

물체 파지점 검출 향상을 위한 분할 기반 깊이 지도 조정

Segmentation-Based Depth Map Adjustment for Improved Grasping Pose Detection

신 현 수¹, 무하마드 라힐 아파잘², 이 성 온[†]

Hyunsoo Shin¹, Muhammad Raheel Afzal², Sungon Lee[†]

Abstract: Robotic grasping in unstructured environments poses a significant challenge, demanding precise estimation of gripping positions for diverse and unknown objects. Generative Grasping Convolution Neural Network (GG-CNN) can estimate the position and direction that can be gripped by a robot gripper for an unknown object based on a three-dimensional depth map. Since GG-CNN uses only a depth map as an input, the precision of the depth map is the most critical factor affecting the result. To address the challenge of depth map precision, we integrate the Segment Anything Model renowned for its robust zero-shot performance across various segmentation tasks. We adjust the components corresponding to the segmented areas in the depth map aligned through external calibration. The proposed method was validated on the Cornell dataset and SurgicalKit dataset. Quantitative analysis compared to existing methods showed a 49.8% improvement with the dataset including surgical instruments. The results highlight the practical importance of our approach, especially in scenarios involving thin and metallic objects.

Keywords: Segmentation, Deep Learning, Robotic Grasping

1. 서 론

딥러닝을 이용한 물체 인식 기술의 발전으로 다양한 로봇 분야에서 물체를 처리하는 유연한 작업들이 요구되고 있다. 특히 대다수의 제조업이 다품종 소량 생산성을 갖게 되며 산업용 로봇은 3차원 비전 센서와 함께 지능화가 이루어지고 있다. 로봇을 이용하여 복잡한 환경 속에서 임의로 배치되어 있는 물체를 파지하기 위해서는 목표 대상의 자세를 추정하는 기술이 필요하다. 이를 위해 딥러닝 기술을 기반으로 알려진 물체의 자세를 추정하는 연구가 많이 진행되고 있다¹⁻⁴. 그러나 물류 현장에서 구분해야 하는 물체의 종류가 많기 때문에 기존 딥러닝 기반 인식 방법을 적용하기 위해서는 시간 소모

가 크고, 새로 추가되는 물체에 대해서는 한계가 있다. 따라서 최근에는 목표 물체의 종류는 파악할 수 없지만, 처음 보는 물체에 대해서도 파지가 가능한 연구가 주목받고 있다^{5,6}. 대표적으로 D Morrison이 제시한 Generative Grasping Convolution Neural Network (GG-CNN)^{6,7}은 깊이 지도를 입력으로 하여, 물체 파지에 대한 품질, 너비, 각도 정보를 각 픽셀마다 출력한다. GG-CNN은 색상 정보 없이 3차원 깊이 지도만을 입력으로 받기 때문에, 센서의 깊이 추정 기법 및 깊이 정밀도에 따라 네트워크의 성능 차이가 나타날 수 있다^{8,9}. 특히, 물체의 두께가 얇거나 검은색 등의 물체 성질에 의해 3차원 깊이 지도가 불안정할 경우 파지점 추출이 실패한다^{6,7}. 따라서 Wang et al¹⁰은 U-Net 구조를 통해 기존보다 높은 해상도를 이용한 파지점 추출 네트워크를 연구하였지만, 목표 물체의 두께가 센서의 깊이 정밀도보다 적을 경우 파지점 추출의 한계가 있다.

본 연구에서는 네트워크 입력인 깊이 지도가 불안정한 경우, 영상 분할(Segmentation)을 통해 물체의 깊이 정보를 조정함으로써 파지점 추출의 성능 향상을 검증한다. 이때, 처음 보는 물체에 대해 영상 분할을 수행하기 위해 제로샷(Zero shot) 영상 분할이 가능한 기법을 적용한다. 그리고 추출된 영역의

