Hongjia Huang

+1 6462985765
hh3043@nyu.edu

EDUCATION

New York University Shanghai Shanghai, China 09/2022 – present
Double Majoring in Computer Science and Mathematics

- **Overall GPA: 3.933** / 4.0
- **Research Interest:** Robotics, Computer Vision, and Physics-Grounded Video Generation
- **Core Course:** NLP (Graduate-Level), Inference & Representation (Graduate-Level), RL (Graduate-Level), Machine Learning, Parallel Computing, Algorithms, Operating Systems and Data Structures; ODE (Honor), PDE, Analysis (Honor), Numerical Analysis (Honor), Linear Algebra (Honor), Linear and Nonlinear Optimization, Probability & Statistics
- **Programming:** Proficient in Python; familiar with C++/C; PyTorch, OpenMP, MPI, CUDA, and LaTeX
- **ORCID Page:** `https://orcid.org/0009-0007-7627-0102`

Awards and Honors
Summa Cum Laude (B.S., NYU Shanghai); Dean's Undergraduate Research Fund (22–23), Dean's Honor List (22–25), Recognition Award (23–25)

RESEARCH

Auxiliary Streams for Video-Diffusion World-Action Models 12/2025 – present
Independent Research; Supervised by Prof. Shengjie Wang (NYUSH) and Prof. Tianyi Zhou (UMD)

- Investigating whether training-time auxiliary structure can improve the policy of a frozen Wan2.2-5B video-diffusion world-action model, particularly its spatial out-of-distribution generalization.
- Added inference-free alignment teachers (VGGT, CoTracker, DINO/REPA) and co-denoised streams routed by a Mixture-of-LoRA-Experts, passing 3D structure from 2D video to the action head at test time.

Physics Informed VLA 06/2025 – present
Independent Research; Supervised by Prof. Furong Huang (UMD) and Prof. Tianyi Zhou (UMD)

- Aiming to integrate trajectory planning into the action prediction task of VLA models such as Gr00t-N1.5.
- Use a 4D encoder that fuses multiple views across multiple timesteps to propagate spatial and temporal information jointly, aiding action prediction via future latent alignment.
- Designing an implicit 2D-to-3D mapping whose past-to-current 3D alignment, guided by the injected past 2D trace, aids action prediction.
- Implementing a cascaded pipeline for trace prediction and action prediction conditioned on that trace.

TraceGen: World Modeling in 3D Trace-Space 06/2025 – 09/2025
Summer Research with Prof. Furong Huang (UMD)

- Addressed the small-data problem in manipulation with a compact 3D "trace-space" of scene-level trajectories, letting a world model predict future motion geometrically instead of in pixel space and learn from cross-embodiment, cross-environment, and cross-task video.
- Contributed to TraceForge, the data pipeline that converts heterogeneous human and robot video into consistent 3D traces, yielding the 123K-video, 1.8M observation–trace–language triplet pretraining corpus.
- Pretraining on that corpus reaches 80% success across four tasks from five target-robot videos at 50–600× the inference speed of video-based world models, and 67.5% real-robot success from five uncalibrated handheld human videos.

Efficient Training for Small Molecular Forcefield 06/2024 – 12/2025
Independent Research; Supervised by Prof. Shengjie Wang (NYUSH) and Prof. Tianyi Zhou (UMD)

- Acquired abundant low-precision molecular dynamics samples and subsampled them via submodular selection to find diverse molecular conformations, which were then recomputed with high-precision methods.
- Demonstrated that training GemNet on datapoints selected via similarity-based submodular functions outperforms random and equal-timestep sampling.
- Implemented graph VAEs to learn geometry-aware features for submodular selection; modified EGNN to predict a coarse molecular forcefield, yielding forcefield-aware features for selection.

Atomic Environment Imaging for Efficient Machine Learning Force Fields 10/2025 – 12/2025
Independent Research; Supervised by Prof. Shengjie Wang (NYUSH)

- Used 3D visualizations of molecules to improve the performance of forcefield prediction models.
- Constructed a general pipeline that takes predefined multiview images of the local structure centered on each atom.
- Implemented additional image encoders for models with direct force prediction.
- Implemented additional coordinate encoders that reconstruct images, for gradient-based force-prediction models that take only coordinates as input.

SELECTED PROJECT

Cost-aware Finetuning of LMs for Chemical Reaction Prediction 11/2024
Final Project for Natural Language Processing with Representation Learning

- Efficiently finetuned pretrained BART-based models for chemical reaction prediction.
- Constructed USPTO50K$_\gamma$, a dataset containing LLM-generated predictions and experimental data.