

계층적 딥러닝 기반 로봇 제어기 설계: 상위 토글 포인트 예측과 하위 디퓨전 모델 결합

이경호, 이민서, 김성연, 조경훈

인천대학교 정보통신공학과

e-mail : {kno022603, minseo6384, 202201569, ckh0923}@inu.ac.kr

Hierarchical Deep Learning-Based Robotic Controller: Toggle-Point Prediction Coupled with a Diffusion Model

Kyoungho Lee, Minseo Lee, Seongyeon Kim, Kyunghoon Cho

Incheon National University

Abstract

We propose a hierarchical learning-based controller for robotic manipulation that explicitly separates discrete contact-toggle decisions from continuous motion synthesis. The high-level network predicts the next toggle event—its action (open/close), the timing inside a short future window, and a target joint configuration—from an RGB observation and a compact history of robot motion. Conditioned on this plan, a low-level diffusion model generates short-horizon joint motions and gripper signals that smoothly approach the target while remaining reactive to visual feedback. To obtain stable supervision and consistent context, demonstrations are resampled along joint-space arc length, aligning history and future windows at fixed distance increments. In PyBullet pick-place tasks with a Franka Panda, the hierarchy reduces toggle-time instability and improves success without sacrificing trajectory smoothness.

* 이경호, 이민서는 공동 제1저자로서 이 논문에 동일하게 기여하였음.

I. 서론

조작(Manipulation) 문제는 본질적으로 이산적 결정과 연속적 제어가 결합되어 있다. 그리퍼를 여닫거나 접촉을 형성·해제하는 순간(본 논문에서는 이를 토글 포인트라 부른다)은 성공과 실패를 가르는 핵심 사건이고, 이 사건들 사이를 잇는 팔의 연속 궤적은 유연하고 안정적이어야 한다. 기존의 단일 정책으로 모든 것을 한 번에 예측하는 접근은 토글 시점의 불안정, 토글 직후의 목표 이탈, 그리고 테스트 단계의 불필요한 샘플 선택 의존을 낳기 쉽다. 이는 상이한 물리적 성격을 가진 두 하위 문제를 억지로 한 모델에서 처리하려는 구조적 한계에서 비롯된다.

본 연구는 이러한 한계를 해소하기 위해 계층적 컨트롤러를 제안한다. 상위 계층은 영상과 최근 동작의 요약을 바탕으로 다음 토글의 액션(열기/닫기), 가까운 미래 내 시점, 그리고 그 시점에서 도달해야 할 목표 관절 상태를 예측하는 계획자(planner) 역할을 맡는다. 하위 계층은 이 계획을 조건으로 확산(Diffusion) 모델 기반의 연속 정책[1–2]이 짧은 지평의 동작 시퀀스를 합성·집행한다. 상위는 “어디에서 접촉을 전환할 것인가(토글)와 그때의 자세는 무엇인가”를 결정하고, 하위는 “그 지점까지 어떻게 부드럽고 견고하게

갈 것인가”를 해결한다. 결과적으로 토클 근방에서의 급격한 실패 모드를 줄이면서도, 연속 구간에서는 데이터 주도적 다양성과 부드러움을 유지할 수 있다.

II. 본론

2.1 계층적 구조의 필요성

조작 과정은 물체와 그리퍼가 만나는 접촉(contact)의 형성과 해제라는 불연속 사건에 의해 국면이 바뀐다. 집기(pick)와 놓기(place)의 경계는 대개 그리퍼의 여단기와 정확히 겹치며, 이 순간에 자세가 조금만 어긋나도 미끄러짐, 충돌, 실패가 발생한다. 반면 그 사이 구간은 비교적 연속적이며 여러 경로가 허용된다. 즉, 문제의 난점과 위험은 토클 근방에 집중되어 있고, 나머지 구간에서는 다양한 해가 존재한다. 따라서 토클의 액션과 시점, 그리고 그때의 목표 자세를 독립적으로 예측하는 상위 모듈과, 그 계획을 따라가는 연속 궤적 합성을 수행하는 하위 모듈로 분해하는 것이 합리적이다. 이 분해는 학습과 실행 모두에서 난점의 국소화와 책임의 명확화를 이끌어 내며, 데이터 사용 효율과 일반화 가능성을 높인다.