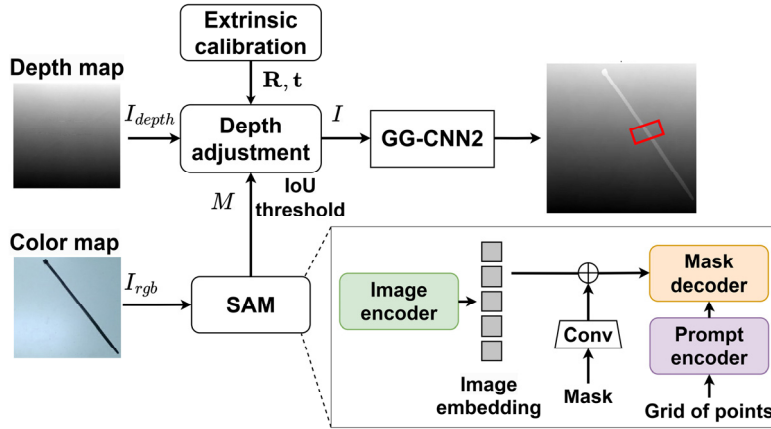
Received : Nov. 28. 2023; Revised : Jan. 8. 2024; Accepted : Jan. 12. 2024

※ This research was supported by the MOTIE (Ministry of Trade, Industry, and Energy) in Korea, under the Fostering Global Talents for Innovative Growth Program (P0008745) supervised by the Korea Institute for Advancement of Technology (KIAT)

1. Postdoctoral Researcher, Department of Electrical and Electronic Engineering, Hanyang University, Ansan, Korea (shs2316@hanyang.ac.kr)

2. Team Leader, Flanders Make, Belgium (raheel379@gmail.com)

† Professor, Corresponding author: Department of Robotics, Hanyang University, Ansan, Korea (sungon@hanyang.ac.kr)



[Fig. 1] Proposed grasping pipeline for the unstable depth map. Segmentation Anything Model (SAM)

깊이 성분을 조정한다. 본 연구에서 제안된 전체 과정을 요약하면 [Fig. 1]과 같다. 얇고 금속성을 띤 물체들이 포함된 데이터셋을 이용하여 제안된 방법으로 3차원 깊이 지도 변화에 따른 GG-CNN/GG-CNN2 성능을 평가한다.

각 장에 대한 구성은 다음과 같다. 2장에서는 GG-CNN/GG-CNN2에 대한 소개와 3차원 깊이 추정 기법과 영상 분할에 대해 설명한다. 그리고 3장에서는 깊이 지도가 불안정한 경우, 파지 추출 성능 향상을 위해 물체 분할을 활용한 깊이 지도 보정을 설명한다. 그리고 4장에서 보정된 깊이 지도를 사용한 GG-CNN/GG-CNN2의 결과를 비교한다. 마지막으로 5장에서 실험 결과를 통한 논의 사항에 대해서 정리한다.

2. 배경

2.1 Generative Grasping Convolution Neural Network (GG-CNN)

D Morrison이 제안한 GG-CNN^[6] 식 (1)과 같이 깊이 지도 I 를 네트워크 f_θ 입력에 사용하여 처음 보는 물체에 대한 평행 그리퍼 파지점의 품질 Q_θ , 각도 Φ_θ , 너비 W_θ 지도를 각각 추정한다. 품질 지도는 각 픽셀마다 파지 성공률에 따라 $[0, 1]$ 범위의 값을 갖고, 각도와 너비는 [Fig. 2]와 같이 각각 $[-\frac{\pi}{2}, \frac{\pi}{2}]$ 와 $[0, 150]$ 를 갖는다.

$$f_\theta(I) = (Q_\theta, \Phi_\theta, W_\theta) \quad (1)$$

네트워크의 출력이 각 픽셀별로 나타나기 때문에 최종 로봇의 파지점은 품질 지도에서 최대 픽셀에 해당하는 값을 사용한다. [Fig. 2(b)]의 파지점 P 은 Depth 카메라를 통해 해당 픽셀에 대한 3차원 좌표를 사용한다.

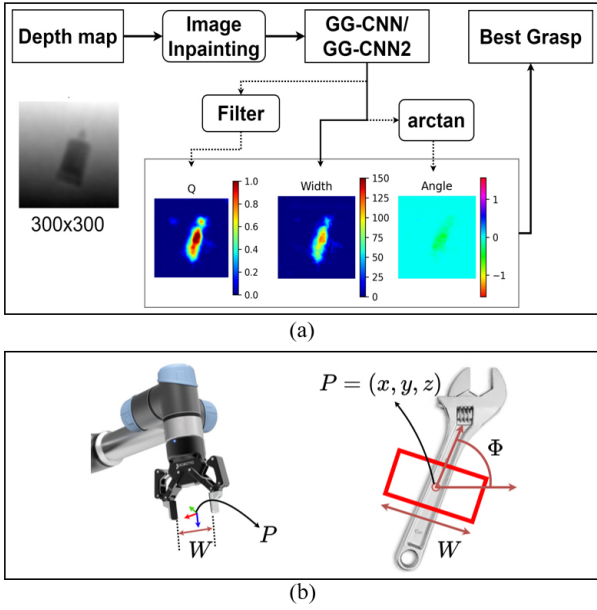
$$\theta = \underset{\theta}{\operatorname{argmin}} L_2(\mathbf{G}_T, f_\theta(I_T)) \quad (2)$$

