

Xiaotong Wu

San Jose, CA | +1-669-669-1122 | xiaotongwu116@gmail.com | linkedin.com/in/xiaotong-wu

RESEARCH INTERESTS

Artificial intelligence and machine learning; computer vision; multimodal learning; robust and efficient deep learning; intelligent autonomous systems.

EDUCATION

San José State University

Aug. 2024 - May 2026

M.S., Interdisciplinary Engineering | San Jose, CA | Advisor: Ahmed Hambaba

Master's project (May 2026): From Pillars to Transformers: Two Fusion Paradigms Based on PointPillars for 3D Object Detection

Changchun University of Technology

Sep. 2018 - Jun. 2022

B.S., Computer Science and Engineering | Changchun, China

PUBLICATIONS

Sanghee Suh, Pavithra M. Girithimmappa, **Xiaotong Wu**, Le Shen, and Ahmed Hambaba. "Pillar-level Multimodal Fusion for Efficient LiDAR-Camera 3D Object Detection." WI-IAT 2026. Published.

MANUSCRIPTS SUBMITTED FOR REVIEW

[1] S. H. Suh, P. M. Girithimmappa, **X. Wu**, L. Shen, and A. Hambaba, "FusionPointPillars: Efficient pillar-wise LiDAR-camera fusion for real-time 3D object detection in autonomous vehicles," *IEEE Trans. Intell. Veh.*, submitted for publication.

[2] S. H. Suh, P. M. Girithimmappa, **X. Wu**, L. Shen, and A. Hambaba, "Adaptive pillar fusion for efficient LiDAR-camera 3D object detection," submitted to the 2026 IEEE 29th International Conference on Intelligent Transportation Systems (ITSC); under review.

RESEARCH EXPERIENCE

Research Assistant | University of Illinois Urbana-Champaign

Jul. 2026 - Present

Illinois iRisk Lab | Advisor: Zhiyu (Frank) Quan

- Proposed LayerGate-BEV, replacing fixed camera-LiDAR fusion coefficients with scene-conditioned gates across four encoder layers; completed the proposal and architecture design.
- Designed baseline-preserving initialization and a nuScenes evaluation plan to study detection, modality reliance, and robustness under LiDAR thinning.

Master's Research | San José State University

2024 - 2026

Advisor: Ahmed Hambaba | FusionPointPillars and BEVFormerFusion

- Designed BEVFormer encoder-decoder and encoder-only fusion architectures and participated in training and validation. Encoder-decoder fusion improved nuScenes validation mAP by 5.0 points and NDS by 3.6 points; mATE decreased from 0.949 to 0.870.
- Contributed to FusionPointPillars architecture, training, and validation of SSF, GMF, and APF. On KITTI Moderate 3D AP40, GMF improved Pedestrian by 3.02 points and Cyclist by 7.12 points; APF improved Cyclist by 7.38 points.
- Benchmarked RTX 3060 inference at approximately 12 FPS for ResNet-50 fusion; a ResNet-18 ablation reached 14.10 FPS with reduced accuracy.

Multimodal Restaurant Recommendation | San José State University

2025

CMPE 256 course research project

- Extracted BERT/ResNet-18 text-image features, constructed Louvain/FAISS community graphs, and trained a multimodal GAT recommender.
- The GAT model achieved validation RMSE of 0.1493 on ratings normalized to [0,1], with early stopping for checkpoint selection.

TECHNICAL SKILLS

Deep Learning & Vision: PyTorch, MMCV, MMDetection3D, CNNs, Transformer, BERT.

Graph Learning & Data: Graph attention networks (GAT), Louvain, FAISS, scikit-learn, NumPy, Pandas.

Programming & Tools: Python, SQL, Git, Docker.