

딥러닝 기반 6D 객체 포즈 추정을 위한 경계선 강화 블록

엽옥영, 박한훈*

부경대학교

yeyuning12@gmail.com, *hanhoon.park@pknu.ac.kr

A Boundary Enhancement Block for Deep Learning-Based 6D Object Pose Estimation

Yuning Ye, Hanhoon Park*

Pukyong National Univ.

요약

본 논문은 RGB 입력 영상으로부터 텍스처가 부족한 객체의 6D 포즈를 추정하는 딥러닝 모델 중 하나인 EPro-PnP의 성능을 개선하기 위해 경계선 강화 블록을 추가하여 객체의 경계선 영역의 특징 정보를 강화하는 방법을 제안한다. LINEMOD 데이터 셋을 사용한 실험을 통해 텍스처가 부족한 객체의 경우 6D 포즈 추정에 유용한 특징 정보가 경계선 영역에 집중되기 때문에 이를 강화함으로써 EPro-PnP의 포즈 추정 정확도가 향상될 수 있음을 확인하였다. 또한, 경계선 강화 블록을 추가하는 위치에 따른 성능 차이를 비교, 분석하였다.

I. 서론

객체 6D 포즈 추정은 객체가 존재하는 월드 좌표계를 기준으로 한 카메라 좌표계의 3D 회전 변환과 3D 이동 변환을 찾는 문제이다. 실용성 면에서 영상 정보를 사용한 방법이 좋으며, 최근에는 딥러닝 기술을 활용하여 단일 시점 RGB 영상으로부터 객체의 6D 포즈를 추정하는 방법에 대한 연구가 주를 이루고 있다. 대표적인 예로, EPro-PnP는 객체 영역 내 각 픽셀의 3차원 좌표를 추정하는 모델에 기존 PnP(Perspective-n-Points) 알고리즘을 대체하는 PnP 헤드(head)를 부착하여 end-to-end 방식으로 객체의 6D 포즈를 추정함으로써, 정확도를 크게 개선하였다[1]. 그러나, 객체 포즈 추정 정확도는 3차원 좌표 추정 모델의 백본(backbone) 네트워크 성능에 의존적이며, 입력 영상으로부터 객체 포즈 추정에 유용한 특징 지도(feature map)를 추출하는 것이 중요하다[2].

본 논문에서는 텍스처가 부족한 객체의 경우 객체 포즈 추정에 유용한 특징 정보가 객체 내외의 경계선(edge or boundary) 부분에 집중된다는 것을 발견하고, 이를 위해 객체 경계 부분의 특징 정보가 잘 부각된 특징 지도가 추출될 수 있도록 EPro-PnP의 백본 네트워크(그림 1)에 그림 2와 같은 경계선 강화 블록을 추가하는 방법을 제안한다. 경계선 강화 블록을 추가하는 위치에 따라 성능이 달라질 수 있으므로, 각 단계(stage) 또는 여러 단계에 추가했을 때의 성능을 비교, 검증한다.

그림 2의 경계선 강화 블록은 입력된 특징 지도의 각 채널에 Sobel 필터를 적용한 후 시그모이드 함수를 통해 가중치 값을 얻고, 이를 각 채널에 곱함으로써, 특징 지도의 경계선 영역은 강화되고, 에지 이외의 영역은 억제된다.

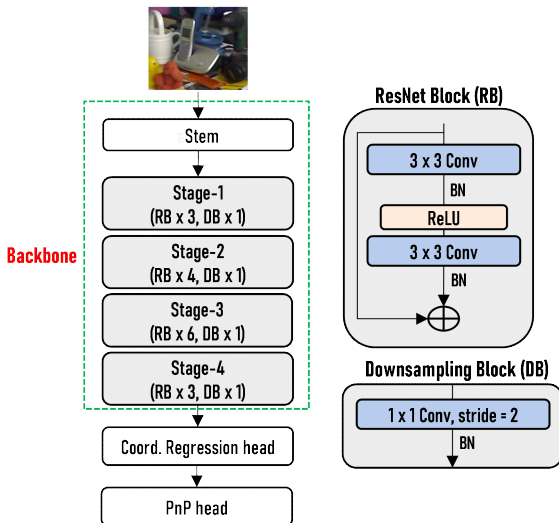


그림 1. 객체 6D 포즈 추정을 위한 EPro-PnP 모델 구조.

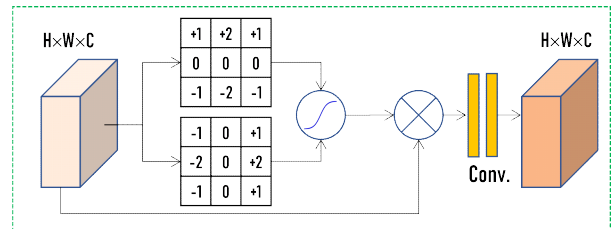


그림 2. 경계선 강화 블록.