식 (2)의 L_2 오차 함수에서 θ 는 네트워크의 가중치를, I_T, \mathbf{G}_T 는 각각 학습 데이터의 입력과 출력을 나타낸다. [Fig. 2(a)]과 같이 입력 이미지 I 의 크기는 300×300 픽셀을 갖으며, 네트워크 파라미터 수는 약 62,000개로 CNN을 기반으로 파지점을 추출하는 다른 네트워크들에^[11,12] 비해서 연산량이 적다. 기존 연구에서는^[5,11,12-14] 처음 보는 물체에 대한 파지점 추정을 위해 먼저 이미지 혹은 3차원 점군 데이터를 통해 파지점 후보를 추출하고, CNN을 통해 파지 순위를 정하였다. 일정 구간별로 샘플링하여 각 패치에서의 파지점 후보들을 구하기 때문에 연산량이 많은 단점이 있었다^[11,12].

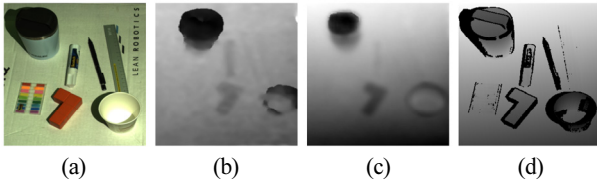
반면에, GG-CNN에서는 입력 데이터의 전체 픽셀에 대한 하나의 파지 품질 지도를 출력하기 때문에 빠른 연산 속도를 갖는다. GG-CNN의 후속 연구인 GG-CNN2^[7]에서는 물체 분할 성능 향상에 초점을 맞춰서 Dilated convolution^[15]을 추가하였다. 또한, 네트워크 학습을 위해 데이터셋을 기존 연구에 사용한 Cornell grasping 데이터셋^[9]뿐만 아니라 ShapeNet^[16]을 기반으로 데이터 수를 크게 증가시킨 synthetic Jacquard 데이터셋^[17]을 추가하였다. 그러나, 얇은 물체나 검은색 등 물체 성질에 의해서 깊이 지도가 불분명할 경우, 파지점 추출이 실패한다^[6].

2.2 3차원 깊이 추정

GG-CNN2의 입력에 사용되는 깊이 지도는 3차원 점군 데이터를 2차원 이미지 상에 투영시켜 생성된다. 따라서 깊이 지도의 해상도는 센서의 3차원 깊이 정밀도에 의존한다. 3차원 깊이 정보를 파악하는 대표적인 기술은 빛이 피사체에 반사되어 돌아오는 시간을 측정하여 거리를 계산하는 Time-of-Flight



[Fig. 2] (a) Generative Grasping Convolution Neural Network (GG-CNN) and (b) definition of robotic grasping pose



[Fig. 3] Comparison of the depth maps using three different types of sensors; (a) color map, (b) Realsense D455 (Stereoscopy), (c) Azure Kinect (ToF), and (d) Zivid One+ M (SL)

(ToF) 방식, 두 개의 카메라에서 대응점에 대한 시차를 활용하는 스테레오(Stereoscopy), 구조광(Structured Light, SL) 패턴을 물체에 투영하고 촬영된 영상으로부터 왜곡된 이미지를 활용하는 SL 방식이 있다. ToF 혹은 스테레오 방식은 SL 방식에 비해 거리 제한이 자유로운 편이나, SL보다 정밀도가 떨어진 다. 세 가지 깊이 추정 방식으로 동일한 거리에서 획득된 깊이 지도는 [Fig. 3]와 같이 다르게 나타난다. 특히, 고가형 SL 방식 센서의 경우, [Fig. 3(d)]의 왼쪽 아래에 있는 두께 약 1.9 mm인 물체에 대해서도 경계선 부분이 검출되는 것을 볼 수 있다. GG-CNN2은 깊이 지도만을 입력 받기 때문에 이러한 센서 별 깊이 분해능 차이는 파지점 추출 성능에 큰 영향을 끼친다.

