

# HAOTIAN ZHANG (CARL)

Information Processing Lab, Department of Electrical & Computer Engineering, University of Washington

• Email: [haotian.carl.zhang@gmail.com](mailto:haotian.carl.zhang@gmail.com) • Tel: +1 (206) 226-0646

## EDUCATION

---

### University of Washington (UW)

*Ph.D. Student, Electrical & Computer Engineering*

Seattle, WA  
2022 (expected)

- Current Major GPA: 4.0/4.0
- Advisor: Prof. Jenq-Neng Hwang (IEEE Fellow); Information Processing Laboratory
- **Research topic:** 2D/3D Object Detection, Tracking; Vision & Language

### University of Washington (UW)

*Master of Science, Applied Mathematics*

Seattle, WA  
June 2021

- Major GPA: 3.9/4.0
- Concurrent Degree

### University of Washington (UW)

*Master of Science, Electrical & Computer Engineering*

Seattle, WA  
Jan. 2019

- Major GPA: 3.86/4.0
- BS-MS Combined Degree (early admission program)

### Shanghai Jiao Tong University (SJTU)

*Bachelor of Science, Micro & Nano Electronics*

Shanghai, China  
June. 2017

- Cumulative GPA: 84.6 / 100, Ranking Top 7
- Honored as the Chun-Tsung Scholar

## HONORS & AWARDS

---

- NeurIPS Young Scholar Award 2022
- University Outstanding Graduates, SJTU 2017
- Chun-Tsung Scholar, supported by the famous Nobel Laureate Physicist Tsung-Dao Lee 2017
- International First Prize for Mathematical Contest in Modeling, MCM 2016, 2017
- Huawei Scholarship for Excellence (5%) 2016
- A-Level Scholarship of Excellence 2014 – 2017
- National Second Prize, Chinese Physics Olympiad, CPhO 2013

## TEACHING EXPERIENCE

---

### University of Washington (UW), Seattle

*Teaching Assistant*, PMP 596 Autumn: *Deep Learning for Big Visual Data* 2021

*Teaching Assistant*, EE 443 Spring: *Design and Application of Digital Signal Processing* 2020

*Teaching Assistant*, EE 442 Winter: *Digital Signal and Filtering* 2019

## RESEARCH EXPERIENCE

---

**Lab research**, Information Processing Lab 2021 - Now

My research mainly based on a monocular vision based autonomous driving framework to perform localization, tracking, and segmentation, supervised by Prof. Jenq-Neng Hwang.

**Graduate Research Assistant**, “Joint Monocular Vehicle 3D Localization, Tracking, and Segmentation”

- An RCNN-based Loc4Trk-Net is proposed to not only generate 2D bounding box and instance masks but also simultaneously predict both the 3D orientation and distance of vehicles in the camera coordinate.
- The proposed JMV3D achieves 1st ranking on the KITTI-MOTS leaderboard. (The runner-up of Track1 KITTI-MOTS at CVPR 5th BMTT challenge, 2020.)
- The proposed framework achieves 1st place on the KITTI-STEP leaderboard. (The winner of Track2 KITTI-STEP at ICCV 6th BMTT challenge, 2021)

**Lab research**, Information Processing Lab 2019 - 2020  
My research mainly based on a monocular vision based autonomous driving framework to perform detection, tracking and localization, supervised by Prof. Jenq-Neng Hwang.  
**Graduate Research Assistant**, “Monocular 3D Localization on Road Scenes”

- An RCNN-based LOCNet is proposed to simultaneously regress both the 3D orientation and distance of vehicles, which can serve as a good initialization for follow-up optimizations.
- A single-frame optimization technique based on the fitness evaluation score (FES) is applied to ensure the object spatial robustness in the 3D localization.
- A 3D TrackletNet Tracker, which takes into account both discriminative CNN appearance features and accurate 3D spatial object information from each frame, is introduced to associate detections across frames.
- A multi-frame optimization technique is incorporated to reduce the impact of unreliable or missing detections and generate more accurate object localization by considering temporal consistency.

**Lab research**, Information Processing Lab 2018 - 2019  
My research mainly based on depth estimation & localization, supervised by Prof. Jenq-Neng Hwang.  
**Graduate Research Assistant**, “Robust Pedestrian Localization from Sky based on Real-time Probabilistic Ground Plane Estimation”.

- Constructed a probabilistic dense depth map, in which the Gaussian modeling is integrated into a monocular SLAM algorithm (SVO, etc.) to estimate per-pixel depths based on the live video stream.
- A fast-smoothing method that takes into account the measurement uncertainty to provide spatial regularity to mitigate the effect of noisy camera pose estimation.
- Achieved parallelizable implementations by estimating depth for every pixel independently using CUDA toolkit.
- Combined advanced Detection (Mask RCNN) & Tracking (TrackletNet) method to localize pedestrians on the ground (mean error < 1m).

