# 방해 요소 환경에서의 강건한 잠재 행동 학습: 변분 정보 병목과 SIGReg의 결합

**신성국[1], 김기응**

**한국과학기술원**

**{skshyn, kekim}@kaist.ac.kr**

# Robust Latent Action Learning under Distractors: Combining Variational Information Bottleneck and SigReg

**요 약**

행동 레이블이 없는 대규모 비디오 데이터를 활용한 로봇 에이전트 사전학습에서 잠재 행동 모델(Latent Action Model, LAM)은 핵심적인 역할을 한다. 그러나 실세계 비디오의 방해 요소(distractor) 환경에서 기존 LAM은 잡음 방해(noise distraction) 및 표현 붕괴(representation collapse) 문제에 취약하다. 본 논문에서는 기존 LAM 아키텍처에 변분 정보 병목(VIB)과 SIGReg를 결합하여 두 문제를 동시에 완화하는 방법을 제안하고, Distracting Control Suite의 walker-run 환경에서 그 유효성을 검증한다.

## 1. 서 론

최근 행동 레이블(action label)이 없는 대규모 비디오 데이터를 활용한 로봇 에이전트 사전학습이 주목받고 있으며, 이를 위한 핵심 기술로 잠재 행동 모델(Latent Action Model; LAM)이 제안되었다[1]. 잠재 행동 모델은 연속된 관측 간의 변화를 잠재 행동(latent action)으로 압축하여 모방 학습에 활용한다.

그러나 기존 잠재 행동 모델(LAM)은 두 가지 독립적인 문제에 취약하다. 첫째, 실세계 비디오에 포함된 움직이는 배경, 카메라 흔들림 등 행동과 상관된 방해 요소(distractor)로 인해 잠재 행동 모델의 역동역학 모델(Inverse Dynamics Model; IDM)이 실제 행동보다 방해 요소의 변화를 더 쉽게 설명하게 되어, 제어와 무관한 잡음을 잠재 행동으로 인코딩하는 잡음 방해(noise distraction) 문제가 발생한다[2]. 둘째, self-supervised 학습 과정의 최적화 불안정성으로 인해 상태 임베딩 및 잠재 행동 임베딩이 특정 상수 값으로 수렴하는 표현 붕괴(representation collapse)[3]가 나타나며, 이 경우 역동역학 모델(IDM)이 의미 있는 잠재 행동을 추출하지 못해 잠재 행동의 품질이 급격히 저하된다.

본 논문에서는 이러한 두 가지 문제를 해결하기 위해 기존 잠재 행동 모델 구조 중 LAOM[2]에서 변분 정보 병목(Variational Information Bottleneck; VIB)[4]과 SIGReg(Sketched Isotropic Gaussian Regularization)[3]를 결합한 개선 방법을 제안한다. 실험을 통해 제안 방법이 방해 요소 환경에서 잠재 행동의 품질과 하위 과제(downstream task) 성능을 유의미하게 향상시킴을 보인다.

## 2. 잠재 행동 모델 아키텍처

1) 이 논문은 2026년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.RS-2020-II200940, 안전한 강화학습 원천 기술 개발 및 자연어 처리에의 응용)

본 논문에서 제안하는 모델은 [1]의 잠재 행동 모델 3단계 학습 파이프라인을 따른다.

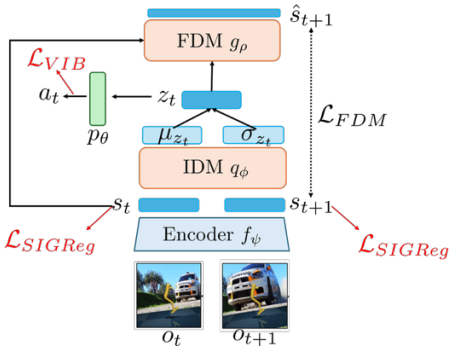


그림 1. LAOM[2]에서 VIB+SIGReg 통합 아키텍처

1. 전체 데이터셋 $\mathcal{D}$를 활용하여 잠재 행동 모델(LAM) 사전학습
2. 사전학습된 잠재 행동 모델로 $\mathcal{D}$ 전체를 잠재 행동으로 재레이블링하고, 이를 바탕으로 행동 복제(Behavioral Cloning; BC) 모델 학습
3. 행동 레이블이 있는 소수 데이터셋 $\mathcal{D}_L$을 활용하여 잠재 행동에서 실제 행동으로의 디코더 Fine-tuning 학습

