Jun-Kun Chen

Ph.D. Candidate in Computer Science | Controllable Generation | Visual World Models

immortalco233@gmail.com | junkun3@illinois.edu | +1 872-772-0116 | Website | Scholar | LinkedIn | GitHub

Research Focus

Ph.D. researcher in **diffusion-based controllable generation** and visual world models, with first/co-first-author papers at NeurIPS, CVPR, and 3DV and two Meta Reality Labs internships. Research spans image/video synthesis, long-horizon generation, multimodal conditioning, and **visual outcome prediction**; current work explores **bidirectional LLM–diffusion reasoning** and multimodal post-training.

Industry Experience

Meta Reality Labs — **May–Aug. 2023; May–Dec. 2025**
Research Scientist Intern (2025 continued part-time Aug.–Dec.) — Zurich, Switzerland / Bay Area, CA

- **2023:** Developed 3D-consistent diffusion editing with multi-view conditioning, structured noise, and self-supervised consistency training; published as *ConsistDreamer* at CVPR 2024.
- **2025:** Scaled training for a separate 3D/video-generation project to **128–256 GPUs** across multi-node clusters, enabling faster model and data iterations.

SpreeAI — **May 2024 – May 2025**
AI Research Intern — Remote, U.S.

- **Dress&Dance** (WACV 2027, under review): developed a multimodal conditioning mechanism, paired-triplet data construction, and multi-stage training for garment, identity, and motion control.
- **Virtual Fitting Room** (NeurIPS 2025): built anchored autoregressive generation for identity-consistent, minute-scale try-on videos from short training clips.

Current Research: Generative World Models

Research direction: *outcome-grounded training of language and visual generation components for computer-use, physical, and robotic settings.*

Computer Use — Action-Conditioned Image-Diffusion World Model

- Built an action-conditioned next-screen predictor on SenseNova MoT using **structured consequence supervision** and branch-specific post-training.
- Preliminary 2,850-case evaluation showed **400 improved next-screen predictions vs. 130 regressions** (2,320 unchanged).

Physics — Video-Diffusion World Model for Physical Events

- Built and post-trained a **Cosmos3 video-diffusion system** predicting 69 future frames from a 49-frame RGB history; completed fixed inference and evaluation suites.
- Implemented **online multimodal-teacher post-training** with sequential Reasoner and diffusion-Generator updates; evaluating physical-consistency gains.

Robotics — Motion-Conditioned Video Diffusion World Model

- Built an initial synthetic **video-diffusion prototype** for robot–object interaction prediction from a scene and prescribed motion.
- Conditioned on robot-only **visual proxies and target-system correspondence examples** for controlled motion transfer.

Selected Publications

1. Virtual Fitting Room: Generating Arbitrarily Long Videos of Virtual Try-On from a Single Image ↗ — **NeurIPS 2025**
Jun-Kun Chen, Aayush Bansal, Minh Phuoc Vo, Yu-Xiong Wang — Project | Early preview (arXiv, Sep. 2025)

2. Dress&Dance: Dress up and Dance as You Like It ↗ — **WACV 2027**, under review
Jun-Kun Chen, Aayush Bansal, Minh Phuoc Vo, Yu-Xiong Wang — Project | Early preview (arXiv, Aug. 2025)

3. HDEdit: Editing Videos and 3D Scenes with Video Diffusion Through Hierarchical Task Decomposition ↗ — **3DV 2026**
Yanming Zhang*, **Jun-Kun Chen***, Jipeng Lyu, Yu-Xiong Wang — Project | arXiv (Mar. 2025)

4. ProEdit: Simple Progression is All You Need for High-Quality 3D Scene Editing ↗ — **NeurIPS 2024**
Jun-Kun Chen, Yu-Xiong Wang — Project | arXiv (Nov. 2024)

5. Instruct 4D-to-4D: Editing 4D Scenes as Pseudo-3D Scenes Using 2D Diffusion ↗ — **CVPR 2024**
Linzhan Mou*, **Jun-Kun Chen***, Yu-Xiong Wang — Project | arXiv (Jun. 2024)

6. ConsistDreamer: 3D-Consistent 2D Diffusion for High-Fidelity Scene Editing ↗ — **CVPR 2024**
Jun-Kun Chen, Samuel Rota Bulò, Norman Müller, Lorenzo Porzi, Peter Kontschieder, Yu-Xiong Wang — Project | arXiv (Jun. 2024)

*Equal contribution. Preferred/publication name: **Jun-Kun Chen** (J.-K. Chen); legal name and earlier papers: Junkun Chen / J. Chen.

