

그리퍼 유연 접촉층의 두께에 따른 VLA 기반 로봇 조작 성능 평가

이강석*, 김민준**, 송영준***, 최동수****

Evaluation of VLA-based Robot Manipulation Performance According to the Thickness of a Flexible Gripper Contact Layer

Kang-Seok Lee*, Min-Jun Kim**, Young Joon Song***, and Dong-Soo Choi****

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (RS-2026-25490371, RS-2026-25516934). 또한, 2026년도 교육부 및 경상북도의 재원으로 경북RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)-(특성화대학)의 결과입니다(2026-rise-15-105).

요 약

본 연구에서는 그리퍼 접촉면에 적용된 유연 소재의 두께가 미세조정된 VLA 정책의 조작 성능에 미치는 영향을 평가한다. 제한된 GPU 환경에서 후보 모델의 실행 성능을 비교하여 Octo-Small을 선정하고, 50개의 시연 데이터로 미세조정한다. 유연 접촉층의 두께는 0, 1, 2 및 3 mm로 설정하고, 조건별 20회의 집기 및 놓기 실험을 수행한다. 실험 결과, 0, 1 및 2 mm 조건의 성공률은 각각 75.0%, 70.0% 및 80.0%이지만, 3 mm 조건에서는 40.0%로 감소하고 평균 파지 시도 횟수도 증가한다. 이러한 결과는 유연 접촉층이 일정 두께 이상 증가하면 그리퍼의 외형과 접촉 조건 변화로 인해 기존 VLA 정책의 조작 성능이 저하될 수 있음을 보여준다.

Abstract

This study evaluates how the thickness of a flexible gripper contact layer affects a fine-tuned Vision-Language-Action (VLA) policy. Octo-Small is selected under a resource-constrained GPU environment and fine-tuned using 50 demonstrations. Pick-and-place experiments are conducted 20 times for each thickness of 0, 1, 2, and 3 mm. The success rates are 75.0%, 70.0%, 80.0%, and 40.0%, respectively, with more grasp attempts observed at 3 mm. These results indicate that excessive contact-layer thickness may degrade VLA-based manipulation performance by altering the gripper's appearance and contact conditions.

Keywords

vision-tactile-language-action, robotic manipulation, end-effector, pick-and-place

- * 국립금오공과대학교 컴퓨터공학전공 학사과정
- ORCID: <https://orcid.org/0009-0006-8028-3996>
- ** 국립금오공과대학교 컴퓨터·AI융합공학과 석사과정
- ORCID: <https://orcid.org/0009-0007-3625-0852>
- *** 국립금오공과대학교 전자공학부 교수(공동교신저자)
- ORCID: <https://orcid.org/0000-0002-9736-602X>
- **** 국립금오공과대학교 컴퓨터공학부 부교수(공동교신저자)
- ORCID: <https://orcid.org/0000-0001-8694-0965>

- Received: Aug. 18, 2026, Revised: Sep. 02, 2026, Accepted: Sep. 05, 2026
- Co-corresponding Author 1: Young Joon Song
Dept. of Electronic Engineering, Kumoh National Institute of Technology,
Republic of Korea
Tel.: +82-54-478-7486, Email: yjsong@kumoh.ac.kr
- Co-corresponding Author 2: Dong-Soo Choi
Dept. of Computer Engineering, Kumoh National Institute of Technology,
Republic of Korea
Tel.: +82-54-478-7538 Email: choids@kumoh.ac.kr

I. 서 론

VLA(Vision-Language-Action) 모델은 시각 정보와 자연어 명령을 바탕으로 로봇의 제어 행동을 생성한다. 최근에는 다양한 로봇과 대규모 조작 데이터로 VLA 모델을 사전학습한 후, 소량의 시연 데이터를 이용해 새로운 로봇과 목표 작업에 맞게 미세조정(Fine-tuning)하는 방식이 주로 활용되고 있다[1].

그러나 시각 정보를 주요 관측값으로 활용하는 VLA 정책은 대상 물체와의 실제 접촉 상태를 직접적으로 인식하는 데 한계가 있다. 이를 보완하기 위해 시각 및 언어 정보와 함께 촉각 정보를 활용하는 VTLA(Vision-Tactile-Language-Action)에 관한 연구가 이루어지고 있다[2][3]. VTLA를 구현하기 위해서는 그리퍼 표면에 촉각 센서나 유연한 접촉층을 부착해야 하므로, 기존 그리퍼의 외형과 접촉 조건을 변화시킬 수 있다. 특히 촉각 인터페이스의 두께가 증가하면 카메라에 관찰되는 그리퍼의 형상뿐만 아니라 물체와의 접촉 위치와 파지 간격도 달라질 수 있어, 기존에 미세조정된 VLA 정책의 성능에 영향을 줄 가능성이 있다.

