International Journal for Modern Trends in Science and Technology
Volume 12, Issue 07, pages 263-270.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue06.html
DOI: https://doi.org/10.5281/zenodo.21651627

Survey on Vision Datasets for Human Action Recognition

Rahul Kumar[1] | Neha Gupta[1] | Akshay Mool[2] | Ankit Yadav[2]

[1]*Department of Computer Science and Engineering, Delhi Technological University, Delhi, India.*
[2]*Department of Information Technology, Delhi Technological University, Delhi, India.*

To Cite this Article
Rahul Kumar, Neha Gupta, Akshay Mool, & Ankit Yadav (2026). Survey on Vision Datasets for Human Action Recognition. International Journal for Modern Trends in Science and Technology, 12(07), 263-270. https://doi.org/10.5281/zenodo.21651627

Article Info
Received: 11 June 2026; Revised: 01 July 2026; Accepted: 05 July 2026.

KEYWORDS

Activity Recognition; Vision Dataset; Transformer; Convolutional Neural Network; Action Former.

ABSTRACT

HAR (Human Action Recognition) is the most challenging task and a key area of research in visual analysis and processing. HAR solves many problems, such as crowd behavior analysis, visual surveillance, sports activity analysis, human interaction and daily activity analysis. In this paper, we discussed various key vision datasets, including their action classes and the number of visual clips available for analysis. This article provides a comparative analysis of the vision dataset and the state of the reference method for accurate activity classification. The main objective of this article is to provide an analysis of key vision datasets, their strengths and limitations and to outline the key points of each dataset

I. INTRODUCTION

Human Action Recognition (HAR) refers to the automated process of identifying and classifying human actions from video or image sequences. The advent of machine learning(ML) and deep learning(DL) in particular, specifically convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has made a significant contribution to the advancement of HAR. The effectiveness of these models, however, depends on the quality and variety of training data. This survey provides an overview to the most notable datasets that have been used in HAR, outlining their usefulness and potential for improvement.

Human activity recognition's major goal is to properly describe human behaviors and interactions using previously unknown data sequences. It is sometimes difficult to reliably distinguish human actions from video data due to a variety of issues, including dynamic backgrounds and low-quality films. Human activity recognition approaches address two key questions:"Which action is performed?(action recognition) and "Where exactly in the video? (localization). The series of pictures are known as frames. Thus, the basic goal of an action recognition(AR) challenge is to interpret incoming video clips and identify subsequent human activities [1].

The technique of detecting and labelling actions based on visual observations and monitoring a sequence of motions by a human being is known as HAR. While multiple persons execute at multiple spots from diverse perspectives, AR attempts to find the action that perfectly matches an action instance. Complex actions that are challenging to recognize can all be broken into smaller ones that are simpler to detect. In certain circumstances, detecting extraneous objects in an image or sequences of images can help offer a better comprehension of human behaviours because it can produce essential details about the event or activity shown in the pictures or motion images When it comes to the sub-activities, the methods utilized to identify them differ greatly.

The activity identification requirement in HAR may include a detailed and comprehensive representation and analysis of action. These activities are recorded utilizing a variety of sensors, including RGB, range, radar, and wearable sensors. These datasets may be divided into RGB and RGB-D (categories based on the information included. Occlusions, illumination difference, perspective variation, annotation as usual, and integration of modalities are the main issues with these dataset [2].

This study is organised as follows. Section II provides a detailed review of the major vision-based HAR datasets, including Kinetics, UCF101, HMDB51, JHMDB, NTU RGB+D, Sports-1M, Activity Net, Something-Something V2, Fine Gym, and PKU-MMD. The data collection methodology, the number of action classes, the dataset sizes, the modalities, their strengths and weaknesses, and the domains of application are discussed for each dataset.In Section III, the comparative analysis of these datasets is discussed with accuracy. Finally, this paper summarizes the key observations, research gaps and future scopes for HAR datasets and benchmarking in Section IV.

Although many HAR datasets are available, researchers often struggle to choose which to use for model development and evaluation due to factors such as scale, modality, action diversity, annotation quality, and benchmarking protocols. For this reason, a thorough evaluation of popular vision-based HAR datasets is important to understand their characteristics, strengths, weaknesses, and applicability to various research problems. This survey will provide such an analysis and serve as a reference guide for researchers involved in video-based human action recognition.

