Yue Li
2003.9
Sun Yat-sen University
School of Intelligent Systems Engineering
Intelligent Science and Technology

☎ +86-13952599359
✉ yue3184@gmail.com
✉ liyue228@mail2.sysu.edu.cn
GitHub Homepage
ly245422

Education

- Sun Yat-sen University School of Intelligent Systems Engineering September 2021 - Present

  Intelligent Science and Technology Undergraduate GPA: 3.6

Campus activity

- Youth League Committee, School of Intelligent Engineering, Sun Yat-sen University Junior year

  Deputy Minister of Organization

  – Organize activities such as site visits, holding lectures, etc. Strong teamwork and leadership skills

Skills

Ielts Score: 6.5

Papers

- Skeleton2Point: Recognizing Skeleton-Based Actions as Point Clouds

  MM ACM Multimedia (under review)

  Homepage: Skeleton2Point

  Supervisor: Mengyuan Liu

  – Background : The existing methods pay much more attention to encoding joints' position with the given time and serial number information, neglecting to model the positional information contained in the 3D coordinate channel itself.

  – Innovation Points : We propose a skeleton-to-point network(Skeleton2Point) to model joints' position relationships in three-dimensional space. This method is the first to regard skeleton joints as point clouds via incorporating the position information of skeletons into point cloud methods, demonstrating the validity of modeling position relationships with 3D coordinates. We devise a novel information transformation module (ITM) to merge the original time and series information and the joint coordinates information, which means transform the skeleton data into point cloud' s form. We also propose a Cluster-Dispatch-based interaction module (CDI) to focus on overall movement trends. Suppose an action sample contains n points and the center is m, all n points in the sample are aggregated by global averaging to get the center point.

  – Result : In comparison with existing methods on NTU-RGB+D 60 and NTU-RGB+D 120 datasets, Skeleton2Point achieves SOTA levels on both joint modality and stream fusion. Especially, on the challenging NTU-RGB+D 120 dataset under the X-Sub and X-Set setting, the accuracies reach 90.63% and 91.86%.

- HDBN: A Novel Hybrid Dual-branch Network for Robust Skeleton-based Action Recognition

  ICME Workshop 2024

  Supervisor: Mengyuan Liu

  – Backgrounds : Current methodologies often lean towards utilizing a solitary backbone to model skeleton modality, which can be limited by inherent flaws in the network backbone.

  – Innovation Points : We advocate for utilizing different network structures to achieve robust skeleton-based action recognition, fully leveraging the structural complementarity between different backbones. We propose a new dual-branch framework called Hybrid Dual-Branch Network (HDBN), which effectively combines GCNs and Transformers. In detail, the skeleton data are inputted to GCN and Transformer backbones for modeling high-level features, which will be effectively combined through a late fusion strategy for more robust skeleton-based action recognition.

  - Result : Extensive experiments on benchmark UAV-Human dataset verify the effectiveness of our HDBN. In this large-scale action recognition datasets, our model outperforms most existing action recognition methods. In the 2024 ICME Grand Challenge, achieving accuracies of 47.95% and 75.36% on two benchmarks of the UAV-Human dataset by outperforming most existing methods.

- Theatergen: Character Management with LLM for Consistent Multi-turn Image Generation

  ECCV 2024 (under review)

  Homepage: TheaterGen

  Supervisor: Xiaodan Liang

  - Background : multi-turn image generation, which is of high demand in real-world scenarios, faces challenges in maintaining semantic consistency between images and texts, as well as contextual consistency of the same subject across multiple interactive turns.
  - Innovation Points : We propose TheaterGen, which is a training-free framework that utilizes a large language model to drive a text-to-image generation model, effectively addressing the issues of semantic consistency and contextual consistency in multi-turn image generation tasks without specialized training. TheaterGen can engage in multi-turn natural language interactions with users to accomplish tasks such as story generation and multi-turn editing. We propose a new benchmark, CMIGBench, to evaluate both the semantic and contextual consistency in multi-turn image generation and demonstrate the superior performance of TheaterGen.
  - Result : Extensive experimental results show that TheaterGen outperforms state-of-the-art methods significantly. It raises the performance bar of the cutting-edge Mini DALL · E 3 model by 21% in average character-character similarity and 19% in average text-image similarity.

- SemiPL: A Semi-supervised Method for Event Sound Source Localization

  ICME Workshop 2024

  Supervisor: Mengyuan Liu

  - Background : Given the limited and chaotic amount of data in the Chaotic World dataset, the existing self-supervised methods are no longer good enough to learn the characteristics of the data in depth.
  - Innovation Points : We propose a semi-supervised method SemiPL. As far as I know, We are the first to explore the application of the semi-supervised method SemiPL on event sound source localization. We also explore the effect of different parameters.
  - We achieve SOTA performances with significant margins on the Chaotic World datasets. In particular, our model achieved an improvement of 12.2% cIoU and 0.56% AUC in Chaotic World compared to the results provided.

- SFMVIT: SlowFast Meet Vit in Chaotic World

  ICME Workshop 2024

  Supervisor: Mengyuan Liu

  - Backgrounds : The task of spatiotemporal action localization in chaotic scenes is a challenging task toward advanced video understanding.
  - Innovation Points : We have introduced the dual-stream spatiotemporal feature modeling network SFMViT, which integrates SlowFast ' s ability to capture temporal features with Vision Transformer ' s capability in complex scene spatiotemporal modeling. We introduce a Confidence Pruning Strategy to find the most suitable anchor number of the instances, which can be used to prune anchors and filter out the optimal anchors while taking into account the efficiency and performance, increasing the mAP by nearly 2%.
  - Result : We achieve SOTA performances, a mAP of 26.62% with significant margins on the Chaotic World datasets.

Awards

- ICME Grand Challenge 2024 March 2024

  Multi-Modal Video Reasoning and Analyzing Competition (MMVRAC)

Track 10: Skeleton-based Action Recognition	Third place
Track 1: Spatio-temporal Action Localization	First place
Track 4: Sound Source Localization	First place