

GIST, 사람과 협업할 줄 아는 세계 최고 성능 AI 로봇 파지 모델 개발

단 한 번의 추론으로 다양한 환경에서 어떤 물체든 정확하고 안정적으로 움켜쥐어... 산업 현장 혁신 기대

- AI융합학과 이규빈 교수팀, 다양한 형태의 프롬프트 입력 지원으로 작업자의 의도에 따라 복잡한 환경에서 미학습 물체도 정확히 파지... 작업자 시선 추적 기반 파지 기능까지
- 가정용 및 서비스 로봇 등 다양한 분야에 활용 기대... 2025년 5월 세계 최대 로봇학술대회 'IEEE International Conference on Robotics and Automation'에서 발표 예정



▲ (왼쪽부터) AI융합학과 이규빈 교수, 노상준 박사과정생, 김종원·강래영·남동우 석박사통합과정(박사과정), 백승혁 한국기계연구원 선임연구원

산업 현장에서 로봇이 물체를 정확하고 안정적으로 잡기 위해서는 정교한 파지(把持, grasp) 기술이 필수적이다. 하지만 기존의 파지 기술은 물체의 종류나 환경 변화에 따라 새로운 모델 학습이 필요하며, 제한된 데이터세트에 의존하는 한계가 있었다.

광주과학기술원(GIST, 총장 임기철)은 AI융합학과 이규빈 교수 연구팀이 단순한 자동화를 넘어, 작업자와의 협업을 고려한 세계 최고 성능의 혁신적인 로봇 파지 모델 'GraspSAM'을 개발했다고 밝혔다.

작업자와의 협업을 고려한 혁신적 AI 모델인 GraspSAM은 점, 박스, 텍스트 등 다양한 형태의 프롬프트 입력을 지원하며, 한 번의 추론만으로 물체의 파지점을 정확히 예측할 수 있도록 설계되었다. 이를 통해 기존 모델의 한계를 극복하고, 학습되지 않은 물체라도 복잡한 환경에서 안정적으로 파지할 수 있는 능력을 갖췄다.

기존의 딥러닝 기반 파지 모델들은 환경과 상황에 따라 별도의 시모델을 학습해야 하는 한계를 지니고 있었다.

연구팀은 이를 해결하기 위해 페이스북 모회사 메타(Meta)가 개발한 범용 이미지 분할 모델인 SAM(Segment Anything Model)을 최초로 로봇 파지 출력이 가능하도록 도입했다.

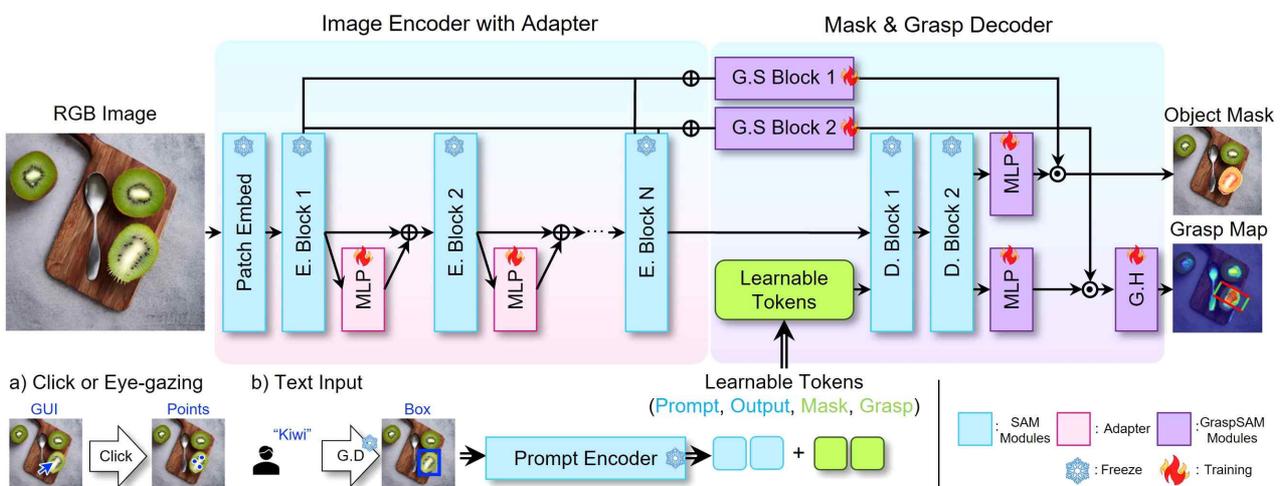
연구팀이 개발한 GraspSAM은 SAM의 강력한 객체 분할 능력을 활용해 최소한의 미세 조정만으로 물체의 파지점을 예측하는 혁신적인 모델이다. 이를 위해 어댑터(Adapter) 기법과 학습 가능한 토큰(Learnable Tokens) 기법을 적용해 SAM을 파지점 추론에 최적화했다.