2.2 상위 네트워크: ResNet 기반 토클-플래너

본 계층의 목적은 (i) 다음 토클의 작동 클래스(open/close), (ii) 토클 시점에서의 목표 관절 상태 $\mathbf{q}^* \in \mathbb{R}^{\text{DoF}}$ 의 동시 추정이다. 상위 네트워크는 관찰 영상과 최근 동작 이력을 입력으로 하여, 장면 맥락과 조작 진행을 통합적으로 해석한다.

RGB 이미지 사정 정의된 크기로 리사이즈·크롭되며, 이미지 인코더로 ResNet 백본을 사용한다. 백본 출력은 글로벌 풀링 후 $f_{img} \in \mathbb{R}^m$ 의 고정 차원 임베딩으로 투영된다. 최근 히스토리는 조인트 공간 아크길이로 균등 간격 샘플링되며, 1D-Convolution을 통해 $f_{hist} \in \mathbb{R}^m$ 로 요약한다. 현재 관절 상태는 MLP를 통해 $f_{curr} \in \mathbb{R}^m$ 로 변환된다.

최종 결합 표현은 위 세 임베딩을 채널 방향을 결합하고 선형 변환 후 LayerNorm(LN)을 적용해 공유표현 \mathbf{h} 을 얻는다:

$$\mathbf{h} = \text{LN}(\mathbf{W}[f_{img} \| f_{hist} \| f_{curr}] + \mathbf{b}) \in \mathbb{R}^m.$$

학습 목표는 공유 표현 \mathbf{h} 위에 2개의 출력을 구한다.

- (1) 이전 분류 $\hat{y}_{act} \in \{0,1\}$ (open/close)
- (2) 토클 시점의 목표 관절 $\hat{\mathbf{q}}^* \in \mathbb{R}^{\text{DoF}}$

학습을 위한 손실은 다음과 같다:

$$\mathcal{L} = \lambda_{act} \mathcal{L}_{CE}(\hat{y}_{act}, y_{act}) + \lambda_{plan} \|\hat{\mathbf{q}}^* - \mathbf{q}^*\|_1.$$

ResNet 기반 시각 피처는 물체 기하와 배치를 안정적으로 포착하고, 아크길이기반으로 이전경로를 처리함으로써 노이즈에 강인하다. 두 출력의 병렬 학습은 토클의 무엇(액션), 어디를 동시에 구속해, 하위 모델이 따라야 하는 명시적 계획 신호를 일관되게 제공한다.

2.3 하위 네트워크: 계획 조건 디퓨전 모델

하위 계층은 상위 네트워크가 산출한 $\hat{\mathbf{q}}^*$ 와 토클 클래스 \hat{y}_{act} 를 조건으로 받아, 짧은 지평 T 동안의 연속 관절 변화열과 그리퍼 신호를 생성한다. 이는 계획-추종(planned-tracking)과 반응형(reactive)제어를 동시에 만족하는 확률적 생성 모델로 구현된다.

현재 관찰 \mathbf{o} (RGB 이미지), 현재 관절 상태 \mathbf{s} , 최근 히스토리 \mathbf{h} , 그리고 계획($\hat{\mathbf{q}}^*$, \hat{y}_{act})가 주어졌을 때, 하위 모델은 짧은 지평의 제어 $\mathbf{u}_{1:T}$ (관절 변화와 그리퍼 신호)를 샘플링 한다:

$$\mathbf{u}_{1:T} \sim \pi_{\theta}(\cdot | \mathbf{o}, \mathbf{s}, \mathbf{h}, \hat{\mathbf{q}}^*, \hat{y}_{act}).$$

우리는 π_{θ} 를 확산 모델로 파라미터화하고 [2], 노이즈 예측 네트워크는 Transformer [3]로 구현하였다. 학습은 표준 노이즈 예측 손실로 수행하며, 추론 시에는 DDIM [4] 샘플러로 효율적으로 제어 시퀀스를 생성한다.

상위 네트워크와 마찬가지로 RGB 이미지, 최근 히스토리 그리고 현재 관절상태는 각각 $f_{img}, f_{hist}, f_{curr}$ 로 임베딩된다. 상위 모듈의 계획($\hat{\mathbf{q}}^*$, \hat{y}_{act})은 각각 ϕ_q, ϕ_{act} 로 투영된다. 최종 조건 벡터는

$$\mathbf{c} = [f_{img} \| f_{hist} \| f_{curr} \| \phi_{plan} \| \phi_{act}]$$

로 구성하고, 확산 모델의 노이즈 예측기 입력 토큰에 가산 바이어스로 주입한다(토큰 임베딩과 시간 임베딩에 각각 선형 결합).

