특집논문 (Special Paper)
방송공학회논문지 제30권 제4호, 2025년 7월 (JBE Vol.30, No.4, July 2025)
https://doi.org/10.5909/JBE.2025.30.4.580
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

GraspLDM 기반 파지 생성에서 객체 스케일의 영향 분석

김 연 지[a], 이 상 민[a], 김 희 원[a]‡

Analysis of the Effect of Object Scale on Grasp Generation Using GraspLDM

Yeonji Kim[a], Sangmin Lee[a], and Heewon Kim[a]‡

요 약

본 연구는 자율 로봇 조작을 위한 확산 기반 모델인 GraspLDM이 객체 크기 변화에 따라 파지 생성을 어떻게 달리 수행하는지를 분석한다. ARNOLD 데이터셋의 bottle 객체를 0.5배, 0.8배, 1.2배, 1.5배로 스케일링한 후, 생성된 3차원 파지점들을 정량적으로 평가하였다. 분석 결과, 객체 크기를 축소할 경우 정규화된 평균 파지 거리가 증가하고 파지점이 객체의 외곽으로 분산되어 일관성이 감소하는 경향을 보였다. 반대로, 객체 크기를 확대하면 파지점이 중심부에 집중되어 파지의 다양성이 저하되었다. 이러한 결과는 GraspLDM이 객체 크기의 변화에 단순히 비례적으로 적응하지 않으며, 축소와 확대 상황에서 서로 다른 한계를 나타냄을 시사한다. 이에 따라 향후 파지 생성 모델의 강인성과 일반화 성능을 향상시키기 위해서는 객체 스케일 정보를 명시적으로 반영하는 접근이 필요하다.

Abstract

This study investigates how object scale affects grasp generation in GraspLDM, a diffusion-based model designed for autonomous robotic manipulation. Using a bottle object from the ARNOLD dataset, we scaled the object by factors of 0.5×, 0.8×, 1.2×, and 1.5× and quantitatively analyzed the resulting 3D grasp points. Our analysis reveals that reducing object size increases normalized grasp distances and a wider dispersion of grasp points toward the object's periphery, indicating decreased consistency. In contrast, enlarging the object causes grasp points to cluster near the center, reducing grasp diversity. These findings suggest that GraspLDM does not adapt proportionally to changes in object scale and exhibits distinct limitations depending on whether the object is reduced or enlarged. Incorporating explicit object scale information may be essential for enhancing the robustness and generalization capabilities of future grasp generation models.

Keyword : Robotic Manipulation, Diffusion Model, Object Scaling

a) 숭실대학교 글로벌미디어학부(Global School of Media, Soongsil University)
‡ Corresponding Author : 김희원(Heewon Kim)
E-mail: hwkim@ssu.ac.kr
Tel: +82-2-820-0679
ORCID: https://orcid.org/0000-0001-7777-9823

※ This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the National Program for Excellence in SW (2024-0-00071), the Graduate School of Metaverse Convergence support program (IITP-2025-RS-2024-00430997), and the Convergence security core talent training business support program (IITP-2025-RS-2024-00426853) supervised by the IITP (Institute of Information & communications Technology Planning & evaluation). This work was supported by the Technology Development Program (RS-2024-00510957) funded by the Ministry of SMEs and Startups (MSS, Korea)
· Manuscript May 26, 2025; Revised July 4, 2025; Accepted July 4, 2025.

Ⅰ. 서 론

최근 로봇 공학 분야에서 물체 조작(Robotic Manipulation)은 핵심적인 연구 주제로 다뤄지고 있다. 특히 미지의 물체에 대한 안정적인 파지점을 찾는 것은 로봇이 다양한 환경에서 효과적으로 작업을 수행하기 위한 과제 중 하나이다[1]. 이러한 배경 속에서, 생성 모델을 활용하여 파지 자세를 합성하려는 연구들이 진행되고 있다[2]. 그 중에서도 GraspLDM[3]은 VAE(Variational Autoencoder)[4]로 학습된 잠재 공간(latent space) 안에서 확산 모델(Diffusion Model)[5]을 기반으로 한 잠재 확산 모델(Latent Diffusion Model)[6]을 활용하여 고품질의 다양한 6 자유도(6-DoF) 파지 자세 생성을 목표로 하는 프레임워크이다. GraspLDM[3]은 Issac Gym 시뮬레이션 기반 학습을 통해 실제 환경에서도 별도의 파인튜닝 없이 우수한 파지 성능을 보이며, 효과적인 파지점 생성이 가능함을 입증하였다.

그러나 GraspLDM[3]과 같은 사전 학습된 모델의 성능은 학습 데이터의 다양성과 특성에 크게 의존할 수 있다. 특정 도메인이나 기하학적 특성을 가진 데이터셋에 기존 모델을 적용할 경우, 예상치 못한 파지점 생성 패턴이나 한계가 나타날 수 있다. 만약 모델이 특정 유형의 객체에 대해 제한된 파지점만을 생성하거나, 중요한 파지 가능 영역을 놓친다면 실제 적용 시 로봇의 작업 유연성과 성공률이 저하될 수 있다[1].

