Binzhu Xie

+86 15656533660 | xbz_nicous@bupt.edu.cn | https://github.com/Nicous20

EDUCATION BACKGROUND

Beijing University of Posts and Telecommunications (BUPT)   09/2020-06/2024 (Expected)
- BSc (Eng) in E-Commerce Engineering with Law (offered jointly by BUPT and Queen Mary University of London)
- GPA: 90.2/100; Ranking: Top 6%

Related Courses: Advanced Mathematics A(I) (91), Linear Algebra (95), Engineering Mathematics (87), Advanced Programming in C(98), Discrete Mathematics(95), Advanced Programming in Java(92), Probability Theory and Mathematical Statistics(96), Data Structures(95), Database(92), Advanced Network Programming(94), Internet Protocols (97), Data Mining(97), Artificial Intelligence and Big Data(97), etc.

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 Mobile Network Technology) 09/2023 - Present
- 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 (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-Present
- 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. And prepared a paper for submission to **CVPR 2024 (Link: https://arxiv.org/pdf/2306.14899.pdf).**

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

Research Assistant, Supervisor: Prof. Guoshun Nan (National Engineering Lab of Mobile 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 (https://github.com/fesvhtr/DocMSU). **.**

Road Disaster Monitoring   **Beijing, China**

Research Assistant, Supervisor: Prof. Ming Wu (BUPT Pattern Recognition and Intelligent System Lab)   07/2022-07/2023
- Used Python tools such as geopandas and shapely for the automated processing of satellite remote-sensing images.
- Studied semi-supervised learning and traditional pattern-matching algorithms for pre- and post-disaster road matching and corresponding target detection.
- Developed an algorithm for road damage monitoring based on change detection and created a high-quality dataset of roads before and after disasters based on current datasets.
- One paper was accepted at **PRCV 2023** (Change-Aware Network for Damaged Roads Recognition and Assessment Based on Multi-temporal Remote Sensing Imageries.).

INTERNSHIPS

***Full Stack Engineering Intern*, China Petroleum Planning Institute, Refining Department**   05/2023-11/2023
- 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.

External Network Maintenance Intern, **Siemens (China), Communications Department** 04/2022-10/2022
- Provided web technology support for the Communications Department, including frontend network maintenance, web page design, and implementation.
- Crawl Siemens webpage using Scrapy and Selenium in Python to detect inappropriate content and incorrect files.

PUBLICATIONS

[1]. 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.**

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

[3]. 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.**

[4]. 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

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