II. VISION-BASED HAR DATASETS

Kinetics Dataset

Kinetics, created by DeepMind, is one of the largest HAR datasets, along with Kinetics-400, Kinetics-600, and Kinetics-700. The number of action classes and YouTube video clips is also increased in each version. The clips are all 10 seconds long and demonstrate actions including "brushing teeth," "playing guitar," and "skiing. The data set represents a wide variety of human activities across different settings, with considerable diversity in background, clothing, and lighting. It supports multi-label classification and is frequently used in transfer learning benchmarks [3]. The HAR research area is gaining interest due to its wide range of applications in the fields of elderly monitoring, suspicious activity monitoring, sports activity, pose estimation, and health monitoring. The wide range of variations in normal human activities adds complexity to the recognition process. The use of automated systems is crucial in facilitating the increased utilisation of cameras. Table 1 provides an overview of the different versions of the Kinetics dataset, along with the number of classes in each version.

Table 1:Kinetics Dataset

Dataset	No. Of Classes	No. of Videos	Duration
Kinetics-400	400	~306,245	10s
Kinetics-600	600	~500,000	10s
Kinetics-700	700	~650,000	10s

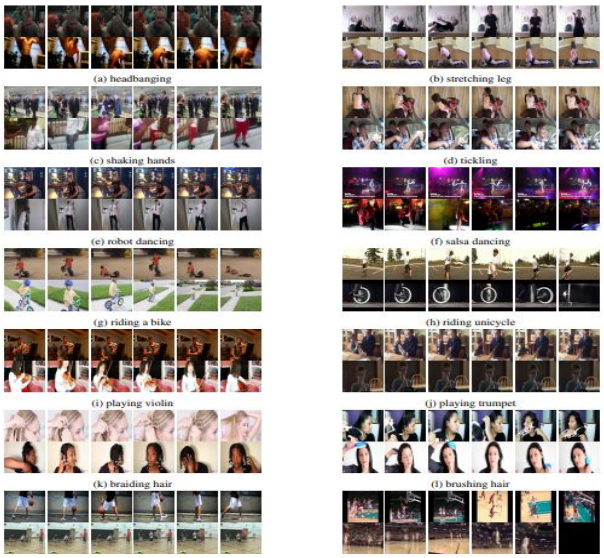

Figure 1: Sample action classes from the Kinetics dataset are shown, and are best viewed in color and

with zoom for clarity. In some instances, a single frame may not be sufficient to accurately identify the action, such as headbanging or to distinguish between similar actions, like 'dribbling basketball' versus 'dunking basketball'[3].

UCF101

UCF101 is a widely cited dataset containing around 13,000 visual clips spread over 101 human action categories, clubbed into five sets: Human-Object Interaction, Body-Motion Only, Human-Human Interaction, Playing Musical Instruments, and Sports. Videos are collected from YouTube, making them realistic and diverse in terms of camera motion, background clutter, and lighting. Each video has a fixed resolution and frame rate, which simplifies preprocessing and benchmarking [4].

This dataset is very useful for recognizing human activity in real environments. Every video in this dataset was gathered from YouTube. It is a more sophisticated form of the UCF 50 dataset. It includes 101 action lessons, including brushing teeth, breaststroke, and boxing speed bag. Over 13,000 films are included in total, and they are further separated into 25 sections. There are four to seven action clips in each group, and they all share similar elements, such as background and point of view [4].

One popular benchmark in the field of HAR research is the UCF101 dataset. The University of Central Florida developed it, and it is renowned for its versatility and applicability to actual video-based action categorization problems. Figure 2 shows the sample action recorded in the UCF 101 dataset.

