CHENHUAN CAI

Mobile: (+41) 076-752-7703, Email: chenhuancai1125@gmail.com
Website: https://caichenhuan.github.io
8051 Zürich, Switzerland

EDUCATION

University of Zurich — Zurich, Switzerland
M.S. in Artificial Intelligence, Faculty of Business, Economics and Informatics — Sep. 2023 - Present
Teaching Assistant for Foundations of Data Science (FDS), Fall 2024

Swiss Federal Institute of Technology in Zurich (ETH Zurich) — Zurich, Switzerland
Visiting Student in Computer Science, enrolled in core courses — Sep. 2023 - Present
Core Courses: Introduction to Machine Learning, Computer Vision, Digital Humans, 3D Vision, etc.

Harbin Institute of Technology, Shenzhen (HIT) — Shenzhen, China
B.S. in Data Science and Big Data Technology — Sep. 2019 - Jun. 2023
GPA: 84.59 / 100
Core Courses: Advanced Mathematics, Advanced Algebra and Geometry, High-level Language Programming, Data Structures, etc.

National University of Singapore Suzhou Research Institute (NUSRI) — Suzhou, China
Exchange Program, Industrial Systems Engineering and Management (ISEM) — Sep. 2022 - Jun. 2023

RESEARCH & INTERNSHIP

3D Vision Research Assistant — Zurich, Switzerland
Advisor: Prof. Marc Pollefeys, ETH Zurich — Mar. 2025 – Present

➤ Project foundation: Built on *NoPoSplat*, a cutting-edge method that reconstructs 3D Gaussian representations from sparse multi-view images without camera poses. Current work targets higher reconstruction quality when views are few and overlap is minimal.
➤ Multi-view extension: Incorporate a Fusion Transformer into NoPoSplat to fuse information from additional viewpoints.
➤ Intrinsic-free reconstruction: Replace the traditional camera-intrinsic embedding with a prediction network, aiming for a pipeline that is completely independent of intrinsic parameters.
➤ Generative prior: Introduce a diffusion model as a generative prior for 3D Gaussians, markedly boosting both the quality and stability of the reconstruction.

Computer Vision Algorithm Engineer — Beijing, China
NetEase — Apr. 2022 – Jun. 2022

➤ Contributed to the visual-detection R&D for NetEase Youdao's "Aladdin" smart desk lamp, leading the finger-on-paper contact module and its feasibility study.
➤ Phase 1: Generated pseudo-labels and applied image preprocessing; evaluated image-classification and YOLO-based object-detection approaches (YOLOv5, YOLOX), ultimately achieving reliable separation with YOLOX under lights-on conditions.
➤ Phase 2: Expanded the dataset, performed extensive data augmentation, and explored tracking-based methods; trained a model that surpassed 80 % accuracy in lights-off scenarios, validated feasibility, and enabled the module's integration into later product releases

Machine Learning Research Assistant — Remote
Advisor: Prof. Ding Jun, McGill University — July. 2022 – Nov. 2022

➤ Worked on biological information and active learning projects, responsible for user interaction, basic analysis, and visualization functions. Built a Dash-based web page with features like file upload/download, Anndata preprocessing, dimension reduction, clustering, and visualization.
➤ Handle the data processing section using Scanpy for cell data processing and analysis, and Velocyto for cellular trajectory inference in lung epithelial cell differentiation.

COMPETITIONS & PROJECTS

Kaggle: Stable Diffusion - Image to Prompts Team Member | *Silver medal (Top 5%)* — Feb. 2023 – May. 2023
Predict text prompts based on generated images.

➤ Dataset creation: We first verified feasibility with the COCO and official SD2-v2 datasets, then built the final training set by pairing POLO labels with 1.8 million images generated by SD2-v2.
➤ Model training: Text labels are encoded into 384-dimensional vectors using a Sentence-Transformer, while images are projected into the same space by ViT or ConvNeXt; the cosine distance between the two vectors is used as the loss.
➤ Model refinement: To raise similarity we apply layer-wise learning rates (higher in lower, task-specific blocks), raise input resolution, re-initialize the final linear layer with a normal distribution, disable dropout, and incorporate adversarial weight perturbation (AWP) for added robustness.
➤ Inference: Each base model merges four cross-validation folds with one full-data model (five sub-models total). Multiple base models are then combined via weight-search ensembling to yield the final predictions.

China Undergraduate Mathematical Contest in Modeling | *2nd Prize (Top 3.3%)* — Sept. 2021
Conduct quantitative analysis and modeling for the data set of chemical experiment results.

➤ We used correlation coefficient matrix, OLS, and goodness of fit methods to study variable correlations.
➤ After training multiple models, they chose XGBoost along with SHAP to investigate the influence of different features.
➤ To optimize the solution and maximize the target value, they utilized an improved Particle Swarm Optimization (PSO) algorithm.

PUBLICATION

Wu, Y.*, **Cai, C.***, Li, H., Bai, Z. & Yan, L. (2025). **STaR: Multi-Granular Spatio-Temporal Reasoning for Long-Form Dense Video Captioning.** *Accepted at the 28th European Conference on Artificial Intelligence (ECAI 2025). (* Equal contribution.)*
Long-form DVC aims to identify multiple event segments in untrimmed videos and generate natural language descriptions for each segment.

➤ Proposes the **STaR**, a framework integrates global, local, and spatial features through a Multi-Granular Spatio-Temporal Reasoning.
➤ Introduces the SSS and CC modules, which enhance spatial position awareness and global context understanding, respectively.
➤ By incorporating a Large Language Model (LLM), STaR improves caption generation quality on the SoccerNet-Caption dataset.

SKILLS

Skills: *Python, Pytorch, Numpy, Pandas, Linux, Machine Learning, Deep Learning, Data Mining, Computer Vision, 3D Vision, etc.*
Language: Chinese (native); English (business)