2.3 영상 분할

본 연구에서는 센서의 깊이 분해능 한계를 극복하기 위해 컬러 지도 내에서 영상 분할을 통해 물체의 깊이를 조정한다. 전통적 영상처리 기법의 경우, 대표적으로 임계치에 의한 영

상 분할, 영역 성장(Region growing) 기법, 엣지 검출 기법 등이 존재한다. 하지만 이러한 전통적 방법들은 노이즈나 외부 환경에 민감한 단점이 있다. 한편, 딥러닝 기술 기반 영상 분할은 기존 기법들에 비해 노이즈에 강건한 성능을 보인다. 대표적으로 Mask R-CNN^[18]의 경우, 학습 데이터에 포함된 계층에 대해서는 높은 분할 성능을 보이지만, 처음 보는 물체에 대한 분할은 학습이 필요하다. 따라서 본 연구에서는 제로샷 객체 분할이 가능한 네트워크를 사용한다. 제로샷 분할이란 학습에 사용되지 않은 새로운 물체를 분할하는 것을 말한다. 이는 큰 규모의 데이터셋에 대해 사전 학습하고 특정 작업에 맞춰 Fine-tuning하는 방식을 갖는다.

3. 방법

3.1 영상 분할 기반 깊이 지도 조정

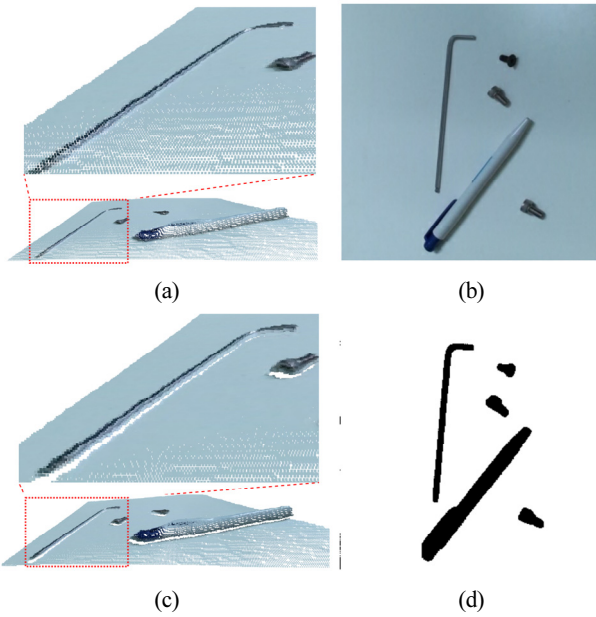
GG-CNN2는 깊이 지도만 입력으로 사용되기 때문에 안정된 파지를 위해서 기존 저자는 전처리 과정으로서 인페인팅을 적용하였다. 이를 통해 영상 노이즈를 제거하여 파지점 추출이 가능하였으나, 얇은 물체나 검은색과 같이 깊이 지도에 나타나지 않는 물체들은 불가능하였다. 따라서, 본 연구에서는 깊이 지도가 불안정한 물체들을 검출하기 위해 식 (3)과 같이 컬러 지도에서 물체들의 객체 분할 마스크 M 를 생성한다. 그리고 식 (4)와 같이 깊이 지도 I_{depth} 에서 마스크 M_{depth} 에 해당하는 픽셀들을 λ 만큼 곱하여 조정한다. M_{depth} 은 카메라의 외부 파라미터를 통해 Depth 카메라 좌표계로 정렬된 마스크 이미지이다. 제안된 방법의 전체 프로세스는 [Fig. 1]과 같다.

