Zifu Zhang

Beihang University, Beijing, China | zifuzhang@buaa.edu.cn | +86 157 7689 6531 | scholar.google

Education

Beihang University, Beijing, China. BS in Electrical Engineering Sept. 2019 – June 2023
- GPA: 3.71/4.0 Ranking: 19/234 (***top 10%***)
- Key Courses: Digital Image Processing, Data Structures and Algorithms, Mathematical Analysis, Linear Algebra.

Beihang University, Beijing, China. MS in Electrical Engineering Sept. 2023 – Jan. 2026
- GPA: 3.84/4.0 Ranking: 6/126 (***top 5%***)
- Key Courses: Matrix Theory, Pattern Recognition, Intelligent Optimization, Frontiers of Artificial Intelligence.

The HKUST, Hong Kong, China. Phd in Arts and Machine Creativity Sept. 2026 – June 2030

Publications

- **Zhang, Z.**, Xu, T., Li, S., Li, S., Zhang, Y., Xu, M., & Wang, Y. Benchmarking and Enhancing VLM for Compressed Image Understanding, ICML, 2026.
- Li, S.(Supervisor), **Zhang, Z.**, Xu, M., Jiang L., Liu, Y., & Zhu, C. Hierarchical Semantic Compression for Consistent Image Semantic Restoration, IEEE Transactions on Image Processing, 2025.
- **Zhang, Z.**, Li, S., Xu, M., Liu, Z., & Xia, J. Machines Serve Human: A Novel Variable Human-machine Collaborative Compression Framework, IEEE Transactions on Image Processing, 2025.(Major Revision)
- Liu, Z.†, **Zhang, Z.**†, Li, S., Xu, M., Xu, T., & Deng, X. FC-FORMER: Efficient Feature Coding for Machines via a Hybrid CNN-Transformer Architecture, ICASSP, 2026, **Oral**.
- **Zhang, Z.**, Li, S., Liu, T., Xu, M., Xu, T., & Guan, Z. Hybrid Single Input and Multiple Output Method For Compressing Features Towards Machine Vision Tasks, ICIP, 2024, **Oral**.
- **Zhang, Z.**, Li, S., Liu, H., Xu, M., & Zhu, C. Continuous Patch Stitching for Block-wise Image Compression, IEEE Signal Processing Letters, 2025.
- Huang, Z.†, **Zhang, Z.**†, Li, S., Xu, M., Jiang, L., & Deng, X. Multi-Hybrid Mamba Network for Efficient Learned Image Compression, IEEE Signal Processing Letters, 2026.
- Li, S.(Supervisor), **Zhang, Z.**, Xu, T., Xu, M., Jiang, L., & Deng, X. m66614: Report of CE 1.1.9: Single Input for Multiple Output Method with VVC Codec for Feature Compression. In ISO/IEC JTC 1/SC 29/WG 4. ***(MPEG International Standard)***

Work Experience

Algorithm Engineer Intern, JingDong(JD) – Beijing, China Aug. 2024 – Dec. 2024
- ***Multimodal Fusion for Psychiatric Disorder Diagnosis***: Proposed a multimodal feature fusion classification framework for Bipolar Disorder diagnosis by jointly leveraging resting-state fMRI (RS-fMRI) and structural MRI (T1w-MRI). The framework was evaluated on the OpenfMRI benchmark and a self-collected clinical dataset, demonstrating consistent improvements in balanced accuracy (BACC) and F1 score over unimodal baselines.

Research Assistant, Tsinghua Air – Beijing, China May 2025 – Dec. 2025
- ***Autoregressive Vision-Language Tokenization for Image Compression***: Investigated the impact of autoregressive vision-language generative models on image compression by establishing a comprehensive benchmark, with analysis of how different vision tokenizer strategies affect rate-distortion performance.
- ***Multimodal Large Language Benchmark and Enhancement for Image Compression***: Investigated the impact of compressed images on multimodal understanding by constructing a benchmark built upon MLLMs, including GPT-5 and Qwen-VL, to evaluate compression-induced degradation on vision-language tasks. Proposed a plug-and-play enhancement adaptor to effectively restore the visual comprehension capability of MLLMs.

Algorithm Engineer Intern, Alibaba – Beijing, China Dec. 2025 – Present
- ***Text-to-Video Generative Model-Driven Video Compression***: Designed a video compression framework leveraging multimodal text-to-video diffusion models to preserve perceptual reconstruction quality under extremely low-bitrate conditions.
- ***MLLM-Based Compression Agent Design***: Developed an intelligent compression agent built upon existing MLLMs, enabling adaptive codec selection tailored to diverse downstream tasks, including human vision reconstruction and machine vision tasks such as object detection and instance segmentation.

Research Experience

Benchmarking and Enhancing MLLMs for Compressed Images Understanding June 2025 – Present

- Establish a comprehensive benchmark to explore the impact of **image compression on MLLM understanding**, including 1 million compressed images and 8 kinds of multimodal VLMs like GPT5 and QwenVL.
- Analyse the performance gap between compressed and uncompressed images, and propose two gaps using open-source MLLMs like InternVL3 and Qwen-VL2.5 models.
- Introduce **a universal lightweight MLLM adaptor** to enhance performance on heavily compressed images across different codecs. Empirically, our adaptor improve by 10%-30% for InternVL3 and Qwen-VL2.5 models.

Visual Applications Leveraging Text-to-Video Generation Models Oct. 2025 – Present

- Improving reconstruction capability with multi-scale binary **spherical quantization tokenizer** in visual auto-regressive (VAR) models for video generation.
- Constructing an **adaptive time-step** video compression model using VACE conditioning in a Wan-based **flow matching** video generation framework and get the SOTA results.
- Proposing a **reinforcement learning** based fine-tuning strategy to further fine-tune the diffusion-based generative model for better video reconstruction quality and compression efficiency

Semantic Image Compression for Human Vision Jan. 2024 – Present

- Build a novel hierarchical semantic compression framework (HSC) based on the StyleGAN architecture. Design a general inversion encoder to address the unidirectional generation problem.
- Developed a bitstream-level semantic editing pipeline that enables intuitive image editing directly from compressed representations, naturally aligned with existing multimodal image editing paradigms.

Semantic Feature Compression for Human-Machine Collaborative Visions Sept. 2023 – Present

- Design a single input and multiple output framework to compress multiscale features for machine vision. Propose a multimodal fusion adapter to fuse machine-vision and human-vision features to reconstruct images.
- The **first collaborative compression paradigm** from the machine-vision perspective, via seamlessly incorporate the diffusion process with control net for human vision.

Honors and Awards

- **National Scholarship** (2025).***(top 3%)***
- **BYD Company Scholarship** (2025).***(top 1%)***
- First-Class Academic Scholarship (2023, 2024) ***(top 15%)***, First-Class Scholarship for Scientific and Technological Innovation (2021, 2022) ***(top 15%)***
- Outstanding Graduates of Beijing Municipality (2026). ***(top 5%)***
- Outstanding Master's Thesis of Beijing Municipality (2026). ***(top 3%)***
- Outstanding Graduates of Beihang University (2023). ***(top 15%)***
- First Prize in the 21st National Undergraduate Mathematics Competition (2021).

Skills

Languages: English (IELTS 6.5, Reading 8, Writing 7), Chinese (native).
Programming Languages: Python / Matlab / C++ / C.
Technical Skills: Linux / Pytorch / Docker / Numpy / Git / CompressAI.
Research Interests: Computer vision, deep generative models, vision language models and learned image/video codecs.