이러한 배경에서, 본 연구는 정적인 파지 생성 모델인 GraspLDM[3]을 ARNOLD[7]가 제공하는 현실적이고 다양한 3D 객체 에셋에 적용하여, 특히 객체의 '절대적 크기' 변화라는 조건 하에서의 일반화 가능성을 탐색하고자 한다. ARNOLD[7]는 언어 기반의 연속 상태 과업 학습을 목표로 설계되었으며, 사실적인 3D 장면과 40가지의 다양한 객체를 포함하고 있어 GraspLDM[3]과 같은 모델이 현실적인 객체의 기하학적 변화에 어떻게 반응하는지 분석하기에 적합한 환경을 제공한다.

본 연구는 이러한 ARNOLD[7] 데이터셋의 객체 중 bottle 카테고리를 대상으로 GraspLDM[3]을 적용할 때, 입력 데이터의 다양한 특성에 맞추어 생성되는 파지점의 특성, 특히 다양성과 분포에 어떠한 영향을 미치는지 분석하는 것을 목표로 한다. 구체적으로, 해당 객체의 특성을 반영해 포인트 클라우드(point cloud) 증강 기법들을 적용하고, 이러한 증강된 데이터를 입력으로 사용했을 때 GraspLDM[3]이 생성하는 파지점들의 변화를 정량적으로 평가하고자 한다. 본 연구는 객체 크기 변화를 통한 입력 데이터 증강이 파지점 생성의 다양성 측면에서 어떠한 변화를 유도하는지를 실험적으로 관찰하고 분석함으로써, 특정 데이터셋에 대한 GraspLDM[3]의 적용 가능성을 높이고 향후 모델 개선을 위한 기초 자료를 제공하는 데 기여하고자 한다. 이는 추후 시뮬레이션 기반의 파지 성공률 검증 및 모델 파인튜닝 연구로 나아가기 위한 중요한 사전 분석 단계로서의 의미를 가진다.

로봇이 실제 환경에서 객체를 인식할 때, 센서의 거리나 객체의 실제 크기 변화, 또는 유사하지만 크기가 다른 객체 등으로 인해 인식되는 객체의 크기는 매우 다양하게 변할 수 있다. 이러한 크기 변화는 파지점 생성 전략에 직접적인 영향을 미칠 수 있는 중요한 요소이다. 본 연구에서는 우선적으로 '크기' 변화에 집중하였으나, 향후 연구에서는 회전 증강을 포함한 다양한 증강 기법의 복합적인 효과를 분석하고, 이를 통해 ARNOLD[7]와 같은 연속적인 작업 환경에서 요구되는 파지점 생성의 강인성을 더욱 향상시킬 계획이다.

Ⅱ. 선행 연구

본 장에서는 GraspLDM[3]의 핵심 구성 요소인 VAE[4]와 잠재 확산 모델(Latent Diffusion Model)[6]에 대해 각각 설명한 후, GraspLDM[3]의 파지 생성 과정에 대해 기술한다.

1. VAE (Variational Autoencoder)

VAE(Variational Autoencoder)[4]는 입력 데이터의 주요 특징을 나타내는 정보를 추출해 잠재 벡터에 넣고, 이 잠재 벡터를 통해 입력 데이터와 유사한 데이터를 생성하는 것을 목표로 하는 모델이다. 이미지 생성에 대한 관심이 높아지며 VAE[4]를 활용해 다양한 이미지를 생성하는 모델[8-11]이 다수 제안되었으며, 이들은 원본과 유사하지만 새로운 데이터를 생성하는 데에 중점을 둔다. 이 과정에서 각 특징이 가우시안 분포를 따른다고 가정하고 잠재 벡터는 각 특징에 대한 평균과 분산값을 구한다. Encoder와 Decoder 구조를 활용하여, Encoder에서 입력 이미지에 대한 잠재 벡터

를 찾아낸다. 다시 이를 Decoder를 통해 복원해 기존 이미지와 비슷하지만 새로운 확률 분포를 찾아내 새로운 이미지를 생성한다.

2. 잠재 확산 모델 (Latent Diffusion Model)

확산 모델(Diffusion Model)[5]은 데이터에 노이즈(noise)를 더한 후, 노이즈(noise)를 점차 제거하며 데이터를 복원해가는 과정을 통해 새로운 데이터를 생성하는 모델이다. 순방향 과정은 고정된 마르코프 연쇄(Markov Chain)를 통해 시간이 지날 때마다 원본 데이터에 점차 노이즈(noise)를 추가하여 순수 노이즈(noise)에 가까운 상태로 만든다. 역방향 과정은 학습된 신경망이 각 단계에서 노이즈(noise)를 예측하고 제거함으로써, 점진적으로 원본 데이터를 복원하는 과정을 통해 새로운 샘플을 생성한다. 확산 모델은 VAE[4]와 마찬가지로 생성 모델[11-16]로 활용되고 있다.

그러나 이러한 확산 모델(Diffusion Model)[5]은 순차적 평가로 인해 추론 비용이 높다는 단점이 있다. 이러한 계산적 부담을 줄이고 확산 모델(Diffusion Model)[5]의 품질과 유연성을 유지하며 제한된 자원에서 학습을 할 수 있도록 잠재 확산 모델(Latent Diffusion Model)[6]이 제안되었다.

