Binzhu Xie

+86 15656533660 / +852 56051423 | bnizhuxie@gmail.com | https://nicous20.github.io/

EDUCATION BACKGROUND

Queen Mary University of London E-Commerce Engineering with Law
- Bachelor of Engineering **GPA:** 3.7/4; **Ranking:** 18/180. 09/202006/2024

Beijing University of Posts and Telecommunication E-Commerce Engineering with Law
- Bachelor of Engineering **GPA:** 88.8/100; **Ranking:** 18/180. 09/2020-06/2024

My interests: Visual Understanding (VLLM), Vision 3D (with XR), LLM Scaling Analysis, Generative AI.

PUBLICATIONS

[1]. Yuyang Jiang, **Binzhu Xie**, Lina Xu, Xiaokang Lei, Shi Qiu, Luwen Yu, Pan Hui. Generative Multi-Sensory Meditation: Exploring Immersive Depth and Activation in Virtual Reality. **ACM MM 2025**

[2]. Sicheng Zhang *, **Binzhu Xie** *, Zhonghao Yan *, Yuli Zhang, Donghao Zhou, Xiaofei Chen, Shi Qiu, Jiaqi Liu, Guoyang Xie, Zhichao Lu. Trade-offs in Image Generation: How Do Different Dimensions Interact? **ICCV 2025**

[3]. Zhenghao Xing *, Hao Chen *, **Binzhu Xie**, Jiaqi Xu, Ziyu Guo, Xuemiao Xu, Jianye HAO, Chi-Wing Fu, Xiaowei Hu, Pheng-Ann Heng. EchoTraffic: Enhancing Traffic Anomaly Understanding with Audio-Visual Insights. **CVPR 2025**.

[4]. **EgoLife Team**. EgoLife: Towards Egocentric Life Assistant. **CVPR 2025.**

[5]. Shi Qiu, **Binzhu Xie,** Qixuan Liu, Pheng-Ann Heng. Creating Virtual Environments with 3D Gaussian Splatting: A Comparative Study. **IEEE VR** (Poster) 2025.

[6]. Shi Qiu, **Binzhu Xie,** Qixuan Liu, Pheng-Ann Heng. Advancing Extended Reality with 3D Gaussian Splatting: Innovations and Prospects. **IEEE AIxVR** (Poster) 2025.

[7]. **Xie, B. ***, Zhang, S. *, Zhou, Z. *, Li, B., Zhang, Y., Hessel, J., Yang, J. and Liu, Z., 2023. FunQA: Towards Surprising Video Comprehension. **ECCV 2024.**

[8]. Du, H. *, Zhang, S. *, **Xie, B. *,** Nan, G., Tao, X., Jiang X. 2023, Uncovering What, Why and How: A Comprehensive Benchmark of Anomaly Causation of Video Anomaly. **CVPR 2024.**

[9]. Du, H., Nan, G., Zhang, S., **Xie, B.**, Fan, H., Cui, Q., Tao, X., Jiang X., 2023, DocMSU: A Comprehensive Benchmark for Document-level Multimodal Sarcasm Understanding. **AAAI 2024.**

[10]. Chen, J., Wu, M., Yan, H., **Xie, B.**, Zhang, C., 2023. Change-Aware Network for Damaged Roads Recognition and Assessment Based on Multi-temporal Remote Sensing Imageries. **PRCV 2023.**

*** Equal Contribution**

WORK EXPERIENCE

Assistant Research, Chinese University of Honkong, CSE Department 09/2024-Present
- Research Direction is computer 3D vision highly related XR and HCI.

***Full Stack Engineering Intern*, China Petroleum Planning Institute, Refining Department** 05/2023-Present
- Establish a database concerning the international TEMA standard and achieve China's first standard localization: use SQLLITE to maintain database files, use Python and C# to develop test programs by connecting db files, and use Python and HTML for front-end and back-end website development.
- Integrated SQL Server database, utilized Python's numpy and pandas for data analysis, and employed matplotlib and pychart for data visualization.
- Analyzed inter-table relationships and variable characteristics according to company needs, and used MySQL to rebuild the database for the backend and computational development engineers.

