

최신 영상 합성 및 Physical AI 기술 소개

KAIST 김재철시대학원 주재걸 교수 연구실 (DAVIAN)

연사: 박민호 박사과정

최신 영상 합성 및 Physical AI 기술 소개

주재걸 교수 연구실에서 수행하는 영상 합성 및 Physical AI 연구

- 이미지/비디오 합성
 - SphereDiff 360° 파노라마 영상 생성.
 - EgoX 3인칭 영상으로부터 1인칭 영상 합성.
 - CA-LoRA 이미지 생성 모델을 활용한 영역 분할 데이터셋 생성.
- Physical AI
 - ACG 로봇 흔들림 제어 알고리즘 개발.
 - 3D HAMSTER 3D Keypoint를 활용한 로봇 조작.
 - PHUMA 비디오 영상으로부터 휴머노이드 행동 데이터셋 구축.

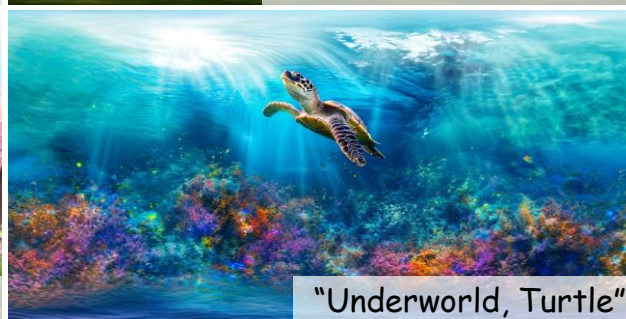
최신 영상 합성 기술 소개

KAIST 김재철시대학원 주재걸 교수 연구실 (DAVIAN)

360° 파노라마 영상 생성

추가 학습 없는 360° 파노라마 이미지/비디오 생성 방법론

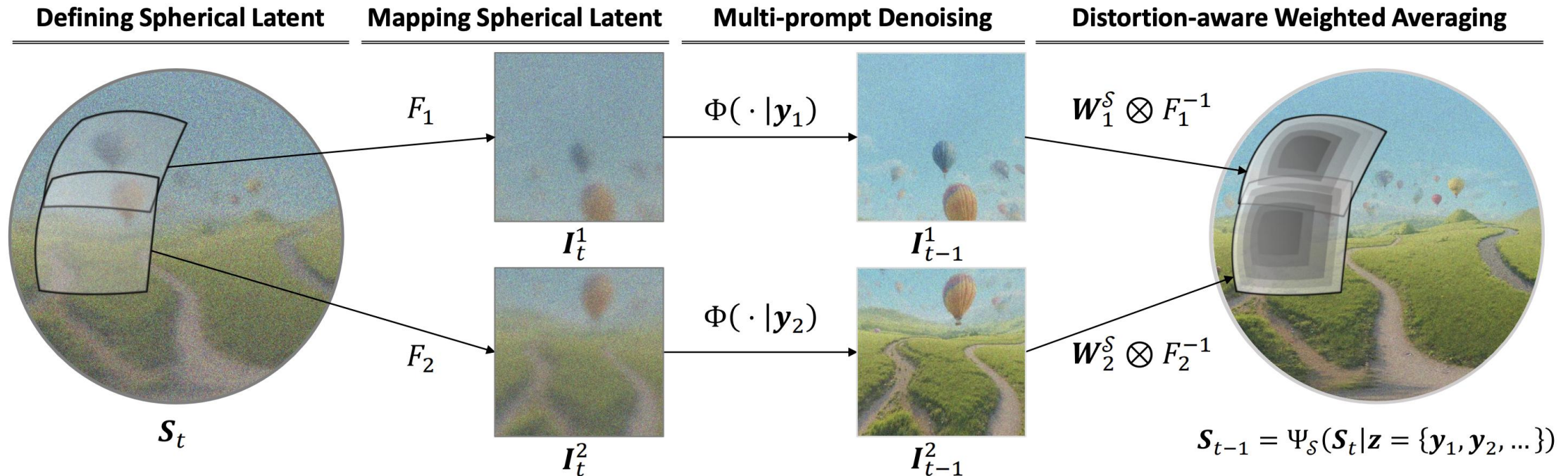
- 일반 이미지/동영상만 생성하던 모델을 활용하여, 추가 학습 없이 AR/VR 기기 등에 활용될 수 있는 360° 파노라마 영상을 생성할 수 있다.