II. 실험 및 결과

2.1. 실험 환경 및 성능 지표

객체 6D 포즈 추정을 위한 벤치마크 데이터 셋인 LINEMOD 데이터 셋 [3]을 사용하여 제안 방법의 성능을 검증하였다. LINEMOD는 15개의 텍스처가 거의 없는 객체에 대한 18,273장의 RGB 영상으로 이루어지며, 각 영상에서의 객체의 포즈 정보 및 3D CAD 모델이 함께 제공된다. 구현은 공개된 소스 코드[4]를 활용하였다. PyTorch를 사용하였으며, 학습은 사전 학습 없이(from scratch) 진행하였으며, 배치 크기는 16, epoch 수는

320, 학습률은 $1e-4$ 로 설정하고, RMSProp 옵티마이저($\alpha=0.99$, $\epsilon=10^{-8}$, $\lambda=0$, $\mu=0$)를 사용하였다. 성능 검증을 위한 지표로는 ADD-0.1d와 2D-Proj-5px를 사용하였다. ADD-0.1d는 추정된 포즈와 ground-truth 포즈에 의해 변환된 객체 3차원 좌표 사이의 거리가 객체 지름의 10% 이내인 입력 영상의 비율을 의미한다. 2D-Proj-5px는 추정된 포즈와 ground-truth 포즈에 의해 투영된 객체의 3차원 좌표 사이의 2차원 투영 오차가 5픽셀 이내인 입력 영상의 비율을 의미한다.

2.2. 실험 결과

그림 3은 제안된 경계선 강화 블록을 그림 1의 각 단계 뒤에 추가했을 때의 특징 지도의 차이를 보여준다. 여기서, s1은 Stage-1 뒤에 추가, s123은 Stage-1, Stage-2, Stage-3 뒤에 모두 추가하는 것을 의미한다. 추가되는 위치에 따라 양상은 달라지지만, 경계선 강화 블록을 추가함으로써 객체의 경계선 영역의 특징 정보가 강화되거나 경계선 이외의 특징 정보가 억제되었다. 이러한 효과는 객체 6D 포즈 추정 결과에서 긍정적인 효과로 나타났으며, 표 1과 2에서 보는 것처럼, 경계선 강화 블록을 추가함으로써 정확도가 향상되었다. 실험 결과, 경계선 강화 블록을 1, 2, 3 단계에 모두 추가한 s123의 성능이 가장 우수했다.

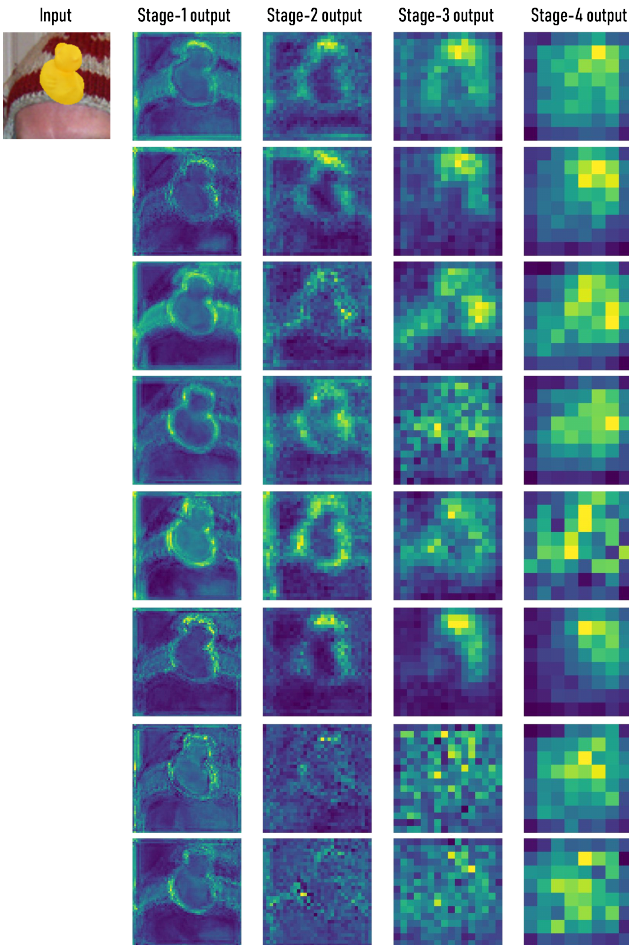


그림 3. 경계선 강화 블록 추가로 인한 특징 지도 변화. 각 행은 차례대로 EPro-PnP와 EPro-PnP에 경계선 강화 블록을 추가한 모델(s1, s2, s3, s4, s12, s123, s1234)들의 특징 지도 결과임.