폴리염화비닐(Poly vinyl chloride) 젤(PVC gel)은 높은 유연성과 대변형 특성을 갖는 고분자 소재로, 소프트 액추에이터, 센서 및 다양한 유연 소자에 활용되고 있다[4][5]. 이러한 특성으로 인해 PVC gel은 향후 그리퍼에 부착되는 유연 촉각 인터페이스의 소재로 활용될 수 있다. 그러나 PVC gel을 적용하기 위해서는 적절한 두께를 결정해야 하며, 이를 위해서는 촉각 정보의 활용에 앞서 gel의 부착 두께 자체가 기존 VLA 정책의 조작 성능에 미치는 영향을 우선적으로 확인할 필요가 있다.

이에 본 연구에서는 향후 VTLA로의 확장을 위한 선행 연구로서, 그리퍼에 부착된 PVC gel의 두께 변화가 미세조정된 VLA 정책의 물체 조작 성능에 미치는 영향을 정량적으로 평가하였다. 사전학습된 VLA 모델을 집기 및 놓기(Pick-and-place) 작업에 맞게 미세조정 후, 추가 학습이나 정책 변경 없이 PVC gel의 두께를 0 mm, 1 mm, 2 mm 및 3 mm로 변화시켰다. 각 조건에서 동일한 작업을 20회씩 수행하고, 작업 성공률, 평균 수행시간 및 파지 시도 횟수를 분석하여 PVC gel의 두께에 따른 조작 성능

을 비교하였다.

II. 실험 방법 및 결과

2.1 VLA 정책 선정

본 연구에 적용할 VLA 정책을 선정하기 위해 NVIDIA GeForce RTX 3060 Ti(8 GiB)가 탑재된 동일한 PC 환경에서 후보 모델의 정책 실행 빈도와 최대 GPU 메모리 사용량을 측정하였다(표 1).

각 모델은 배치 크기 1에서 10회의 예열 후 30회 반복 추론하였으며, 전처리, 모델 추론, 후처리를 포함한 전체 처리시간의 중앙값을 이용해 정책 실행 빈도를 산출하였다. JAX(자동 미분과 JIT 컴파일을 지원하는 Python 기반 수치 연산 프레임워크) 기반의 Octo 및 π 계열 모델은 GPU 메모리 사전 할당을 비활성화한 동일한 조건에서 실행하였다.

표 1. VLA 정책별 실행 성능 비교

Table 1. Comparison of execution performance of VLA policies

Model	Quantization	Policy rate [Hz]	Peak GPU memory [GiB]
Octo-small	None	13.88	0.549
Octo-base	None	10.32	0.828
OpenVLA-7B	INT4 NF4	1.94	4.41
π_0 -base	None	N/A	N/A
π_0 -FAST-base	None	N/A	N/A
$\pi_{0.5}$ -base	None	N/A	N/A

비교 결과, 정상적으로 추론이 수행된 모델 중 Octo-Small이 13.88 Hz로 가장 높은 정책 실행 빈도를 나타냈다. Octo-Base와 OpenVLA-7B는 각각 10.32 Hz와 1.94 Hz를 기록하였다. 반면, π_0 -Base, π_0 -FAST-Base 및 $\pi_{0.5}$ -Base는 8 GiB GPU 환경에서 모델 가중치를 불러오는 과정에서 메모리 부족이 발생하여 추론을 수행하지 못하였다. 이에 본 연구에서는 단일 집기 및 놓기 작업의 미세조정과 실시간 실행에 적합한 모델로 Octo-Small을 선정하였다.

2.2 폴리염화비닐 젤 두께별 조작 성능 평가

실험 환경은 협동로봇(UR5e)과 그리퍼(OnRobot

RG2)로 구성하였다(그림 1). 로봇의 전체 작업공간을 관찰하기 위해 외부 카메라(Intel RealSense D455)를 로봇 측면에 주 카메라로 설치하였으며, 조작 물체와 그리퍼 주변의 근접 영상을 획득하기 위해 손목 카메라(oCam-1CGN-U-T2)를 그리퍼에 장착하였다.