360° 파노라마 영상 생성

추가 학습 없는 360° 파노라마 이미지/비디오 생성 방법론

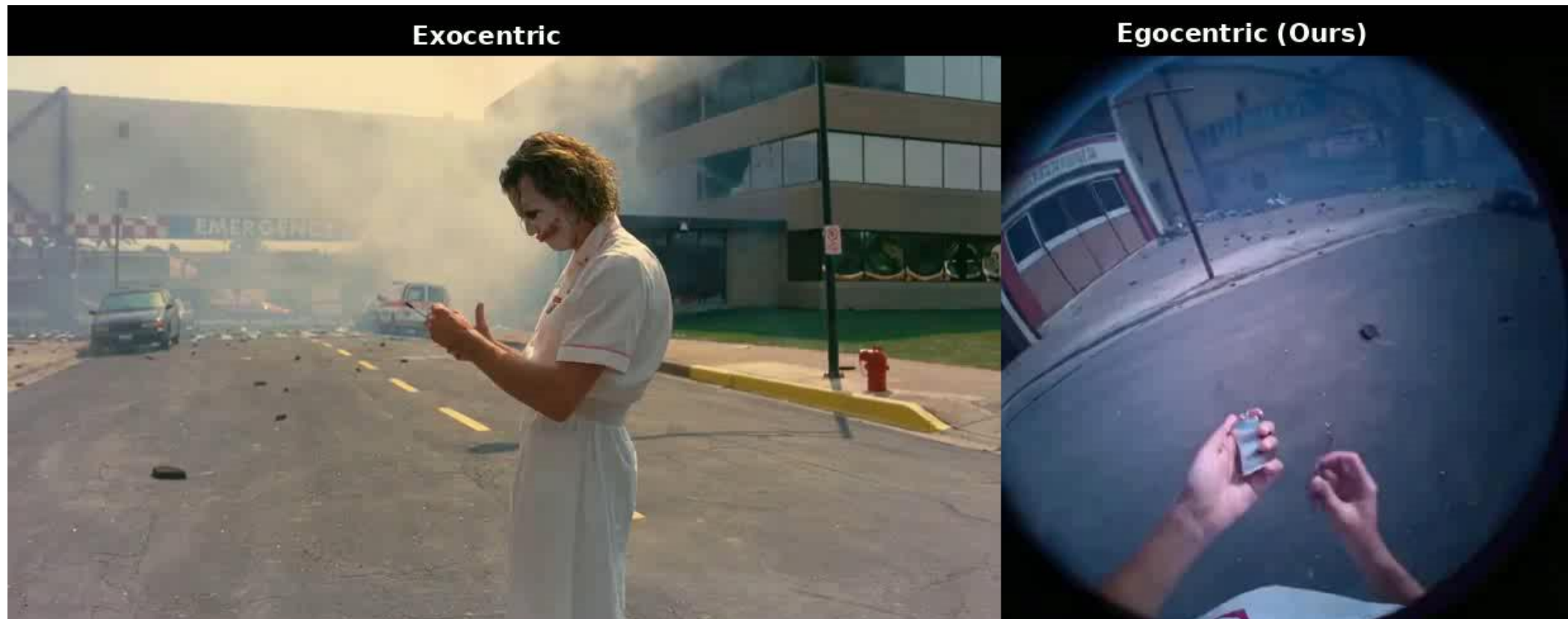
- 4단계 프레임워크:
 1. 구면 잠재 표현 정의.
 2. 구면 잠재 표현을 평면으로 매핑.
 3. 사전학습된 디퓨전 모델과 여러 프롬프트를 활용하여 노이즈 제거.
 4. 왜곡을 고려한 가중 합을 통한 구면으로의 재매핑.



3인칭 영상으로부터 1인칭 영상 합성

임의의 3인칭 영상을 1인칭으로 합성할 수 있는 학습 방법 및 모델 공개

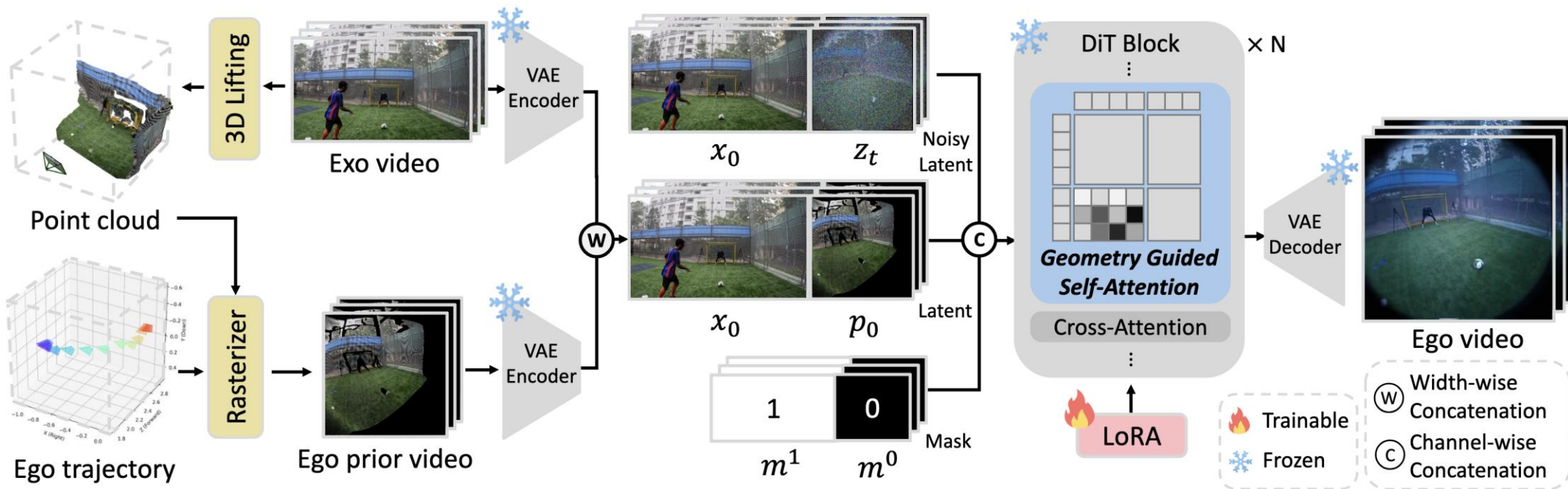
- 3인칭 영상으로부터 3D 정보를 추출 및 활용하여 1인칭 영상으로 합성할 수 있다.



