

Taewoong Kang

Graduate School of AI, KAIST • Seoul, South Korea

✉ keh0t0@kaist.ac.kr  [KEH0T0](https://github.com/KEH0T0)  [Google Scholar](https://scholar.google.com/citations?user=KEH0T0)  keh0t0.github.io

Research Interest

My research builds **video world models** that are spatially and temporally consistent and physically plausible, with a focus on **humans and their interactions with objects**. I have worked on **video generation** [C1, C3], **human appearance and motion modeling** [C2, C4-C6, C8], and **3D representations and neural rendering** [C4, C5, C7], and am now extending these to **world-action models** that learn human-object interaction from egocentric video for Physical AI [O1, O2].

Keywords: World Foundation Models, Video Generation, Human Modeling, Human-Object Interaction, 3D/4D Representations, Neural Rendering, Physical AI

Education

Korea Advanced Institute of Science and Technology (KAIST)

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

Seoul, South Korea

Mar. 2024 - Present (Expected Feb. 2029)

- GPA: 3.86 / 4.3

Korea University

B.E. in Electrical Engineering

Seoul, South Korea

Mar. 2018 - Feb. 2024

- GPA: 4.01 / 4.5 (Major GPA: 4.24 / 4.5)

Ongoing Project

[O1] Dexterous World-Action Models from Human Video

Taewoong Kang, et al.

In collaboration with NVIDIA Research

TL;DR: Teleoperated data for multi-fingered hands is scarce, while egocentric video of human-object interaction is abundant. We pretrain a world-action model on large-scale human manipulation video and transfer it to robot hands through a physics-verified human-to-robot paired dataset, aiming for a policy that performs new dexterous tasks from a single human demonstration.

[O2] In-Context Self-Improvement for World-Action Models

Taewoong Kang, et al.

TL;DR: We train a large-scale world-action model to improve from its own trial and error in context: conditioned on a few recent attempts and their outcomes, the policy corrects its next actions without gradient updates, and can additionally take a human demonstration video to learn new tasks in one shot.

[O3] Real-to-Sim Task Generation for Dexterous Manipulation

TL;DR: We convert real-world scenes and objects into simulation to generate a large and diverse set of manipulation tasks for training general-purpose dexterous hands.

Publications

(* equal contribution)

Under Review

[U1] **Recall & Roll: Benchmarking How Embodied Agents Remember, Act, and Update Memory**

Taewoong Kang, et al.

In collaboration with Qualcomm AI Research

TL;DR: R&R-Bench is a controlled evaluation of how embodied agents recall past experience and track progress across navigation and manipulation. We find that visual and textual memory have complementary strengths and introduce R&R Agent, which combines both and achieves the best memory-QA accuracy and full-pipeline success.

Conference Paper

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

Taewoong Kang*, Kinam Kim*, Dohyeon Kim, Minho Park, Junha Hyung, Jaegul Choo

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

TL;DR: Repurposes a pretrained video diffusion model to generate a first-person video from a single third-person video, keeping the synthesized view spatially and temporally consistent with the source scene and the human actions in it.

[C2] **AHS: Adaptive Head Synthesis via Synthetic Data Augmentation**

Taewoong Kang*, Hyojin Jang*, Sohyun Jeong*, Jaeseong Lee, Seunggi Moon, Gihwi Kim, Hoon Jin Jung, Jaegul Choo

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

TL;DR: Uses synthetic data augmentation to train a head synthesis model that adapts to diverse identities, poses and expressions, improving human appearance fidelity in real-world head generation.

[C3] **SphereDiff: Tuning-free Omnidirectional Panoramic Image and Video Generation via Spherical Latent Representation**

Minho Park*, Taewoong Kang*, Jooyeol Yun, Sungwon Hwang, Jaegul Choo

AAAI Conference on Artificial Intelligence ([AAAI](#)), 2026

Oral presentation (acceptance rate: 17.6%)

TL;DR: Generates 360° panoramic images and videos with pretrained diffusion models and no tuning, using a spherical latent representation that keeps the output geometrically consistent across the full sphere.

[C4] **TexAvatars: Hybrid Texel-3D Representations for Stable Rigging of Photorealistic Gaussian Head Avatars**

Jaeseong Lee*, Junyeong Ahn*, Taewoong Kang, Jaegul Choo

International Conference on 3D Vision ([3DV](#)), 2026

TL;DR: Combines texel-space and 3D representations for Gaussian head avatars, giving stable rigging under large expressions and poses for photorealistic human appearance and motion.

[C5] **SurFhead: Affine Rig Blending for Geometrically Accurate 2D Gaussian Surfel-based Head Avatars**

Jaeseong Lee*, Taewoong Kang*, Marcel Böhler, Min-Jung Kim, Sungwon Hwang, Junha Hyung, Hyojin Jang, Jaegul Choo

International Conference on Learning Representations ([ICLR](#)), 2025

TL;DR: Builds head avatars from 2D Gaussian surfels with affine rig blending, recovering accurate geometry alongside photorealistic rendering of a deforming human head.

[C6] **Zero-Shot Head Swapping in Real-World Scenarios**

Taewoong Kang*, Sohyun Jeong*, Hyojin Jang*, Jaegul Choo

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

TL;DR: Swaps entire heads, including hair and head shape, in real-world images without per-identity fine-tuning, blending the new head naturally with the body and scene.

[C7] **VEGS: View Extrapolation of Urban Scenes in 3D Gaussian Splatting using Learned Priors**

Sungwon Hwang*, Min-Jung Kim*, **Taewoong Kang**, Jayeon Kang, Jaegul Choo

European Conference on Computer Vision ([ECCV](#)), 2024

TL;DR: Uses learned priors to render urban scenes from viewpoints far outside the training trajectory with 3D Gaussian Splatting, improving neural rendering quality under view extrapolation.

[C8] **Expression Domain Translation Network for Cross-domain Head Reenactment**

Taewoong Kang*, Jeongsik Oh*, Jaeseong Lee, Sunghyun Park, Jaegul Choo

IEEE International Conference on Acoustics, Speech and Signal Processing ([ICASSP](#)), 2024

TL;DR: Translates facial expressions across domains so a driving human video can reenact heads from a different visual domain, transferring human motion beyond the training domain.

Preprints

[P1] **Injecting Inductive Bias to 3D Gaussian Splatting for Geometrically Accurate Radiance Fields**

Sungwon Hwang, Soohyun Ryu, Jaeseong Yoon, Jaeseong Lee, Junha Hyung, Min-Jung Kim, **Taewoong Kang**, Joohyun Park, Jaegul Choo, Suyoung Yeon

[P2] **GauShot: One-shot Gaussian Head Avatars via Expressive Diffusion Reenactment Prior**

Jaeseong Lee, **Taewoong Kang**, Nikita Drobyshev, Junho Kim, Jaegul Choo, Gayoung Lee, Jin-Hwa Kim

Honors and Awards

- **Sukrim Scholarship**, Korea University Fall 2022
- **Award Certificate**, Korea University Spring 2022, Fall 2022, Spring 2023, Fall 2023

Skills

- **Programming languages:** Python, C/C++, MATLAB
- **Machine learning tools:** PyTorch, scikit-learn, NumPy

Reference

- [Jaegul Choo](#), Ph.D. Advisor, KAIST AI • jchoo@kaist.ac.kr