

BASIC INFORMATION

Name: Yuan Yao **Gender:** Male **Age:** 27
Tel: (+86)18813093272 **Email:** yuanyao17@mails.ucas.ac.cn **Homepage:** <https://yuanyao.github.io>

EDUCATION

University of Chinese Academy of Sciences (UCAS), Beijing, China Sep. 2017 – Present
Ph.D. in Electronic and Communication Engineering
Supervisor: Prof. Qixiang Ye
Core Courses: Digital Image Processing, Pattern Recognition, Machine Learning Methods and Applications

Beijing Jiaotong University (BJTU), Beijing, China Sep. 2013 – Jul. 2017
B.Eng in Electronics Science and Technology
Core Courses: Calculus, Probability Theory, Linear Algebra, Signals and Systems, Data Structure

RESEARCH INTERESTS

Computer Vision Object Detection/Localization, Video Classification/Retrieval
Machine Learning Self-supervised Learning, Semi-/Weakly supervised Learning

EXPERIENCE

University of Chinese Academy of Sciences (UCAS), Beijing, China
Pattern Recognition and Intelligent System Development Laboratory (PriSDL)

Weakly Supervised Object Detection (WSOD) Sep. 2021 – Present
The redundant and near-duplicate proposals used in previous WSOD methods decrease the detection efficiency and introduce ambiguity to feature representation. To avoid these effects from dense proposals, a sparse proposal evolution approach is built upon the visual transformer, which advances WSOD from the two-stage “enumerate-and-select” pipeline with dense proposals to an end-to-end “seed-and-refine” pipeline with sparse proposals. The proposed method outperforms the state-of-the-art end-to-end methods by 8.1% AP50 on COCO dataset. It is also ~60 times faster than the two-stage methods.

Weakly Supervised Object Localization (WSOL) Apr.2021 – Sep.2021
When introducing visual transformer to WSOL, the semantic-agnostic attention map brings semantic confusions to the final localization results. In order to make the activation maps are more discriminative to categories, a group of category-related tokens is defined to encodes the semantic information of each category. Correspondingly, a group of semantic-aware attention maps are generated to replace the semantic-agnostic attention map for localization. The proposed method achieves the state-of-the-art performance on ILSVRC 2012 dataset.

Self-supervised Video Representation Learning Jun.2019 – 2021.04
The temporal resolution (long-short term) characteristics are not yet fully explored, which limits representation capabilities of learned models. So, a novel self-supervised method, referred to as video playback rate perception, are proposed. It is consisted of a discriminative perception model for classifying fast-forward rates and a generative perception model for reconstructing slow-down videos. The cooperative learning between these two models further facilitates the network's understanding of motion. Experiments show that the proposed method outperforms state-of-the-art self-supervised models in the same period.

Campus Intelligent Video Surveillance System based on 5G Networks Mar.2018 – Jun.2019
In order to meet the real-time requirements of object detection and tracking in multi-channel video, in cooperation with Huawei 5G Lab, a cross-camera video surveillance system demonstration platform is built by optimizing the video cooperative transmission model of 5G network, realizing the object detection and tracking framework based on deep network, and improving the discriminativeness of classification features for person re-identification.

PUBLICATIONS AND PATENTS (Google Scholar)

- [1] M. Liao, F. Wan, Y. Yao, Z. Han, J. Zou, Y. Wang, B. Feng, P. Yuan, Q. Ye, "End-to-End Weakly Supervised Object Detection with Sparse Proposal Evolution," *ECCV* 2022.
- [2] C. Liu, Y. Yao, D. Luo, Y. Zhou, Q. Ye, "Self-supervised Motion Perception for Spatio-temporal Representation Learning," *IEEE Trans. Neural Networks Learn. Syst. (TNNLS)*, 2022.
- [3] Y. Yao, F. Wan, W. Gao, etc., Q. Ye, "TS-CAM: Token Semantic Coupled Attention Map for Weakly Supervised Object Localization," *IEEE Trans. Neural Networks Learn. Syst. (TNNLS)*, 2022. (To be published)
- [4] Y. Yao, C. Liu, D. Luo, Y. Zhou, Q. Ye, "Video Playback Rate Perception for Self-Supervised Spatio-Temporal Representation Learning," *IEEE CVPR*, 2020.
- [5] C. Liu, F. Wang, Y. Yao, X. Zhang, Q. Ye, "Orthogonal Decomposition Network for Pixel-wise Binary Classification," *IEEE CVPR*, 2019.
- [6] Q. Ye, Y. Yao, F. Wan, Y. Zhang, J. Jiao, Z. Han, "A Self-Supervised Video Spatio-Temporal Representation Learning Method based on Frame Rate Perception," China Invention Patent, CN202010281494.0

TECHNICAL SKILLS AND AWARDS

Programming Languages Python, MATLAB, C/C++
Deep Learning Frameworks PyTorch, Caffe
Tools L^AT_EX, MS Office, Linux, PyCharm, NumPy, Matplotlib, Scikit-learn
Awards Merit Student of UCAS, 2020. The Second Prize Scholarship of UCAS, 2018-2021.