WEI WU

Personal Website: https://weiwu-ww.github.io
E-mail: wuw.9902@gmail.com

EDUCATION

University of Science and Technology of China **September 2020 - Present**
Ph.D. Student
Advisor: Prof. Zheng-Jun Zha, Prof. Qibin Sun

Sydney University **July 2019 - October 2019**
Research Intern
Advisor: Prof. Dacheng Tao

University of Science and Technology of China **September 2016 - July 2020**
Bachelor GPA Rank Top5%
Major: Electronic Information Engineering

RESEARCH INTERESTS

Computer Vision
Multi-modal Foundation Models
Multi-modal Representation Learning
Image and Video Retrieval

INTERNSHIP EXPERIENCES

Ant Research **December 2023 - Present**
Advisor: Yujun Shen, Kecheng Zheng
We constructed 100M image-text pairs dataset and developed pre-training strategies for the vison-language foundation model (CLIP) to boost its fine-grained alignment and long text understanding.

Alibaba Tongyi Lab **October 2022 - November 2023**
Advisor: Deli Zhao, Yu Liu
We developed a method for adapting CLIP to few-shot classification tasks, achieving 18.73% accuracy improvement, which is also synergistic with existing tuning methods.

TECHNICAL SKILLS

Computer Languages	Python, C, Matlab
Software & Tools	PyTorch, LaTeX, Matlab

PUBLICATIONS

(* indicates equal contribution)

[NeurIPS 2024] Wei Wu, Kecheng Zheng, Shuailei Ma, Fan Lu, Yuxin Guo, Yifei Zhang, Wei Chen, Qingpei Guo, Yujun Shen, Zheng-Jun Zha, **LoTLIP: Improving Language-Image Pre-training for Long Text Understanding**.

[ECCV 2024] Kecheng Zheng*, Yifei Zhang*, Wei Wu, Fan Lu, Shuailei Ma, Xin Jin, Wei Chen, Yujun Shen, **DreamLIP: Language-Image Pre-training with Long Captions**.

[ICCV 2023] Kecheng Zheng*, Wei Wu*, Ruili Feng, Kai Zhu, Jiawei Liu, Deli Zhao, Zheng-Jun Zha, Wei Chen, Yujun Shen, **Regularized Mask Tuning: Uncovering Hidden Knowledge in Pre-trained Vision-Language Models**.

[PAIN 2022] Jihong Fang*, Wei Wu*, Jiawei Liu, Sicheng Zhang, **Deep Learning Guided Postoperative Pain Assessment in Children**.

[CVPR 2022] Wei Wu*, Jiawei Liu*, Kecheng Zheng, Qibin Sun, Zheng-jun Zha, **Temporal Complementarity-Guided Reinforcement Learning for Image-to-Video Person Re-Identification**.

[CVPR 2021] Jiawei Liu, Zheng-jun Zha, Wei Wu, Kecheng Zheng, Qibin Sun, **Spatial-Temporal Correlation and Topology Learning for Person Re-Identification in Videos**.

[Preprint] Qinying Liu, Wei Wu, Kecheng Zheng, Zhan Tong, Jiawei Liu, Yu Liu, Wei Chen, Zilei Wang, Yujun Shen, **TagAlign: Improving vision-language alignment with multi-tag classification**.

RESEARCH EXPERIENCE

▶ **Long text-image pre-training** 2024*@Ant Research*

- Boosted the long text understanding ability of VLM, improving retrieval accuracy by 17.3%;
- Conducted VLM pre-training on self-constructed 100M long text-image pairs dataset;
- Constructed an open-source benchmark of long text-image retrieval.

▶ **Improving image-language alignment with generated texts** 2024*@Ant Research*

- Utilized MLLMs to generate texts based on 30M images for VLM pre-training;
- Designed a loss to prompt the alignment between local regions of images with texts;
- Achieved superior or comparable performance to 400M pre-trained CLIP with 30M data.

▶ **Exploring multi-label learning based on CLIP** 2023*@Ant Research*

- Enhanced CLIP for fine-grained language-image alignment using multi-label learning;
- Improved zero-shot semantic segmentation mIoU by 5.2% over existing methods.

▶ **Transferring CLIP to downstream tasks with binary masks** 2023*@Alibaba Tongyi Lab*

- Transferred CLIP to few-shot classification tasks with learnable network parameters masking;
- Improved few-shot classification accuracy by 18.7% over zero-shot CLIP;
- Demonstrated compatibility with existing parameter-efficient tuning methods.

▶ **Reinforcement learning for image-to-video person re-identification** 2022*@USTC*

- Treated image-to-video person re-identification as a point-to-set matching problem;
- Dynamically selected multiple consecutive frames in the video to match with an image;
- Improved retrieval accuracy by 22.6% on the iLIDS-VID dataset.

▶ **Graph convolution network for person re-identification** 2021*@USTC*

- Constructed cross-temporal multi-scale graphs based on the inherent relationships between human body parts and semantic information;
- Boosted the local and global features of pedestrians with graph convolution network.

HONORS AND AWARDS

BYD Outstanding Research Scholarship	2024
Outstanding Student Scholarship	2017-2019
Outstanding League Cadres	2018
Outstanding Volunteer	2018