

Kinam Kim

Graduate School of AI, KAIST • Daejeon, South Korea

✉ kinamplify@kaist.ac.kr  [kinam0252](https://github.com/kinam0252)  kinam0252.github.io | Updated: July 6, 2026

Research Interest

I am deeply interested in AI research, particularly in **Vision-Language-Action (VLA) models** [O1, U1, C2] and **Video Diffusion Models** [C1, C3-C4, P1]. With experience in Robotics and Generative Video Models, I aim to contribute to the advancement of real-world applicability through innovative AI technologies.

Keywords: Vision-Language-Action (VLA) Models, Video Diffusion Models

Education

KAIST, Graduate School of AI

M.S/Ph.D Integrated (Advisor: Prof. Jaegul Choo)

- GPA: 4.0 / 4.3

Daejeon, South Korea

Feb. 2025 - Present

KAIST, School of Computing

B.S. in Computer Science

- GPA: 3.9 / 4.3

Daejeon, South Korea

Mar. 2019 - Feb. 2024

Samsung High School

IT Diploma

Asan, South Korea

Mar. 2016 - Dec. 2018

Publications

Ongoing Project

[O1] Energy-Based Vision-Language-Action Models

Kinam Kim*, Minho Park*, Hojoon Lee, Jaegul Choo

Ongoing research project, 2025

TL;DR: We develop energy-based VLA models that incorporate a System-2 thinking process atop end-to-end policies. By integrating System-1 and System-2 behaviors through inference-time optimization, our approach aims to achieve reasoning-driven action generation that surpasses current state-of-the-art flow-based VLA models.

Under Review

[U1] Object-Centric Residual RL for Zero-Shot Sim-to-Real VLA Enhancement

Kinam Kim, Namiko Saito, Heecheol Kim, Katsushi Ikeuchi, Jaegul Choo, Yasuyuki Matsushita
under review (Conference on Robot Learning, CoRL 2026)

TL;DR: We propose an object-centric residual reinforcement learning framework that augments a base VLA policy with a residual policy trained in simulation. Our approach transfers zero-shot to the real world, substantially improving success rates on manipulation tasks such as pick-and-place and drawer closing.

Conference Paper

[C1] EgoX: Egocentric Video Generation from a Single Exocentric Video

Kinam Kim*, Taewoong Kang*, Dohyeon Kim, Jaegul Choo

The IEEE/CVF Conference on Computer Vision and Pattern Recognition ([CVPR](#)), 2026

TL;DR: We explore generating egocentric video from single exocentric video for VR/AR applications. This capability allows third-person videos to be transformed into first-person perspectives, enabling users to step into the scene and experience it as if they were the subject.

[C2] ACG: Action Coherence Guidance for Flow-based VLA models

Minho Park*, **Kinam Kim***, Junha Hyung, Hyojin Jang, Hoiyeong Jin, Jooyeol Yun, Hojoon Lee, Jaegul Choo
IEEE International Conference on Robotics and Automation ([ICRA](#)), 2026

TL;DR: We introduce Action Coherence Guidance (ACG), a training-free test-time guidance method that reduces noise-induced jerks, pauses, and jitter in VLA model policies. We show that ACG improves action coherence and significantly boosts success rates in RoboCasa, DexMimicGen, and real-world SO-101 manipulation tasks.

[C3] Temporal In-Context Fine-Tuning for Versatile Control of Video Diffusion Models

Kinam Kim*, Junha Hyung*, Jaegul Choo

The Thirty-Eighth Annual Conference on Neural Information Processing Systems ([NeurIPS](#)), 2025

TL;DR: We propose Temporal In-Context Fine-Tuning (TIC-FT), which inserts noisy buffer frames to adapt pretrained video diffusion models for controllable generation. We show that TIC-FT works with as few as 10–30 samples and improves both fidelity and efficiency over existing methods.

[C4] Spatiotemporal Skip Guidance for Enhanced Video Diffusion Sampling

Junha Hyung*, **Kinam Kim***, Susung Hong, Min-Jung Kim, Jaegul Choo

The IEEE/CVF Conference on Computer Vision and Pattern Recognition ([CVPR](#)), 2025

TL;DR: We propose Spatiotemporal Skip Guidance (STG), a training-free sampling method that enhances video diffusion models by simulating a weak model through selective layer skipping. We show that STG improves sample quality without sacrificing diversity or motion, unlike conventional CFG or autoguidance.

Preprints

[P1] Cross-Frame Representation Alignment for Fine-Tuning Video Diffusion Models

Sungwon Hwang*, Hyojin Jang*, **Kinam Kim**, Minho Park, Jaegul Choo

arXiv preprint, 2025

TL;DR: We propose Cross-frame Representation Alignment (CREPA), a regularization method that leverages neighboring frames to improve semantic consistency in video diffusion models. We show that CREPA boosts both visual fidelity and cross-frame coherence when fine-tuned with parameter-efficient methods like LoRA.

[P2] Overcoming Camera-Facing Gaze Bias in EG3D Scene Generation

Kinam Kim, Juhyun Lee, Youngdo Lee

arXiv preprint, 2023

TL;DR: We address the gaze-following problem in 3D GANs by integrating a pretrained gaze estimation model (L2CS-Net) into the EG3D training pipeline. We show that this fusion preserves realistic image quality while successfully correcting gaze direction on the FFHQ dataset.

Work Experience

NVIDIA

Research Intern

Starting Jul. 2026

Microsoft Research Asia

Research Intern

Tokyo, Japan

Feb. 2026 - Jul. 2026

KAIST DAVIAN Lab

Undergraduate Research Intern

Daejeon, South Korea

Aug. 2023 - Mar. 2025

ThetaOne Korea

Natural Language Processing AI Developer

Seoul, South Korea

Jun. 2023 - Dec. 2023

KAIST Siit Lab

Undergraduate Research Intern

Daejeon, South Korea

Jun. 2020 - Dec. 2020

Teaching Assistant

KAIST AI Mentoring Program

AI Mentor and Lead Program Coordinator

Daejeon, South Korea

Jun. 2020 - Dec. 2020

Reference

Jaegul Choo

Ph.D. Advisor (KAIST AI)

Seongnam, South Korea

Mar. 2021 - Present

• E-mail: jchoo@kaist.ac.kr

• [website](#)