3인칭 영상으로부터 1인칭 영상 합성

임의의 3인칭 영상을 1인칭으로 합성할 수 있는 학습 방법 및 모델 공개

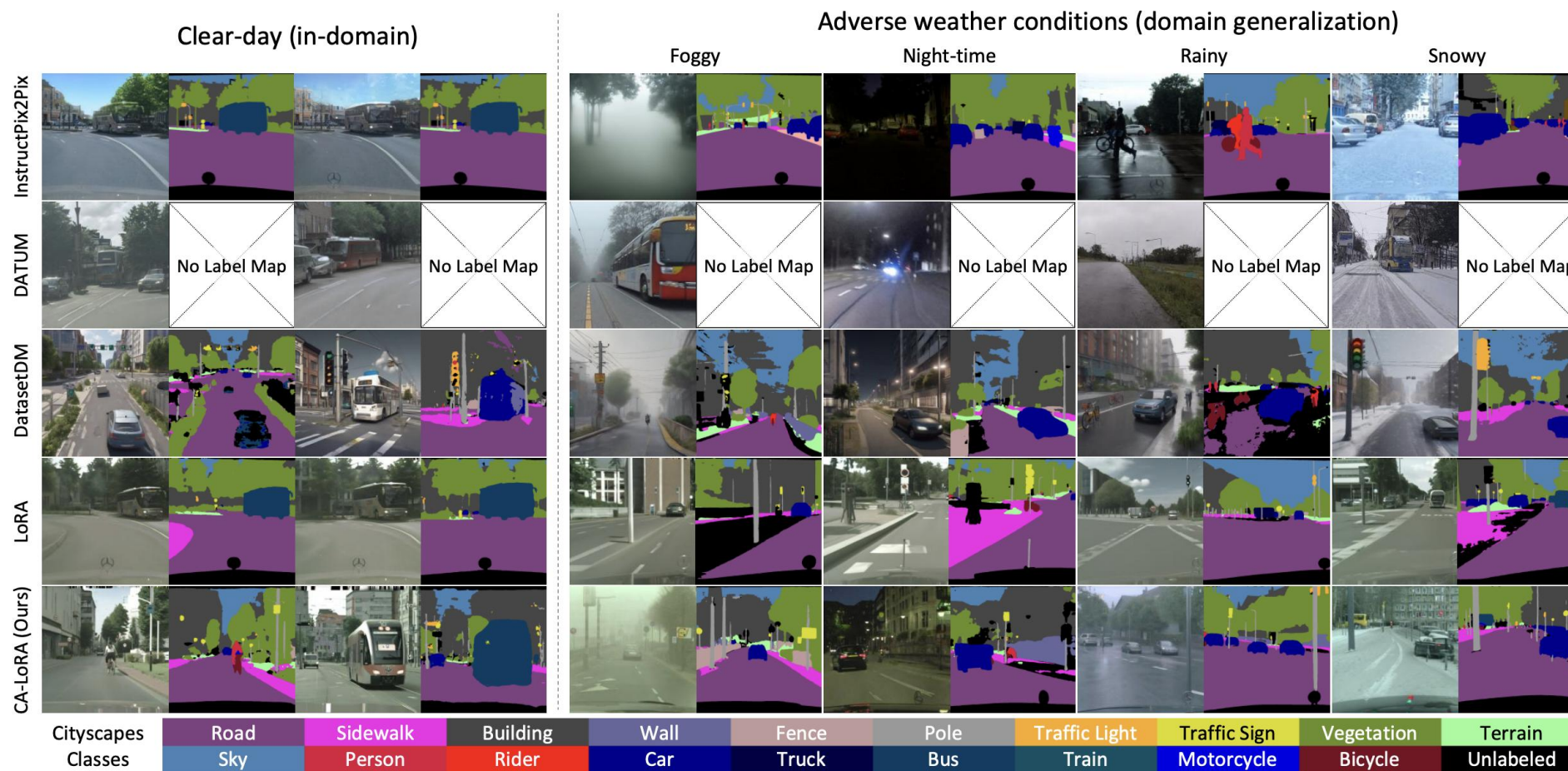
1. 공개된 대규모 3인칭-1인칭 paired 데이터셋 (EgoExo) 활용하여 학습.
2. 3인칭 비디오를 3D로 재건한 뒤, 1인칭에서 볼 각도로 투영하여 힌트로 제공.
3. 3D 상에서 가까이 있는 픽셀에 더 많은 가중치가 걸리도록 self-attention layer를 변형.



이미지 생성 모델을 활용한 영역 분할 데이터셋 생성

최신 이미지 생성 모델의 잠재 표현을 적극 활용하여 영역 분할 데이터셋 생성 및 성능 향상.