여기서 $\mathcal{D} = \mathcal{D}_L \cup \mathcal{D}_U$로 정의하며, $\mathcal{D}_U$는 행동 레이블이 없는 대규모 데이터셋, $\mathcal{D}_L \subset \mathcal{D}$는 소수의 행동 레이블이 제공되는 데이터셋을 나타낸다. 본 논문은 이 파이프라인을 그대로 유지하되, 방해 요소 환경에서 잠재 행동의 품질이 이후 모든 단계의 성능을 결정한다는 점에 주목하여 **1단계 잠재행동 모델 사전학습을 개선**하는 것을 핵심 기여로 삼는다.

이를 위해 본 논문에서 제안하는 모델은 LAOM [2]의 아키텍처를 기반으로 하며, 공유 인코더 $f_\psi$, 역동역학

모델(Inverse Dynamics Model; IDM) $q_\phi$, 순동역학 모델(Forward Dynamics Model; FDM) $g_\rho$로 구성된다. 인코더는 연속된 관측 $o_t, o_{t+1}$를 상태 임베딩 $s_t = f_\psi(o_t)$, $s_{t+1} = f_\psi(o_{t+1})$로 변환하고, IDM은 두 임베딩으로부터 잠재 행동 $z_t$를 추출하며, FDM은 $s_t$와 $z_t$로부터 다음 상태 임베딩 $\hat{s}_{t+1}$을 예측한다. 여기에 서론에서 언급한 두 가지 문제를 해결하기 위해 다음의 구성 요소를 추가로 도입한다.

먼저, 기존 LAOM[2]의 역동역학 모델(IDM)이 $z_t$를 결정론적으로 출력하는 것과 달리, 본 논문에서는 평균 $\mu_{z_t}$와 로그 분산 $\log \sigma_{z_t}^2$를 출력하고 reparameterization trick으로 $z_t = \mu_{z_t} + \sigma_{z_t} \odot \epsilon$, $\epsilon \sim \mathcal{N}(0, \mathbf{I})$를 샘플링하는 변분 추론 형태를 채택한다. 이를 통해 KL divergence 페널티로 $z_t$가 제어와 무관한 정보를 과도하게 인코딩하지 않도록 압축을 유도한다.

또한, LeJEPA [3]에서 제안된 SIGReg(Sketched Isotropic Gaussian Regularization)를 인코더 출력에 적용한다. SIGReg는 배치 내 상태 임베딩의 경험적 특성 함수를 $\mathcal{N}(0, \mathbf{I})$의 특성 함수에 일치시켜 임베딩 분포가 등방성 가우시안을 따르도록 유도하며, stop-gradient 및 EMA 인코더 없이도 표현 붕괴를 방지할 수 있어 $\mathcal{L}_{\text{FDM}}$의 타겟 임베딩에도 역전파를 허용한다.

## 3. 최종 목적식

본 절에서는 제안 모델의 학습 목적식을 구성하는 세 가지 손실 함수를 기술한다. 이하에서 $a_t$는 $\mathcal{D}_L$에서 제공되는 실제 행동, $q_\phi(z_t \mid s_t, s_{t+1})$는 역동역학 모델(IDM)이 출력하는 잠재 행동의 변분 사후분포로 $\mathcal{N}(\mu_\phi, \sigma_\phi^2)$로 모델링되며, $p_\theta(a_t|z_t)$는 1단계 잠재 행동 모델 사전학습 시 $\mathcal{D}_L$의 레이블을 활용하여 잠재 행동으로부터 실제 행동을 예측하는 supervision 디코더로, 1단계 학습에서만 활용된다.

