Yunbin Tu

(+86) 15010751895 | tuyunbin22@mails.ucas.ac.cn | https://tuyunbin.github.io/

Research Interests

- **Vision-and-Language**: video captioning, change captioning, image captioning, multi-modal machine translation.
- **Multimedia content analysis**: multi-modal fusion, text-based information retrieval from video.

Education

University of Chinese Academy of Sciences *Sep. 2022 – Present*
Ph.D. in Computer Applied Technology, Advisor: Prof. Li Su and Prof. Liang Li (ICT, CAS) Beijing, China

Kunming University of Science and Technology *Sep. 2019 – Jun. 2022*
M.S. in Pattern Recognition and Intelligent Systems, Advisor: Prof. Zhengtao Yu Kunming, China

Hangzhou Dianzi University *Sep. 2014 – Jun. 2018*
B.S. in Automation, Advisor: Prof. Chenggang Yan Hangzhou, China

Honors & Awards

- National Scholarship for Doctoral Students, Ministry of Education of the People's Republic of China, 2024.
- Diversity and Inclusion Award, Annual Meeting of the Association for Computational Linguistics (ACL), 2024.
- Merit Student, University of Chinese Academy of Sciences, 2023, 2024.
- Yunnan Merit Student, Yunnan Province, 2021, 2022.
- National Scholarship for Master's Students, Ministry of Education of the People's Republic of China, 2021.
- Student Travel Grant, ACM International Conference on Multimedia (ACM MM), 2017.

Academic Services

- PC member of conferences: AAAI 2023/2024/2025, ACL 2023/2024, EMNLP 2023/2024, ACM MM 2023/2024, CVPR 2025, NACCL 2025.
- Reviewer of journals: IEEE TIP (IF:10.8), IEEE TCSVT (IF:8.3), IEEE TETCI (IF:5.3), Artificial Intelligence Review (IF:10.7), Neurocomputing (IF:5.5), CVIU (IF:4.3), Multimedia Systems (IF:3.5), International Journal of Multimedia Information Retrieval (IF: 3.6).

Selected Publications

- Citations = 769; h-index = 13; i10-index = 14; as of December 14, 2024.
- Check my google scholar profile for the detailed list of publications.

[1] **Yunbin Tu**, Liang Li, Li Su, Qingming Huang. Query-centric Audio-Visual Cognition Network for Moment Retrieval, Segmentation and Step-Captioning, **AAAI**, 2025.
[2] **Yunbin Tu**, Liang Li, Li Su, Chenggang Yan, Qingming Huang. Distractors-Immune Representation Learning with Cross-modal Contrastive Regularization for Change Captioning, **ECCV**, 2024.
[3] **Yunbin Tu**, Liang Li, Li Su, Zheng-Jun Zha, Chenggang Yan, Qingming Huang. Context-aware Difference Distilling for Multi-change Captioning, **ACL** (main conference, long paper), 2024.
[4] **Yunbin Tu**, Liang Li, Li Su, Zheng-Jun Zha, Qingming Huang. SMART: Syntax-calibrated Multi-Aspect Relation Transformer for Change Captioning, IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), 2024 (IF:20.8).
[5] **Yunbin Tu**, Liang Li, Li Su, Zheng-Jun Zha, Chenggang Yan, Qingming Huang. Self-supervised Cross-view Representation Reconstruction for Change Captioning, **ICCV**, 2023.
[6] **Yunbin Tu**, Liang Li, Li Su, Junping Du, Ke Lu, Qingming Huang, Viewpoint-Adaptive Representation Disentanglement Network for Change Captioning, IEEE Transactions on Image Processing (**TIP**), 2023 (IF:10.8).
[7] **Yunbin Tu**, Liang Li, Li Su, Ke Lu, Qingming Huang, Neighborhood Contrastive Transformer for Change Captioning, IEEE Transactions on Multimedia (**TMM**), 2023 (IF:8.4).

[8] **Yunbin Tu**, Chang Zhou, Junjun Guo, Huafeng Li, Shengxiang Gao, Zhengtao Yu, Relation-aware attention for video captioning via graph learning, Pattern Recognition (**PR**), 2023 (IF:7.5).
[9] **Yunbin Tu**, Liang Li, Li Su, Shengxiang Gao, Chenggang Yan, Zhengjun Zha, Zhengtao Yu, Qingming Huang, I^2Transformer: Intra- and Inter-relation Embedding Transformer for TV Show Captioning, IEEE Transactions on Image Processing (**TIP**), 2022 (IF:10.8).
[10] **Yunbin Tu**, Chang Zhou, Junjun Guo, Shengxiang Gao, Zhengtao Yu, Enhancing the Alignment between Target Words and Corresponding Frames for Video Captioning, Pattern Recognition (**PR**), 2021 (IF:7.5).
[11] **Yunbin Tu**, Liang Li, Tingting Yao, Jiedong Lou, Shengxiang Gao, Zhengtao Yu, Chenggang Yan, Semantic Relation-aware Difference Representation Learning for Change Captioning, **Findings of ACL** (Long), 2021.
[12] **Yunbin Tu**, Liang Li, Chenggang Yan, Shengxiang Gao, Zhengtao Yu, R^3Net: Relation-embedded Representation Reconstruction Network for Change Captioning, **EMNLP** (Long), 2021.
[13] Chengang Yan, **Yunbin Tu**, Xingzheng Wang, Yongbing Zhang, Xinhong Hao, Yongdong Zhang, Qionghai Dai, STAT: spatial-temporal attention mechanism for video captioning, IEEE Transactions on Multimedia (**TMM**), 2019 (IF:8.4, ACM MM 2017 extension).
[14] **Yunbin Tu**, Xishan Zhang, Bingtao Liu, Chenggang Yan, Video Description with Spatial-Temporal Attention, **ACM MM**, 2017.