$$M = S(I_{rgb}) \quad (3)$$

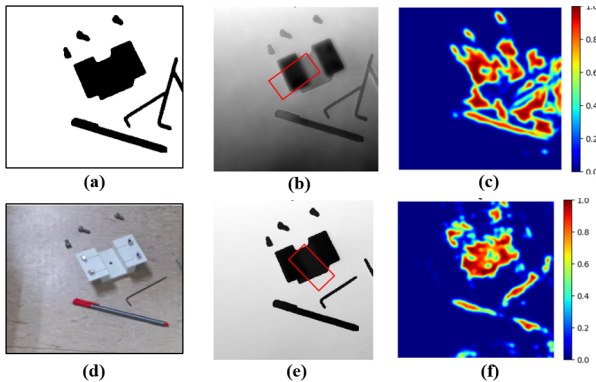
$$I = \lambda M_{depth} \otimes I_{depth} \quad (4)$$

이미지의 불연속성 혹은 임계점 등을 통해 특정 물체를 분할하는 전통적인 방법들은 환경에 민감한 단점이 있다. 반면에, 딥러닝 기반 영상 분할은 노이즈에 강건하다. Meta사에서 공개한 Segmentation Anything Model (SAM)은 파운데이션 모델을 적용하여 레이블이 지정되지 않은 물체들도 영상 분할이 가능하다^[9].

약 1천백만 장의 데이터셋 (SA-1B)를 통해 학습하여, 높은 성능의 제로샷 영상 분할을 갖는다. SAM은 [Fig. 1]과 같이 세 가지 요소로 구성되어 있다. 본 연구에서는 처음 보는 물체를 자동으로 분할하기 위해 Prompt encoder 입력으로 32x32 전체 그리드 점들을 사용한다. 그리고 Mask decoder의 Intersection over Union (IoU)의 임계점을 설정하여 모든 물체에 대한 분할



[Fig. 4] (a,c) Pointcloud before and after adjustment, (b) color image and (d) segmentation mask using SAM

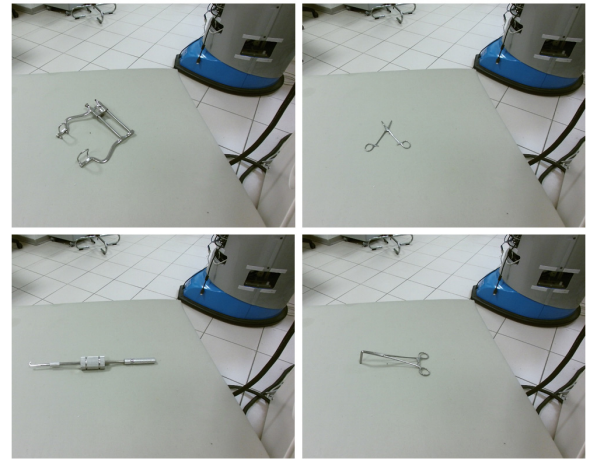


[Fig. 5] (a) Segmentation mask using SAM, (b,c) depth map and grasping quality map with $\lambda=1.05$, (d) color map, and (e,f) depth map and grasping quality map with $\lambda=1.4$

마스크를 구하였다. 이를 통해 처음 보는 물체를 분할하고, 깊이 지도와의 정렬을 통해 해당 영역의 깊이 성분을 조정할 모습은 [Fig. 4]와 같다. 여기서 λ 는 경험적으로 1.3 이상 일 경우, [Fig. 5(e)]와 같이 물체와 평면 사이의 깊이 차이가 커져 파지점이 불안정하다. 따라서, 본 연구에서는 λ 를 1.05로 설정하였다.

3.2 데이터셋 학습

본 연구에서는 Cornell grasping 데이터셋 뿐만 아니라 얇은 물체에 대한 파지점 추출을 위해 SurgicalKit 데이터셋²⁰⁾을 추



[Fig. 6] Examples of images from the SurgicalKit dataset

가적으로 학습하였다. SurgicalKit 데이터셋은 [Fig. 6]과 같이 Kinect 2를 통해 얇은 수술 도구를 촬영하여 320장의 원본 데이터를 임의의 회전, 확대 등을 통해 증강시켜 9,920장의 데이터셋으로 구성되었다. 데이터셋에 사용된 수술용 도구들은 반사성이 강한 금속 재질로 구성되어 깊이 지도가 불안정하다. 파지점들은 [Fig. 3(b)]에서 보이듯이 각 물체마다 4개의 점으로 나타나며, 이는 손실 계산 시 파지점의 위치, 각도와 너비로 사용된다. 즉, Cornell 데이터셋과 유사한 형식을 갖는다.