구체적으로, 잠재 확산 모델(Latent Diffusion Model)[6]은 학습된 AutoEncoder의 잠재 공간에 확산 모델(Diffusion Model)[5]을 적용하는 방식이다. 이러한 잠재 확산 모델에서는 먼저 이미지를 저차원의 공간으로 압축하는 Auto Encoder를 학습시킨 후, 확산 모델은 저차원 잠재 공간 내에서 학습되고 작동하여 고품질의 샘플을 생성한다. 생성된 잠재 표현은 AutoEncoder의 Decoder를 통해 다시 원본 공간으로 복원된다.

3. GraspLDM

GraspLDM[3]은 앞서 설명한 VAE[4]와 잠재 확산 모델(Latent Diffusion Model)[6]의 강점을 결합하여, 객체에 대한 다양하고 안정적인 6 자유도(6-DoF) 파지 자세를 합성하는 것을 목표로 한다.

GraspLDM[3]의 주요 구성 요소 및 작동 방식은 그림 1과 같다. 포인트 클라우드(point cloud) Encoder에서 입력된 객체의 포인트 클라우드(point cloud)를 처리하여 객체의 형상 정보를 압축한 잠재 벡터를 형성한다. Grasp Encoder는 파지 자세와 해당 객체의 형상 잠재 벡터를 입력받아 조건부 파지 벡터로 인코딩한다. Grasp Decoder는 파지 잠재 벡터와 형상 잠재 벡터를 바탕으로 원본 파지 자세를 복원하는 과정을 통해 새로운 파지 자세를 생성한다.

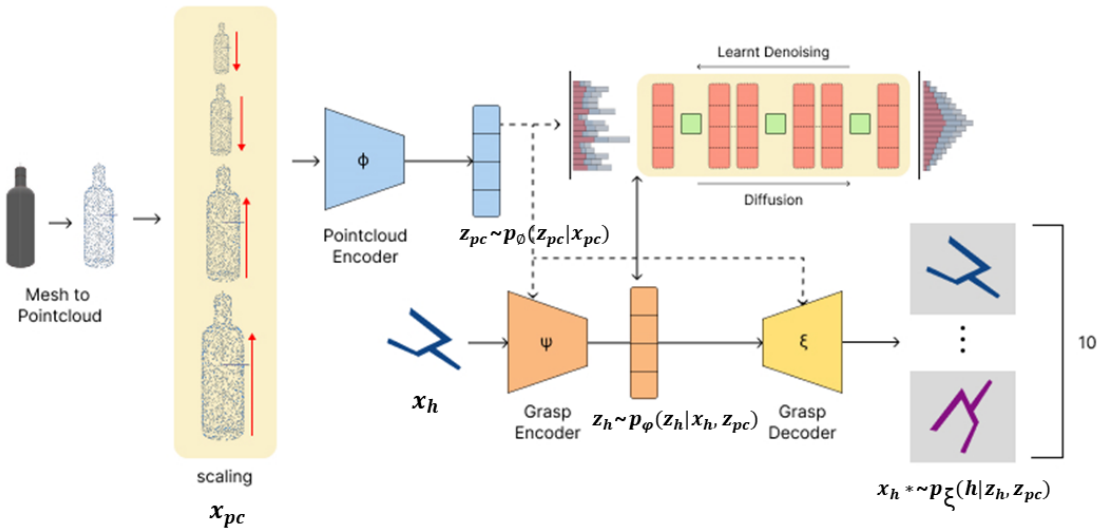

그림 1. 본 연구의 분석 파이프라인
Fig. 1. Analysis Pipeline of this study

III. 제안 방법

1. 실험 목적

특정 객체의 포인트 클라우드(point cloud) 데이터에 GraspLDM[3] 모델을 적용할 때, 객체 크기 변화에 따라 변화되는 입력 데이터에 맞추어 생성되는 파지점의 다양성 및 분포 특성에 미치는 영향을 분석하는 것을 목표로 한다.

2. 분석 파이프라인

3D 객체 메쉬(mesh)로부터 포인트 클라우드(point cloud)를 추출하고, 이를 원본 포함 총 5가지 크기(예: 0.5배, 0.8배 축소, 원본, 1.2배, 1.5배 확대)로 스케일링한다. 각 스케일별 포인트 클라우드(point cloud)에 GraspLDM[3] 모델을 적용하여 파지 자세를 10개 생성하고, 생성된 파지 자세들로부터 파지점을 추출하여 크기 변화가 파지 특성에 미치는 영향을 분석한다. 그림 1은 이러한 전체 과정을 도식화하여 보여준다.