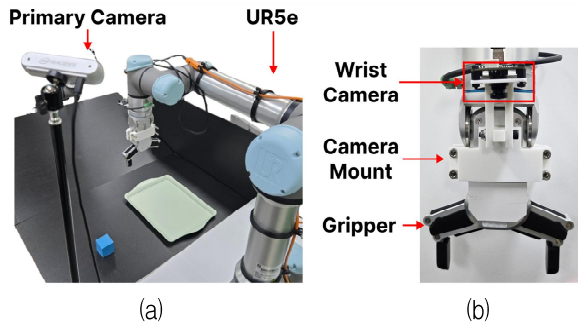


그림 1. 실험 환경
(a) 전체 실험 환경 (b) 그리퍼 구성

Fig. 1. Experimental setup
(a) overall experimental environment (b) gripper configuration

로봇의 행동 정책으로는 사전학습된 Octo-Small 모델을 사용하였다. 모델에는 주 카메라와 손목 카메라에서 획득한 RGB 영상과 자연어 명령을 입력하였다. 집기 및 놓기 작업에 대한 정책을 학습하기 위해 자체 수집한 50개의 시연 데이터로 모델을 미세조정하였다. 이때 사전학습된 기반 신경망(Backbone)은 고정하고, 행동 출력부(Action head)의 매개변수만을 학습하는 Head-only 방식을 적용하였다. 또한 조도와 노출 변화에 대한 강건성을 높이기 위해 0.5의 확률로 영상 밝기를 원본의 0.8~1.2배 범위에서 무작위로 조정하였다.

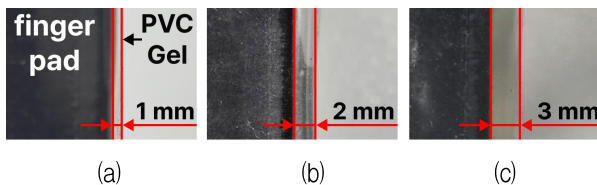


그림 2. 그리퍼에 부착된 폴리염화비닐 젤
(a) 1 mm (b) 2 mm (c) 3 mm

Fig. 2. PVC Gel attached to the gripper fingertip
(a) 1 mm (b) 2 mm (c) 3 mm

폴리염화비닐 젤의 두께에 따른 조작 성능 변화

를 평가하기 위해 그리퍼 접촉부에 두께가 서로 다른 젤을 부착하였다. 젤이 부착되지 않은 상태를 0 mm의 기준 조건으로 설정하고, 1 mm, 2 mm 및 3 mm 두께의 젤을 부착한 조건을 구성하였다(그림 2). 모든 실험 조건에서 젤 두께를 제외한 로봇과 카메라 위치, 대상 물체, 목표 위치, 자연어 명령을 동일하게 유지하였다.

환경 변화가 실험 결과에 미치는 영향을 줄이기 위해 0 mm, 1 mm, 2 mm 및 3 mm 조건의 순서로 각 5회씩 실험하는 과정을 1개 사이클로 구성하고, 이를 총 4회 반복하였다. 이에 따라 각 두께 조건에서 동일한 집기 및 놓기 작업을 20회씩 수행하여 총 80회의 실험을 진행하였다. 로봇이 대상 물체를 정상적으로 파지하여 목표 위치까지 운반한 후, 물체 전체를 목표 영역 내부에 놓은 경우를 성공으로 판정하였다. 작업 수행시간은 성공한 시행을 대상으로 첫 번째 제어 행동이 실행된 시점부터 목표 위치에서 그리퍼가 개방된 시점까지로 정의하였다. 또한 각 시행에서 그리퍼의 닫힘(Close) 동작 횟수를 기록하고, 조건별 평균값을 파지 시도 횟수(Grasp attempts)로 산출하였다.

표 2. 폴리염화비닐 젤 두께에 따른 조작 성능

Table 2. Manipulation performance according to PVC gel thickness

Thickness	Success rate	Task time [s]	Grasp attempts
0 mm	75%	57.26 ± 14.95	4.45
1 mm	70%	61.68 ± 9.98	8.3
2 mm	80%	56.36 ± 11.67	3.9
3 mm	40%	58.40 ± 14.99	9.75

실험 결과, 0 mm 조건에서는 총 20회의 실험 중 15회가 성공하여 75.0%의 성공률이 관찰됐다. 1 mm와 2 mm 조건의 성공률은 각각 70.0%(14회)와 80.0%(16회)였으며, 3 mm 조건에서는 40.0%(8회)로 크게 감소하였다.

성공한 작업의 평균 수행시간은 0 mm, 1 mm, 2 mm 및 3 mm 조건에서 각각 57.26 ± 14.95 s, 61.68 ± 9.98 s, 56.36 ± 11.67 s 및 58.40 ± 14.99 s로 나타났다. 모든 조건의 평균 수행시간이 약 56~62 s 범위에 분포하여, 젤 두께에 따른 뚜렷한 차이는 관찰되지 않았다.