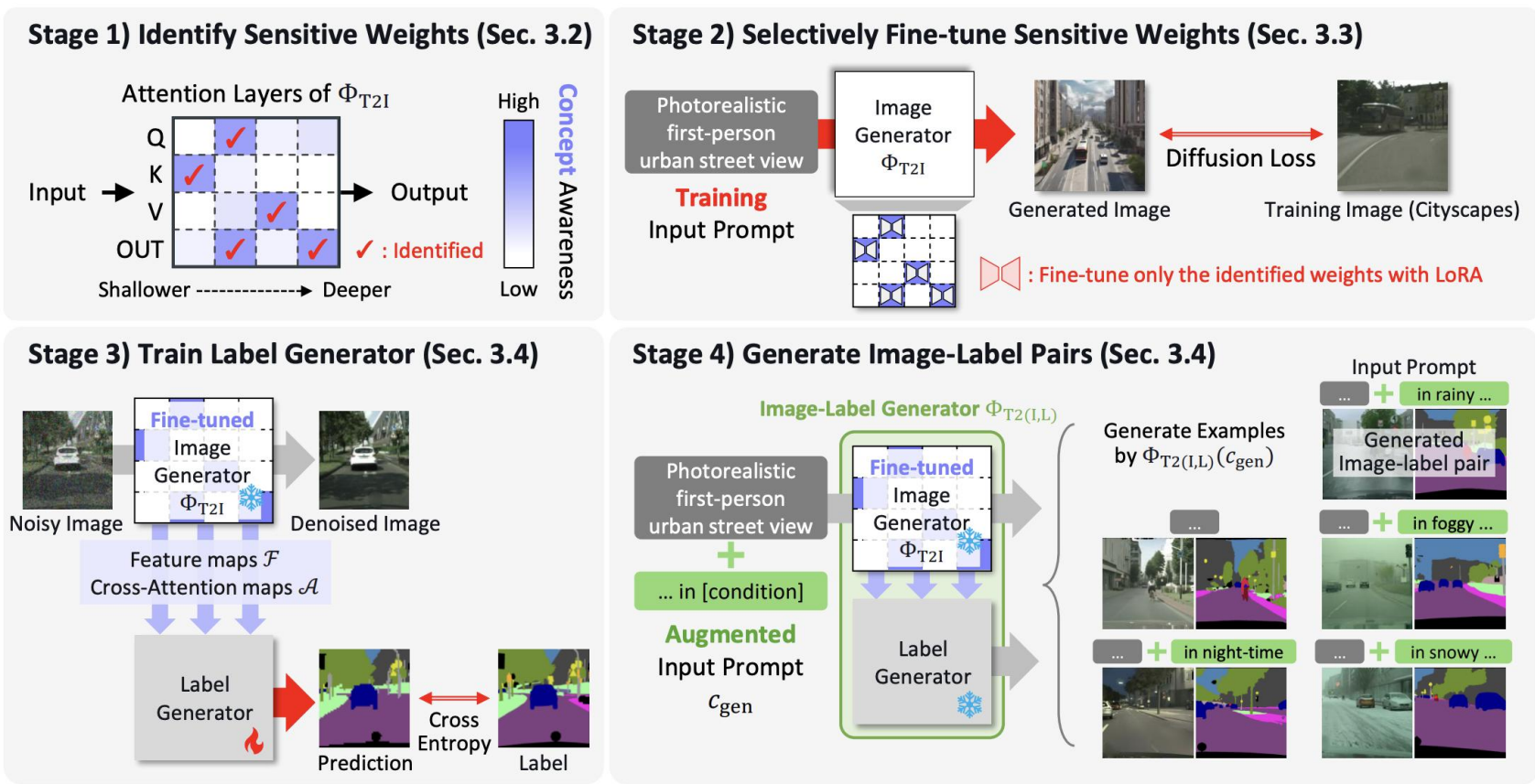
- 이미지만 생성하는 것이 아닌, 영역 분할에 사용할 수 있는 레이블 동시 생성.



이미지 생성 모델을 활용한 영역 분할 데이터셋 생성

최신 이미지 생성 모델의 잠재 표현을 적극 활용하여 영역 분할 데이터셋 생성 및 성능 향상.

- 이미지와 레이블을 동시 생성하기 위한 프레임워크 구축.
- 이미지의 스타일을 다양화 시키면서도 자율주행 시 필요한 카메라 시점을 유지하는 학습 방법론 제안.



