Haosheng Tan

su21822@alumni.bristol.ac.uk | 0788 0909 024

Research Interest

- Embodied Multimodality Models: *Embodied-VQA, Spatial Vision Understanding*
- Multi-agent Collaboration System on LLMs: *Code Generation, Software Development, Data Analytics and Visualization*
- Data-Driven DeepLearning

Education

University of Bristol, MSc in Data Science   Sept 2021 – Nov 2022

- Score: 72/100 (Distinction)
- **Related Courses:** Statistical Computing and Empirical Methods(*score 78*)Data Analysis(text and visual), Artificial Intelligence (statistical methods, reinforcement learning, *score 80*) , Deep Learning(NLP,*score 80*), Cloud Computing(AWS)

South China Agricultural University, BA in Communication Engineering   Sept 2017 – June 2021

- GPA: 4.0/5.0 (top 5%)
- **Related Courses:** Machine Learning, Calculus, Linear Algebra, C Programming language and C++.

Work Experience

Data Engineer, Kuaishou Technology – Beijing, China   July 2023 – June 2024

- Department of Data Platform - Core Data Asset Team
- Goal of Team: Generating **data assets** which are general tags or features of an object (users, products, streaming) to serve multiple sub-businesses of the whole company.
- Responsible for: Data aggregation and fusion for app installation and activation signals for users, generating the general tag **App preference** for each user. This tag is mainly used for the company's advertising and marketing department.
- Specific Tasks: Data Generation with **Hive SQL**, Data Analysis, and **Feature Selection** using Machine Learning models before feeding into the recommendation system

Research Assistant, University of Glasgow – College of Science and Engineering, UK   July 2024 – Now

- Research Team Goal - Time series prediction for wind power energy sector
- Responsible for: Assisting the task of **Dynamic Covariance Modelling** for correlation-based statistical learning model.

Research Experience

Modelling the covariance matrix of UK regional electricity demand   June 2022 – Sept 2022
University of Bristol – Position: Research Project Owner

- Goal: The goal of this project is to model the covariance of electricity demand in 4 UK regions using GAM including other covariates. Gradient Boosting for multiple distribution parameters is implemented for fitting.
- Responsible for: Constructing GAM framework and performing Gradient Boosting

Covid-19 Single-cell Differential Analysis with Explainable Machine Learning Models   Feb 2022 – June 2022
University of Bristol – Position: Research Group Member

- Goal: Find the most effective features from Covid-19 medical data that can differentiate patients in severity
- Responsible for: T cells analysis and visualization using Umap, Building and training dense neural network for cell classification, Feature explanation of neural network, Constructing pipeline.

Deep convolutional neural network for license plate detection and recognition   Oct 2020 – Mar 2021
South China Agricultural University – Position: Research Project Owner

- Goal: Developing robust license plate detection and recognition system via convolutional neural networks
- Responsible for: Data Collection and Labeling, Data Cleaning, Building up U-Net and a Modified VGG-16 Network using TensorFlow Keras Framework, Model Training, Hyperparameters Tuning
- Result: Achieving 92% accuracy compared to traditional image processing and can be adopted in poor illuminated scenarios.

Citrus Pest Detection Using SVM(Competition Project)   May 2020 – Sept 2020
South China Agricultural University – Position: Team Leader

- Goal: Identify pest-infected citrus from images captured by drone camera in the farm and classify citrus pests
- Responsible for: Image pre-processing, Vision Feature Selection(SIFT and ORB), SVM training for pest detection
- Result: Achieving over 80% accuracy. This project won the first prize at the *1st Artificial Intelligence Innovation Contest held by the College of Electronic Engineering, South China Agricultural University*

Publications

Data Renovation for Computer Vision Datasets   Expected Date: May 2025
Haosheng Tan, Jiaheng Wei, Zirui Pang, Yuhan Pu
Expected Conference: NeurIPS 2025

Awards

1st Artificial Intelligence Innovation Contest (2020): *First Prize*
Scholarship(2018,2019,2020): *Third Class*
Intelligent Mobile in College of Artificial Intelligence and Electronic Engineering (2018): *Best Theory Prize*

Skills

Languages: Python, SQL, C++, R
Deep Learning: Transformer, UNet, ViT, Diffusion Model, PyTorch, TensorFlow
Data Warehouse: HiveSQL, Spark, Kafka, Redis for API
Data Analysis: Textual and Visual
Statistical Learning: Generalized Linear Model, Dynamic Covariance Modelling, Gradient Boosting, Diffusion
AWS Cloud Computing: AWS academic certificate

Language

English: IELTS: Overall: 7.0 Listening: 7.5 Reading: 7.5 Writing: 6.5 Speaking: 6.0
Mandarin: Local
Cantonese: Local