평균 파지 시도 횟수는 0 mm, 1 mm, 2 mm 및 3 mm 조건에서 각각 4.45회, 8.30회, 3.90회 및 9.75회로 나타났다. 2 mm 조건은 가장 높은 성공률과 가장 적은 파지 시도 횟수를 보인 반면, 3 mm 조건은 가장 낮은 성공률과 가장 많은 파지 시도 횟수를 나타냈다. 따라서 0~2 mm 범위에서는 젤 두께 증가에 따른 일관된 성능 저하가 확인되지 않았으나, 3 mm 조건에서는 성공률 감소와 파지 시도 횟수 증가가 동시에 관찰되었다. 이는 VLA의 낮은 입력 해상도로 인해 1~2 mm의 두께 변화는 뚜렷한 시각적 차이로 반영되지 않은 반면, 3 mm 조건에서는 그리퍼 외형과 접촉 위치의 변화가 모델 입력에 유의미하게 반영되었기 때문으로 추정된다.

III. 결 론

본 연구에서는 향후 VTLA로의 확장을 위한 선행 연구로서, 그리퍼에 부착된 폴리염화비닐 젤의 두께가 미세조정된 VLA 정책의 조작 성능에 미치는 영향을 평가하였다. 집기 및 놓기 작업에 맞게 미세조정된 Octo-Small 정책을 추가 학습 없이 유지한 상태에서 젤의 두께를 0, 1, 2 및 3 mm로 변화시키고, 작업 성공률, 수행시간 및 파지 시도 횟수를 비교하였다.

실험 결과, 0, 1 및 2 mm 조건의 성공률은 각각 75.0%, 70.0% 및 80.0%로 나타나, 0~2 mm 범위에서는 두께 증가에 따른 일관된 성능 저하가 관찰되지 않았다. 반면 3 mm 조건에서는 성공률이 40.0%로 감소하고 평균 파지 시도 횟수가 9.75회로 증가하였다. 성공한 시행의 평균 수행시간은 모든 조건에서 약 56~62 s 범위로 나타나 두께에 따른 뚜렷한 차이를 보이지 않았다. 이러한 결과는 일정 수준 이상의 젤 두께가 적용될 경우, 추가 학습 없이 기존 VLA 정책을 그대로 사용하는 과정에서 조작 성능이 저하될 가능성을 보여준다.

폴리염화비닐 젤의 두께 변화는 그리퍼의 시각적 외형뿐만 아니라 유효 파지 간격, 접촉 위치 및 물체와의 상호작용 조건을 함께 변화시킨다. 따라서 본 연구의 결과만으로 성능 변화가 시각적 변화와 물리적 변화 중 어느 요인에 의해 발생했는지를 구분하기는 어렵다. 또한 단일 작업과 대상 물체를 대

상으로 조건별 20회의 실험을 수행하였다는 점에서 결과를 일반화하는 데 한계가 있다. 향후에는 젤이 부착된 상태에서 추가로 미세조정된 정책과의 비교, 시각적·물리적 요인을 분리한 실험 및 다양한 물체와 작업에 대한 반복 평가가 필요하다. 이후 폴리염화비닐 젤 기반 촉각 정보를 VLA 정책의 입력으로 통합하여, 촉각 정보의 추가가 조작 성능과 강건성에 미치는 영향을 분석할 예정이다.

References

- [1] M. J. Kim, et al., "OpenVLA: An Open-Source Vision-Language-Action Model", Proc. The 8th Conference on Robot Learning, Proceedings of Machine Learning Research, Munich, Germany, Vol. 270, pp. 2679-2713, Nov. 2024.
- [2] Z. Cheng, Y. Zhang, A. Tang, K. Wang, W. Zhang, H. Li, H. Zhang, and L. Song, "OmniVTLA: Vision-Tactile-Language-Action Models With Semantic-Aligned Tactile Sensing", IEEE Robotics and Automation Letters, Vol. 11, No. 9, pp. 10146-10153, Sep. 2026. <https://doi.org/10.1109/LRA.2026.3710060>.
- [3] J. Bi, K. Y. Ma, C. Hao, M. Z. Shou, and H. Soh, "VLA-Touch: Enhancing-Vision-Language-Action Model With Dual-Level Tactile Feedback", IEEE Robotics and Automation Letters, Vol. 11, No. 7, pp. 8487-8494, Jul. 2026. <https://doi.org/10.1109/LRA.2026.3692345>.
- [4] Y. Li, H. Zhu, Z. Zhang, L. Zhu, X. Zhang, M. Guo, Y. Li, and M. Hashimoto, "Plasticized Polyvinyl Chloride: From Material Properties to Flexible Applications", Advances in Colloid and Interface Science, Vol. 337, Art. no. 103384, Mar. 2025. <https://doi.org/10.1016/j.cis.2024.103384>.
- [5] Y. Li, M. Guo, and Y. Li, "Recent Advances in Plasticized PVC Gels for Soft Actuators and Devices: A Review", Journal of Materials Chemistry C, Vol. 7, No. 42, pp. 12991-13009, Nov. 2019. <https://doi.org/10.1039/C9TC04366G>.