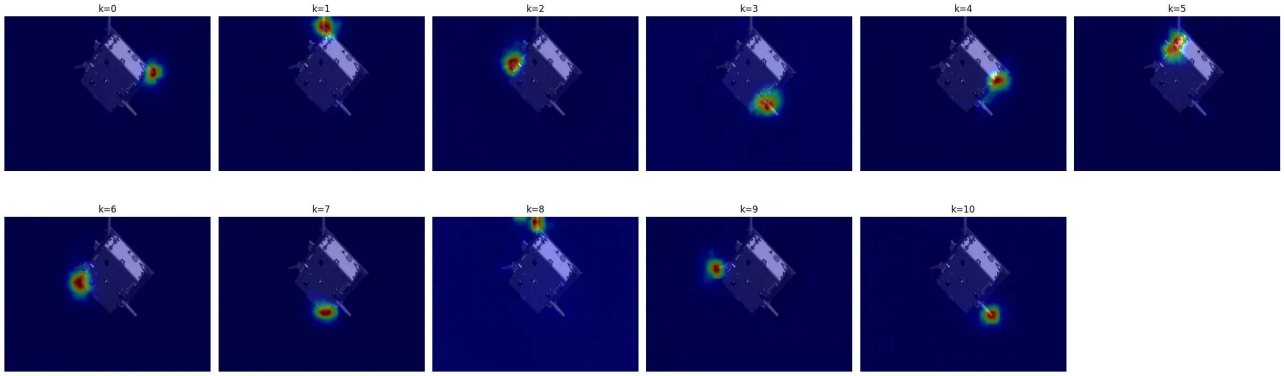

Fig. 3. Heatmap 추정 결과

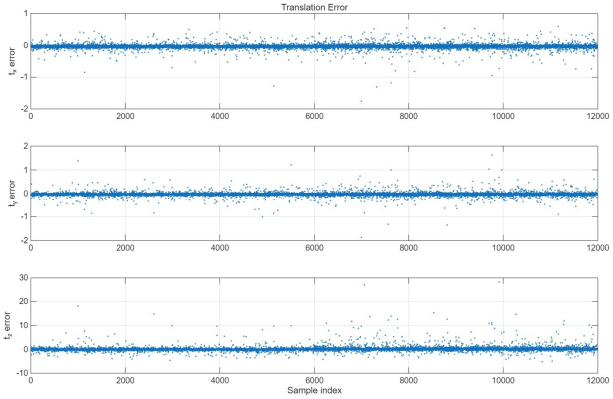

Fig. 4. 상대 위치 벡터 에러

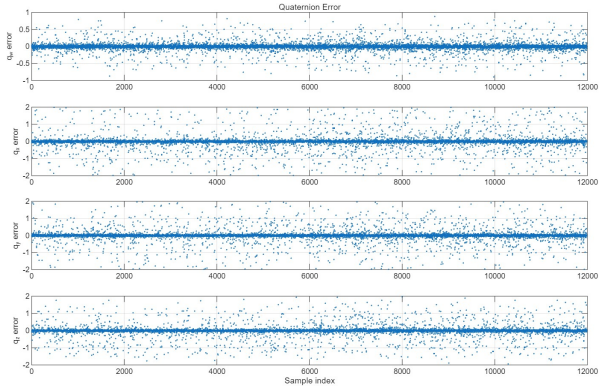

Fig. 5. 쿼터니언 에러

운 것을 확인할 수 있다. 학회 전까지 추가적인 연구 및 모델 학습 등을 진행하여 모델의 전체적인 pose 추정 성능을 향상시킬 예정이다. 또한 domain gap 문제를 해결하여 실제사진으로 pose 추정을 진행하고 최종적인 모델 성능을 도출할 예정이다.

후 기

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (RS-2025-00560240).

참고문헌

1) Park Tae Ha, Märtens Marcus, Jawaid Mohsi, Wang Zi, Chen Bo, Chin Tat-Jun, Izzo Dario, and D'Amico Simone, "Satellite Pose Estimation Competition 2021: Results and Analyses," *Acta Astronautica*, Vol. 204, 2023, pp. 640-665

2) Wang Wenhai, Xie Enze, Li Xiang, Fan Deng-Ping, Song Kaitao, Liang Ding, Lu Tong, Luo Ping, and Shao Ling, "Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 568-578.

3) Ren Shaoqing, He Kaiming, Girshick Ross, and Sun Jian, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *Preceedings of the 28th International Conference on Neural Information Processing Systems (NIPS 2015)*, 2015, pp.91-99

4) Lin Tsung-Yi, Dollár Piotr, Girshick Ross, He Kaiming, Hariharan Bharath, and Belongie Serge., "Feature Pyramid Networks for Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp.2117-2125.

5) Xu Yufei, Zhang Jing, Zhang Qiming, Tao Dacheng, "ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation," *Advances in Neural Information Processing Systems (NeurIPS 2022)*, 2022, pp.38571-38584.