Liu He

919-928-2396 | riverliuhe71@gmail.com | Portfolio (arking1995.github.io) | Linkedin

SUMMARY

Research Engineer / Applied Scientist focused on multimodal image-video generative AI, Video/Image Diffusion Model post-training, MLLM post-training, synthetic data, and scalable evaluation. Industry experiences cover building video/image understanding, video/image generation at Amazon. Ph.D. in CS with top-tier publications across ECCV, ICCV, ICLR, NeurIPS, EMNLP, COLM, etc.

CORE EXPERIENCES

Video/Image Generator Post-training	DM-based Image/video generation; physical-coherent I2V; hallucination reduction.
MLLM Post-training	SFT; DPO; GRPO; synthetic CoT data; hallucination reduction; quantization.
Agentic System	Agentic multimodal auto-evaluation; Agentic Tool-Use; long-context multimodal RAG.

INDUSTRY

Applied Scientist at **Amazon** *2024 - Present, Bellevue, WA*

Video / Image Generation for Ads and Campaigns at Amazon

- Built synthetic image editing and generation workflows for broad product categories using models (e.g. Qwen-Image-Edit).
- Advanced minute-long product-in-use video generation with >**90%** successful rate by Wan2.2 and agentic auto-auditing.
- Post-Training (**GRPO**) improves open-sourced video generation stability on complex human-product interactions.

Smart Video Description for Video Auto-Captioning and Alerting at Ring

- Led full-stack post-training of video MLLM, spanning data recipe, SFT, quantization, evaluation, guardrail.
- Elevated video understanding across sub-3s clips and long videos up to 5 mins to **>95%** accuracy while reducing hallucinations.
- Accelerated post-launch inference **2X** faster through quantization and model distillation while preserving task performance.

MLLM-based Grocery Defect Smart-Detection at Amazon Fresh

- Post-trained MLLM-based image understanding and referring-segmentation models across diverse grocery categories.
- Supported onsite system testing and deployment across FCs across US, EU, IN for Amazon Fresh and Whole Foods Market.

RESEARCH

PhysAlign — Physical-Coherent I2V by Aligning 3D Representation. *2026*

- Developed 3D-representation alignment framework to improve physical consistency in I2V generation, targeting motion, gravity, interactions, and fluid behavior.
- Advanced **Wan2.2** on multiple I2V and physical alignment metrics by training on 1.5K synthetic physical-simulated videos.

LongPerceptualThoughts - Synthetic CoT for Multimodal Reasoning *2025*

- Enabled **"Aha"** moment of Long-CoT synthetic data generation distilled from ***DeepSeek-R1*** for visual perception.
- Improved ***Qwen2.5-VL-7B-Instruct*** by **DPO** and **SFT**, **+3.4**pt across 5 vision-centric benchmarks, **+2**pt on MMLU-Pro.

DocAgent - Agentic Multimodal Long-Context Document Understanding *2025*

- Projected tree-structured outline enabling agents to identify relevant sections for long-context multimodal understanding.
- Developed an interactive reading interface for RAG of multimodal contents. Boost GPT-4o understanding Acc by **26.6%**.

Ulti3D - Advancing MLLMs by 3D Visual Instruction Data Generation. *2024*

- Proposed an unlimited VQA data generator focusing camera-object relation, keeping photorealistic image quality.
- Released ***Ultimate3D*** dataset (240K) and benchmark (7K), improved ***LLaVA-1.6*** to outperform ***GPT-4o/Claude-V3.5*** by **33.4%** on camera-object relation prediction accuracy.

Kubrick - Video Generation by MLLM Agent Collaborations *2024*

- Proposed multi-modal LLM agentic workflow for 3D generation, simulation, and animation given multi-modal prompts.
- Designed **multi-agent reflection** and **collaboration** for complex instructions of 3D Engine (Blender) tool usage.

RefineATF - Refine Generative Artifacts by Semantic Alignment. *2024*

- Innovated training-free artifact localization via cross-attention over SD feature priors for personalized image generation.
- Outperformed AnyDoor, and other baselines on generative artifact refinement; built GenArtifactBench for artifact refinement.

Additional Generative Modeling Research

- DocDiff: DDPM for controllable document layout generation using CLIP-guided prompts; ICDAR 2023 Oral.
- COHO/GlobalMapper: GAT, MAE for scalable arbitrary-shaped urban layout synthesis; ECCV 2024 Oral and ICCV 2023.

HIGHLIGHTED PUBLICATIONS

Yuan, Y., …, **He, L**. .. (2026) OptiWorld: Optimal Control for Video World Generation under Physical Constraints. (Under Review)

Xiong, Z., Song, Y., **He, L.**, …(2026) *PhysAlign*: Physics-Coherent Image-to-Video Generation through Feature and 3D Representation Alignment. (Under Review)

He, L., … (2026) Advancing Multimodal LLMs by Large-Scale 3D Visual Instruction Dataset Generation. In ***WACV2026 Award Finalists*** (**Project**)

He, L., Song, Y., ... (2025). Kubrick: Multimodal Agent Collaborations for Video Generation. In ***CVPR 2025 AI4CC Workshop***. (**Project**)

Song, Y.**, He, L.,** ... (2025). Refine-by-Align: Refinement of Generative Artifacts for Personalized Image Generation. In ***ICLR 2025***. (**Project**)

Sun, L., **He, L.**, … (2025). DocAgent: An Agentic Framework for Multi-Modal Long-Context Document Understanding. In ***EMNLP 2025*** Main.

Liao, Yuan., Elflein, S., **He, L.**, (2025). LongPerceptualThoughts: Distilling System-2 Reasoning for System-1 Perception. In ***COLM 2025*** (**Project**)

Hua, H., Zeng, Z., Song, Y., Tang, Y., **He, L.**,…(2025) MMIG-Bench: Towards Comprehensive and Explainable Evaluation of Multi-Modal Image Generation Models. In ***NeurIPS 2025***. (**Project**)

He, L., & Aliaga, D. (2024, **Oral**). COHO: Context-Sensitive City-Scale Hierarchical Urban Layout Generation. In ***ECCV 2024 Oral***. **(Project)**

He, L., & Aliaga, D. (2023). GlobalMapper: Arbitrary-Shaped Urban Layout Generation. In ***ICCV 2023*****. (Project)**

He, L., Lu, Y., Corring, J., Florencio, D., Zhang, C. (2023, **Oral**). Diffusion-Based Document Layout Generation. In ***ICDAR 2023***. **(Project)**

He, L., Shan, J., Aliaga, D.(2023). Generative Building Feature Estimation from Satellite Images. *IEEE Transactions on Geoscience and Remote Sensing.*

EDUCATION

Purdue University, *Ph.D. in Computer Science* *2019 – 2024, West Lafayette, IN*

- Honors: 2024 Purdue *Merit Recognition Award* ($1500, for high-profile research)

University of North Carolina at Chapel Hill, *Master's* *2017 – 2019, Chapel Hill, NC*

Wuhan University, *Bachelor's* **(**Top: **1%)** *2013 – 2017, Wuhan, P.R.C.*