**변분 정보 병목 손실.** 잠재 행동 학습의 목표를 변분 정보 병목(Variational Information Bottleneck; VIB) 관점[4]에서 형식화하면 $\max_{z_t} I(a_t; z_t) - \beta \cdot I(s_t, s_{t+1}; z_t)$로 표현된다. 첫 번째 항은 $z_t$가 실제 행동 $a_t$에 대한 정보를 보존해야 함을, 두 번째 항은 $z_t$가 방해 요소를 포함한 관측 전체에 과도하게 의존하지 않도록 압축을 유도한다. 여기서 $\beta$는 잠재 행동의 압축 강도를 의미하는 하이퍼파라미터다. 각 상호정보량$(I(a_t; z_t), I(s_t, s_{t+1}; z_t))$을 tractable한 형태로 근사하면, 변분 정보 병목 손실은 다음과 같이 유도된다:

$$\mathcal{L}_{VIB} = -\mathbb{E}_{q_\phi(z_t|s_t, s_{t+1})}[\log p_\theta(a_t|z_t)] + \beta \cdot D_{KL}(q_\phi(z_t|s_t, s_{t+1}) || \mathcal{N}(0, \mathbf{I}))$$

LAOM[2]은 소수의 행동 레이블($a_t$)이 있는 데이터 $\mathcal{D}_L$을 활용하므로 첫 번째 항을 프록시 없이 직접 최적화할 수 있으며, $q_\phi$가 가우시안 분포인 경우 KL divergence는 closed-form으로 계산된다. 인코더 $f_\psi$는 별도의 손실 없이 $\mathcal{L}_{\text{VIB}}$를 포함한 전체 목적식을 통해 통합 학습된다.

**표현 붕괴 방지 손실.** Self-supervised 학습 과정에서 인코더 $f_\psi$의 출력인 상태 임베딩 $s_t$가 특정 상수 값으로 수렴하는 표현 붕괴 문제를 방지하기 위해, LeJEPA[3]는 SIGReg(Sketched Isotropic Gaussian Regularization)를 도입한다. SIGReg는 배치 내 상태 임베딩 $\{s_i\}_{i=1}^N$의 경험적 특성 함수(empirical characteristic function)를 표준 정규분포 $\mathcal{N}(0, \mathbf{I})$의 특성 함수에 일치시킴으로써, 임베딩 분포 전체가 등방성 가우시안을 따르도록 유도하여 특정 상수로의 수렴을 억제한다:

$$\mathcal{L}_{SIGReg} = || \frac{1}{N} \sum_{i=1}^{N} e^{i\omega^T s_i} - e^{-\frac{1}{2}||\omega||^2} ||^2$$

여기서 $\omega$는 랜덤 주파수 샘플 벡터, $s_i$는 인코더 $f_\psi$의 출력 상태 임베딩, $N$은 배치 크기를 나타낸다. 또한 LeJEPA [3]에서 보인 바와 같이 SIGReg는 stop-gradient 및 EMA 인코더 없이도 표현 붕괴를 방지할 수 있으며, 이를 본 논문에서는 $\mathcal{L}_{\text{FDM}}$의 타겟 임베딩에도 역전파를 허용하는 근거로 활용한다.

**순동역학 모델 손실.** 순동역학 모델(FDM)은 인코더 $f_\psi$가 추출한 상태 임베딩 $s_t$와 잠재 행동 $z_t$로부터 잠재 공간에서 다음 상태 임베딩을 예측한다. 기존 LAOM[2]에서는 타겟 임베딩에 stop-gradient 연산자 $\text{sg}(\cdot)$ 또는 EMA 인코더를 적용하였다. 그러나 본 논문에서는 SIGReg를 정규화 항으로 도입함으로써 임베딩의 붕괴를 직접적으로 억제할 수 있으므로, stop-gradient 없이 다음과 같이 순동역학(FDM) 손실을 구성한다.

$$\mathcal{L}_{FDM} = \left|\left| g_\rho(s_t, z_t) - s_{t+1} \right|\right|^2$$

이를 통해 타겟 임베딩 $s_{t+1} = f_\psi(o_{t+1})$에도 역전파가 가능해져 인코더 $f_\psi$가 보다 풍부한 학습 신호를 받을 수 있으며, 표현 붕괴 방지는 $\mathcal{L}_{SigReg}$가 전담한다.