3. 데이터 준비 및 스케일 변환

GraspLDM[3]은 입력 데이터를 포인트 클라우드(point cloud) 형태로 요구하므로, ARNOLD[7] 데이터셋 내 객체들의 메쉬(mesh)를 모두 포인트 클라우드(point cloud) 형태로 변환하였다. 스케일링을 적용하기 전, 각 원본 포인트 클라우드(point cloud) 형태를 구성하는 모든 점들의 산술평균 위치를 계산하여 해당 지점을 객체의 중심으로 정의하였다. 모든 포인트는 이 계산된 중심점을 기준으로 원점으로 평행이동하였다. 이러한 중심점 정렬 과정은 크기 변환의 기준점을 일관되게 설정하여 객체의 위치 변화 없이 순수하게 크기 변화에 대해서만 관찰할 수 있으며, 객체의 원본 형태와 비율은 유지하면서 절대적인 크기만을 변경한다. 중심점 정렬 후, 그림 2에서와 같이, ARNOLD[7] 데이터셋에서 선정된 8개의 다른 모양을 가진 원본 bottle 객체

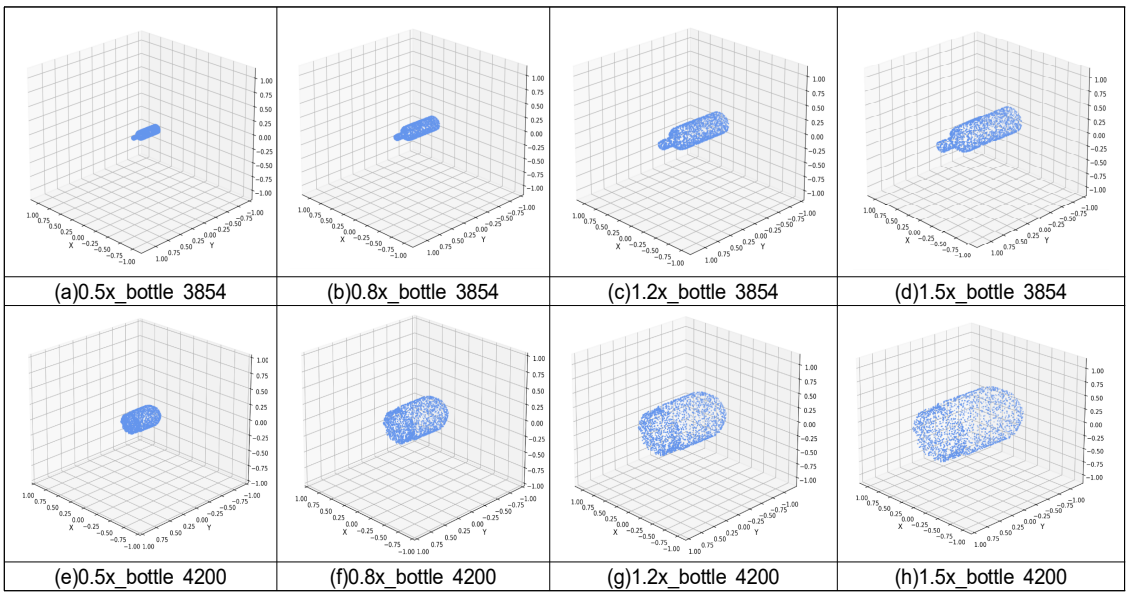

그림 2. 포인트 클라우드 증강 결과
(a): 3854 bottle 객체 0.5배 축소 (b): 3854 bottle 객체 0.8배 축소 (c): 3854 bottle 객체 1.2배 확대 (d): 3854 bottle 객체 1.5배 확대 (e): 4200 bottle 객체 0.5배 축소 (f): 4200 bottle 객체 0.8배 축소 (g): 4200 bottle 객체 1.2배 확대 (h): 4200 bottle 객체 1.5배 확대
Fig. 2. Point cloud augmentation results
(a): Bottle object 3854 scaled to 0.5x (b): Bottle object 3854 scaled to 0.8x (c): Bottle object 3854 scaled to 1.2x (d): Bottle object 3854 scaled to 1.5x (e): Bottle object 4200 scaled to 0.5x (f): Bottle object 4200 scaled to 0.8x (g): Bottle object 4200 scaled to 1.2x (h): Bottle object 4200 scaled to 1.5x

에 대해 각각 0.5배, 0.8배, 1.2배, 1.5배의 스케일 팩터(scale factor)를 적용하여 총 32개의 포인트 클라우드(point cloud) 데이터를 증강하였다. 원본 객체를 포함하여 총 40개의 객체를 GraspLDM[3]의 입력으로 사용하여 파지 생성 실험을 진행하였다.

4. GraspLDM[3] 기반 파지 생성

본 연구의 목적은 기존 모델이 학습하지 않은 데이터에 대해 생성하는 파지 자세를 분석하는 것이므로, 제공된 사전 학습 모델을 추가 학습 없이 그대로 활용하여 기존 ARNOLD[7]의 데이터셋의 bottle 객체들과 증강한 데이터들에 대한 파지 생성을 진행하였다.

각 객체 인스턴스에 대한 파지를 생성할 때, GraspLDM[3]의 추론 단계는 1,000번으로 설정하였으며, 객체당 10개의 파지 자세를 생성하도록 하였다.

GraspLDM[3]으로부터 생성된 각 파지 자세는 4x4 변환 행렬로 표현된다. 본 연구에서는 이 행렬의 평행이동 성분 ($m_{1\ 4}$, $m_{2\ 4}$, $m_{3\ 4}$)을 파지점의 3차원 위치 좌표로 정의하여 사용하였다.