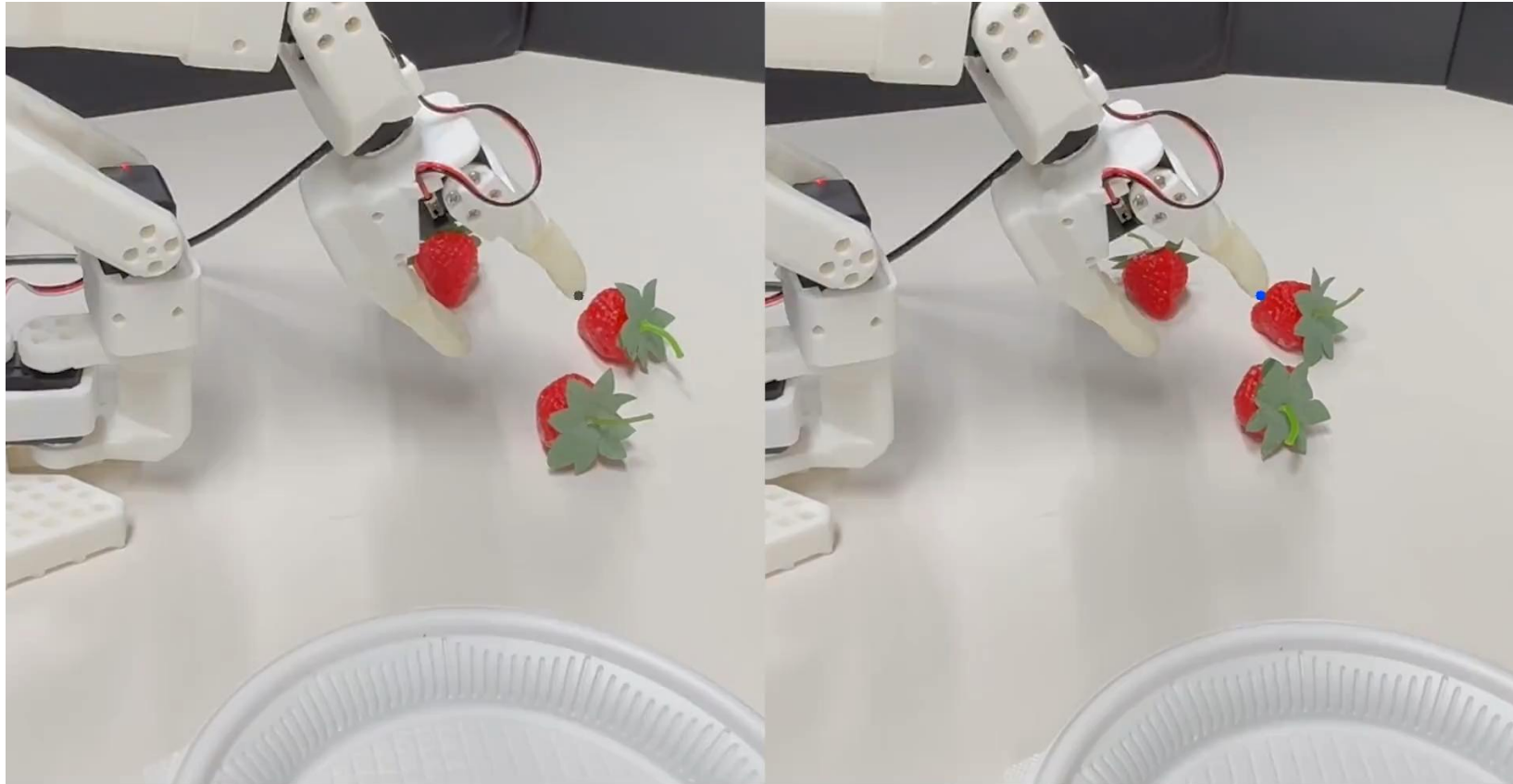
Physical AI 기술 소개

KAIST 김재철 AI대학원 주재걸 교수 연구실 (DAVIAN)

로봇 흔들림 제어 알고리즘 개발

Flow-based VLA에서 추가 학습 없이 사용할 수 있는 떨림 개선 알고리즘 개발.

- 저렴한 로봇(e.g., SO-101)은 로봇 조작을 할 때 떨림이 굉장히 심한데, 이를 완화하는 알고리즘 개발.



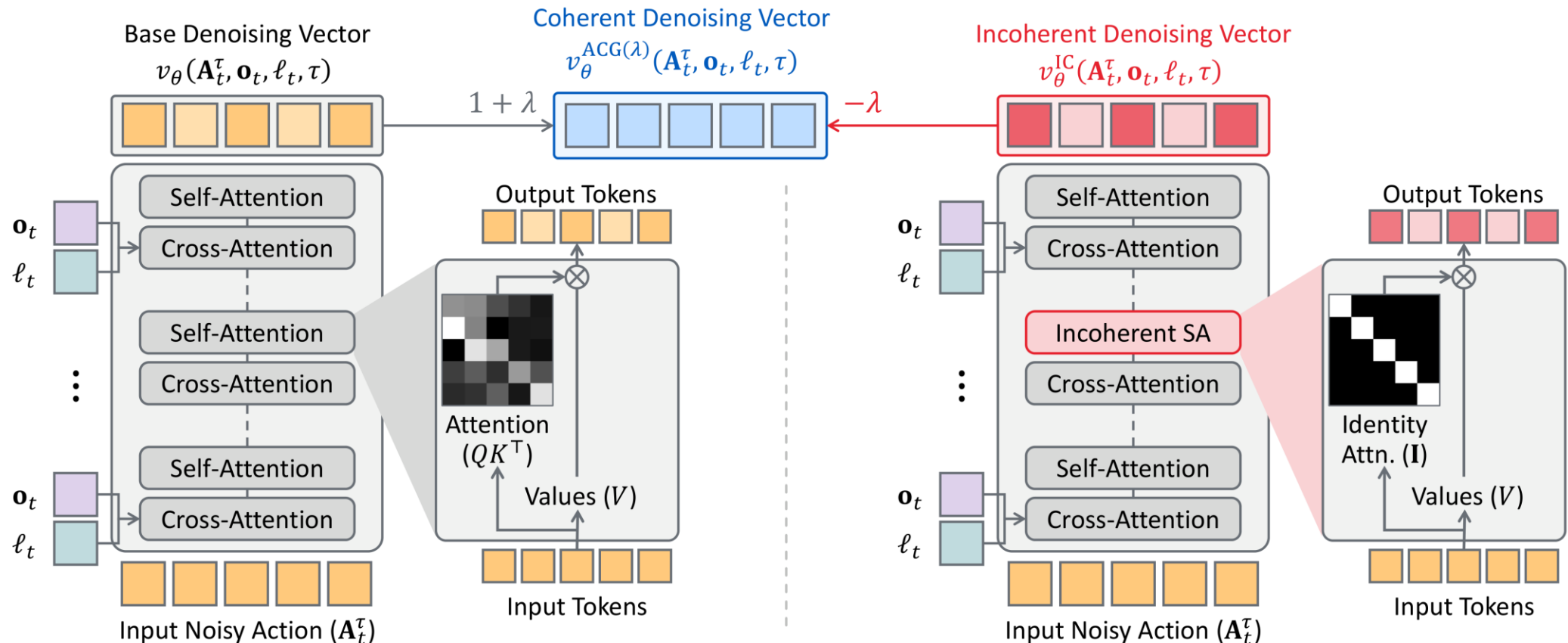
Vanilla GR00T-N1-2B

w/ ACG (Ours)

로봇 흔들림 제어 알고리즘 개발

Flow-based VLA에서 추가 학습 없이 사용할 수 있는 떨림 개선 알고리즘 개발.

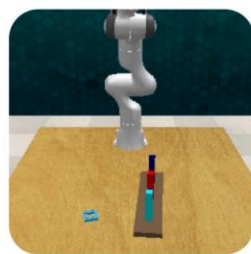
- 더욱 불안정한 방식으로 로봇 동작을 생성하는 디노이징 방향을 만들고, 이를 반면교사 삼아서 로봇 동작을 생성.



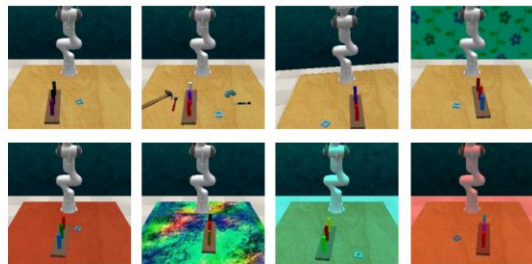
3D Keypoint를 활용한 로봇 조작

일반화 성능 향상을 위한 3D Keypoint를 생성 및 활용 방법론 제안.