**Lab research**, Information Processing Lab 2018 - 2019  
My research mainly based on using the traffic sign constraints to perform bundle adjustment to mitigate the drifting effect of current Visual Odometry system, supervised by Prof. Jenq-Neng Hwang.  
**Graduate Research Assistant**, “Bundle Adjustment for Monocular Visual Odometry Based on Detections of Traffic Signs”

- Proposed a traffic-sign based joint Bundle Adjustment (BA), which includes the constraints imposed from traffic signs to reduce the accumulated drift error on top of the standard local BA process.
- Experimental results by applying our traffic sign feature-based BA module show an improved vehicular localization accuracy compared with the state-of-the-art baseline VO method.

**Research project**, The Joint Center for Aerospace Technology Innovation (JCATI) 2017 - 2018  
The project mainly based on visual odometry and autonomous system, supervised by Prof. Jenq-Neng Hwang.  
**Graduate Research Assistant**, “Camera Self-Calibration for Drone Autopilot”.

- Presented an autonomous quadrotor system that is able to perform landing on small platform in both outdoor and indoor environment (GPS-denied).
- Obtained more accurate, scale-aware drone pose estimation and navigation from Semi-Visual Odometry (SVO).
- Formed autonomous landing with no more than 5 centimeters repeated error.

**The Chun-Tsung Program**, SJTU 2016 - 2017  
Granted by Hui-Chun Chin and Tsung-Dao Lee Chinese Undergraduate Research Endowment (CURE)  
**Chun-Tsung Scholar**, “Study on Carbon- Based Interconnects in Nanoscale Integrated Circuits”.

- Supported by the famous Nobel Laureate Physicist Tsung-Dao Lee.
- Only 15 students among the all in SJTU are chosen to participate this program each year.

**Lab research**, Center for Microwave and RF Technologies (CMRFT) 2014 - 2017  
My research mainly based on electromagnetic and microwave, supervised by Prof. Junfa Mao.  
**Undergraduate Research Assistant**, “Study on Carbon-Based Interconnects in Nanoscale Integrated Circuits”

- Simplified equivalent single-conductor models for carbon-based Interconnects.
- Developed Fast algorithms for evaluating delay and crosstalk performance.
- Considered both electrical and thermal effects on Multi-physics modeling.

## **INTERSHIP EXPERIENCE**

---

**Microsoft Research, Redmond** – Deep Learning Team

2022

We present GLIPv2, a grounded VL understanding model, that serves both localization tasks (e.g., object detection, instance segmentation) and Vision-Language (VL) understanding tasks (e.g., VQA, image captioning), working with Pengchuan Zhang, Xiaowei Hu, Yen-Chun Chen, Liunian Li, Xiyang Dai, Lijuan Wang, Lu Yuan, Jenq-Neng Hwang and Jianfeng Gao.

**Research Intern – Deep Learning**, “GLIPv2: GLIPv2: Unifying Localization and Vision-Language Understanding”

- The model unification not only simplifies the previous multi-stage VLP procedure but also achieves mutual benefits between localization and understanding tasks.
- Experimental results show that a single GLIPv2 model (all model weights are shared) achieves near SoTA performance on various localization and understanding tasks.
- Show strong zero-shot and few-shot adaption performance on open-vocabulary object detection tasks and superior grounding capability on VL understanding tasks
- Microsoft Research Blog Post: [Object Detection in the Wild via Grounded Language Image Pre-training](#).

**Microsoft Research, Redmond** – Deep Learning Team

2021

My research project is to present a grounded language-image pre-training (GLIP) model for learning object-level, language-aware, and semantic-rich visual representations, supervised by Pengchuan Zhang, Jianwei Yang, Chunyuan Li, Lu Yuan, Lei Zhang, and Jianfeng Gao.

**Research Intern – Deep Learning**, “GLIP: Grounded Language-Image Pre-training”

- When directly evaluated on COCO and LVIS (without seeing any images in COCO during pre-training), GLIP achieves 49.8 AP and 26.9 AP, respectively, surpassing many supervised baselines.
- After fine-tuned on COCO, GLIP achieves 60.8 AP on *val*, and 61.5 AP on test-dev, surpassing previous SoTA.
- When transferred to 13 downstream object detection tasks, a few-shot GLIP rivals with a fully supervised Dynamic Head.

**Microsoft AI & Cloud, Redmond** - Visual Document Intelligence Team

2020

My research project is to investigate pre-training models with additional visual modalities by involving image embeddings in the pre-training steps, supervised by Yijuan Lu, Dinei Florencio and Cha Zhang.

**Research Intern – Computer Vision**, “Visual LayoutLM: Involving Visual Features in the Pre-training Stage of LayoutLM”

- Adopted a simple yet powerful Transformer model as the backbone and extends it to take both visual and text embedded features.
- Designed the Masked Language Model (MLM) and Image-Text Matching (ITM) to jointly model interactions between language, layout, and rich visual information.
- Visual-LayoutLM model has shown its potential to outperform the original LayoutLM and other SOTA models in several document understanding tasks.