물체의 파지 성공 유무는 추정된 직사각형이 물체 추정 평가 지표에 사용되는 IoU의 25% 이내와 정답과 30도 이내를 기준으로 정하였다. 네트워크 학습에는 Positive 파지점만 사용하였다.

4. 실험 및 결과

4.1 정량적 비교

본 연구에서 제안된 방식의 파지점 검출 성능을 검증하기 위해, 두 가지 데이터셋에 대한 성능 결과를 기존 방식과 비교하였다. 실험에 사용된 PC 환경은 Intel Core i9-9900K CPU @ 3.60GHz와 NVIDIA GeForce RTX 2080 Ti를 갖는다. [Table 1]과 같이 Cornell 데이터셋의 경우, 제안된 방법을 적용할 경우 기

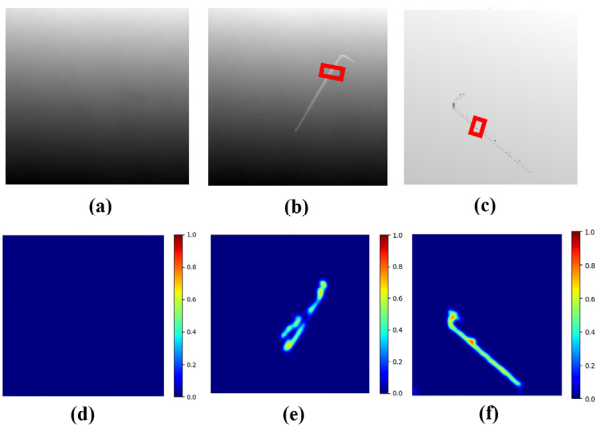
[Table 1] Comparative accuracy [%] results on several grasping datasets

Method	Dataset	
	Cornell	Surgical
GG-CNN	73.0	-
GG-CNN2	84.0	28.2
Refined GG-CNN2	93.0	78.0

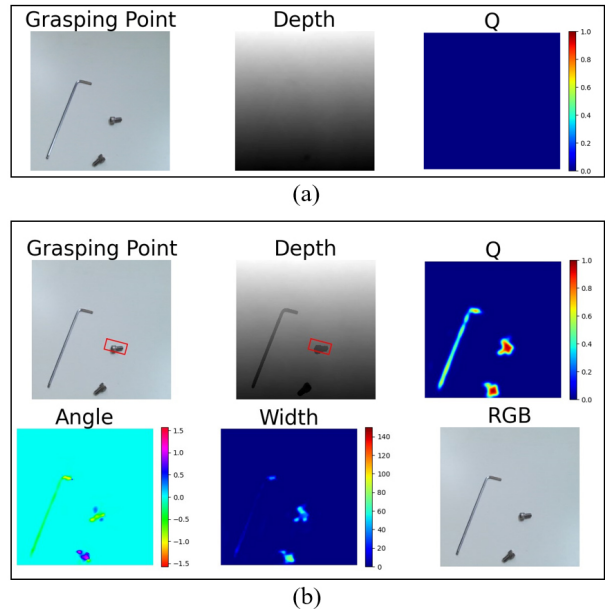
존 GG-CNN2 보다 약 10.7% 증가시켰다. 이를 통해 Cornell 데이터셋에서도 불안정한 깊이 지도에 의해 파지점 추출 성능이 저하된 것을 알 수 있다. 또한, 얇은 물체가 포함된 SurgicalKit 데이터셋에 대해서 기존 방식은 약 28%로 굉장히 낮은 추정치를 갖는다. 반면, 제안된 방법은 처음 보는 물체에 대해서 물체 분할을 통해 깊이 지도의 깊이를 조정함으로써, 동일한 데이터셋에 78%로 약 1.7배 증가하였다.