5. 파지 특성 정량화 지표 및 분석 방법

이렇게 추출된 파지점들의 위치를 기반으로, 객체의 크기 변화가 객체 중심으로부터 파지점까지의 평균 거리와 해당 거리 분포의 표준 편차에 어떠한 영향을 미치는지 정량적으로 분석하였다. 모든 거리는 객체의 크기 변화 효과를 보정하기 위해 해당 스케일 팩터(scale factor)로 정규화하여 비교하였다.

Ⅳ. 실험 및 결과

본 절에서는 앞서 제시한 실험 설계에 따라 수행한 축소와 확대를 포함하는 두 가지 실험 결과를 보고한다. 생성한 10개 파지 자세에 대한 파지점의 평균적인 좌표를 기반으로 각 파지점과 객체 원점 사이 거리들의 평균 거리와 위거리 값들의 표준 편차를 구하였다. 평균 거리와 표준 편차는 각각 스케일링에 사용한 스케일 팩터(scale factor) 값으로 나누어 객체 크기 변화를 고려한 정규화를 진행한 값으로 변환해 사용하였다.

1. Scale을 축소했을 경우 (원본 대비 0.5, 0.8배 한 경우)

표 1에서와 같이, 정규화된 평균 파지 거리는 대부분의 객체에서 원본에 대비해 큰 폭으로 증가하였다. 이는 객체가 절반 크기로 작아졌음에도 생성된 파지점들이 객체의 새로운 중심으로부터 상대적으로 멀리 위치하고 있음을 의

표 1. 10개 생성 파지의 평균, 표준 편차 정량적 비교 (축소 시)
Table 1. Comparison 10 grasps of the mean and standard deviation (when reduced)

bottle id	index	Original Distance	scaled to 0.5x	rate of change	scaled to 0.8x	rate of change
3854	average	0.215464	0.403201	+87.13%	0.255332	+18.50%
	deviation	0.030478	0.709063	+2226.47%	0.045049	+47.81%
3933	average	0.220666	0.389390	+76.46%	0.231771	+5.03%
	deviation	0.053191	0.055028	+3.45%	0.046556	-12.47%
3934	average	0.240417	0.316451	+31.63%	0.223620	-6.99%
	deviation	0.052632	0.059467	+12.99%	0.027442	-47.86%
3990	average	0.214794	0.362277	+68.66%	0.221371	+3.06%
	deviation	0.057816	0.058426	+1.06%	0.048471	-16.16%
4084	average	0.210450	0.279558	+32.84%	0.232778	+10.61%
	deviation	0.040778	0.081500	+99.86%	0.053585	+31.41%
4118	average	0.201026	0.322101	+60.23%	0.204614	+1.78%
	deviation	0.061136	0.051091	-16.43%	0.034387	-43.75%
4200	average	0.205041	0.326320	+59.15%	0.238580	+16.36%
	deviation	0.061136	0.040699	-33.43%	0.046496	-23.95%
4233	average	0.175308	0.306647	+74.92%	0.257013	+46.61%
	deviation	0.054659	0.065872	+20.51%	0.046927	-14.15%

표 2. 10개 생성 파지의 평균, 표준 편차 정량적 비교 (확대 시)
Table 2. Comparison 10 grasps of the mean and standard deviation (when enlarged)

bottle id	index	Original Distance	scaled to 1.2x	rate of change	scaled to 1.5x	rate of change
3854	average	0.215464	0.169080	-21.53%	0.120058	-44.28%
	deviation	0.030478	0.046815	+53.60%	0.036802	+20.75%
3933	average	0.220666	0.154410	-30.03%	0.123312	-44.12%
	deviation	0.053191	0.043886	-17.49%	0.034314	-35.49%
3934	average	0.240417	0.196877	-18.11%	0.123735	-48.53%
	deviation	0.052632	0.030471	-42.11%	0.031358	-40.42%
3990	average	0.214794	0.131256	-38.89%	0.120053	-44.11%
	deviation	0.057816	0.035053	-39.37%	0.034023	-41.15%
4084	average	0.210450	0.177278	-15.76%	0.100917	-52.05%
	deviation	0.040778	0.020867	-48.83%	0.019584	-51.97%
4118	average	0.201026	0.135885	-32.40%	0.106455	-47.04%
	deviation	0.061136	0.026301	-56.98%	0.033100	-45.86%
4200	average	0.205041	0.150866	-26.42%	0.131072	-36.08%
	deviation	0.061136	0.053108	-13.13%	0.041732	-31.74%
4233	average	0.175308	0.143053	-18.40%	0.096624	-44.88%
	deviation	0.054659	0.026750	-51.06%	0.010904	-80.05%

미한다. 즉, 객체의 축소된 크기에 비해 예상 파지점이 상대적으로 더 외곽에 위치하게 되었음을 의미한다. 정규화된 표준 편차 역시 대부분의 객체에서 증가하는 경향을 보였으며, 특히 일부 객체에서는 그 증가 폭이 컸다. 이는 평균적인 파지점을 중심으로 넓게 분산되어 파지점들이 생성되었음을 의미하며, 결과적으로 파지 위치가 넓게 퍼져 위치하고 있음을 의미한다.

