Gene Chou

gene@cs.cornell.edu · genechou.com · github.com/gene-chou

EDUCATION

Cornell University 2023 - 2027
PhD student in computer science
Advisors: Noah Snavely and Bharath Hariharan

Princeton University 2018 - 2022
BSE in computer science, minor in applied mathematics
Honors: magna cum laude, departmental GPA: 3.96 / 4.0

EXPERIENCE

Apple AIML | *Cupertino, CA* 4/2026 -
Fast and interactive video world models.

Google | *San Francisco, CA* 5/2025 - 4/2026
Spatially-grounded, city-scale, and navigable video generation using Street View data.

Netflix Eyeline Studios | *Los Angeles, CA* 1/2025 - 5/2025
Real-time, high-resolution 4D reconstruction.

Adobe Research | *San Jose, CA* 5/2024 - 1/2025
Generated 3D-consistent videos from unposed, wide-baseline images such as internet photos.

Backflip AI | *San Francisco, CA* 7/2023 - 9/2023
Scaled up 3D asset generation at a GenAI startup.

Disney Research Imagineering | *Glendale, CA* 2/2023 - 7/2023
Combined large vision models and RL for human-robot interaction and sim-to-real transfer.

Princeton Computational Imaging Lab | *Princeton, NJ* 3/2022 - 2/2023
Explored generalization and generation of neural scene representations, advised by Prof. Felix Heide.

Multimedia Technology Lab at Academia Sinica | *Taipei, Taiwan* 6/2020 - 8/2021
Improved object detection robustness and efficiency, advised by Prof. Hong-Yuan Mark Liao.

ACADEMIC ACTIVITES

Teaching Assistant
Computer Vision (2025); Deep Learning (2024); Algorithms and Data Structures (2019 - 2022)

Service
Reviewer for CVPR, ICCV, ECCV, SIGGRAPH, TPAMI

Awards
NSF Graduate Research Fellowship (2023 - 2028)

PUBLICATIONS

FastFrames: Video Generation at Interactive Latency
G. Chou, G. Yang, F. Luan, L. Mescheder, W. Dong, S. Richter, V. Koltun *preprint 2026*
FastFrames enables real-time world models and is 8x to 37x faster than previous state-of-the-art.

CityRAG: Stepping Into a City via Spatially-Grounded Video Generation *NeurIPS 2026 Spotlight*
G. Chou, C. Herrmann, K. Genova, B. Deng, S. Peng, B. Hariharan, J. Zhang, N. Snavely, P. Henzler
CityRAG generates minutes-long, spatially-grounded videos that are simulations of real locations, trained on Street View. It supports arbitrary trajectories, weather, and dynamic object configurations.

FlashDepth: Real-time Streaming Video Depth Estimation at 2K Resolution
G. Chou, W. Xian, G. Yang, M. Abdelfattah, B. Hariharan, N. Snavely, N. Yu, P. Debevec *ICCV 2025*
FlashDepth predicts depth of streaming high-res videos at 24 FPS on an A100 GPU.

Generating 3D-Consistent Videos from Unposed Internet Photos
G. Chou, K. Zhang, S. Bi, H. Tan, Z. Xu, F, Luan, B. Hariharan, N. Snavely *CVPR 2025*
We propose the task of generating videos from sparse, unposed internet photos, and train a 3D-aware video model without any 3D annotations such as camera parameters.

MegaScenes: Scene-Level View Synthesis at Scale
J. Tung*, **G. Chou***, R. Cai, G. Yang, K. Zhang, G. Wetzstein, B. Hariharan, N. Snavely *ECCV 2024*
MegaScenes is a scene-level dataset containing 100K SfM reconstructions and 2M registered images. We validate its effectiveness in training generalizable models for single image novel view synthesis.

Generalist YOLO: Towards Real-Time End-to-End Multi-Task Visual Language Models
HS. Chang, CY. Wang, R. Wang, **G. Chou**, HY. Liao *WACV 2025*
Builds on YOLOR to jointly train multiple vision and vision-language tasks. Fast and lightweight.

Thin On-Sensor Nanophotonic Array Cameras *SIGGRAPH Asia 2023*
P. Chakravarthula, J. Sun, X. Li, C. Lei, **G. Chou**, M. Bijelic, J. Froesch, A. Majumdar, F. Heide
Recovers images in broadband using a single flat metasurface optic and probabilistic deconvolution.

Diffusion-SDF: Conditional Generative Modeling of Signed Distance Functions
G. Chou, Y. Bahat, F. Heide *ICCV 2023*
Performs diffusion on the latent space of neural SDFs to conditionally generate 3D objects.

GenSDF: Two-Stage Learning of Generalizable Signed Distance Functions
G. Chou, I. Chugunov, F. Heide *NeurIPS 2022 Spotlight*
Reconstructs neural SDFs from raw input point clouds of over a hundred unseen object classes.