***Computer Vision Algorithm Intern*, Perfect World, Sustainable Development Department** 02/2023-05/2023
- Responsible for the scientific research work of the AIGC field: Studied and reproduced classic image generation models and multimodal models, such as GAN, DDPM, DDIM, and CLIP.
- Explored Stable Diffusion and DALLE-2 models. By reproducing controlNet, it provides companies with a variety of IP image virtual persons and publishes them on its own Xiaohongshu platform.
- Created a guide for AIGC mid-journey usage and scientific research support for the cultural and creative competition east of the Forbidden City in Dongcheng District, Beijing.

RESEARCH EXPERIENCE

Uncovering What, Why and How: A Benchmark for Causation Understanding of Video Anomaly **Beijing, China**

Research Assistant, Supervisor: Prof. Guoshun Nan (National Engineering Lab of Network Technology) 09/2023–05/2024
- Researched all benchmarks in the Anomaly Detection field including but not limited to UCF-Crimes, ShanghaiTech, CUHK Avenue, UBnormal, etc., analyzed and designed novel anomaly detection dataset – CUVA for causation understanding.
- Designed and implemented importance curve annotation pipeline. Optimized the importance curve by downstream tasks like Video Captioning, Entailment, and Grounding with using CLIP to measure the similarity between sentences and frames.
- Designed a multimodal model as the novel metrics (MMEval) for evaluating the free-text consistency of VLMs' responses. For this, I chose one VLM as the foundation model and utilized natural language prompts to guide VLMs in specifying the task types in CUVA. Additionally, I utilized curve labels to help VLMs focus more on segments of anomalies within the video.

- Designed a prompted-based method – Anomaly Guardian. Designed soft prompt by CLIP and BERT encoder and hard prompt by multi-round conversations with VLMs.
- One paper was accepted at **CVPR 2024** (Link: https://github.com/fesvhtr/CUVA).

FunQA: Towards Surprising Video Comprehension (GitHub: https://funqa-benchmark.github.io/) **Beijing, China**

Research Assistant, Supervisor: Prof. Ziwei Liu (MMLab@NTU) 04/2023-07/2024

- Designed FunQA benchmark to evaluate LLM capabilities for counter-intuitive videos. Researched all metrics of NLG evaluation and completed the codebase, including BLEU-4, ROUGH-L, CIDEr, and BLEURT. Additionally, I designed new metrics to evaluate the performance of LLMs' free-text responses based on GPT-4, forming a complete FunQA benchmark.
- Proposed the prompt-based method – FunMentor - an agent that refines a VLM's answer through multi-turn dialogues. Then I conducted extensive experiments that proved our method brings VLM performance on FunQA to state-of-the-art levels.
- Reproduced the video-captioning model (mPLUG, GIT, etc) and instruction-based VLMs (video-chat, video-chatgpt, Otter, etc) in the VQA field and finetuned Otter on the FunQA benchmark, and conducted audio-related ablation experiments.
- Based on FunQA and Otter, and jointly with **PAZHOU LAB** to hold a " Talk about the Video" competition with a prize of 1 million yuan.
- One paper was accepted at **ECCV 2024 (Link: https://arxiv.org/pdf/2306.14899.pdf).**

DocMSU: Multimodal Sarcasm Information Recognition Benchmark, Multimodal Information Recognition Dataset and 1Model Construction **Beijing, China**

Research Assistant, Supervisor: Prof. Guoshun Nan (National Engineering Lab of Network Technology) 02/2023-07/2023

- Utilized the Scrapy framework to write a web crawler that collected a significant amount of text and image data for processing and annotation to create the dataset.
- Reproduced code from numerous papers, familiarizing with deep learning models like Bert, ViT, Swin, and GCN.
- Modified SwinTransformer by using BERT and ResNet as encoders for text and images and created multi-modal embedding for pixel-level alignment. I incorporated a binary cross entropy loss function for detection and YoloX head network for localization during inference. This led to state-of-the-art performance on both tasks of localization and detection.
- One paper was accepted at **AAAI 2024.**

HONORS & AWARDS

Xiaomi Corporate Scholarship 10/2022
BUPT First-Class Scholarship 10/2021
Silver Award, 14th & 16th BUPT Programming Contest 11/2020 &11/2022

PROFESSIONAL SKILLS

Language: C++, Java, Python (including libraries such as PyTorch, Numpy, Pandas), C/C++, HTML/CSS/JS, SQL

Technologies/Frameworks: Machine Learning and Deep Learning, LLMs, Data Statistics, Full-stack Development