Sixun Dong 董思勋

Email: dongsx@shanghaitech.edu.cn Mobile: +86-136-4924-4805
Github: https://github.com/Ironieser

About Me

I am a third-year graduate student at ShanghaiTech University, supervised by Prof. Shenghua Gao. Before that, I received my Bachelor's degree in 2020 from Dalian University of Technology. My research interests lie in video understanding and weakly supervised learning, including human activity recognition and video representation learning. I am also focusing on multi-modal learning.

Education

July 2024	Master of Computer Science at **ShanghaiTech University**, China
(expected)	**Major**: Computer Vision & Deep Learning
July 2020	Bachelor Degree in Process Equipment and Control Engineering (Major)
	Dalian University of Technology, China
July 2020	Bachelor Degree in Computer Science (Dual Degree)
	Dalian University of Technology, China

Academic Experience

† = Co-first author

present **Improve video understanding ability of large visual language modal.**
Sixun Dong, et al., Advisor: Prof. Shenghua Gao ShanghaiTech University
Oct. 2023 Working on video understanding by leveraging pre-trained large language models. To address the limitations of large visual language models in understanding temporal information in videos, we collected instruction data and utilized supervised fine-tuning techniques on large language models to enhance their video comprehension capabilities. This project is currently in progress.

May. 2023 **Weakly Supervised Video Representation Learning with Unaligned Text for Sequential Videos.**
[CVPR 2023] [Paper] [Code] ShanghaiTech University
Sixun Dong†, Huazhang Hu†, Dongze Lian, Weixin Luo, Yicheng Qian, Shenghua Gao.
Apr. 2022 Worked on weakly supervised sequential video understanding where the accurate time-stamp level text-video alignment is not provided. By borrowing ideas from CLIP, we aggregated frame-level features for video representation and encoded the texts corresponding to each action and the whole video, respectively.
· Proposed a novel weakly supervised video representation learning pipeline with unaligned text for sequential videos.
· Designed a multiple granularity contrastive learning loss that uses the fact that video actions happen sequentially in the temporal domain to generate pseudo frame-sentence correspondence.
· Extensive experiments on video sequence verification and text-to-video matching showed the effectiveness of our proposed approach.

Apr. 2022 **TransRAC: Encoding Multi-scale Temporal Correlation with Transformers for Repetitive Action Counting.**
[CVPR 2022 Oral] [Paper] [Code] ShanghaiTech University
Huazhang Hu†, **Sixun Dong**†, Yiqun Zhao, Dongze Lian, Zhengxin Li*, Shenghua Gao*.
Sep. 2021 Worked on repetitive action counting(RAC). Specifically, the previous works focus on performing RAC in short videos, which is tough for dealing with longer videos in more realistic scenarios, such as interruption during the actions or inconsistent action cycles.
· Collected a new repetitive action counting dataset with fine-grained annotations.
· Encoded multi-scale temporal correlation with transformers that can consider both performance and efficiency.
· Designed a density map regression-based method to predict the action period.
· Our approach yielded better performance with sufficient interpretability and achieved SoTA results.

Academic Community Service

Reviewer of CVPR 2023, ICCV 2023, ACM MM 2023 and CVPR 2024.

Internship Experience

Present **Co-Speech Gesture and head motion generation**

Leader of Algorithm DGene, Algorithm Intern, Digital Human Algorithm Department

Nov. 2023 Working on enhancing realistic co-speech gesture generation. The existing methods often overlook the relationship between gesture and head motion in digital avatar synthesis. My approach involves aligning co-speech gestures with head poses to significantly improve the generation quality of digital avatars. This project is currently in progress.

Oct. 2023 **Human Body Reconstruction and Anthropometric Measurements Based on Multi-view Camera Systems**

Leader of Algorithm DGene, Algorithm Intern, Digital Human Algorithm Department

Aug. 2023 Worked on human anthropometric measurements, which is a crucial requirement for customized digital avatars. Based on single or multi-view photographs provided by users, we accurately reconstruct 3D human bodies and provide precise body measurement information, including limb lengths and the circumference measurements of the chest, waist, and hips. In the process of business integration, I also take responsibility for organizing and summarizing project documentation to ensure on-time project delivery.

· In response to real-world business scenarios, our proposed algorithmic framework takes multiple RGB images as input and generates a precise parameterized model of the human body, providing accurate body measurement results.

· We have achieved the generation of a pose-controllable parameterized human body model with a relative measurement error of less than **7%** within **3 minutes**. Our subsequent work will focus on facial expression reconstruction and the generation of high-quality texture maps.

Aug. 2023 **Audio Driven Talking Head Video Generation**

Leader of Algorithm Transsion Holdings, Algorithm Intern, Audio-Video Generation Department

Apr. 2023 Worked on talking head video generation. Existing methods suffer from several issues, including low realism, poor temporal consistency, inaccurate lip synchronization, and limited generalization. Inspired by diffusion models and video generation, we designed algorithmic optimization strategies.

· Proposed optimization strategies for training and model architecture in the audio-driven 2D talking head video generation, combining relevant techniques from the field of video understanding.

· Incorporated image post-processing techniques such as CodeFormer and Gaussian image blending algorithms designed for facial restoration.

· Fine-tuned the model for specific scenarios, which resulted in quality significantly superior to the current SoTA models in both industrial and academic contexts.

Technical Skills

Programming: Python, Pytorch, C/C++, Linux, Git

Research Topics: Video understanding; Weakly supervised learning; Contrastive learning; Video generation;