Bowen Jing

Arthur12137 linkedin github Manchester, UK 07751859915 Personal Page Google Scholar

EDUCATION AND PROFESSIONAL EXPERIENCE

Tsinghua University Feb. 2025 – Aug. 2025
Research Intern, Institute for AI Industry Research (AIR), Beijing, China

University of Manchester: ***MSc.*** *Advanced Computer Science* - ***Grade :*** *Distinction (76)* Sept. 2023 - Sept. 2024

University of Manchester: ***BSc.*** *Computer Science* 2020 - 2023

PHD RESEARCH INTEREST SUMMARY

My long-term research goal is to develop autonomous agents capable of accurate perception, goal-driven decision-making, and continual learning in uncertain, real-world environments. I am particularly interested in leveraging diffusion models and other generative frameworks for structured world modeling. In the short term, I aim to explore how these models can be used for realistic video generation—producing physically plausible, temporally consistent data that can serve as high-quality training material for corner cases in autonomous driving.

PROJECT EXPERIENCE

StyleDrive: Driving-Style Aware Benchmarking for End-to-End Autonomous Driving [link]:

- **Developed StyleDrive**, the **first real-world large-scale dataset** for personalized end-to-end autonomous vehicles.
- **Constructed** over 30,000 real-world driving scenarios enriched with fine-grained annotations for A,N,C driving styles, derived through a hybrid pipeline integrating HD map topology, motion heuristics, and fine-tuned Video-LLaMA3.
- **Designed** a multi-stage annotation framework combining objective behavior analysis and subjective VLM-based reasoning, with human-in-the-loop validation for high-quality, interpretable style labels.
- **Introduced** a novel evaluation metric—**Style-Modulated PDMS (SM-PDMS)**—to assess not only safety and feasibility, but also alignment with user-intended driving styles.
- **Benchmarked** four representative E2EAD architectures (AD-MLP, TransFuser, WoTE, DiffusionDrive) under style conditioning, demonstrating substantial improvements in behavior alignment and human trajectory closeness.
- **Validated** the effectiveness of style modeling through both closed-loop evaluation and L2 trajectory error against human demonstrations, with style-aware DiffusionDrive achieving the best overall performance.

SignBart: Gloss-Free Sign Language Translation for Human-Robot Interaction(submitted to ICRA 2026):

- **Proposed** a novel gloss-free sign language translation system tailored for real-world human-robot interaction, addressing scalability and deployment issues in robotic vision-language understanding.
- **Designed** a spatiotemporal visual encoder called **CSIFE-ConvNeXt**, which integrates multi-branch large kernel convolutions and inter-feature enhancement strategies, achieving 58% parameter reduction while maintaining translation accuracy.
- **Introduced TTT-mBART**, the first adaptation of Test-Time Training to sign language translation; this linear-complexity, self-supervised module enables dynamic adaptation to unseen distributions during inference.
- **Achieved state-of-the-art performance** on PHOENIX-2014T and CSL-Daily datasets, outperforming GFSLT and SignBERT+ on BLEU-4 and ROUGE metrics, with significant gains in long-form semantic coherence and test-time adaptability.
- **Conducted ablation studies** demonstrating that each component (CSIFE-ConvNeXt, TTT-mBART, and data augmentation) contributes positively to both performance and computational efficiency.
- **Targeted real-world deployment** in assistive service robots through an efficient, gloss-free, end-to-end SLT pipeline with high generalization to user and environment variability.

MInM: Mask Instance Modeling for Visual Representation Learning (submitted to AAAI 2026):

- **Proposed** a novel self-supervised framework (MInM) that leverages **instance-aware semantic masks** from SAM2 to replace random masking in masked image modeling (MIM), enabling more meaningful and efficient representation learning.
- **Integrated** MInM into the MAE pipeline without altering model architecture, resulting in **faster convergence and lower reconstruction loss** on ImageNet-1K pretraining.
- **Demonstrated** improved generalization on multiple downstream tasks using pretrained ViT-B backbones, including **object detection (Pascal VOC 2007, MS COCO)**.
- **Achieved consistent gains** in object-centric understanding, especially for complex categories like *cat*, *dog*, and *sofa*, outperforming MAE and SimMIM baselines.
- **Positioned** MInM as a lightweight yet effective drop-in enhancement for semantic-aware pretraining, with promising applications in medical imaging, robotics, and autonomous driving.

VDiff-SR: Multi-Scale Visual Autoregressive Guided Diffusion for Real-World Image Super-Resolution:

- **Co-authored a NeurIPS submission** proposing a hybrid diffusion framework (VDiff-SR) for image super-resolution by integrating coarse-to-fine multi-scale structural priors from a pretrained Visual AutoRegressive (VAR) model.
- **Designed and implemented** two novel modules: the Condition-Gated Unit (CoGU) for efficient feature conditioning and the Cross-Scale Prior-Aligned Attention (CSPA) for hierarchical structural alignment across scales.