표 1. 경계선 강화 블록의 위치에 따른 ADD-0.1d 결과

	-	s1	s2	s3	s4	s12	s123	s1234
ape	53.14	52.38	58.48	57.33	51.71	57.81	52.48	57.71
benchvise	88.17	91.66	88.55	90.49	87.20	90.98	92.34	88.55
camera	65.49	70.69	74.02	68.53	63.82	70.29	76.67	70.88
can	75.10	79.92	76.28	81.00	75.00	80.31	83.56	78.54
cat	58.38	69.46	68.46	69.46	59.38	70.26	73.15	70.86
driller	78.39	83.15	78.99	82.85	81.86	85.23	84.94	80.77
duck	60.28	53.71	61.97	56.53	48.26	51.46	64.60	64.88
eggbox	97.56	98.97	98.31	98.97	98.50	99.06	98.87	98.87
glue	80.12	82.82	84.36	84.27	85.71	86.29	87.26	87.55
puncher	62.32	74.31	64.13	66.51	49.29	70.88	77.83	64.80
iron	87.54	89.89	90.19	89.58	90.19	90.60	90.50	89.27
lamp	88.87	94.43	92.61	93.47	94.05	92.90	95.30	91.65
phone	63.83	72.99	74.50	70.92	68.65	73.56	76.96	72.14
Avg.	73.78	78.03	77.76	77.69	73.36	78.43	81.11	78.19

표 2. 경계선 강화 블록의 위치에 따른 2D-Proj-5px 결과

	-	s1	s2	s3	s4	s12	s123	s1234
ape	97.52	97.90	97.24	97.90	97.71	97.43	97.90	97.71
benchvise	96.70	94.67	96.99	96.90	95.54	96.90	96.51	95.83
camera	95.10	97.16	97.94	97.75	96.18	97.84	98.04	97.25
can	93.11	94.59	94.69	94.00	94.69	95.47	96.46	94.59
cat	98.20	98.60	98.30	98.70	98.70	98.80	98.40	98.70
driller	90.39	90.29	90.39	91.48	92.47	92.37	94.05	89.49
duck	98.40	98.40	98.69	98.59	98.40	97.93	98.40	98.50
eggbox	98.59	99.06	98.78	99.25	98.97	98.97	98.78	98.69
glue	93.63	95.27	95.56	94.69	97.39	96.33	95.95	96.04
puncher	98.86	98.95	99.14	99.05	98.57	98.67	98.86	98.95
iron	91.52	92.54	93.56	92.85	92.95	93.56	93.67	91.83
lamp	90.88	93.28	91.84	93.67	92.99	92.13	93.86	92.32
phone	92.26	94.33	95.00	94.71	94.24	94.90	96.22	93.77
Avg.	95.01	95.77	96.01	96.12	96.06	96.25	96.70	95.67

III. 결론

본 논문에서는 RGB 입력 영상으로부터 end-to-end 방식으로 객체의 6D 포즈를 추정하는 딥러닝 모델인 EPro-PnP의 성능을 개선하기 위해 영상으로부터 특징 정보를 추출하기 위해 사용된 백본 네트워크에 경계선 강화 블록을 추가하는 방법을 제안하였다. 제안된 방법은 텍스처가 부족한 객체에 대해 경계선 영역의 특징 정보를 강화함으로써 6D 포즈 추정 정확도 향상에 기여할 수 있었다. 그러나, 경계선 강화 블록을 추가하는 위치에 따라 정확도 향상 정도가 달라질 수 있으며, 가장 우수한 성능을 보여주는 위치를 실험적으로 검증하였다.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) Grant by the Korean Government through the MSIT under Grant 2021R1F1A1045749.

참고 문헌

- [1] H. Chen, P. Wang, F. Wang, W. Tian, L. Xiong, and H. Li, "EPro-PnP: generalized end-to-end probabilistic perspective-n-points for monocular object pose estimation," Proc. of CVPR, pp. 2771-2780, 2022.
- [2] Y. Ye and H. Park, "FusionNet: an end-to-end hybrid model for 6D object pose estimation," Electronics, vol. 12, no. 19, 4162, 2023.
- [3] S. Hinterstoisser, C. Cagniard, S. Ilic, P. Sturm, N. Navab, P. Fua, and V. Lepetit, "Gradient response maps for real-time detection of textureless objects," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, pp. 876-888, 2012.
- [4] <https://github.com/tjiiv-cprg/EPro-PnP>