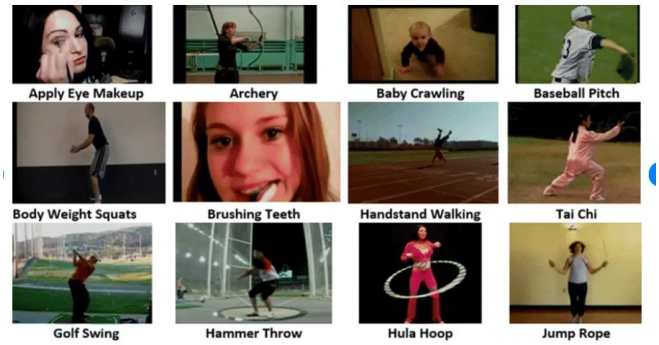

Figure 2:UCF 101 Sample frame of Actions Performed**[4]**

HMDB51

The majority of the visual clips in the HMDB51 dataset are taken from motion clips, although some are also sourced from public available databases and online video libraries such as YouTube. There are 6766 visual clips with resolutions of 320 × 240. HMDB51 contains 51 different actions to understand human activity, interaction and behavior [5]. The HMDB51 single-view dataset, which helps researchers to understand the different action sequences with a frame rate of 30 fps.

The main aim of this dataset is to accurately recognize human motion. Visual clips of this dataset are cluttered, and the background is not static. This dataset is not suitable for real-time action recognition. Compared to today's scenario, this dataset is relatively small in volume, which can lead to overfitting in the deep learning model.

HMDB51 comprises 6,849 video clips divided into 51 action categories such as "clap," "drink," and "kiss." These videos are sourced from digitized movies and public web videos, which introduce challenges such as camera motion, occlusions, and scene transitions. The dataset includes three splits for training/testing and is annotated with metadata, including body-part visibility and camera motion [5].Figure 3 illustrates example frames that showcase activity sequences from the HMDB51 dataset.

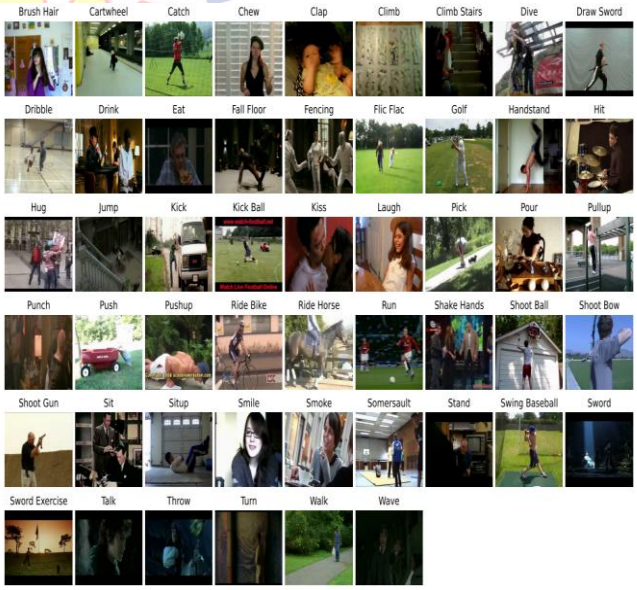

Figure 3: Sample frames illustrating several sorts of human actions from the HMDB51 Dataset [5].

JHMDB

The Brown University team of researchers published the Joint-annotated Human Motion Data Base (JHMDB) [6], a completely annotated data collection for human postures and activities. This dataset contains nearly 920 visual clips with 21 action categories, like running,

picking, brushing hair, pouring, throwing, shooting the bow, etc.

The JHMDB dataset includes a variety of annotations, making it helpful for action identification and posture estimation studies. There are also annotations of joint positions, which are important points on the human body for detecting posture and motion. Puppet masks are also included, where the human segmentation masks are used to separate an individual from the background. Optical flow annotations provide information about motion between consecutive frames, helping models understand motion dynamics. In addition, action descriptors are used to specify the activity classes at the video level of each clip. Meta labels with further details, such as camera moves, perspective, and other features that form part of the scene, are also included in the collection.