하위 모델은 계획 신호로 전역 목표를 공유하면서, 확산 모델의 멀티모달 표현력을 통해 국소적 시각 변화와 접촉 잡음에 탄력적으로 대응한다. 상·하위 계층의 역할 분리는 조작에서 나타나는 불연속 구간(토클 등)과 그 외의 연속 구간이라는 이질적 특성을 구조적으로 반영하여, 전체 제어 파이프라인의 안정성과 일반화 성능을 동시에 높인다.

III. 실험 및 결과

실험은 PyBullet[5] 환경에서 Franka Panda 로봇으로 수행하였다. 과제는 머그컵 잡기 → 중간 배치 → 재잡기 → 최종 배치의 네 단계로 구성된다. 모든 에피소드는 컵이 옆으로 눕혀진 초기 자세에서 시작하며, 핵심은 적절한 접촉/토글 지점을 정확히 예측하고 이를 기준으로 동작을 계획·실행하는 데 있다. 상위 네트워크는 실시간 추론을 위해 경량화된 ResNet 기반으로 설계하여 각 시점의 토글(열기/닫기) 결정과 목표 관절 자세를 산출한다. 하위 확산 모델은 이 목표를 조건으로 짧은 지평의 관절 변화를 반복적으로 샘플링하고 집행한다. 또한 하위 모델에서는 매 주기 K개의 후보 궤적을 생성한 뒤, 충돌 위험이 낮고 예측된 접촉 지점과의 일치도가 높은 후보를 선택해 실행하는 best-of-K 규칙을 적용하였다.

그림 1은 제안한 방법으로 예측된 접촉/토글 지점(초록색 원)과 해당 지점으로부터 생성된 후보 궤적들을 보여준다. 궤적은 그리퍼 기준의 엔드이펙터 경로로 시각화했으며, 이들 중 실제로 선택되어 실행된 경로는 파란색으로 강조하였다. 선택된 파란색 경로가 다른 생성된 경로 대비 접촉/토글 지점을 향함을 알 수 있다.

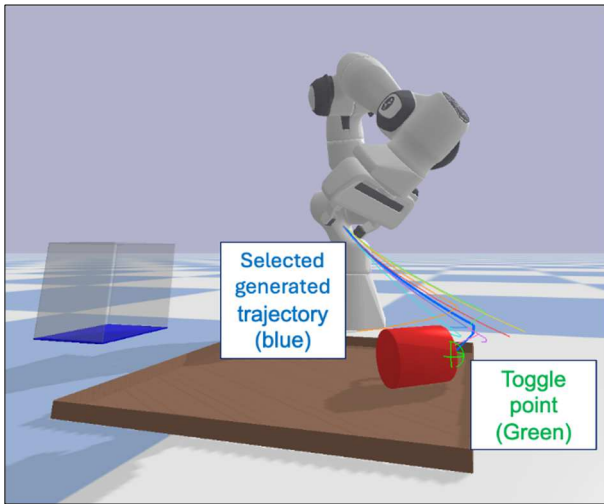


그림 1. 상위 모듈이 예측한 토글 지점(초록)과, 이를 조건으로 하위 확산 모델이 생성한 후보 경로들 중 선택된 궤적(파랑).

그림 2는 제안한 계층 제어의 머그컵 조작 시나리오를 완수하는 과정을 단계별로 보여준다. 먼저 (1) 상위 모듈이 눕혀진 머그에 대해 최초의 접촉/토글 지점을 예측하고, 하위 모듈이 그 지점을 향해 접근한다. 이어

서 (2) 예측 지점에 도달하자 그리퍼를 닫아 물체를 안정적으로 파지한다. 이후 (3) 다음 토글—실제론 내려놓기—지점을 향해 추종 경로를 실행하며 이동하고, (4) 해당 지점을 재예측해 미세 조정하며 접근한다. 이어 (5) 안정적인 파지 상태를 유지한 채 최종 배치 위치로 이동한 뒤, (6) 목표 위치에 정확히 배치함으로써 임무를 완료한다.