Education

University of Illinois Urbana-Champaign — **Expected Dec. 2026**
Ph.D. Candidate in Computer Science; Advisor: Prof. **Yu-Xiong Wang** — Urbana, IL

Tsinghua University — **2017 – 2021**
B.Eng. in Computer Science, Yao Class (IIIS); Major GPA: 3.94/4.0 — Beijing, China

Technical Skills

Models, Tools & Evaluation: PyTorch; Python; C++; controllable image/video diffusion; visual world models; multimodal conditioning and post-training; Transformers; vision-language models; data curation; benchmark design; quantitative and human evaluation.

Selected Research Contributions

Long-Horizon & Multimodal Video Generation — VFR and Dress&Dance

- Developed a **multimodal conditioning mechanism** for garment, text, layout, pose, identity, and motion controls; reduced FVD by 20–59% versus a Kling 1.6 try-on-and-animation pipeline across Internet and captured evaluations.
- Built anchored autoregressive generation for **minute-scale, temporally consistent video** from short training clips; led all three VBench consistency/smoothness metrics on 30–50 s hard-motion evaluations against FramePack and Kling 2.0.

General-Purpose Diffusion Editing — HDEdit and ProEdit

- Designed **hierarchical task decomposition** that separates instruction fulfillment from content preservation, combining LLM-based scheduling with interpolation-based refinement for training-free video and scene editing.
- Developed tuning-free preservation control and a **progressive editing workflow** with preview and edit-strength control, enabling high-fidelity geometry and appearance editing.

Consistent Scene and 4D Generation — ConsistDreamer and Instruct 4D-to-4D

- Created context-rich multi-view conditioning, **3D-consistent structured noise**, and self-supervised warping consistency for high-fidelity diffusion editing across views.
- Extended 2D diffusion to long-term dynamic scenes through **anchor-aware attention** and optical-flow-guided sliding windows.

Geometric Visual Representations — NeuralEditor and PointTree

- Built point-based radiance-field and relaxed k-d-tree representations for editable scene geometry and **transformation-robust point-cloud encoding**.

Additional Publications

7. SceneCraft: Layout-Guided 3D Scene Generation ↗ — **NeurIPS 2024**

8. Contrastive Learning Relies More on Spatial Inductive Bias Than Supervised Learning: An Empirical Study ↗ — **ICCV 2023**

9. NeuralEditor: Editing Neural Radiance Fields via Manipulating Point Clouds ↗ — **CVPR 2023**

10. PointTree: Transformation-Robust Point Cloud Encoder with Relaxed K-D Trees ↗ — **ECCV 2022**

11. Is Self-Supervised Contrastive Learning More Robust Than Supervised Learning? ↗ — **ICML 2022 Workshop**

12. TorchDrug: A Powerful and Flexible Machine Learning Platform for Drug Discovery ↗ — **arXiv preprint, 2022**

13. RNNLogic: Learning Logic Rules for Reasoning on Knowledge Graphs ↗ — **ICLR 2021**

14. Pre-training Assisted Scene-aware Dialogue Generation ↗ — **DSTC8 Workshop at AAAI 2020**

Earlier Experience

Mila – Quebec AI Institute — **Feb. – Aug. 2020**
Research Intern; knowledge-graph reasoning with Prof. Jian Tang — Montreal, Canada

Baidu, Natural Language Processing Department — **July 2019 – Jan. 2020**
Research / Development Intern; scene-aware dialogue generation (DSTC8 top-3 system) — Beijing, China

Competitions, Leadership & Coursework

Additional Competition Distinctions: CTSC 2017 (7th); CTSC 2016 Gold (8th); NOI Winter Camp 2016 Gold (3rd).
Leadership & Problem Setting: Problem setter for NOI, CTSC, ICPC, and CCPC events; Vice Chair, Tsinghua Student Association of Algorithms and Competitions (2018–2019).
Selected Coursework (Generative Models): Deep Generative Models (A+); Generative Models for Biomedicine and Life Sciences (A).
Vision & 3D: 3D Vision (A+); Deep Learning for Computer Vision (A); Computer Vision (A); Computational Photography (A+).
Learning: Meta-Learning (A+); Machine Learning (A).