International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Artificial Intelligence Enabled Methods for Human Action Recognition using Surveillance Videos

Boddupally Janaiah
Research Scholar, Osmania University, Department of CSE, Hyderabad, India
Assistant Professor, Department of CSE, MVSR Engineering College
Osmania University, Hyderabad, India
janaiah_cse@mvsrec.edu.in

Dr. Suresh Pabboju
Professor, Chaitanya Bharathi Institute of Technology
Department of IT
Hyderabad, India
psuresh_it@cbit.ac.in

Abstract— Computer vision applications have been attracting researchers and academia. It is more so with cloud computing resources enabling such applications. Analysing video surveillance applications became an important research area due to its widespread applications. For instance, CCTV camera are used in public places in order to monitor situations, identify any theft or crime instances. In presence of thousands of such surveillance videos streaming simultaneously, manual analysis is very tedious and time consuming task. There is need for automated approach for analysis and giving notifications or findings to officers concerned. It is very useful to police and investigation agencies to ascertain facts, recover evidences and even exploit digital forensics. In this context, this paper throws light on different methods of human action recognition (HAR) using machine learning (ML) and deep learning (DL) that come under Artificial Intelligence (AI). It also reviews methods on privacy preserving action recognition and Generative Adversarial Networks (GANs). This paper also provides different datasets being used for human action recognition research besides giving an account of research gaps that help in pursuing further research in the area of human action recognition.

Keywords- Human Action Recognition, Machine Learning, Deep Learning, Artificial Intelligence, Generative Adversarial Network

I. INTRODUCTION

Computer vision applications are rapidly increasing in various domains. Particular video processing applications play crucial role in domains like security and video surveillance. In such applications, it is important to protect privacy of people. Video analytics can obtain required details without disclosing identity of people unless it is necessity governed by laws prevailing. Human activity recognition is an application that needs to detect human actions while preserving privacy. Existing literature on this has been reviewed and research gaps are identified. Many privacy preserving action recognition methods are found in the literature. They are explored in [1], [2], [3], [4] and [5] to mention few. These approaches do have action recognition and also preserving privacy. They used different ML and DL models that are associated with artificial intelligence (AI). There are many approaches found in the literature using Generative Adversarial Network (GAN) architectures. They are explored in [4], [79], [80], [81], [82] and [83]. These models are promising as they do have support for data augmentation. There are many ML and deep learning methods such as [1], [2], [3], [4] and [5], to mention few, found in the literature. They used AI enabled methods but they lack in privacy preserving approaches. They also lack in the usage of GAN based architecture.

Figure 1: Actions from HMDB dataset

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

As explored in [36] HMDB dataset is most widely used publicly available dataset. It has 51 actions available for research. An excerpt of the actions is provided in Figure 1. Review of literature has resulted in many useful insights. First, human action recognition is one of the computer vision applications that has become more useful in the contemporary era. Second, automatic recognition of human actions and notifying authorities concerned saves time and effort besides helping in evidence recovery in investigations and digital forensics. Third, it is important to have privacy preserving approach to human action recognition to ensure the privacy of people is not lost in certain applications. GAN based approaches became more attractive due to their power to have data augmentation and improve quality in the training process. Our contributions in this paper are as follows.

1. Literature review is made on different methods used for human action recognition from videos. It includes AI based methods encapsulating machine learning, deep learning and GAN based approaches.
2. Research gaps in the existing approaches have been identified to ascertain the need for further research in the area of human action recognition.
3. Summary of the datasets widely used in vision based human action recognition research is provided.

The remainder of the paper is structured as follows. Section 2 focused on privacy preserving HAR methods. Section 3 reviews on the GAN based approaches. Section 4 reviews on diversified benchmark datasets used in the current research. Section 5 presents latest research findings. Section 6 provides the research gaps identified while Section 7 concludes the paper and gives directions for future work.