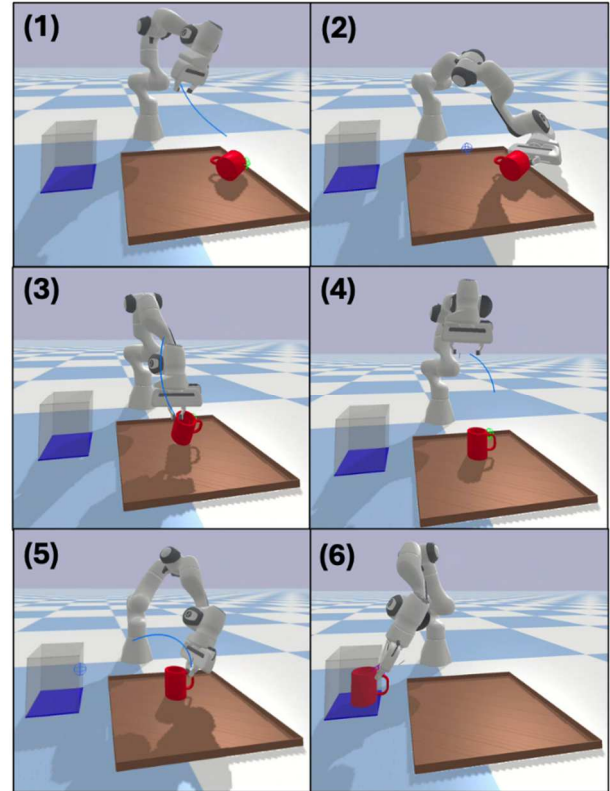


그림 2. 제안한 계층 제어에 따른 머그컵 조작의 단계별 실행 결과. (1) 최초 접촉/토글 지점 예측 및 접근, (2) 예측 지점에서 그리퍼 닫기(파지), (3) 다음 토글/배치 지점을 향한 이동, (4) 배치 지점 재예측 및 미세 접근, (5) 파지 상태로 최종 목적지로 이동, (6) 목표 위치에 정확히 배치하여 임무 완료.

계층 구조를 쓰지 않고 단일 확산 모델만으로 제어하는 비교모델과 제안한 계층 모델을 동일 조건에서 정량 비교하였다(표1). 머그컵 pick-place 시나리오를 각 200회 실행하여 작업 성공률을 측정했다. 비교모델은 특히 집기/놓기 전후의 중간 토글 지점에서 실패가 집중되는 반면, 제안 모델은 상위 모듈이 토글 시점과 목표 자세를 명시적으로 제공하여 사건 전후의 접근 이탈이 더욱 안정적이었다.

	제안된 모델	비교모델
임무성공률	91%	84%

표 1. 임무성공률 비교

IV. 결론 및 향후 연구 방향

본 연구는 토글(접촉) 사건을 상위에서 예측하고 하위에서 확산 모델로 추종·집행하는 계층형 제어기를 제안했지만, 보다 다양한 시나리오에서의 체계적 검증이 필요하다. 단일 머그업·고정된 조명·단순 배치 조건을 넘어, 복잡한 상황/임무에서의 성능을 평가할 필요가 있다. 또한 “이벤트” 자체—즉 그리퍼 개폐와 접촉 전후의 미세 접근/이탈—를 더 안정적으로 수행하기 위한 장치가 요구된다. 이러한 확장과 검증을 통해, 제안한 계층형 구조가 다양한 조작 임무에서도 안정적이고 재현성 있는 이벤트 수행을 달성함을 보일 수 있을 것이다.

사사

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임(IITP-2025-RS-2023-00259678)

참고문헌

- [1] J. Ho, A. N. Jain, and P. Abbeel, “Denoising Diffusion Probabilistic Models,” *Advances in Neural Information Processing Systems* (NeurIPS), vol. 33, 2020.
- [2] C. Chi, Z. Xu, S. Feng, E. Cousineau, Y. Du, B. Burchfiel, R. Tedrake, and S. Song, “Diffusion Policy: Visuomotor Policy Learning via Action Diffusion,” *The International Journal of Robotics Research* (IJRR), early access, 2024.
- [3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention Is All You Need,” in *Proc. Adv. Neural Inf. Process. Syst.* (NeurIPS), vol. 30, pp. 5998–6008, 2017.
- [4] J. Song, C. Meng, and S. Ermon, “Denoising Diffusion Implicit Models,” *International Conference on Learning Representations* (ICLR), 2021.
- [5] E. Coumans and Y. Bai, “PyBullet: A Python Module for Physics Simulation for Games, Robotics and Machine Learning,” available at pybullet.org, 2016–present.