

Zitong Qi / 齐梓桐

Computer Vision & Multimodal Learning

Ningbo University · B.S. in Mathematics and Applied Mathematics · 2024-2028
Seeking computer vision / multimodal AI opportunities · Open to research collaboration

I study video understanding, vision-language models, and multimodal representation learning, with an emphasis on reproducible evaluation and deployable AI systems.

RESEARCH & EXPERIENCE

Research in Progress | Multimodal Clinical Prediction from Brain Imaging

Mar 2026-May 2026

- Investigated how to recover regional atrophy and clinical semantics missing from a frozen 3D MRI encoder without retraining the backbone.
- Developed a lightweight framework that aligns frozen 3D MRI representations, AAL regional volumes, and BioClinicalBERT clinical embeddings in a shared space.
- Used SigLIP-style contrastive learning and designed a nocog protocol to reduce cognitive-score leakage; ran systematic ablations across encoders, tasks, and modalities.

Intelligent Video Analytics R&D Internship | R&D Engineer

Jan 2026-May 2026

- Addressed high false-alarm rates and limited semantic understanding across anomaly detection, unknown-event recognition, and event description.
- Built a cascade for anomaly classification, open-set recognition, and language-based event explanation using VideoMAE v2, CLIP window features, and Attention / Top-K MIL.
- Fine-tuned Qwen3.5-VL with 4-bit QLoRA and structured prompts; completed system evaluation on a public benchmark.
- Developed a Flask monitoring interface with multi-stream MJPEG, ROI configuration, a rule engine, and an end-to-end alert workflow.

SELECTED PROJECTS

DCIC | Scene Text Image Forgery Analysis

48-hour challenge · Mar 2026

- Decomposed a tightly coupled task covering authenticity classification, forged-region localization, and causal explanation into independently tunable branches.
- Designed a decoupled cascade: Qwen3.5-9B + LoRA for authenticity classification and reasoning, plus ResNet34-UNet for pixel-level forged-region localization.
- Delivered the complete training-to-submission pipeline on time, validating the VLM-reasoning + segmentation approach under a tight deadline.

Tencent Kaiwu | Reinforcement Learning Competition

National Third Prize · Mar 2026-May 2026

- Owned feature, network, memory, and reward design for a survival-game and path-planning agent in a complex, partially observable map.
- Designed a three-head PPO Actor-Critic that separates movement / flash actions and estimates survival, collection, and exploration values independently.
- Used hierarchical attention with global / short-term memory to improve long-horizon planning.
- Designed staged multi-objective rewards, dense feedback for escape behavior, and curiosity-driven exploration to address sparse credit assignment.

CONTACT

[github.com/fingercd](#)
Ningbo, China

RESEARCH FOCUS

Video understanding

Long-range modeling · weak supervision · open-set recognition

Vision-language & alignment

Multimodal representation · parameter-efficient tuning · medical imaging

Reliable, deployable AI

Ablation · leakage-aware evaluation · web systems

TECHNICAL TOOLKIT

Methods

Contrastive learning · MIL · PPO · LoRA / QLoRA · evaluation

Models & frameworks

PyTorch · Transformers · VideoMAE · CLIP · Qwen-VL · BioClinicalBERT

Engineering

Python · Flask · video streaming · ROI / rules · Git

EDUCATION & HONORS

Ningbo University

B.S. Mathematics and Applied Mathematics, in progress
Undergraduate Second-Class Scholarship
MCM/ICM: Meritorious Winner and Honorable Mention
Huawei Ascend AI Innovation Competition, Zhejiang Silver Award
Tencent Kaiwu RL Competition, National Third Prize

INTERESTS

Fitness · hiking · cycling
Collaborative deduction games