- 학습과 테스트 간 분포 차이가 굉장히 클 경우에도 잘 작동하는 로봇 제어 방법론 개발.



(a) Colosseum



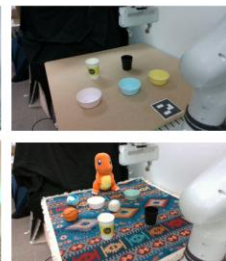
Evaluation variations



(b) Franka Panda



Button pressing



Pouring



Pick-and-place

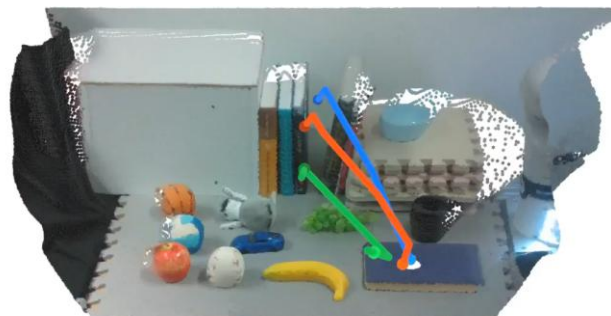
● 3D HAMSTER (Ours)

● RoboBrain 2.5

● Gemini 3.0 Pro



Pick up the blue car and place it in the basket



Pick up the lying book and insert it into the gap between the other books

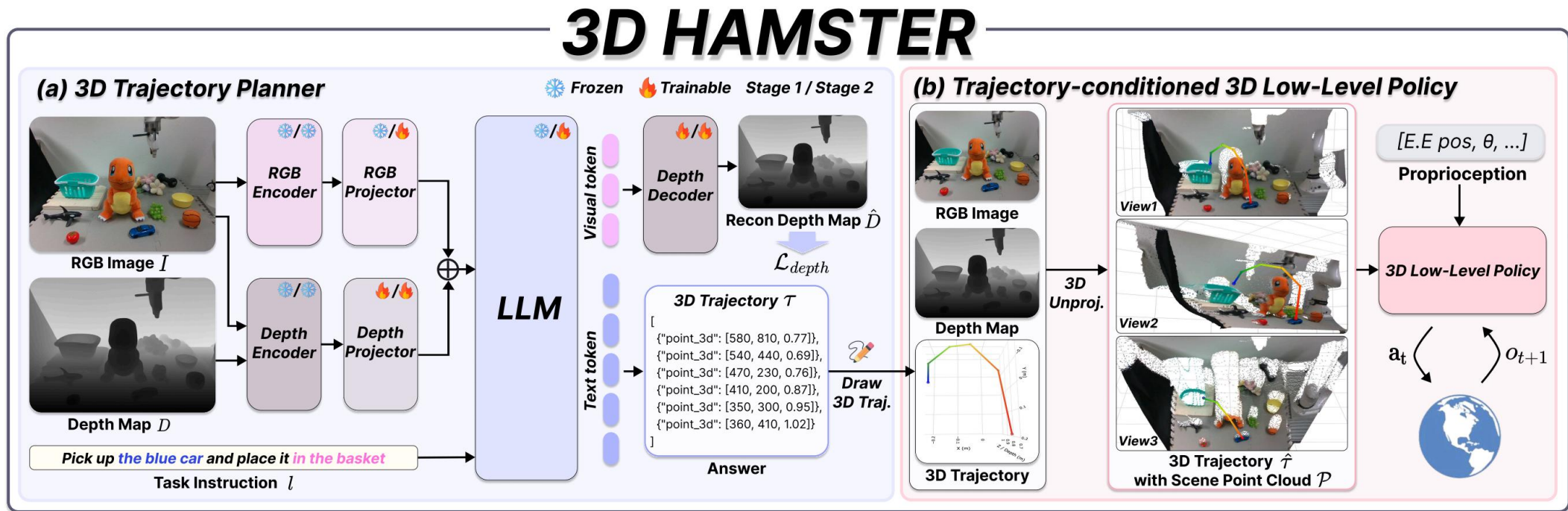


Pick up the banana and place it on the white mat

3D Keypoint를 활용한 로봇 조작

일반화 성능 향상을 위한 3D Keypoint를 생성 및 활용 방법론 제안.

