Zhenzhi Wang

CONTACT INFORMATION

Room 1011, Computer Science Building, Xianlin Avenue 163
Qixia District, Nanjing, Jiangsu, China, 210023

(+86) 13390815189
zhenzhiwang@outlook.com

RESEARCH INTERESTS

Computer Vision: action recognition, temporal action detection/segmentation, temporal grounding.
Machine Learning: cross-modal video understanding, multi-modal pretraining

EDUCATION

School of Artificial Intelligence, Nanjing University — Nanjing, China
M.Sc. Candidate in MCG Lab — Sep. 2019 – Jun. 2022 (expected)
Supervisor: Prof. Limin Wang

School of Physics, Nanjing University — Nanjing, China
B.Sc., Major in Physics (Overall GPA: 87.8/100) — Sep. 2015 – Jun. 2019

PUBLICATION

TOP-TIER CONFERENCES

Zhenzhi Wang, Ziteng Gao, Limin Wang, Zhifeng Li, Gangshan Wu. Boundary-Aware Cascade Networks for Temporal Action Segmentation. European Conference on Computer Vision (ECCV'20), Glasgow, United Kingdom, 2020.

Yixuan Li, Lei Chen, Runyu He, **Zhenzhi Wang**, Gangshan Wu, Limin Wang. MultiSports: A Multi-Person Video Dataset of Spatio-Temporally Localized Sports Actions. IEEE International Conference on Computer Vision (ICCV'21), Virtual, 2021.

PRE-PRINTS

Zhenzhi Wang, Limin Wang, Tianhao Li, Gangshan Wu. Negative Sample Matters: A Renaissance of Metric Learning for Temporal Grounding. *arXiv preprint: 2109.04872*, 2021.

OTHERS

Zhenzhi Wang, Liyu Wu, Zhimin Li, Jiangfeng Xiong, Qinglin Lu. Overview of Tencent Multi-modal Ads Video Understanding Challenge. ACM Multimedia Grand Challenge, Chengdu, China, 2021.

RESEARCH EXPERIENCE

Negative Sample Matters: A Renaissance of Metric Learning for Temporal Grounding
Advisor: Prof. Limin Wang — Dec. 2020 - Mar. 2021

- We tackle temporal grounding task from a metric-learning view instead of from the commonly-used detection/regression view. We present Dual Matching Network (DMN) to directly model the relations between language queries and video moments in a joint embedding space. This new framework enables fully exploiting negative samples from two new aspects: constructing negative pairs from a dual matching scheme and mining negative pairs across different videos.
- DMN achieves the SOTA performance in three temporal grounding datasets (TACoS, Charades-STA and ActivityNet-Captions) with less training costs than the strong baseline 2D-TAN by sharing visual features between sentences. By combining it with human tubelet generators, DMN-based method outperforms strong opponents with powerful multi-modal pretrain models and achieves the SOTA performance in spatio-temporal grounding dataset HC-STVG.

MultiSports: A Multi-Person Video Dataset of Spatio-Temporally Localized Sports Actions
Advisor: Prof. Limin Wang — Oct. 2020 - Nov. 2020

- We propose a multi-person, fine-grained and large-scale video dataset, named *MultiSports*, to advance the future research in spatio-temporal action localization task, which contains high-quality annotations for 62 sports actions in 4 popular sports. Our dataset distinguishes from previous datasets in 1) multiple concurrent actions, 2) fine-grained actions and less characteristic backgrounds, 3) more action categories and instances and 4) high-resolution video records.
- We design a novel error analysis for Video-mAP which decouples the error types and comprehensively analyzes the shortcomings of current methods.
- Contribution: I organize the paper and finish the text part of the paper.

Boundary-Aware Cascade Networks for Temporal Action Segmentation — Jun. 2019 - Dec. 2019
Advisor: Prof. Limin Wang

- We present a framework, called BCN, to generally boost existing temporal action segmentation methods, which has two components: (1) stage cascade for boosting segmentation accuracy for hard frames (e.g., near action boundaries); and (2) local barrier pooling which uses boundary information for more smooth prediction and less over-segmentation errors.
- BCN outperforms the existing state-of-the-art methods by large margin on three challenging datasets: 50Salads, GTEA and Breakfast dataset. The code is available at https://github.com/MCG-NJU/BCN.

Academic Service

- Track organizer of ICCV2021 Workshop DeeperAction on localized-and-detailed understanding of human actions in videos. Our track *MultiSports Challenge* focuses on fine-grained spatio-temporal action localization.
- One of the key organizers of ACM-MM 2021 Grand Challenge Multi-modal ads video understanding during my intern at Tencent.
- Journal reviewer: IEEE TCSVT, Neurocomputing.

Contests

Human-centric Spatio-Temporal Video Grounding Challenge May 2021 - Jun. 2021
In CVPR21 Workshop Person in Context

- We get the **1st place** in the challange (CNY 20000). I am the first author of the challenge paper and did a 15-min oral presentation to introduce our method in the PIC workshop on June 21, 2021.
- We extract tube-level features by Slowfast on linked tubes based on human bounding boxes predicted by Faster R-CNN. Then we use a 2d-map proposal representation similar to 2D-TAN and enhance the feature representation to be more discriminative by multi-modal contrastive learning. The additional contrastive loss benefits the final Recall@IoU or mIoU metrics a lot by utilizing cross-video negative pairs.

Internship

Multi-modal Temporal Video Structuring Mar. 2021 - Present
Tencent Data Platform, Shenzhen. Advisor: Dr. Zhifeng Li

- As one of the organizers of Tencent Advertising Alogrithm Competition 2021, I build baseline models and comprehensive documents of the provided baseline for the competition participants.
- As the first author, I write the overview paper (4 pages) for this challenge and also release the extended version (8 pages) on the Arxiv. The extended paper which introduces our proposed 'Multi-modal Ads Video Structuring' task and 'Tencent Ads Video Structuring' dataset will be submitted to a top-tier computer vision conference later.

Honors and Awards

- **Excellent Graduate Students, awardee** (10/44) Nanjing University 2020
- **2nd Award, YiBao-Payment Scholarship** (3/44) Nanjing University 2020
- **1st Award, Scholarship for Graduate Students** (Top 20%) Nanjing University 2019 & 2020
- **Outstanding Bachelor Graduates** (26/185) Nanjing University 2019
- **3rd Award, People's Scholarship** (35/185) Nanjing University 2016 & 2018
- **NJU-IHEP Scholarship for Excellent Student** (6/185) Institute of High Energy Physics 2018
- **Excellent League Members, awardee** (9/185) Nanjing University 2017 & 2018
- **2nd Award, People's Scholarship** (20/185) Nanjing University 2017
- **Excellent Bachelor Students, awardee** (8/185) Nanjing University 2016

Skills

- Programming: Python, PyTorch, LaTeX, C++
- Languages: Mandarin (native), English (CET-6 542, TOEFL 103 (S 23))