NTU RGB+D

The NTU RGB+D dataset [7] is a comprehensive standard dataset created by Nanyang Technological University for experiments in 3D activity analysis and human action detection. It is considered one of the largest databases for evaluating DL (Deep Learning) and skeleton-based action identification techniques. The data was gathered using the Microsoft Kinect v2 sensor and contains about 56,000 visual samples from 40 people performing 60 action classes. These activities are not only everyday human activities, such as eating, drinking, reading, writing, and dressing, but also health-related activities, such as coughing and falling, and interaction-related activities, such as hugging, hand-shaking, kicking, and punching. Some action sequence represented in Figure 4. The collection contains four distinct data modalities: RGB clips, depth sequence information, infrared (IR) frames, and 3D skeletal data. Among them, the skeletal modality is the most commonly used in research because it provides concise and informative representations of human body motion. Each skeleton sequence frame contains 25 body joints represented by three-dimensional coordinates (x, y, z), and the dataset permits tracking up to 2 humans concurrently in a single frame. The data were collected from three separate camera perspectives and under numerous setup conditions to ensure viewpoint and environmental variability. The dataset offers two standard evaluation methods for assessing the effectiveness of action recognition algorithms: cross-subject (CS) assessment and cross-view (CV). In the Cross-Subject methodology, the training and test sets include a variety of subjects to assess the model's generalization across people. The Cross-View approach separates the training and test sets by camera view to assess robustness to view variability [7].

The NTU RGB+D dataset has become a standard benchmark for developing cutting-edge ML and DL techniques, especially Graph Convolutional Networks (GCNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Transformer-based architectures, due to its size, diversity, and multimodal nature. ST-GCN, 2s-AGCN, CTR-GCN, MSG3D, and PoseC3D are well-known models that have been assessed on this dataset. Later, NTU RGB+D 120, an expanded version of the dataset, was released to further enhance the intricacy and variety of action recognition tasks. This expanded dataset includes 114,480 video samples from 120 activity categories, performed by 106 individuals across a greater variety of camera configurations [8] and activities, as shown in Figure 5. Applications including smart tracking, healthcare tracking, robotics, recognition of gestures, interaction between humans and computers, and smart environment systems make substantial use of the NTU RGB+D datasets [8].

Figure 4:Sample visual images of NTU RGB-D Dataset **[7]**.

Figure 5:Sample visual images of NTU RGB-D 120 Dataset **[8]**.

Sports 1M Dataset

The sports-1 M[9] dataset is one of the largest and is commonly used for recognizing human actions in sports video clips. It includes nearly 1 million YouTube videos categorized into 487 unique sports-action classes, making it an excellent tool for training and testing deep learning models for activity detection. The data applies to a variety of sports, including football, basketball, volleyball, swimming, boxing, wrestling, skiing and tennis. The video clips are obtained from real-life situations and exhibit a range of variations in camera angles, lighting conditions, background noise, motion blur, and obstructions, making the detection process more complex and realistic. The large volume and variety of Sports-1M enable researchers to uncover robust spatiotemporal patterns and improve the generalizability of action identification models[9].

Activity Net Dataset

The ActivityNet Dataset [10] is one of the most comprehensive standards for interpreting human actions in visual clips. Experts from the Activity Net Alliance introduced it to advance the fields of event detection, temporal action localization, action identification, and video captioning. ActivityNet emphasizes untrimmed, real-world clips, making it more representative of real-world scenarios in monitoring, sports evaluation, robotics, and multimedia retrieval systems than previous datasets that focused on brief, trimmed video snippets featuring a single activity. ActivityNet spans 200 human action categories and includes over 20k video clips gathered from web sources and organized into a hierarchical categorization [10].

Playing an instrument, cooking, working out, gardening, swimming, and participating in sporting events are just a few of the many every day, recreational, and athletic activities depicted in movies. The dataset is far larger and more difficult than many earlier action recognition datasets, with a total length of 849 hrs. ActivityNet's emphasis on temporal annotations, in which each video contains exact start and finish times indicating when an activity occurs, is one of its unique features. Due to the untrimmed nature of the clips, a single visual clip may feature several actions, and a substantial portion of the video may contain unrelated content. Because of this feature, the dataset is especially useful for temporal action recognition and localization, where the goal is to both identify the activity and pinpoint its timing in the video timeline. To facilitate standardized evaluation, the dataset is separated into three groups: training, testing, and validation. ActivityNet, with its huge scale, numerous activity categories, and realistic video material, remains one of the most essential datasets for developing research in human action identification, video interpretation, and temporal event detection. All the visual clips in this dataset were composed from social video-sharing platforms. This dataset is separated into three subsets based on application domain, including untrimmed video classification, trimmed video classification, and activity detection on all untrimmed video clips.

