

Hongyang Du

11 East Manning St, Providence, RI | 240-413-5747 | hongyangdu182@gmail.com | Website | LinkedIn | Github

EDUCATION

Brown University M.S. Computer Science (AI/ML Track)	Aug 2025 – May 2027 GPA 4.0/4.0
University of Maryland, College Park B.S. Computer Science (Honors), Minor in Mathematics	Aug 2021 – May 2025 GPA 3.8/4.0

PROFESSIONAL EXPERIENCE

iFlytek-Machine Learning Engineer Intern Agriculture MLLM Team	June 2024 – Aug 2024
<ul style="list-style-type: none">• Pretrained a 2B Chinese LLM from scratch based on the LLaMA2 architecture (e.g., RoPE positional encoding, RMSNorm, SwiGLU). Employed SentencePiece Unigram tokenizer to enhance Chinese linguistic modeling.• Combined ResNet hierarchical classification and YOLOv10 disease-spot detection to classify over 1,000 crop diseases and pests. Constructed RAG-enhanced prompts using visual clues and a knowledge graph.• Finetuned the model using LoRA and reinforcement learning for domain-specific adaptation. Build a Chinese-centric pretraining pipeline with support for FSDP, gradient checkpointing, and customized preprocessing for Chinese text.	

VISION REPRESENTATION

3D Consistent Video Diffusion Model ICML 2026	June 2025 – Feb 2026
<ul style="list-style-type: none">• Developed a novel self-supervised framework that leverages a Geometry Foundation Model to distill 3D structural priors, addressing the long-standing challenge of temporal inconsistency and spatial drift in Video Diffusion Models.• Introduced a training-free pipeline to automatically construct dense preference signals; by applying Direct Preference Optimization (DPO), the model effectively aligns video generation with physical geometric laws• Significantly outperformed SOTA baselines in 3D consistency and motion coherence, achieving superior physical plausibility on complex real-world benchmarks with minimal data requirements.	

Implicit Neural Representation for Video Compression WACV 2026	May 2024 – May 2025
<ul style="list-style-type: none">• Conducted empirical study of video compression methods based on INRs. Developed a unified benchmarking platform to evaluate models across metrics PSNR, MS-SSIM, bitrate, training time, and decoding speed on HD datasets.• Proposed a novel architecture RNeRV, which introduced a Weight Token Masking mechanism and HyperNetwork decoupling to achieve flexible bitrate control and efficient model deployment. Demonstrated +1.27% average improvement in visual quality under equal training budgets and enhanced real-time inference throughput.• Built the XINC (eXplainable INR Compression) tool to visualize neuron-level activation and contribution maps across layers and time steps, improving interpretability and facilitating model optimization for video streaming.	

Computational Geometry Algorithms and Visualization Tools SoCG 2024	Sep 2023 – May 2024
<ul style="list-style-type: none">• Explored Hilbert Metric Space data structures and machine learning algorithms, contributing to developing a potential new data storage paradigm in Hilbert metric coordinates.• Engineered algorithms for generating complex geometric objects using Lua and C++. Enhanced the Ipe extensible drawing editor with efficient tools for Euclidean, Hilbert, and Funk geometric shapes.	

VLM & AGENTS

Agent Skill Dependency-Aware Structural Retrieval Preprint	Apr 2026
<ul style="list-style-type: none">• Structured massive agent skill libraries as an offline dependency graph and retrieved small, relevant skill bundles at inference (seeding + PageRank under a token budget) instead of loading the full library.• Reduced context tokens while improving downstream reward for LLM-based agents operating over large skill sets.	
Self-Evolving Vision Language Models From Zero Data Preprint	March 2025 – June 2026

<ul style="list-style-type: none">• Proposed a novel training paradigm that eliminates the need for human-annotated data ("zero data") by designing an innovative three-role collaborative pipeline: Proposer, Coder, and Solver• Empowered the model to autonomously construct visually complex scenes via code generation; implemented Group Relative Policy Optimization (GRPO) with role-specific reward mechanisms for sequential, closed-loop training.• Drive a comprehensive performance leap across base models (e.g., Qwen3-VL, MIMO-VL) with significant gains in multimodal understanding, mathematical reasoning, and coding capabilities.	
---	--

NeurIPS 2025 [Poster], CVPR 2025 Workshop [Oral]

- Designed VideoHallu dataset for hallucination detection in synthetic video understanding; proposed 2 novel metrics (temporal consistency, semantic mismatch) to quantify alignment between captions and dynamic video scenes.
- Implemented robust hallucination detection framework integrating frame-level caption matching, embedding similarity (CLIP/BLIP), and cross-frame alignment; enhanced via GRPO post-training with curriculum learning on video QA datasets incorporating counterintuitive commonsense and physics reasoning.
- Conducted a systematic survey of cutting-edge vision-language models (e.g., GPT-4V, Gemini, Qwen-VL), detailing their model architectures, instruction alignment mechanisms, and evaluation protocols.

PUBLICATIONS

Graph of Skills: Dependency-Aware Structural Retrieval for Massive Agent Skills

D. Liu*, Z. Li*, H. Du, et al. · Preprint

Paper | GitHub

A Cookbook of 3D Vision: Data, Learning Paradigms, and Application

H. Du*, Z. Li*, D. Liu*, et al. · CVPR 2026 Workshop

GitHub

MM-Zero: Self-Evolving Multi-Model Vision Language Models From Zero Data

Z. Li*, H. Du*, C. Huang*, et al. · Preprint

Paper | GitHub

VideoGPA: Distilling Geometry Priors for 3D-Consistent Video Generation

H. Du*, J. Ye*, X. Cong*, et al. · ICML 2026

Paper | Web | GitHub

How to Design and Train Your Implicit Neural Representation for Video Compression

M. Gwilliam, R. Zhang, N. Padmanabhan, H. Du, et al. · WACV 2026

Paper | Web | GitHub

VideoHallu: Evaluating and Mitigating Multi-modal Hallucinations for Synthetic Videos

Z. Li*, X. Wu*, Y. Qin, H. Du, et al. · NeurIPS 2025

Paper | Web | GitHub

A Survey of State of the Art Large Vision Language Models: Alignment, Benchmark, Evaluations and Challenges

Z. Li*, X. Wu*, H. Du, et al. · CVPR 2025 Workshop (Oral)

Paper | GitHub

Ipelets for the Convex Polygonal Geometry

N. Parepally, A. Chatterjee, A. Gezalyan, H. Du, et al. · SoCG 2024

Paper | Software

TECHNICAL SKILLS & AWARD

Languages: Python, Shell, Java, C++, Bash, Go, Rust, SQL, JS/HTML/CSS, OCaml, Lua, Racket, React, Matlab**Dev Stack:** Spring Boot, Flask, Django, FastAPI, Node.js, RESTful APIs, Redis, MongoDB, Docker, Flutter, Next.js**AI/ML Stack:** PyTorch, TensorFlow, Transformers, scikit-learn, OpenCV, LangChain, LlamaIndex, Haystack, NumPy**Focus Areas:** Self-supervised Learning, Representation Learning, Reinforcement Learning, Contrastive Learning, 2D/3D Vision, Vision-Language Models, Transformers, Diffusion Models**Awards:** Robert Ma Scholarship, Break Through Tech Scholarship, Dean's List (7 semesters), AIME Qualifier (Top 5%), AMC 12 Top 1% Honor Roll of Distinction, TN State Math Competition 1st Place