Liu He

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

PROFESSTIONAL

Applied Scientist at **Amazon Ring AI**   *01/2025 - present, Bellevue, WA*

- Launched **Smart Video Description** feature by finetuning MLLM video perception and understanding ability.
- Improving long video MLLM reasoning ability and computing efficiency in real world domain.

EDUCATION

Purdue University (*Ph.D. in Computer Science*, GPA: **3.94/4.0**)   *06/2019 – 12/2024, West Lafayette, IN*

- RA in CGVLab (Computer Graphics and Visualization Lab) since 2019
- Honors: 2024 Purdue *Merit Recognition Award* ($1500, for high-profile research)

University of North Carolina at Chapel Hill (*M.A. in Geography*)   *08/2017 – 05/2019, Chapel Hill, NC*

- RA in Remote Sensing and Ecological Modeling Lab

Wuhan University (*B.E. in Electr. Info. Sci. and Tech.*, GPA: **3.75/4.0**, Top: **2%**)   *09/2013 – 06/2017, Wuhan, P.R.C.*

- Honors: *Microsoft* Scholarship (**1/248**); Best graduation thesis (Top: **1%**)

INTERNSHIP

Advancing MLLMs by 3D Visual Instruction Data Generation (Project: Ulti3D)   *06/2024 – 09/2024*, Amazon

- Proposed an unlimited VQA data generator focusing camera-object relation, keeping photorealistic image quality.
- Provided *Ultimate3D* dataset (240K) and benchmark (7K) for finetuning and evaluation of camera-object relation perception.
- Improved *LLaVA-1.6* and *Llama3.2-Vision* to outperform *GPT-4o/Claude-V3.5* by **33.4%** on prediction accuracy.

Video Generation by MLLM Agent Collaborations (Project: Kubrick)   *03/2024 – 05/2024*, Baidu Research USA

- 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 tool usage (Blender, etc.).

Diffusion-Based Document Layout Generation (Project: DocDiff)   *05/2022 – 08/2022*, Azure AI, Microsoft Research

- Proposed *DDPM* with **Transformer** backbone for document layout synthesis on multilingual verticals.
- Discovered the user guidance by prompts obtained by pre-trained **CLIP** for controllable generation given real document images.

RESEARCH

Multimodal Reasoning by Synthetic CoT Data Generation (Project)   *10/2024 – present*

- 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.

Agentic Multimodal Long-Context Document Understanding   *12/2024 – present*

- Proposed a structured, tree-formatted outline of documents to help agents identify relevant sections efficiently
- Developed an interactive reading interface for RAG of multimodal contents. Boost GPT-4o understanding Acc by **26.6%**.

Refine Generative Artifacts by Semantic Alignment (Project: RefineATF)   *05/2024 – 09/2024*, Adobe Research (Remote)

- Innovated an automatic artifact localization method by cross-attention on training-free Stable Diffusion feature priors.
- Outperformed SOTAs (Paint-by-Example, AnyDoor, etc.) on generative artifacts refinement for broad image personalization.
- Provided a comprehensive benchmark (*GenArtifactBench*) for generative artifacts detection and refinement.

Scalable Urban Layout Synthesis (Project: GlobalMapper, Project: COHO)   *06/2021 – 03/2024*, Purdue Univ.

- Established graph-based canonical **Vector Quantized** representation for arbitrary-shaped urban layout with scalable hierarchy.
- Introduced Transformer/Graph Attention Network (*GAT*)/Masked Autoencoder (*MAE*) of infinite 3D urban layout synthesis.

Globalwise Styled-Controlled Building Modeling by Staged GANs (Project)   *06/2019 – 05/2021*, Purdue Univ.

- Designed staged *GANs* for large-scale building segmentation with extreme upsampling refinement **(10x)**.
- Utilized learned priors as style control to generate footprints with plausible instance-level metric. Beat SOTA by at least **15%**.

HIGHLIGHTED PUBLICATIONS

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

He, L. Xiao, Z., … (2025) Advancing Multimodal LLMs by Large-Scale 3D Visual Instruction Dataset Generation. (In Review) (**Project**)

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

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.*

Kamath, H. G., Singh, M., Malviya, N., Martilli, A., **He, L.**, Aliaga, D., ... & Niyogi, D. (2024). GLObal Building heights for Urban Studies (UT-GLOBUS) for city-and street-scale urban simulations: Development and first applications. *Scientific Data*, 11(1), 886.

Patel, P., Kalyanam, R., **He, L.**, Aliaga, D., & Niyogi, D. (2023). Deep Learning based Urban Morphology for City-scale Environmental Modeling. *PNAS Nexus*, pgad027.

Zhang, X., Ma, W.,Varinlioglu, G.,Rauh, N.,**He, L.**, & Aliaga, D.(2022). Guided pluralistic building contour completion. *The Visual Computer*, 1-12.

Bhatt, M., Kalyanam, R., Nishida, G., **He, L.**, May, C., Niyogi, D., & Aliaga, D. (2020). Design and Deployment of Photo2Building: A Cloud-based Procedural Modeling Tool as a Service. In *Practice and Experience in Advanced Research Computing* (pp. 132-138).

Wang, L., Huang,Y., Shan,J.,& **He, L.**(2018).MSNet: Multi-Scale Convolutional Network for Point Cloud Classification. *Remote Sensing*,10(4), 612.

SKILL

Languages: C++ | Python | C | JAVA| Matlab | R   **Libraries: Pytorch** | **OpenCV | OpenGL** | Qt | Tensorflow | Pthread
Tools & OS: Linux | Git | LLVM | Google Cloud | **Unity** | Google Earth Engine