2. Scale을 확대했을 경우 (원본 대비 1.2, 1.5배 한 경우)

표 2에서와 같이 객체의 크기가 원본 대비 1.2배, 1.5배 확대되었을 때, GraspLDM[3]이 객체를 축소했을 때와는 구분되는 일관적인 반응을 보였다. 정규화된 평균 파지 거리는 객체가 커짐에 따라 원본 대비 대부분 감소하는 경향을 보였다. 이는 파지점들이 커진 객체 크기에 비례하여 바깥쪽으로 확장되지 않고, 오히려 객체 크기 대비 상대적으로 더 객체 중심부에 가깝게 형성되었음을 의미한다. 정규화된 표준 편차 역시 대부분 객체에 대해 일관되게 감소하였다. 이는 생성된 파지점들이 평균 위치 주변으로 좁은 영역에 집중되어 생성되었음을 의미한다. 즉, 파지점이 다양하기보다 일관성 있게 위치하고 있음을 시사한다.

3. 생성 파지 샘플링 개수를 늘린 경우

GraspLDM[3]은 샘플링 수에 따라 생성 결과가 달라질 수 있는 특성을 가진다. 따라서 각 객체당 10개의 파지 자세를 생성한 기존 결과와 파지 샘플링 수를 20개로 늘려 생성한 결과를 비교하는 실험을 추가로 진행하였다.

정량적 비교에 앞서, 그림 3과 같이 생성된 파지 결과를 시각화하였다. 샘플링 수를 늘려도 10개의 파지 자세를 생성한 기존 결과와 동일하게 GraspLDM[3]은 객체 크기 변화에 대해 일관된 반응 패턴을 보였다.

표 3과 표 4와 같이 10개의 파지 자세를 생성한 결과와 20개의 파지 자세를 생성한 결과가 정량적으로도 비슷한 경향을 보이는 것을 확인하였다. 즉, 객체 크기가 작아지면

그림 3. 생성된 파지 시각화
Fig. 3. Visualization of generated grasps

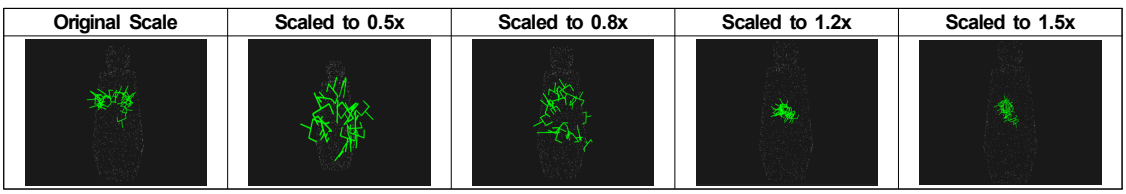

표 3. 20개 생성 파지의 평균, 표준 편차 정량적 비교 (축소 시)
Table 3. Comparison 20 grasps of the mean and standard deviation (when reduced)

bottle id	index	Original Distance	scaled to 0.5x	rate of change	scaled to 0.8x	rate of change
3854	average	0.190015	0.356416	+87.57%	0.243748	+28.28%
	deviation	0.060111	0.709063	+1079.59%	0.045049	-25.06%
3933	average	0.190570	0.391550	+105.46%	0.258115	+35.44%
	deviation	0.042106	0.072003	+71.00%	0.048613	+15.45%
3934	average	0.186697	0.255248	+36.72%	0.260237	+39.39%
	deviation	0.030878	0.083140	+169.25%	0.035630	+15.39%
3990	average	0.167049	0.252029	+50.87%	0.242789	+45.34%
	deviation	0.034939	0.080126	+129.33%	0.031468	-9.93%
4084	average	0.190807	0.235616	+23.48%	0.247280	+29.60%
	deviation	0.034759	0.071457	+105.58%	0.042825	+23.21%
4118	average	0.182119	0.154575	-15.12%	0.267268	+46.75%
	deviation	0.031734	0.056398	+77.72%	0.040468	+27.52%
4200	average	0.201334	0.230058	+14.27%	0.256859	+27.58%
	deviation	0.044994	0.093210	+107.16%	0.047697	+6.01%
4233	average	0.159413	0.172866	+8.44%	0.261311	+63.92%
	deviation	0.043365	0.076049	+75.37%	0.042690	-1.56%

표 4. 20개 생성 파지의 평균, 표준 편차 정량적 비교 (확대 시)
Table 4. Comparison 20 grasps of the mean and standard deviation (when enlarged)

bottle id	index	Original Distance	scaled to 1.2x	rate of change	scaled to 1.5x	rate of change
3854	average	0.190015	0.126861	-33.24%	0.097912	-48.47%
	deviation	0.060111	0.032507	-45.92%	0.025997	-56.75%
3933	average	0.190570	0.144098	-24.39%	0.099690	-47.69%
	deviation	0.042106	0.051156	+21.49%	0.025028	-40.56%
3934	average	0.186697	0.121403	-34.97%	0.110963	-40.57%
	deviation	0.030878	0.033607	+8.84%	0.035456	+14.83%
3990	average	0.167049	0.150896	-9.67%	0.113623	-31.98%
	deviation	0.034939	0.035687	+2.14%	0.029524	-15.50%
4084	average	0.190807	0.135622	-28.92%	0.107328	-43.75%
	deviation	0.034759	0.041404	+19.12%	0.031405	-9.65%
4118	average	0.182119	0.157779	-13.36%	0.121190	-33.46%
	deviation	0.031734	0.049148	+54.87%	0.028844	-9.11%
4200	average	0.201334	0.117709	-41.54%	0.109975	-45.38%
	deviation	0.044994	0.034866	-22.51%	0.035348	-21.44%
4233	average	0.159413	0.143967	-9.69%	0.090604	-43.16%
	deviation	0.043365	0.038356	-11.55%	0.028245	-34.87%

파지점이 넓게 분산되고 외곽에 형성되는 경향과 객체 크기가 커지면 파지점이 객체 중심으로 밀집되는 현상이 유지되었다.