* **SAM(Segment Anything):** Meta AI(구 페이스북)에서 개발한 범용적인 이미지 분할 모델로서 이 모델의 가장 큰 특징은 이름 그대로 어떤 것이라도 분할(Segment Anything)할 수 있다는 점이다. 기존의 이미지 분할 모델들은 특정 데이터셋이나 특정한 객체에 최적화되어 있는 경우가 많았지만, SAM은 사전 학습된 거대한 데이터셋을 기반으로 하기 때문에, 새로운 이미지나 본 적 없는 객체에도 훌륭한 성능을 보인다.

* **어댑터(Adapter)와 학습 가능한 토큰(Learnable Tokens) 기법:** 이 두 가지 개념은 대형 모델을 효율적으로 미세 조정(fine-tuning)하는 방법에서 많이 쓰인다. 각자의 역할은 조금 다르지만, 최대한 적은 자원으로 성능을 높이는 목표는 같다. 먼저 어댑터는 대형 모델의 파라미터를 거의 고정한 채, 작은 네트워크(모듈)만 추가해 학습하는 방식이다. 기존 모델을 크게 변경하지 않고, 새로운 데이터를 학습할 수 있다. 학습 가능한 토큰(Learnable Tokens)은 기존 입력 데이터(텍스트, 이미지 패치 등)와 함께 학습 가능한 임베딩 토큰을 추가해 모델에 전달한다. 문맥이나 추가 정보를 강조하고 싶을 때 사용한다.

GraspSAM은 프롬프트 기반 입력을 지원하며 사용자가 제공하는 간단한 점, 박스, 텍스트 입력을 통해 다양한 환경·물체·상황에 즉각적으로 적응할 수 있도록 설계되었다.

이를 통해 로봇이 보다 다양한 물체를 손쉽게 잡을 수 있으며, 단 한 번의 연산만으로 파지점을 예측해 산업 현장에서의 적용 범위를 획기적으로 확장했다.



▲ **GRASPSAM 전체 파이프라인** : 한 장의 RGB 이미지와 다양한 형태의 프롬프트(점, 박스 또는 언어)를 입력받아 파지 대상 물체의 마스크와 해당 물체 파지를 위한 파지맵을 출력하는 네트워크.

GraspSAM은 유명한 파지 벤치마크 데이터셋인 '**Grasp-anything**'과 '**Jacquard**'에서 **최고 수준(SOTA, State of the Art)***의 성능을 달성했다. 또한, 실험 결과 복잡한 실제 환경에서도 로봇이 안정적으로 파지 작업을 수행할 수 있다는 것을 확인했다.

* **SOTA(State of the Art)**: 주로 인공지능(AI)과 머신러닝(ML) 분야에서 자주 사용되며, SOTA 모델은 일반적으로 벤치마크 데이터셋에서 최고 성능을 기록하거나 특정 과제에서 가장 효율적이고 정확한 결과를 제공하는 모델을 의미한다.

특히, 눈동자 추적 기술과 결합해 **작업자의 시선에 맞춰 파지 작업을 수행하는 기**능까지 성공적으로 구현함으로써 **산업 현장은 물론 다양한 분야에서 폭넓은 활용** 가능성을 제시했다. 보다 자세한 관련 정보는 GraspSAM 공식 웹사이트*에서 확인할 수 있다.

* <https://gistailab.github.io/graspsam/>

이규빈 교수는 "**GraspSAM 모델은 로봇과 사용자의 직관적인 상호작용을 가능**하게 하며, 복잡한 환경에서도 탁월한 파지 능력을 발휘해 산업 현장뿐만 아니라 **가정용 로봇, 서비스 로봇 등 다양한 분야에 폭넓게 활용될** 것으로 기대한다"고 밝혔다.

GIST AI 융합학과 이규빈 교수가 지도하고, 노상준 박사과정생이 수행한 이번 연구는 산업통상자원부와 과학기술정보통신부의 지원을 받았다. 연구 결과는 로봇 분야 세계 최고 권위 학회인 IEEE ICRA(International Conference on Robotics and Automation)에서 2025년 5월 발표될 예정이다.

논문의 주요 정보

1. 논문명, 저자정보

- 저널명 : ICRA (International Conference on Robotics and Automation)
- 논문명 : GraspSAM: When Segment Anything Model meets Grasp Detection
- 저자 정보 : 노상준(제1저자, GIST AI융합학과), 김종원(제2저자, GIST AI융합학과), 남동우(제3저자, GIST AI융합학과), 백승혁(제4저자, 기계연구원), 강래영(제5저자, GIST AI융합학과), 이규빈(교신저자, GIST AI융합학과)