

HAODONG DUAN

✉ dhd.efz@gmail.com · ☎ +86 18211165536 · 🏠 [HomePage](#)

EDUCATION

Peking University, Beijing 2015 – 2019

GPA 3.77/4.00, rank 1st in Data Science students
Undergraduate in Data Science, Yuanpei College

Chinese University of HongKong, HongKong 2019 – Present

Working on *Video Understanding*, supervised by [Dahua Lin](#)
Ph.D. candidate in Information Engineering

RESEARCH PROJECTS

Triplet Representation for Human Body, [Paper](#), [Dataset](#) 2018 – 2019

Design a triplet representation named TRB (as well as its estimation method) to represent 2D human body, which includes both human pose and shape information. The representation can be used in human shape editing.

Omni-sourced Webly-supervised Video Recognition, [Paper](#), [Dataset](#), [Code](#) 2019 – 2020

Propose a framework (**OmniSource**) for webly supervised video recognition, which can utilize various kinds of web medias, including images, trimmed videos, untrimmed videos for trimmed video recognition. Achieve 83.6% Top-1 Accuracy on Kinetics400 with SOTA algorithm and our framework.

Mitigating Unwanted Bias in Action Recognition 2020

Demonstrate that the deep learning based video recognition models are biased towards factors like scene or objects. Establish a new quantitative benchmark to evaluate such bias and propose to mitigate such bias with adversarial training and diversified web data.

Skeleton-based Action Recognition with 3D-CNN, [Paper](#), [Dataset](#), [Code](#) 2020 – 2021

Devise a novel 3D-ConvNet based paradigm (**PoseC3D**: 2D keypoint heatmaps → 3D heatmap volumes → 3D-CNN recognizer) for skeleton-based action recognition. PoseC3D outperforms previous skeleton-based action recognition approaches by a considerable margin across various benchmarks (NTURGB+D, Kinetics, *etc.*).

Efficient Video Recognition for Untrimmed Videos 2021

Propose an efficient framework for untrimmed video recognition, which samples frames from frame candidates to form one representative clip. The framework (w. R50 backbone) can achieve 82.2% Top-1 Accuracy with 1-clip testing (the computational cost is only 52 GFLOPs / video).

Video Self-supervised Learning via Ranking-based Transformation Recognition 2021

Show the great potential of RecogTrans (recognizing transformations applied to video clips) video SSL by introducing a unified Ranking-based formulation. The proposed method significantly outperforms previous RecogTrans approaches on action recognition (UCF Top1 +6%) and video retrieval (UCF R@1 +20%).

Skeleton-based Action Recognition with GCN 2021 – 2022

Design a novel GCN-based architecture for skeleton-based action recognition, which doesn't rely on any pre-defined skeleton topology. The proposed model achieves great recognition performance which surpasses previous state-of-the-arts significantly. Using official annotations provided, we achieve 89.6%, 91.4% Top-1 on two NTURGB+D 120 splits, 40.3% Top-1 on Kinetics-Skeleton.

PUBLICATIONS

Haodong Duan, Kwanyee Lin, Sheng Jin, Wentao Liu, Chen Qian, Wanli Ouyang
TRB: A Novel Triplet Representation for Understanding 2D Human Body (ICCV 2019)

Haodong Duan, Yue Zhao, Yuanjun Xiong, Wentao Liu, Dahua Lin
Omni-sourced Webly-supervised Learning for Video Recognition (ECCV 2020)

Haodong Duan, Yue Zhao, Kai Chen, Dahua Lin, Bo Dai
Revisiting Skeleton-based Action Recognition (CVPR 2022)

Haodong Duan, NanXuan Zhao, Kai Chen, Dahua Lin
TransRank: Self-supervised Video Representation Learning via Ranking-based Transformation Recognition (CVPR 2022)

Jintao Lin, **Haodong Duan**, Kai Chen, Dahua Lin, Limin Wang
OCSampler: Compressing Videos to One Clip with Single-step Sampling (CVPR 2022)

OPENSOURCE PROJECTS

The main contributor and maintainer of [MMAction](#) and [MMAction2](#).

PROFESSIONAL ACTIVITIES

Conference Reviewer for ICCV2021, AAAI2022, CVPR2022, and ECCV2022.

LANGUAGE SKILLS

- TOEFL iBT test: 104pt (Reading: 30, Listening: 28, Speaking: 20, Writing: 26)
- GRE test: 322pt (Verbal: 152, Quantitative: 170)