## **LEADERSHIP ACTIVITIES**

---

**Student Graduate Speaker**

July 2017

- Graduate Commencement speech, School of Electronic, Information and Electrical Engineering (SEIEE), Shanghai Jiao Tong University.

**UW-SJTU Summer Intensive Program, Student Captain**

July 2015

- Served as the captain for a class of more than 50 students. An interview was posted once on UW homepage ([video URL](#)).

**SEIEE Students' Union, Vice President**

April 2014

- Elected to be the Associate Chair in the freshman year, organized various activities for the college and school.
- Owned a sense of responsibility, constantly up of spirit of enterprise, practical spirits and realistic attitude.

## **PUBLICATIONS**

---

- [1] **Haotian Zhang\***, Pengchuan Zhang\*, et al, “**GLIPv2: Unifying Localization and Vision-Language Understanding**”, accepted to NeurIPS 2022, New Orleans, USA. (Spotlight.)
- [2] Liunian Li\*, Pengchuan Zhang\*, **Haotian Zhang\***, et al., “**GLIP: Grounded Language Image Pre-training**”, accepted to CVPR 2022, New Orleans, USA. (Oral & Best Paper Finalist.)
- [3] **Haotian Zhang**, et al., “**U3D-MOLTS: Unified 3D Monocular Object Localization, Tracking and Segmentation**”, accepted to 6<sup>th</sup> BMTT Challenge Workshop, ICCVW 2021. (1<sup>st</sup> ranking on KITTI-STEP leaderboard.)
- [4] **Haotian Zhang**, Haorui Ji, Aotian Zheng, Ren-Hung Hwang, Jenq-Neng Hwang, “**6D-LOCNet++: 6DoF Localization Network for Monocular 3D Object Detection**”, accepted to IJCV, 2021.
- [5] **Haotian Zhang**, Aotian Zheng, Haorui Ji, Ren-Hung Hwang, Jenq-Neng Hwang, “**Monocular 3D Localization on Road Scenes**”, accepted to 2<sup>nd</sup> AVV Workshop, ICCV 2021.
- [6] **Haotian Zhang**, Yizhou Wang, Jiarui Cai, Hung-min Hsu, Haorui Ji, Jenq-Neng Hwang, “**LIFTS: LiDAR and Monocular Image Fusion for Multi-Object Tracking and Segmentation**”, accepted to 5<sup>th</sup> BMTT Challenge Workshop, CVPRW 2020, Seattle, USA. (1<sup>st</sup> ranking on KITTI-MOTS leaderboard.)
- [7] Jiarui Cai, Yizhou Wang, **Haotian Zhang**, Hung-min Hsu, Chengqian Mao, Jenq-Neng Hwang, “**IA-MOT: Instance-Aware Multi-Object Tracking with Motion Consistency**”, accepted to 5<sup>th</sup> BMTT Challenge Workshop, CVPRW 2020, Seattle, USA.
- [8] Longyin Wen\*, **Haotian Zhang**, Jenq-Neng Hwang, et al., “**VisDrone-MOT2019: The Vision Meets Drone Multiple Object Tracking Challenge Results**”, accepted to ICCVW 2019, 27 October, Seoul, Korea.
- [9] **Haotian Zhang**, Gaoang Wang, Jenq-Neng Hwang, “**Eye in the Sky: Drone-Based Object Tracking and 3D Localization**”, accepted to ACM Multimedia 2019, 21 - 25 October 2019, Nice, France. (Oral & Highlight.)
- [10] **Haotian Zhang**, Yanting Zhang, Jie Yang, Jenq-Neng Hwang, “**Adaptive Local Bundle Adjustment using single camera and traffic signs**”, accepted to IEEE ICIP 2019, 22 - 25 September, Taipei, Taiwan.
- [11] Gaoang Wang, Yizhou Wang, **Haotian Zhang**, Renshu Gu, Jenq-Neng Hwang, “**Exploit the Connectivity: Multi-Object Tracking with TractletNet**”, accepted to ACM Multimedia 2019, 21 - 25 October 2019, Nice, France.
- [12] **Haotian Zhang**, Yanting Zhang, Gaoang Wang, Jie Yang, and Jenq-Neng Hwang, “**Bundle Adjustment for Monocular Visual Odometry Based on Detections of Traffic Signs**”, accepted to IEEE Transaction on Vehicular Technology (TVT).
- [13] **Haotian Zhang**, Min Tang, and Jun-fa Mao, “**Electron-Thermal Simulation of Graphene Nanoribbon Interconnects**”, accepted to IEEE AECS 2017 Conference (Oral & Best Student Paper).
- [14] **Haotian Zhang** and Min Tang, “**The High Frequency Effects on Multilayer Graphene Ribbon THz Resonators**”, accepted to IEEE 87th ARFTG Microwave Measurement Conference.