Jinpeng Lu

+86 131-1299-2677 jinplu@mail.ustc.edu.cn Website Scholar

Video Understanding & Generation | Representation Learning | Medical AI

Research Profile & Core Strengths

My research focuses on **video understanding and generation**, **representation learning**, and **medical AI**, spanning dynamic-state modeling and evaluation in world models, transferable representations for VLM/VLA/agent systems, and 3D medical multimodal perception.

My work turns open-ended capability questions into reproducible benchmarks, representation analysis, and efficient model designs, including WRBench world-state evaluation, VeloxSeg, DINOv3 3D transfer evaluation, and H2ASeg.

Experienced with PyTorch / DDP / CUDA / Linux training and evaluation pipelines, data construction, custom model implementation, ablation analysis, visualization, and English academic manuscript writing.

Education

University of Science and Technology of China Sep. 2025 – Present
M.S. in Information and Communication Engineering, expected Jun. 2028
Advisor: Zhiwei Xiong; interests: multimodal foundation models, video understanding and generation, world models, medical multimodal intelligence.

Harbin Institute of Technology, Shenzhen Sep. 2021 – Jun. 2025
B.S. in Data Science and Big Data Technology; GPA: 3.728/4, Rank: 17/172, Top 10%
Core training: mathematics, statistics, numerical computing, data engineering; research training: PET/CT 3D segmentation, generative signal enhancement.

Representative Works

Video Understanding & Generation

Current World Models Lack a Persistent State Core

arXiv preprint, First Author **2026.06**

- **Task:** Test whether video world models keep event-consistent state across visible, hidden, and re-observed phases under viewpoint changes.
- **Method:** Built **WRBench** with camera motion as an **observability intervention**, separating camera execution, visible consistency, re-observation support, and returned-state correctness.
- **Result:** Evaluated **23 models / 9,600 videos** with **2,547 human verdicts**; clear frames and camera control still failed to ensure endpoint binding.

Representation Learning

Does DINOv3 Set a New Medical Vision Standard?

arXiv preprint, Co-first Author **2025.09**

- **Task:** Evaluate natural-image-pretrained DINOv3 as a frozen medical-vision encoder / visual prior.
- **Method:** Owned the **3D medical-imaging track**; co-built a 2D/3D classification, segmentation, and registration benchmark with slice-wise features and pseudo-3D adapters.
- **Result:** Found strong CT-classification and cardiac-MRI registration baselines, but EM/PET-CT/WSI domain-shift failures and non-monotonic scaling.

Medical AI

VeloxSeg: JL-Guided Efficient 3D Medical Segmentation

ICLR 2026, First Author [CCF-A] **2026.04**

- **Task:** Address volumetric complexity, multimodal complementarity, and deployment efficiency in PET/CT and MRI 3D segmentation.
- **Method:** Designed a lightweight CNN-Transformer for multi-scale cross-modal context, JL-guided local 3D features, and training-time texture-prior transfer.
- **Result:** Reached **62.51 / 56.48 Dice** on AutoPET-II / Hecktor2022 with **1.66M parameters**, improving GPU/CPU efficiency and memory.

H2ASeg: Hierarchical PET/CT Tumor Segmentation

MICCAI 2024, First Author [CCF-B] **2024.10**

- **Task:** Frame 3D PET/CT tumor segmentation as multimodal volumetric perception combining PET metabolic localization with CT boundary cues.
- **Method:** Built hierarchical cross-modal fusion for local/global PET-CT interaction and target-aware feature reweighting.
- **Result:** Reached **60.03 / 59.69 Dice** on AutoPET-II / Hecktor2022; best-baseline gains: **+2.70 pp / +4.9 pp**.

Core Methods & Engineering

Research: Video understanding and generation, world-model evaluation and modeling, VLM/VLA/agent representation learning, medical AI, PET/CT and MRI 3D segmentation
Engineering: Python, PyTorch, CUDA, DDP, Shell/Bash, LaTeX; training/evaluation pipelines, experiment reproduction, ablation studies, custom model implementation
Tools: Linux, Git, Docker, Weights & Biases / TensorBoard, vision, multimodal, and medical-imaging data-processing toolchains

Writing: English manuscript writing for ICLR / MICCAI / Medical Image Analysis submissions, experiment analysis, and reproducibility documentation

Full Research Outputs

Video Understanding & Generation **2026 – Present**

[1] **Current World Models Lack a Persistent State Core**, *First Author* **arXiv 2026**

Representation Learning **2025 – Present**

[1] **VLA-Trace: Diagnosing Vision-Language-Action Models through Representation and Behavior Tracing**, *Co-author* **arXiv 2026**

[2] **Pelican-Unify 1.0: A Unified Embodied Intelligence Model**, *Co-author* **arXiv 2026**

[3] **Does DINOv3 Set a New Medical Vision Standard?**, *Co-first Author* **arXiv 2025**

Medical AI **2023 – 2026**

[1] **Johnson-Lindenstrauss Lemma Guided Network for Efficient 3D Medical Segmentation**, *First Author* **ICLR 2026 [CCF-A]**

[2] **H2ASeg: Hierarchical Adaptive Interaction and Weighting Network for Tumor Segmentation in PET/CT Images**, *First Author* **MICCAI 2024 [CCF-B]**

[3] **AttriMIL: Revisiting Attention-based Multiple Instance Learning for WSI Classification**, *Co-author* **Medical Image Analysis 2025 [CCF-C] [SCI Q1]**

[4] **IPGPhormer: Interpretable Pathology Graph-Transformer for Survival Analysis**, *Co-author* **arXiv 2025**

[5] **A Localization-to-Segmentation Framework for Automatic Tumor Segmentation in Whole-Body PET/CT Images**, *Co-author* **arXiv 2023**

Additional Project

Generative Sensor Signal Enhancement Jun. 2023 – Jun. 2024

Lead, National Undergraduate Innovation Program

- Reproduced and compared GAN, diffusion, and flow-based models; designed conditional guidance strategies to improve reconstruction quality and perceptual fidelity, producing a reproducible comparison pipeline and experiment report.