**통합 목적식.** 세 손실 함수를 통합한 최종 목적식은 다음과 같다:

$$\mathcal{L}_{total} = \mathcal{L}_{FDM} + \lambda \cdot \mathcal{L}_{VIB} + \gamma \cdot \mathcal{L}_{SigReg}$$

여기서 $\lambda$는 변분 정보 병목 손실의 가중치, $\gamma$는 표현 붕괴 방지의 강도를 조절하는 하이퍼파라미터다. 인코더 $f_\psi$, IDM $q_\phi$, FDM $g_\rho$는 $\mathcal{L}_{\text{total}}$을 통해 end-to-end로 통합 학습되며, 세 손실의 결합을 통해 방해 요소 환경에서의 잡음 방해 및 표현 붕괴 문제를 동시에 완화한다.

## 4. 실험 결과

본 논문에서는 Distracting Control Suite [5]의 walker-run 환경에 배경 비디오 방해 요소(distractor)만을 적용한 설정을 사용한다. 구체적으로, 배경 방해 요소는 DAVIS2017[6] 데이터셋에서 무작위로 샘플링된 자연 영상 클립을 walker 관절 영역 외 배경 픽셀에 합성하는 방식으로 생성된다. 데이터셋은 LAOM[2]의 설정을 따라 전문가 정책으로부터 수집된 오프라인 데이터로 구성되며, 행동 레이블이 없는 $\mathcal{D}_U$ 5,000개의 trajectory와 레이블이 제공되는 $\mathcal{D}_L$ 125개의

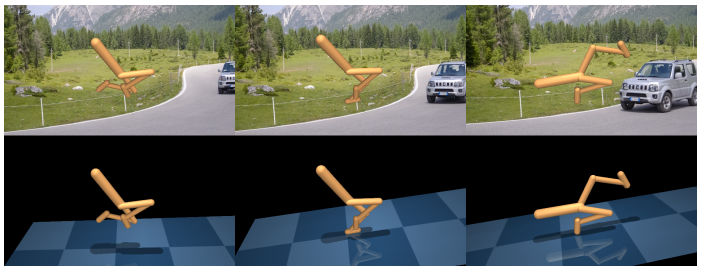

그림 2. 위) walker-run 환경에 방해 요소(distractor)가 삽입된 환경. 아래) 방해 요소(distractor)가 없는 walker-run 환경

trajectory로 이루어진 전체 데이터셋 $\mathcal{D} = \mathcal{D}_L \cup \mathcal{D}_U$를 활용한다.

모델 평가는 2절에서 기술한 3단계 학습 파이프라인을 모두 거친 최종 행동 모델을 대상으로 수행한다. DCS의 walker-run 환경에서 해당 행동 모델을 에피소드당 최대 1,000 스텝 동안 시뮬레이션하여 보상(return)을 수집한다. 평가 시, 10개 에피소드의 누적 보상(return) 평균값을 산출하고 이를 정규화하여 성능 지표로 보고한다. 정규 보상 범위는 [0, 1]이며, 높을수록 성능이 우수함을 의미한다.

표 1 두 방법의 walker-run 환경에서 성능 비교

| Method | Normalized Return($\pm$ Std) |
|---|---|
| LAOM[2] | $0.4770 \pm 0.1825$ |
| LAOM | |
| +VIB | $\mathbf{0.5318} \pm 0.0525$ |
| +SigReg | |

표 1에 제시된 실험 결과에 따르면, LAOM 아키텍처에 변분 정보 병목(VIB)과 SIGReg(Sketched Isotropic Gaussian Regularization)를 결합한 제안 방법이 기존 LAOM 대비 walker-run 환경에서 소폭의 성능 향상을 달성하였다. 이는 방해 요소 환경에서 변분 정보 병목(VIB)를 통한 잠재 행동 압축과 SIGReg를 통한 상태 임베딩 정규화가 실질적인 효과를 지님을 시사한다.