4.2 정성적 비교

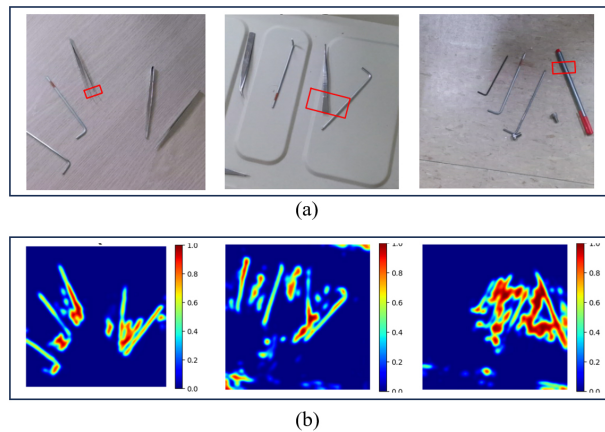
정성적 평가를 위해 두 가지 실험을 평가하였다. 첫 번째 실험은 깊이 정밀도가 다른 두 가지 센서를 이용하여 제안된 방법의 성능을 평가한다. 두 번째 실험은 물체가 놓인 평면의 재질 및 패턴 변화에 강건성을 평가한다. 첫 번째 실험에서는 비교적 낮은 깊이 분해능을 갖는 저가형 센서 Azure Kinect와 정밀한 Zivid One+ M 카메라를 사용되었다. Zivid 센서의 경우, 깊이 방향 분해능이 약 1 mm 이내의 정밀도를 갖기 때문에 얇고 금속성을 띤 물체에 대해서도 [Fig. 7(c)], [Fig. 7(f)]와 같이 기존 네트워크를 통해서 파지점 추출이 가능하다. 동일한 물체에 대해 Azure Kinect는 [Fig. 7(a)], [Fig. 7(d)]에서 나타나듯이 GG-CNN2의 품질 지도 Q_{θ} 가 불안정하여 파지점을 추정할 수 없다. 제안된 방법은 [Fig. 7(b)]와 같이 깊이 지도 조정을 통해 물체의 파지점을 구할 수 있다. 또한 금속성 물체의 크기가 작을 경우, 영상 분할 또한 불안정할 수 있지만, SAM은 [Fig. 8]과 같이 안정적인 결과를 갖는다. 두 번째 실험에서는 학습되지 않은 물체를 배경에 상관없이 분할하기 위해 세 가지 다른 재질 및 패턴을 갖는 환경에서 구성하였다. 실험 결과, [Fig. 9]와 같이 배경 변화에 강건한 결과를 갖는다.



[Fig. 7] Comparison of the depth maps and grasping quality maps of the GG-CNN2 model, (a,d) Azure Kinect, (b,e) our results with Azure Kinect, (c,f) Zivid One+ M



[Fig. 8] Comparison of the results of robotic grasping pose using (a) GG-CNN2 and (b) GG-CNN2 with mask using SAM



[Fig. 9] Comparison of the (a) grasping pose in the color map and (b) grasping quality maps of the GG-CNN2 model across three different environments

5. 결론

본 연구에서는 물체의 파지점을 추출하는 GG-CNN2의 성능을 향상시키기 위해 컬러 지도에서 물체를 분할하고 해당 영역에 대해 깊이 지도를 조정하였다. 영상 분할에 사용된 SAM은 학습되지 않은 물체에 대해서 마스크 생성이 가능하다. 본 연구에서 제시된 방법과 기존 방법을 사용하여 Cornell 데이터셋과 얇은 수술 도구가 포함된 SurgicalKit 데이터셋에 대해 성능을 비교하였다. 특히, SurgicalKit 데이터셋의 경우 기존 GG-CNN2를 사용했을 때보다 49.8% 증가하였다. 결과적으로, 제안된 방법을 통해 얇고 금속성을 띤 물체들의 파지점 추출이 가능하였다.