Figure 6:Visual sequence examples from the Activity Net dataset **[10]**.

Something-Something V2

The Something-Something V2 dataset is a huge reference dataset designed for human activity identification and video analysis, with a focus on modeling temporal linkages and human-object interactions [11]. The dataset consists of approximately 22,000 visual clips across 174 action classes, making it one of the largest and most challenging datasets for fine-grained activity detection.

Unlike typical action detection datasets, which often allow models to exploit scene context or object appearance, Something-Something V2 requires a grasp of the temporal evolution of human-object interactions. These videos were captured using a crowdsourcing system, in which people performed prescribed actions with common items[11] .

Some instances of activity types consist of "putting some thing onto something," "shifting something out of somet hing," and "covering up something with something". Some activity of the dataset is represented in Figure **7**.

The dataset contains 168,913 training video clips, 24,777 validation clips, and 27,157 test visual clips. Because of its emphasis on temporal reasoning and motion dynamics, Something-Something V2 has become a standard benchmark for evaluating modern video comprehension models, including 3D CNNs, Transformer-based architectures, and self-supervised learning techniques.

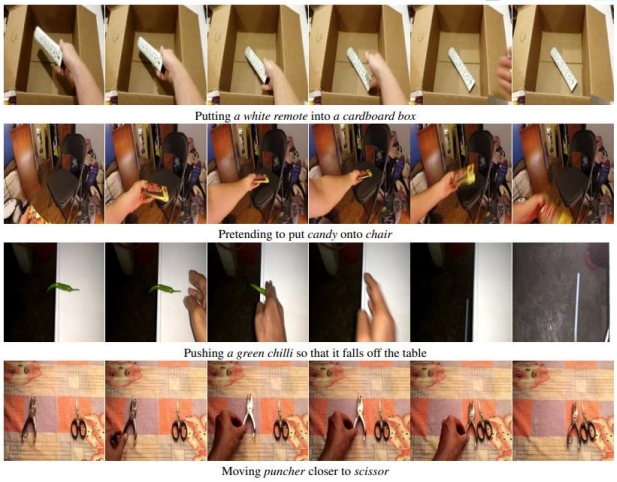

Figure 7:Description of activity performed in dataset **[11]**.

Fine Gym

FineGym [12] is a huge structured visual dataset designed for fine-grained human activity interpretation in sports, notably gymnastics. Shao et al.[12] established a dataset to overcome the shortcomings of traditional activity-detection datasets, which largely focus on coarse-grained activity categorization and often fail to discern subtle differences between comparable activities. FineGym is derived from actual gymnastics contesting video clips and has comprehensive temporal annotations at multiple semantic levels, making it a challenging benchmark for fine-grained action identification, temporal action localization, and activity parsing. The collection contains roughly 708 hours of high-quality gymnastics video clips, with over 95% available in HD 720p or 1080p. These videos are taken from worldwide gymnastics tournaments and show a variety of perspectives, athlete looks, and execution styles. Unlike typical action detection datasets, where background context is important for classification, FineGym has a highly consistent background, pushing models to focus on small movement sequences and body dynamics [12].

PKU MMD

Authors [13] developed this huge dataset at Peking University's Institute of Computer Science and Technology, with funding from Microsoft Research Asia. It centred on multimodal activity analysis and the identification of prolonged, ongoing sequence activities. They used Kinect v2.0 to capture visual clips for this dataset. This dataset is divided into two phases: short and long video clips. Around 1076 long videos, with a duration of 3 to 4 mins, and these actions performed by almost 66 different people in 51 activity classes. These 51 classes are further divided into 41 activity classes for the daily routine and 10 interaction-based classes[13].

PKU-MMD has the benefit of being multimodal, since it delivers four complementary data streams: RGB videos, depth maps, infrared (IR) sequences, and 3D skeleton joint coordinates. Figure 8 describes the analysis of the PKU MMD dataset. This multimodal architecture enables investigators to explore modality-specific and multimodal fusion {Citation}detection.

