Bowen Jing

Arthur12137 linkedin github Manchester, UK 07751859915 Personal Page Google Scholar

EDUCATION AND PROFESSIONAL EXPERIENCE

Tsinghua University Feb. 2025 – Oct. 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

Bowen Jing seeks to pursue a PhD in **Autonomous Systems**, focusing on **generative and world-model-based intelligence**. His research integrates **Vision-Language-Action (VLA)** modeling, **diffusion models**, and **simulation-driven learning** to unify perception, reasoning, and control in embodied agents. His long-term goal is to enable autonomous systems that can *learn, adapt, and reason safely* in uncertain and dynamic environments. His research portfolio currently has over **50 citations on Google Scholar**.

RESEARCH EXPERIENCE

StyleDrive: Driving-Style Aware Benchmarking for End-to-End Autonomous Driving (AAAI 2026 Oral Acceptance) [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)**.
- **Bridged** the Vision-Language-Action paradigm to real-world autonomous driving by conditioning action generation on linguistic style descriptors, demonstrating how language-guided modulation enhances perception-to-control reasoning.
- **Positioned** StyleDrive as an embodied VLA benchmark, where vision encoders, language-based style priors, and trajectory decoders jointly enable personalized and interpretable driving behaviors.

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.
- **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.
- **T**his system is a direct implementation of a **Vision-Action** model. Its modular design (perception backbone + GRU policy head) provides a clear framework for integrating the **Language** modality, directly aligning with the PhD's goal of building collaborative VLA agents.

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.
- **Positioned** SignBart as a foundation for embodied collaborative agents, where multimodal translation (vision → language) serves as the perceptual front-end for subsequent action reasoning and cooperative decision-making.

CounterScene: Counterfactual Diffusion for Adversarial Closed-Loop Traffic Simulation (CVPR 2026 Submission) [link]:

- **Developed CounterScene**, a counterfactual diffusion framework for **adversarial yet physically consistent** traffic simulation.

- **Introduced causal coherence constraints** and modeled trajectory-level **geometric intersections** as causal triggers for localized counterfactual interventions on agents' control variables.
- **Designed** a differentiable **counterfactual guidance mechanism** that steers diffusion sampling toward physically plausible and causally aligned outcomes.
- **Achieved** superior trade-offs between realism, controllability, and causal interpretability across **nuScenes** and closed-loop simulation benchmarks.
- **Positioned** CounterScene as a principled framework bridging **counterfactual reasoning** and **generative diffusion modeling** for safety-critical autonomous driving evaluation.

VDiff-SR: Multi-Scale Visual Autoregressive Guided Diffusion for Real-World Image Super-Resolution(CVPR 2026 Submission):

- **Co-authored a CVPR 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.

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).
- **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.

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.

KEY AI & DEEP LEARNING SKILLS

Programming & Frameworks: Python | PyTorch | TensorFlow | JAX | NumPy | CUDA | OpenCV | ONNX
Generative & Foundation Models: Diffusion Models| Transformers (ViT, LLaMA, BERT) | VAE | RAG | Prompt Engineering
Embodied & Autonomous Systems: CARLA | nuPlan | nuScenes | ROS | Isaac Gym | Gymnasium
Learning & Optimization: Contrastive Learning | Self-Supervised Learning | Causal Inference | Reinforcement Learning (D4PG, PPO)
Simulation & Evaluation: Closed-loop Evaluation | Behavior Cloning | Scenario Generation | Counterfactual Simulation
Tools & Research Practices: Linux | Weights&Biases | HuggingFace | Git | HPC Training | Dataset Design | Academic Writing

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 Model Security: Threats and Defenses** Link

HONORS AND AWARDS

- Academic Representative (2023–2024), ACS Artificial Intelligence
- Team Leader for multiple University Team Coursework Projects (COMP23412, COMP61332)