

# Xinyu Yang

Shanghai Jiao Tong University, P.R. China

✉ hanyuezhuohua24@gmail.com | 🌐 xinyuyang.me/ | 📱 Hanyuezhuohua | 🎓 Xinyu Yang

## Education

### Shanghai Jiao Tong University

Bachelor of Engineering in Computer Science, ACM Honors Class

Shanghai, China

Sept. 2019 - Present

- **ACM Honors Class** is an elite CS program for students ranked in the top 5% with aspirations in research.
- Overall GPA: **90.95 / 100**, Ranking: **1 / 30**
- Awarded **national scholarship** twice

## Research Interests

Efficient Machine Learning, Machine Learning and Systems, Efficient AI4Science.

## Publications

\* indicates equal contributions

### Variational Inference for Training Graph Neural Networks in Low-Data Regime through Joint Structure-Label Estimation

KDD 2022

Danning Lao\*, **Xinyu Yang\***, Qitian Wu and Junchi Yan

### FlatFormer: Flattened Window Attention for Efficient Point Cloud Transformer

CVPR 2023

Zhijian Liu\*, **Xinyu Yang\***, Haotian Tang, Shang Yang and Song Han

### BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation

ICRA 2023

Zhijian Liu\*, Haotian Tang\*, Alexander Amini, **Xinyu Yang**, Huizi Mao, Daniela Rus and Song Han

### Relational Out-of-Distribution Generalization

Preprint

**Xinyu Yang\***, Huaxiu Yao\*, Xinyi Pan, Shengchao Liu, Pang Wei Koh and Chelsea Finn

### Multi-domain Long-Tailed Learning By Augmenting Disentangled Representations

Preprint

Huaxiu Yao\*, **Xinyu Yang\***, Allan Zhou and Chelsea Finn

## Honors & Awards

National Scholarship of P.R. China (**Top 0.2%, nationwide**)

2021, 2022

SenseTime Scholarship (**Top 30 undergraduate in the field of AI, nationwide**)

2021

Han-Ying-Ju-Hua Scholarship (**15 winners each year, Shanghai Jiao Tong University**)

2021

Fan Hsu-Chi Scholarship (**10 winners each year, Shanghai Jiao Tong University**)

2020

Academic Excellence Scholarship (**Top 5%, Shanghai Jiao Tong University**)

2020, 2021, 2022

Zhiyuan Honorary Scholarship

2019, 2020, 2021

## Research Experience

### Stanford University

Stanford, CA, USA

Research Intern at IRIS, advised by Prof. Chelsea Finn

Mar. 2022 - Present

#### • Multi-Domain Long-Tailed Learning

Real-world data often involves multiple domains with distinct imbalanced class distributions. We study this multi-domain long-tailed learning problem and introduce **TALLY** that generalizes well across all domains and classes. **TALLY** empowers augmentation to balance examples over domains and classes by decomposing and reassembling example pairs.

#### • Relational Out-of-Distribution Generalization

In distribution shifts, some test domains only correlate with a few training domains. In this work, we introduce ROOG to harness the power of domain relations derived from domain meta-data. ROOG tailors the generalization process to each test domain by reweighting the training domains according to the given domain relation.

#### • Out-of-Distribution Type Detection

Out-of-distribution detection is critical to ensuring the reliability of machine learning systems. However, prior works only focus on detecting semantic shifts, ignoring other shift types. In this work, we aim to develop an end-to-end framework that detects semantic/covariate/concept shifts and further adapts pre-trained models to changing environments by selectively fine-tuning according to the detected shift types.

### • Efficient Point Cloud Transformer

For 3D point cloud transformers, their latency lags behind sparse convolutional models though achieving on-par performance. In this work, we present **FlatFormer** to close this latency gap by trading spatial proximity for better computational regularity. FlatFormer delivers **SOTA** results on Waymo Open Dataset with **4.6×** speedup over (transformer-based) SST and **1.4×** speedup over (sparse convolutional) CenterPoint. It is the **first** point cloud transformer achieving real-time inference on edge GPUs.

### • Multi-Sensor Fusion in Autonomous Driving System

Multi-sensor fusion is essential in autonomous driving system. In this work, we propose **BEVFusion**, an efficient multi-task multi-sensor fusion framework that unifies multi-modal features in the shared BEV representation space. On nuScenes, It achieves **1.3%** higher mAP on 3D object detection and **13.6%** higher mIoU on BEV map segmentation, with **1.9×** lower computation cost.

### • 3D Panoptic Segmentation

In this project, we apply SPVCNN to panoptic segmentation with an instance branch predicting the offsets to the instance center. Built upon the MeanShift clustering algorithm, SPVCNN++-Panoptic regresses the instance centers of 'things' points to predict instance labels. It won **2nd** place in nuScenes panoptic challenge while achieving **SOTA** panoptic quality in two particle reconstruction tasks in physics with data from HGAL.

### • 3D BEV Map Segmentation Benchmark

In most autonomous drive benchmarks (e.g. nuScenes), their official split is specific for detection task, where training and test scenes may include same regions from distinct timestamps. This leads to data leakage in 3D BEV map segmentation task. Therefore, we re-split scenes for this task and propose a unified and easy-to-use BEV map segmentation benchmark on nuScenes.

## Shanghai Jiao Tong University

Undergraduate Researcher at *ThinkLab*, advised by *Prof. Junchi Yan*

Shanghai, China

Jun. 2021 - Aug. 2022

### • Data-Efficient Graph Learning through Structure-Label Joint Estimation

In real-world scenarios where complete input graph structure and sufficient node labels may not be achieved easily, GNN models encounter severe performance degradation. To address this problem, we propose **WSGNN**, a flexible probabilistic generative framework which harnesses variational inference to solve weakly supervised graph learning in a label-structure joint estimation manner. WSGNN achieves **SOTA** results on both node classification and link prediction over real-world benchmarks.

## Selected Projects

### Mx\* Compiler

Assembly Language, Code Generation and Optimization, LLVM IR, ANTLR

Java

Jan. 2021 - Jun. 2021

- Developed a compiler for C-and-Java-like language (Mx\*) to NASM with **10k+ lines of codes**, which performance is **close to GCC-02**.
- Implemented optimization algorithms such as sparse conditional constant propagation, function inline and static single assignment.

### Traffic Signal Controller

Reinforcement Learning, DQN, City-scale Road Network

Python

Jan. 2021 - Jun. 2021

- Coursework of "Machine Learning" (31st place in *KDD CUP 2021 - City Brain Challenge*)
- Designed an agent that coordinates the traffic signals to maximize number of vehicles served while maintaining an acceptable delay.

### RISC-V CPU

Computer Architecture, Tomasulo, FPGA Programming

Verilog

Nov. 2020 - Dec. 2020

- Designed a RISC-V CPU that supports RV32I Base Integer Instruction Set V2.0 (2.1-2.7).
- Implemented a **Tomasulo's algorithm** with load/store buffer that is more efficient than a five-stage pipelined one.

## Teaching Experience

**Teaching Assistant** *Compiler Design and Implementation*

Fall 2021

**Teaching Assistant** *Principle and Practice of Computer Algorithms*

Summer 2021

**Teaching Assistant** *Data Structures*

Spring 2021

## Skills

### Programming

Python, C++, JAVA, Golang, Verilog

### Tools and Frameworks

Git, PyTorch, TensorFlow, Docker, Vivado, Markdown, LaTeX

### Languages

Mandarin (native), English (TOEFL: 109)