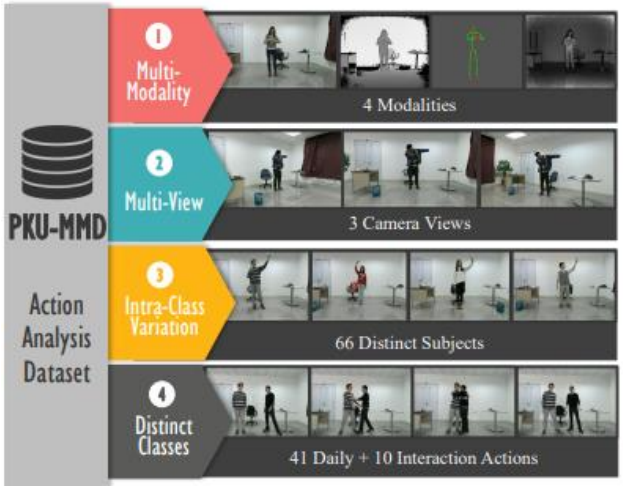

Figure 8:Description of PKU MMD Dataset [13].

III. COMPARATIVE ANALYSIS OF VISION DATASETS

Various authors evaluate these vision-based datasets, and a proposed methodology is proposed to correctly classify each action class. Error! Reference source not f ound. describes the HAR Dataset, the state of the reference model, and the performance achieved by various researchers.

Table 2 Comparative Analysis of Vision Dataset for Human Activity Detection

Year	Dataset	Visual-Clips	Model	Performa-nce
2011	HMDB51	6766	FTP-Uniformer V2[14]	91.5
2012	UCF101	13320	AV-MaskEnhancer[15]	98.8
2013	JHMDB	920	ConvNext+BiLSTM+Temporal Convolution Network[16]	83.38
2014	Sports-1M	1M	P3D-ResNet[17]	87.4
2015	Activity-Net	20000	Intern Video[18]	95.9
2016	NTU RGB D	56880	LLM-Action Recognizer[19]	95.0(CS) 98.4(CV)
2017	PKU-MMD-I	1076	Swimtrans Net[20]	98.4
2018	Something-something V2	220847	3D-CNN+Transformer[21]	73.9
2019	Kinetics-700	650317	InternVideo2[18]	85.9
2020	FineGym	29000	Joint Learning[22]	91.8

CS: Cross Subject, CV: Cross View

P3D ResNet: Pseudo-3D Residual Network

IV. FUTURE SCOPE AND CONCLUSION

This study shows how much research has been successfully done on the vision dataset. This article evaluates several vision datasets available for activity classification and the number of action classes in each dataset. We elaborate on the key dataset for activity classification, including how it was collected, the environment in which it was collected, the number of clips and action classes, and its limitations. This study also investigates the performance of key datasets with methods. This study provides a comparative analysis of the key datasets for HAR. Real-time activity identification is also a challenging task. In the future, sensor-based hybrid datasets and real-time activity.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] V. Sharma, M. Gupta, A. K. Pandey, D. Mishra, and A. Kumar, "A Review of Deep Learning-based Human Activity Recognition on Benchmark Video Datasets," Applied Artificial Intelligence, vol. 36, no. 1, p. 2093705, Dec. 2022, doi: 10.1080/08839514.2022.2093705.

[2] T. Singh and D. K. Vishwakarma, "Video benchmarks of human action datasets: a review," Artif Intell Rev, vol. 52, no. 2, pp. 1107–1154, Aug. 2019, doi: 10.1007/s10462-018-9651-1.

[3] W. Kay et al., "The Kinetics Human Action Video Dataset," May 19, 2017, arXiv: arXiv:1705.06950. doi: 10.48550/arXiv.1705.06950.

[4] K. Soomro, A. R. Zamir, and M. Shah, "UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild," Dec. 03, 2012, arXiv: arXiv:1212.0402. doi: 10.48550/arXiv.1212.0402.

[5] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "HMDB: A large video database for human motion recognition," in 2011 International Conference on Computer Vision, Nov. 2011, pp. 2556–2563. doi: 10.1109/ICCV.2011.6126543.

[6] H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M. J. Black, "Towards Understanding Action Recognition," presented at the Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 3192–3199. Accessed: May 15, 2026. [Online]. Available: https://openaccess.thecvf.com/content_iccv_2013/html/Jhuang_Towards_Understanding_Action_2013_ICCV_paper.html

[7] A. Shahroudy, J. Liu, T.-T. Ng, and G. Wang, "NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 1010–1019. doi: 10.1109/CVPR.2016.115.