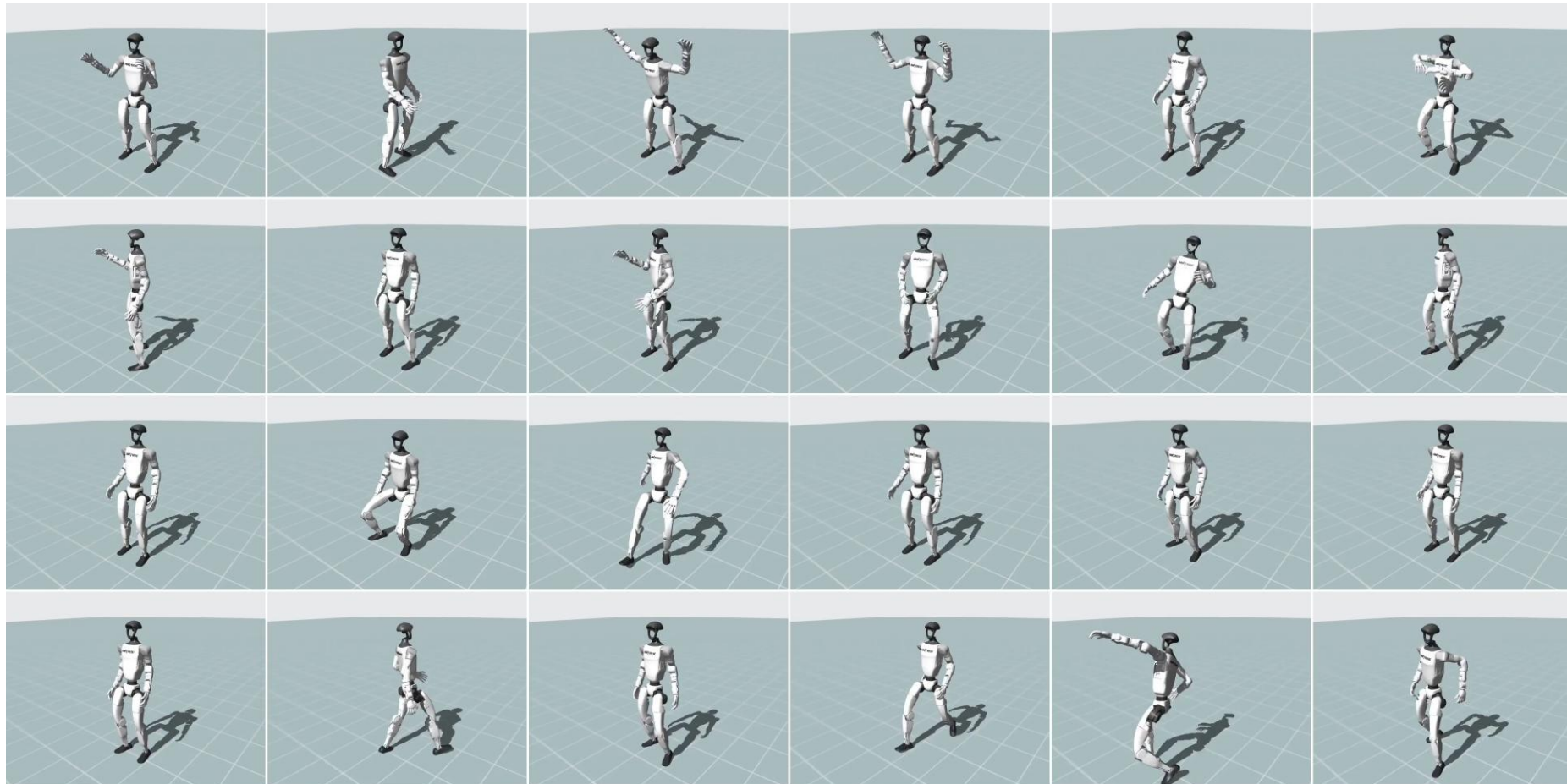
- 3D trajectory를 VLM(LLM)으로부터 생성하도록 한 뒤, 이를 저수준 정책에 힌트로 제공.
- VLM의 일반화 성능을 최대한 유지한채로 3D trajectory를 출력하도록 사후 학습을 진행하였음.



비디오 영상으로부터 휴머노이드 행동 데이터셋 구축

다양한 사람 비디오로부터 모션을 추출하여 이를 휴머노이드 행동 학습에 활용.

- 구축된 휴머노이드 행동 데이터셋으로 예시.