4. 스케일별 후처리 보정 적용

GraspLDM[3]은 입력 객체의 포인트 클라우드(point cloud) 형태를 객체 중심으로 정규화한 상태에서 학습된다. 이로 인해 테스트 시점에 크기가 변형된 객체가 주어지면, 학습 당시와 유사한 형태의 포인트 클라우드(point cloud) 형태로 인식하여 유사한 잠재 표현을 형성하고, 결과적으로 크기 변화에 비례하지 않는 파지 위치를 출력하는 경향을 보인다. 이러한 한계를 사후적으로 보정하기 위해, 생성된 파지의 위치를 각 객체의 스케일 팩터(scale factor)로 나누는 후처리 보정을 적용하였다. 이는 GraspLDM[3]이 반영하지 못한 객체 크기 차이를 사후적으로 조정하여, 이상적으로 기대되는 파지 위치에 근접시키기 위한 시도이다.

그 결과, 전체 실험 중 약 40.6%의 경우에서만 보정을 적용함으로써 오차가 감소하는 것으로 나타났다. 축소된 객체에서는 오히려 오차가 증가했고, 확대된 객체에서만

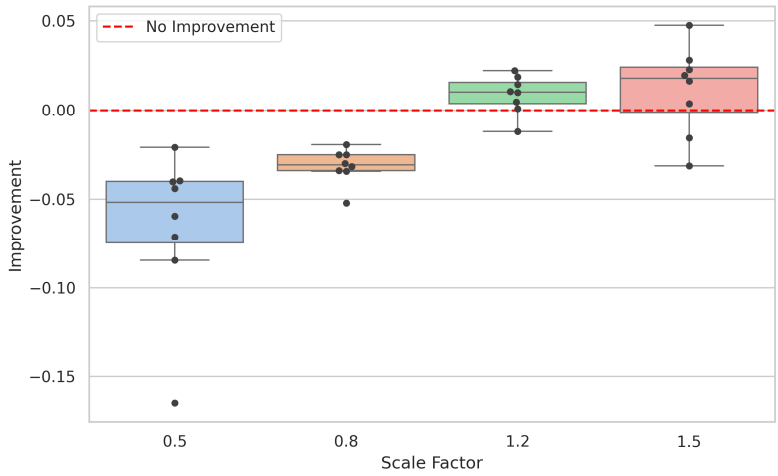

그림 4. 스케일별 후처리 보정 적용 시 오차 개선 비교
Fig. 4. Comparison of error improvement when applying post-processing correction by scale

제한적으로 보정 효과가 나타났다. 이는 GraspLDM[3]이 객체 크기에 대해 단순 비례적으로 반응하지 않으며, 후처리 보정 역시 제한적인 효과만을 가진다는 점을 시사한다.

V. 결 론

본 연구는 파지 생성 모델인 GraspLDM[3]이 객체 크기 변화에 반응하는 방식을 분석하고 정량화하는 데 중점을 두었다. 특히, ARNOLD[7]의 bottle 객체들을 대상으로 한 실험을 통해 GraspLDM[3]이 절대적인 크기 변화에 대해 다음과 같은 두 가지 주요 반응 패턴을 보이는 것을 확인하였다.

첫째, 객체의 크기가 원본 대비 작아질 경우, 대다수 객체에서 파지점은 객체 크기 대비 상대적으로 외곽에 형성되고 그 위치의 일관성 또한 저하되는 경향이 나타났다.

둘째, 객체의 크기가 원본 대비 커질 경우, 대다수 객체에서 파지점이 객체 크기 대비 상대적으로 더 중심부에 가깝게, 그리고 매우 좁은 범위 내에 밀집되어 생성되었다.

이러한 결과는 GraspLDM[3]이 크기 변화에 대해 단순히 비례적으로 반응하는 것을 넘어, 크기 변화에 대응하는 데에 한계가 있음을 시사한다. 오히려 축소와 확대 시 서로 다른 양상의 한계점을 보인다는 새로운 이해를 제공한다. 또한 GraspLDM[3]이 입력 객체의 크기 변화에 민감하게 반응하며, 단순한 위치 기반 후처리 보정은 제한적으로 성능을 개선하는 것을 정량적으로 확인하였다. 이는 파지 생성 모델의 강인성 및 일반화 능력에 대한 실증적 분석 결과로, 향후 모델 개선 및 실제 로봇 환경 적용 시 객체 크기를 고려해야 함을 시사한다.