[8] J. Liu, A. Shahroudy, M. Perez, G. Wang, L.-Y. Duan, and A. C. Kot, "NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding," IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 10, pp. 2684–2701, Oct. 2020, doi: 10.1109/TPAMI.2019.2916873.

[9] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-Scale Video Classification with Convolutional Neural Networks," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA: IEEE, Jun. 2014, pp. 1725–1732. doi: 10.1109/CVPR.2014.223.

[10] F. Caba Heilbron, V. Escorcia, B. Ghanem, and J. Carlos Niebles, "ActivityNet: A Large-Scale Video Benchmark for Human Activity Understanding," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 961–970. Accessed: May 31, 2026. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2015/html/Heilbron_ActivityNet_A_Large-Scale_2015_CVPR_paper.html

[11] R. Goyal et al., "The 'Something Something' Video Database for Learning and Evaluating Visual Common Sense," in 2017 IEEE International Conference on Computer Vision (ICCV), Venice: IEEE, Oct. 2017, pp. 5843–5851. doi: 10.1109/ICCV.2017.622.

[12] D. Shao, Y. Zhao, B. Dai, and D. Lin, "FineGym: A Hierarchical Video Dataset for Fine-grained Action Understanding," Apr. 14, 2020, arXiv: arXiv:2004.06704. doi: 10.48550/arXiv.2004.06704.

[13] C. Liu, Y. Hu, Y. Li, S. Song, and J. Liu, "PKU-MMD: A Large Scale Benchmark for Continuous Multi-Modal Human Action Understanding," Mar. 28, 2017, arXiv: arXiv:1703.07475. doi: 10.48550/arXiv.1703.07475.

[14] H. Lu, A. A. Salah, and R. Poppe, "Improving the generalization of ViTs for action understanding with VLM pre-training," Pattern Recognition, vol. 179, p. 113794, Nov. 2026, doi: 10.1016/j.patcog.2026.113794.

[15] X. Diao, M. Cheng, and S. Cheng, "AV-MaskEnhancer: Enhancing Video Representations through Audio-Visual Masked Autoencoder," Dec. 20, 2023, arXiv: arXiv:2309.08738. doi: 10.48550/arXiv.2309.08738.

[16] J. Shin, Md. A. M. Hasan, Md. Maniruzzaman, S. Nishimura, and S. Alfarhood, "Video-Based Human Activity Recognition Using Hybrid Deep Learning Model," CMES, vol. 143, no. 3, pp. 3615–3638, 2025, doi: 10.32604/cmes.2025.064588.

[17] Z. Qiu, T. Yao, and T. Mei, "Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks," Nov. 28, 2017, arXiv: arXiv:1711.10305. doi: 10.48550/arXiv.1711.10305.

[18] Y. Wang et al., "InternVideo2: Scaling Foundation Models for Multimodal Video Understanding," Aug. 14, 2024, arXiv: arXiv:2403.15377. doi: 10.48550/arXiv.2403.15377.

[19] H. Qu, Y. Cai, and J. Liu, "LLMs are Good Action Recognizers".

[20] H. Chen and X. Yue, "Swimtrans Net: a multimodal robotic system for swimming action recognition driven via Swin-Transformer," Front Neurorobot, vol. 18, p. 1452019, Sep. 2024, doi: 10.3389/fnbot.2024.1452019.

[21] S. Kang, H. Huo, J. Xu, A. Mei, and C. Zhang, "Using Multi-Layer Bidirectional Distillation to Enhance Local and Global Features for Action Recognition," Sensors (Basel), vol. 25, no. 22, p. 6849, Nov. 2025, doi: 10.3390/s25226849.

[22] M. C. Leong, H. L. Tan, H. Zhang, L. Li, F. Lin, and J. H. Lim, "Joint Learning on the Hierarchy Representation for Fine-Grained Human Action Recognition," in 2021 IEEE International Conference on Image Processing (ICIP), Sep. 2021, pp. 1059–1063. doi: 10.1109/ICIP42928.2021.9506157.