References

- [1] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes," *arXiv:1711.00199*, 2018, DOI: 10.15607/RSS.2018.XIV.019.
- [2] B. Tekin, S. N. Sinha, and P. Fua, "Real-time seamless single shot 6d object pose prediction," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, pp. 292-301, 2018, DOI: 10.1109/cvpr.2018.00038.
- [3] S. Peng, Y. Liu, Q. Huang, H. Bao, and X. Zhou, "Pvnet: Pixel-wise voting network for 6dof pose estimation," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, pp. 4556-4565, 2019, DOI: 10.1109/cvpr.2019.00469.
- [4] H. Han, W. Wang, X. Han, and X. Yang, "6-DoF grasp pose estimation based on instance reconstruction," *Intelligent Service Robotics*, Nov., 2023, DOI: 10.1007/s11370-023-00489-z.
- [5] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. Ojea, and K. Goldberg, "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," *Robotics: Science and Systems*, 2017, DOI: 10.15607/RSS.2017.XIII.058.
- [6] D. Morrison, P. Corke, and J. Leitner, "Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach," *Robotics: Science and Systems*, 2018, DOI: 10.15607/RSS.2018.XIV.021.
- [7] D. Morrison, P. Corke, and J. Leitner, "Learning robust, real-time, reactive robotic grasping," *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 183-201, Jun., 2019, DOI: 10.1177/0278364919859066.
- [8] S. Dodge, and L. Karam, "Understanding how image quality affects deep neural networks," *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, Lisbon, Portugal, pp. 1-6, 2016, DOI: 10.1109/QoMEX.2016.7498955.
- [9] J. Park, H. Kim, Y. Tai, M. Brown, and I. Kweon, "High quality depth map upsampling for 3D-TOF cameras," *2011 International Conference on Computer Vision*, Barcelona, Spain, pp. 1623-1630, 2011, DOI: 10.1109/ICCV.2011.6126423.
- [10] S. Wang, X. Jiang, J. Zhao, X. Wang, W. Zhou, and Y. Liu, "Efficient fully convolution neural network for generating pixel wise robotic grasps with high resolution images," *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dali, China, pp. 474-480, 2019, DOI: 10.1109/ROBIO49542.2019.8961711.
- [11] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *The International Journal of Robotics Research*, vol. 34, no.4-5, pp. 705-724, Mar., 2015, DOI: 10.1177/0278364914549607.
- [12] L. Pinto and A. Gupta, "Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours," *2016 IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, pp. 3406-3413, 2016, DOI: 10.1109/ICRA.2016.7487517.
- [13] J. Redmon and A. Angelova, "Real-Time Grasp Detection Using Convolutional Neural Networks," *2015 IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, WA, USA, pp. 1316-1322, 2015, DOI: 10.1109/ICRA.2015.7139361.
- [14] Z. Wang, Z. Li, B. Wang, and H. Liu, "Robot grasp detection using multimodal deep convolutional neural networks," *Advances in Mechanical Engineering*, vol.8, Sept., 2016, DOI: 10.1177/1687814016668077.
- [15] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv:1511.07122*, 2016, DOI: 10.48550/arXiv.1511.07122.
- [16] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, "Shapenet: An information-rich 3d model repository," *arXiv:1512.03012*, 2015, DOI: 10.48550/arXiv.1512.03012.
- [17] A. Depierre, E. Dellandrea, and L. Chen, "Jacquard: A large scale dataset for robotic grasp detection," *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, pp. 3511-3516, 2018, DOI: 10.1109/IROS.2018.8593950.
- [18] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," *International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 2980-2988, 2017, DOI: 10.1109/ICCV.2017.322.
- [19] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. Berg, W. Lo, P. Dollár, and R. B. Girshick, "Segment anything," *arXiv:2304.02643*, pp. 4015-4026, Apr., 2023, DOI: 10.48550/arXiv.2304.02643.
- [20] J. Kim, O. Nocentini, M. Bashir, and F. Cavallo, "Grasping Complex-Shaped and Thin Objects Using a Generative Grasping Convolutional Neural Network," *Robotics*, vol. 12, no. 2, Mar., 2023, DOI: 10.3390/robotics12020041.



신 현 수

2015 한양대학교 ERICA 전자시스템공학과 (학사)
2017 한양대학교 ERICA 전자공학과(석사)
2019~2020 Visiting Scholar, KU Leuven
2023 한양대학교 ERICA 전자공학과(박사)
2023~현재 한양대학교 ERICA융합원 박사후 연구원

관심분야: 로봇 비전, 핸드아이 캘리브레이션, 센서 융합



무하마드 라힐 아파잘

2012 Department of Mechatronics Engineering, Air University(석사)
2018 School of Mechanical and Aerospace Engineering, Gyeongsang National University(박사)
2019~2020 Research Team Leader, KU Leuven
2020~현재 Team Leader, Flanders Make

관심분야: 햅틱, 재활로봇



이 성 운

1997 서울대학교 기계설계학과(학사)
1999 포항공과대학교 지능기계정보학(석사)
2008 동경대학교 기계공학과(박사)
2015~2023 한양대학교 ERICA 전자공학부 교수
2023~현재 한양대학교 ERICA 로봇공학과 교수

관심분야: 의료 영상, 영상기반제어, 딥러닝기반 물체인식