- **Incorporated an inverse learning strategy** to bridge latent priors and noise prediction, enabling tighter coupling of denoising steps with global semantic structures.
- **Outperformed** existing SOTA methods on real-world datasets (RealSR, RealSet5), achieving best-in-class perceptual metrics (CLIP-IQA and MUSIQ), with robust performance across both reference-based and no-reference evaluation protocols.
- **Validated model efficacy** via comprehensive ablation studies across all modules, demonstrating consistent performance gains and highlighting the critical role of multi-scale prior fusion.

End-to-End Autonomous Driving System [link]: Technologies: Python, CARLA, PyTorch, Middle Fusion Mar. 2024–Sept. 2024

- **Deployed** an advanced multi-modal autonomous driving model by fusing camera data and LiDAR data through middle fusion techniques with channel attention, augmented by a GRU-based network for dynamic waypoint prediction. This integration enabled proactive vehicle responses to environmental changes, including moving entities.
- **Rigorous experimental evaluations** involving various ResNet architectures and ablation studies, enhancing the precision of component impact assessments on overall system efficacy.
- **Collected** novel training data and DVS data in CARLA simulator, enriching training under complex scenarios.
- **Compared** with other SOTA model
- **Achieved and substantiated** exceptional model performance metrics, with a Driving Score (DS) of 29.79, Route Completion (RC) of 46.3%, and an Infraction Score (IS) of 0.74, through extensive scenario-based evaluations.
- **Attained an academic distinction** with a score of 76, reflecting high scholastic achievement and expertise in the field.

Comparative Analysis of Deep Learning and Traditional Computer Vision in Robotic Vision [link]: Technologies: ResNet, Vision Transformers (ViT), Bag-of-Visual-Words (BoVW), SIFT, K-Means, SVM April 2024–May 2024

- **Conducted a comprehensive comparative study** between traditional handcrafted feature pipelines and modern deep learning-based vision architectures, targeting image classification for robotic perception tasks.
- **Implemented and optimized** multiple models including ResNet and Vision Transformers (ViT), trained from scratch and fine-tuned with ImageNet21k weights, to evaluate learned feature representations.
- **Engineered a classical BoVW pipeline** from the ground up using SIFT descriptors, K-Means clustering for visual vocabulary generation, and SVM classification, offering a strong baseline for feature-based vision.
- **Designed and curated noise-augmented datasets** derived from CIFAR-100 and Caltech-101, injecting synthetic Gaussian noise at three levels (low/medium/high) to rigorously assess model robustness under degraded conditions.
- **Analyzed model behavior** through confusion matrices, class-wise accuracy breakdowns, and training dynamics, developing insight into how attention-based architectures handle inter-class similarity compared to local-descriptor-based methods.
- **Demonstrated end-to-end ML proficiency**, including dataset augmentation, model training, performance benchmarking, and critical interpretation of quantitative and qualitative results.

Sentence-Level Relation Extraction [link]: Technologies: PA-LSTM, Bi-LSTM, C-GCN, BERT, RoBERTa, SpanBERT

- **Implemented and analyzed** a suite of neural architectures for sentence-level relation extraction, comparing sequential (PA-LSTM, Bi-LSTM), structural (C-GCN), and pretrained transformer-based models (BERT, RoBERTa, SpanBERT).
- **Designed a novel entity marker injection strategy** for BERT-based models, improving over standard entity masking by enabling explicit modeling of entity-pair interactions via learned token embeddings.
- **Engineered multiple experimental conditions** on three benchmark datasets (TACRED, Re-TACRED, TACREV), including:
  - Constructing controlled subsets of sentences with varying lengths (short: $\leq$15 tokens, medium: 16–30, long: $>$30), revealing transformer robustness in syntactically complex cases.
  - Building reduced training sets (10%, 25%, 50%) to benchmark few-shot capabilities.
- **Conducted rigorous ablation studies** across entity representation types (e.g., hidden state pooling vs. special token concatenation), and input granularity, to disentangle the sources of performance gains in pretrained models.
- **Led a 3-person team**, defining milestones and managing deliverables, while contributing key technical insights.
- **Achieved top-5 performance in class**, scoring 92.75% with a well-justified analysis report, praised for its methodological rigor and clarity of comparison across paradigms.

PRE-PRINT ARTICLES

- **A Comprehensive Guide to Explainable AI: From Classical Models to LLMs** Link
- **Exploring Multimodal Embeddings for Text and Impact on Language Processing** Link
- **Multimodal Embeddings for Representation Learning** Link
- **Deep Learning and Machine Learning, Advancing Big Data Analytics and Management: Generative Models** Link

ACADEMIC REFEREES

Prof. Robert Stevens – MSc Dissertation supervisor Email
Dr. Oliver Rhodes – BSc Dissertation advisor Email
Dr. Riza Batista-Navarro – Text Mining course mentor Email

HONORS AND AWARDS

- Academic Representative (2023–2024), ACS Artificial Intelligence
- Team Leader for multiple university team coursework projects (COMP23412, COMP61332)
- University of Manchester Kilburn Entry Scholarship (awarded for achieving A*A*A* at entry)