참 고 문 헌 (References)

[1] Xie, Zhen, Xinquan Liang, and Canale Roberto. “Learning-based robotic grasping: A review.” Frontiers in Robotics and AI, Vol.10, 2023.
doi: https://doi.org/10.3389/frobt.2023.1038658

[2] Wolf, Rosa, et al. “Diffusion Models for Robotic Manipulation: A Survey.”, 2025.
doi: https://doi.org/10.48550/arXiv.2504.08438

[3] Barad, K. R., Orsula, A., Richard, A., Dentler, J., Olivares-Mendez, M., & Martinez, C. “Graspldm: Generative 6-dof grasp synthesis using latent diffusion models.”, IEEE Access, 2024.
doi: https://doi.org/10.1109/ACCESS.2024.3492118

[4] Kingma, Diederik P., and Max Welling. “Auto-encoding variational bayes.”, 2013.
doi: https://doi.org/10.48550/arXiv.1312.6114

[5] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. “Denoising diffusion probabilistic models.”, Advances in neural information processing systems 33, pp.6840-6851, 2020.
doi: https://doi.org/10.48550/arXiv.2006.11239

[6] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B., “High-resolution image synthesis with latent diffusion models.”, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022.
doi: https://doi.org/10.48550/arXiv.2112.10752

[7] Gong, R., Huang, J., Zhao, Y., Geng, H., Gao, X., Wu, Q., ... & Huang, S. “Arnold: A benchmark for language-grounded task learning with continuous states in realistic 3d scenes.” Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023.
doi: https://doi.org/10.48550/arXiv.2304.04321
[8] Razavi, Ali, Aaron Van den Oord, and Oriol Vinyals. “Generating diverse high-fidelity images with vq-vae-2.” Advances in neural information processing systems Vol. 32, 2019.
doi: https://doi.org/10.48550/arXiv.1906.00446
[9] Liu, Kaikai, Renjun Shuai, and Li Ma. “Cells image generation method based on VAE-SGAN.” Procedia Computer Science, Vol.183, pp.589-595, 2021.
doi: https://doi.org/10.1016/j.procs.2021.02.101
[10] Zhang, Chenrui, and Yuxin Peng. “Stacking VAE and GAN for context-aware text-to-image generation.” IEEE Fourth International Conference on Multimedia Big Data (BigMM), 2018.
doi: https://doi.org/10.1109/BigMM.2018.8499439
[11] Zhang, X., Zhao, W., Lu, X., & Chien, J. “Text2layer: Layered image generation using latent diffusion model”, 2023.
doi: https://doi.org/10.48550/arXiv.2307.09781
[12] Mei, K., Delbracio, M., Talebi, H., Tu, Z., Patel, V. M., & Milanfar, P., “CoDi: conditional diffusion distillation for higher-fidelity and faster image generation”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.9048-9058, 2024.
doi: https://doi.org/10.48550/arXiv.2310.01407
[13] Pinaya, W. H., Tudosiu, P. D., Dafflon, J., Da Costa, P. F., Fernandez, V., Nachev, P., ... & Cardoso, M. J. “Brain imaging generation with latent diffusion models”, MICCAI Workshop on Deep Generative Model, Switzerland, pp.117-126, 2022.
doi: https://doi.org/10.48550/arXiv.2209.07162
[14] Xiang, J., Yang, J., Huang, B., & Tong, X. ,“3d-aware image generation using 2d diffusion models.”, Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2383-2393, 2023.
doi: https://doi.org/10.48550/arXiv.2303.17905
[15] Ma, Y., Yang, H., Wang, W., Fu, J., & Liu, J.,“Unified multi-modal latent diffusion for joint subject and text conditional image generation”, 2023.
doi: https://doi.org/10.48550/arXiv.2303.09319
[16] Lee, Sangmin, Sungyong Park, and Heewon Kim. “DynScene: Scalable Generation of Dynamic Robotic Manipulation Scenes for Embodied AI.” Proceedings of the Computer Vision and Pattern Recognition Conference, 2025. https://openaccess.thecvf.com/content/CVPR2025/papers/Lee_DynScene_Scalable_Generation_of_Dynamic_Robotic_Manipulation_Scenes_for_Embodied_CVPR_2025_paper.pdf

저 자 소 개

김 연 지

- 2021년 ~ 현재 : 숭실대학교 글로벌미디어학부 학사
- ORCID : https://orcid.org/0009-0000-1410-0158
- 주관심분야 : 컴퓨터비전, 로보틱스

이 상 민

- 2018년 ~ 2024년 : 숭실대학교 글로벌미디어학부 학사
- 2024년 ~ 현재 : 숭실대학교 미디어학과 석사과정
- ORCID : https://orcid.org/0009-0007-1713-5197
- 주관심분야 : 컴퓨터비전, 머신러닝, 로보틱스

김 희 원

- 2008년 ~ 2014년 : 서울대학교 전기·정보공학부 학사
- 2017년 ~ 2023년 : 서울대학교 전기·정보공학부 박사
- 2023년 ~ 현재 : 숭실대학교 글로벌미디어학부 조교수
- ORCID : https://orcid.org/0000-0001-7777-9823
- 주관심분야 : 컴퓨터비전, 머신러닝, 로보틱스