II. PRIVACY PRESERVING APPROCHES FOR HUMAN ACTION RECOGNITION

Preserving privacy while recognising human actions in videos has its benefits in certain applications where what is done is more important than who has done it. This section covers the state of the art on privacy preserving approaches to human action recognition. Rajput et al. [1] proposed a cloud based service for HAR. They defined image obfuscation method for privacy preserving approach in action recognition. It is cloud based service that can be used on demand. It can be integrated with other applications as well. It is secure means of doing things and scalable as it is cloud based service. Wang et al. [2] proposed a methodology based on phase correlation and deep learning in order to work with coded aperture videos rather than traditional camera videos. They investigated on the privacy preserving approaches and incorporated privacy in their methodology. Yan et al. [3] on the other hand explored image processing techniques for HAR. It mainly focused on the image segmentation that occlude human targets for recognition. It also has extraction of privacy related data from RGB to see that the privacy is not lost with transforming the information for non-disclosure of identification. Wu et al. [4] proposed GAN based framework with deep learning for HAR and visual privacy. Their method includes action-attribute correlation in order to achieve better performance and visual privacy. Action distribution and privacy distribution are considered in order to discriminate privacy and preserve it which achieving HAR. Dai et al. [5] explored their research with low resolution cameras and proposed a method based on image processing. In fact, their algorithm is based on time series pixel wise data in order to ensure privacy and also HAR without the use of ML approaches.

Reference/Year	**Techniques**	**Advantages**	**Limitations**	**Dataset**
Rajput et al. [1], 2020	Deep CNN and RGB-D sensors	Security, efficient recognition of human actions, scalability	No data augmentation technique is used.	UTD-MHAD
Wang et al. [2], 2019	Deep learning and phase correlation	Privacy preserving HAR.	A good first step using coded aperture videos. Needs further improvement.	UCF and NTU datasets
Yan et al. [3], 2020	Image segmentation	Privacy preserving HAR.	No ML approaches are used.	UCF101 and HMDB51
Wu et al. [4]	GAN and deep learning	Visual privacy and HAR	Privacy leakage risk analysis is to be improved.	A-HMDB51
Dai et al. [5], 2020	Time series based pixel wise HAR algorithm	Privacy, HAR with low resolution cameras	No ML approaches are used.	Simulated room scenario based actions
Imran et al. [8], 2020	CNN	Accuracy in violent AR, privacy, security and memory saving	No data augmentation technique is used.	Hockey Fight, , Violent Flows and Movies datasets
Liu and Zhang [11]	Coupled CNNs	Privacy, HAR	No data augmentation technique is used.	PL-interaction and DogCentric
Hao et al. [12], 2020	Neural Network and SVM	Privacy, HAR	Deep learning is not incorporated.	Wireless vision dataset
Liu et al. [17], 2020	Visual shielding	Fall detection and privacy	Degree of privacy preservation is not considered.	Combined CS dataset

Xu and Pombo [25], 2019	IoT assisted monitoring method	Privacy, HAR	Deep learning is not exploited.	Video collected using IoT sensors

Table 1: Summary of privacy preserving human action recognition methods based on ML and DL

Imran et al. [8] proposed a CNN based method for HAR with privacy and security. For security encryption is used. For privacy, privacy preserving approach such as transformation of identifiable information is employed. Its novelty is that, they considered the violent activity recognition from three different datasets that include movies, violent flows and hockey flight incidents. Liu and Zhang [11] investigated on low resolution videos for HAR and privacy. They exploited CNN by making a partially coupled CNN model that could provide better performance. Their method has provision for privacy and HAR. Hao et al. [12] proposed a methodology for HAR and prserve privacy as well using a neural network linked to WiFi and SVM. They exploited wireless vision dataset for the empirical study. They could identify human activity and location with faster convergence. Liu et al. [17] proposed method based on the notion of visual shielding for fall detection from videos. Their method also ensures privacy of humans by considering privacy preserving approach. It makes use of combined CS dataset that brings diversity of fall activities. Xu and Pombo [25] proposed an IoT technology based method that captures video and then data is analysed to know human actions. It also sees that the privacy of humans is not lost. It is evaluated with live videos captured using IoT technology.