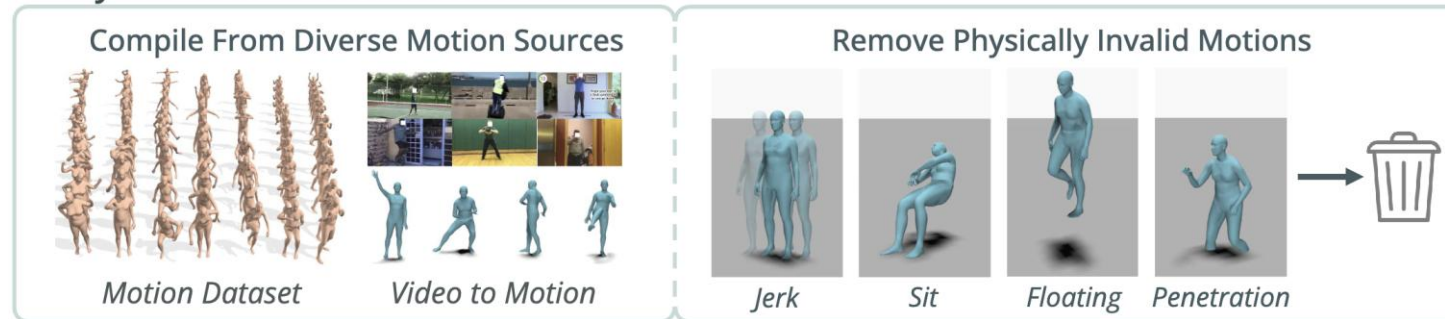


비디오 영상으로부터 휴머노이드 행동 데이터셋 구축

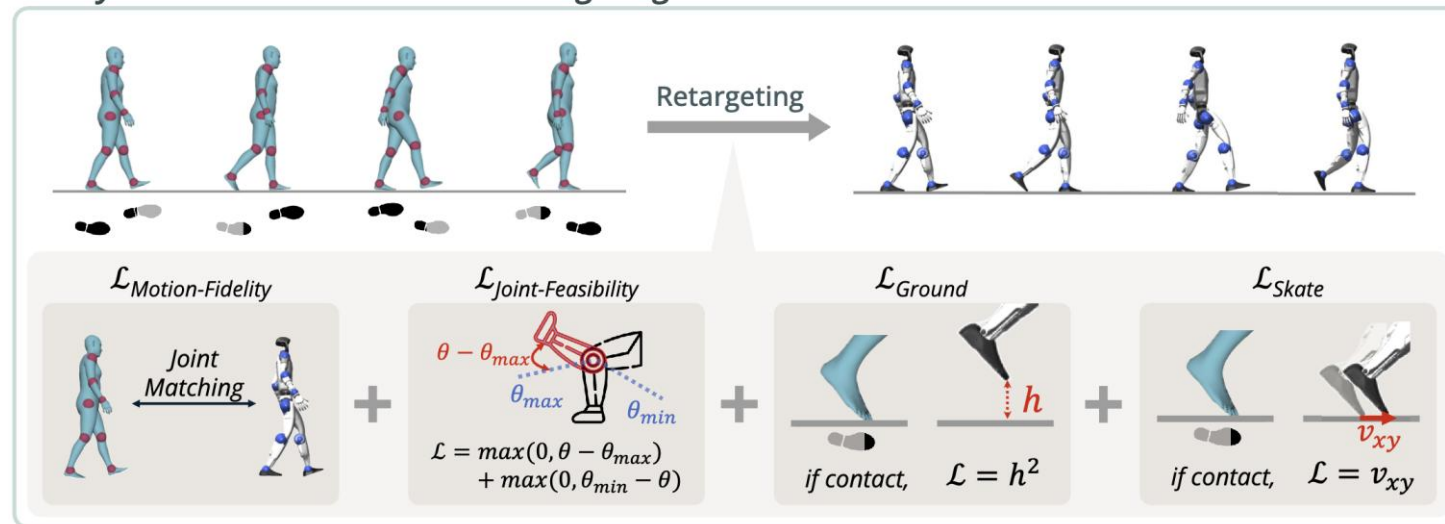
다양한 사람 비디오로부터 모션을 추출하여 이를 휴머노이드 행동 학습에 활용.

- 비디오 모션은 굉장히 노이즈가 많고 물리적으로 올바르지 못한 데이터가 많은데, 이를 정제하여 휴머노이드 학습 용 데이터셋을 구축함.

1. Physics-Aware Motion Curation



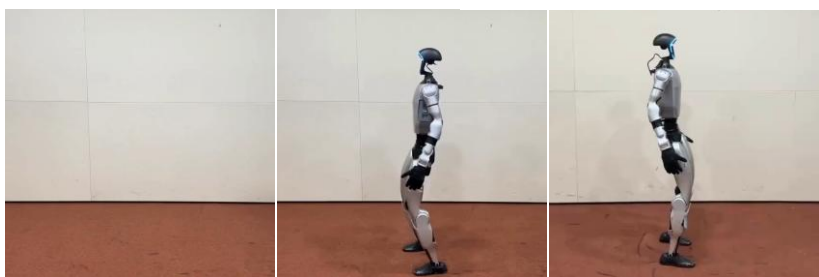
2. Physics-Constrained Motion Retargeting



마치며

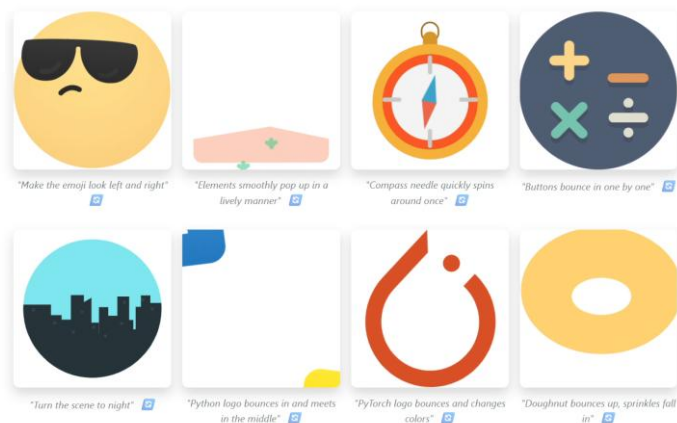
KAIST 주재걸 교수 DAVIAN 연구실 (Computer Vision)

- 그 외에도 더욱 다양한 영상 합성 및 Physical AI 연구를 수행하고 있습니다. 😊

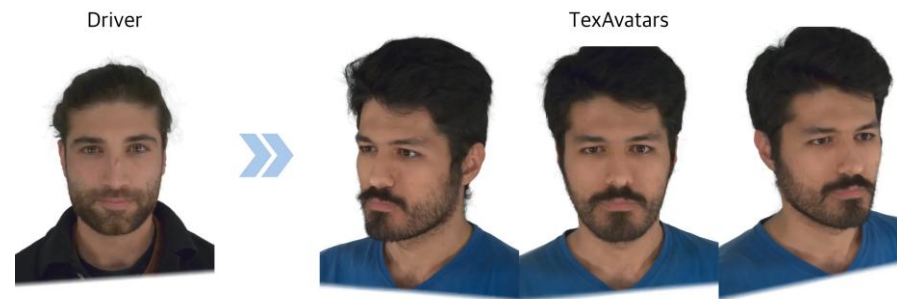


G1 walking G1 turning G1 push perturbation

RL (FlashSAC, RSS'26)



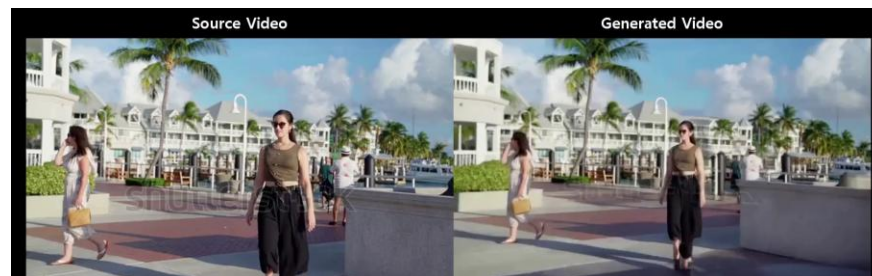
Vector Graphic (Vector Prism, CVPR'26)



Graphics (TexAvatars, 3DV'26)



Object Insertion (InsertAnywhere, preprint)



Camera Control (InfCam, preprint)

Condition



Video Generation (TIC-FT, ICLR'26)