다만 성능 향상의 폭이 제한적인 것은, 배경 방해 요소가 잠재 행동 학습에 미치는 영향이 단일 정규화 기법만으로 완전히 제거되기 어렵다는 점에서 기인한다. 그럼에도 불구하고 본 결과는 두 가지 측면에서 의미 있는 발전 가능성을 보여준다. 첫째, VIB의 KL divergence 페널티는 기존 LAOM에서 발생하는 잡음 방해 문제, 즉 역동역학 모델(IDM)이 제어와 무관한 배경 변화를 잠재 행동으로 인코딩하는 현상을 억제하는 방향으로 작용한다. 둘째, SIGReg는 상태 임베딩이 특정 상수로 수렴하는 표현 붕괴 문제를 명시적으로 방지함으로써, 인코더가 보다 풍부한 학습 신호를 받을 수 있는 기반을 마련한다.

향후 잠재 행동의 시각화 등 정성적 분석을 병행하면 VIB(Variational Information Bottleneck) 및 SIGReg(Sketched Isotropic Gaussian Regularization)가 각각 잡음 방해와 표현 붕괴 억제에 실질적으로 기여함을 보다 직관적으로 보여줄 수 있다. ablation study 및 추가 백본 구조, 추가 환경과 데이터 조합을 통해 제안 방법의 효과가 더욱 뚜렷하게 나타날 것으로 기대된다.

## 5. 결 론

본 논문에서는 LAOM 아키텍처에 변분 정보 병목(Variational Information Bottleneck; VIB)와 SIGReg (Sketched Isotropic Gaussian Regularization)를 결합하여 방해 요소 환경에서의 잡음 방해 및 표현 붕괴 문제를 해결하는 방법을 제안하였다. DCS(Distracting Control Suite)의 walker-run 환경에서의 실험을 통해 제안 방법이 방해 요소 존재 하에서 잠재 행동의 품질과 하위 과제 성능을 향상시킴을 검증하였다.

향후 연구로는 두 가지 방향의 확장을 고려할 수 있다. 첫째, 제안 방법을 대규모 비디오 데이터 기반의 generalist policy 사전학습에 접목함으로써 배경 편향 없이 강건한 잠재 행동을 학습하고, 레이블이 부족한 실세계 로봇 환경에서도 효과적으로 적용 가능한 범용 정책 학습으로 나아갈 수 있다. 둘째, 최근 LeWM [7]과 WAM[8]의 등장으로 JEPA 기반의 로봇 월드 모델과 월드 행동 모델(WAM)이 로봇 환경에서 주목받고 있는 만큼, 본 논문에서 제안한 SIGReg 기반의 표현 안정화 기법을 LeJEPA 기반 잠재 행동 World Action Model로 확장하는 것을 고려할 수 있다. 이를 통해 방해 요소에 강건한 잠재 행동 표현과 안정적인 World Model 학습을 동시에 달성할 수 있을 것으로 기대된다.

## 참고문헌


[1] D. Schmidt and M. Jiang, "Learning to Act without Actions," in Proc. Int. Conf. Learn. Representations (ICLR), 2024.

[2] A. Nikulin et al., "Latent Action Learning Requires Supervision in the Presence of Distractors," in Proc. 42nd Int. Conf. Mach. Learn. (ICML), 2025.

[3] R. Balestriero and Y. LeCun, "LeJEPA: Provable and Scalable Self-Supervised Learning without Heuristics," arXiv:2511.08544, 2025.

[4] A. A. Alemi, et al., "Deep Variational Information Bottleneck," in Proc. Int. Conf. Learn. Representations (ICLR), 2017.

[5] A. Stone, et al., "The Distracting Control Suite - A Challenging Benchmark for Reinforcement Learning from Pixels," arXiv:2101.02722, 2021.

[6] J. Pont-Tuset, F. Perazzi, S. Caelles, P. Arbelaez, A. Sorkine-Hornung, and L. Van Gool. The 2017 davis challenge on video object segmentation. arXiv:1704.00675, 2017.

[7] L. Maes, Q. Le Lidec, D. Scieur, Y. LeCun, and R. Balestriero, "LeWorldModel: Stable End-to-End Joint-Embedding Predictive Architecture from Pixels," arXiv:2603.19312, 2026.

[8] S. Ye, Y. Ge, et al., "World Action Models are Zero-shot Policies," arXiv:2602.15922, 2026.