III. GENERATIVE ADERSARIAL NETWORK BASED METHODS

GAN is widely used in different real world applications. This section throws light on fundamentals of GAN and its constituents. It is designed to achieve adversarial learning where there is non-cooperative game between generator (G) and discriminator (D). As shown in Figure 2, generator is designed to produce new data samples. On the other hand, the discriminator non-cooperatively examines the samples generated by G to know whether they are real or fake.

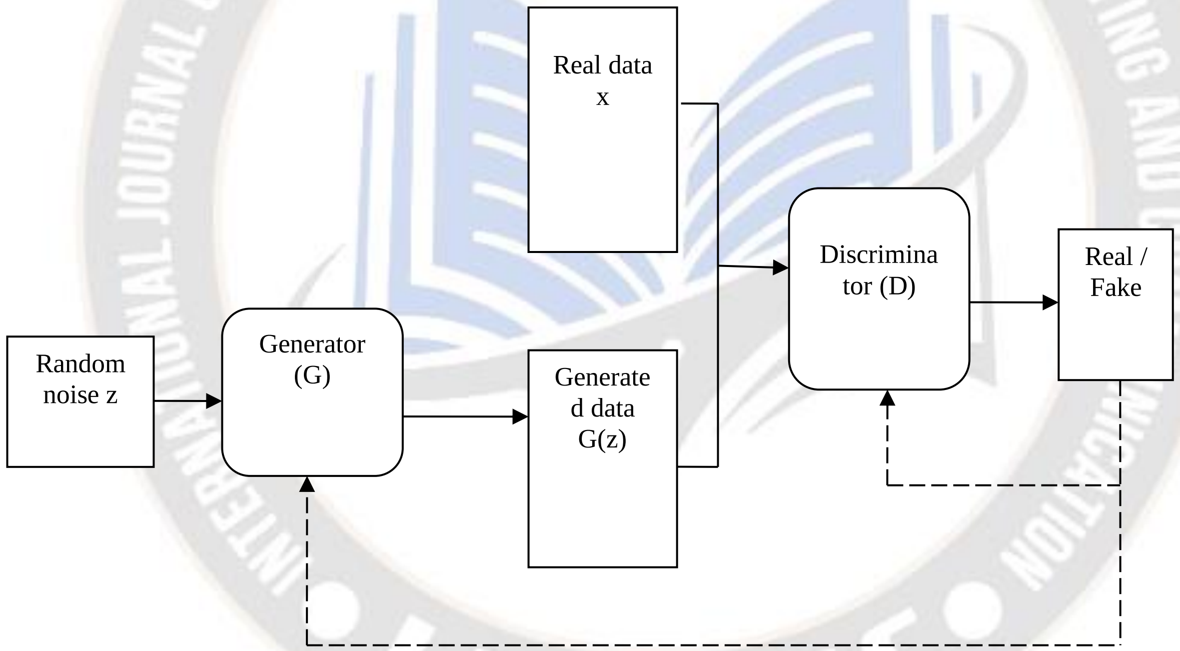

Figure 2: Overview of GAN architecture

Both G and D can be implemented using deep learning techniques. Their hyperparameters can be tuned by using backpropagation algorithm. Loss function is used by D to determine the correctness of generated samples as in Eq. 1.

$$V(D,\theta^{(D)}) = -E_{x\sim p_r(x)}[\log D(x)] - E_{z\sim p_g(z)}[\log(1-D(g(z)))] \quad (1)$$

In the same fashion, G uses a loss function as expressed in Eq. 2.

$$V(G,\theta^{(G)}) = E_{z\sim p_g(z)}[\log(1-D(g(z)))] \quad (2)$$

In the process of non-cooperative game where G tries to maximize $V(G)$, $(\theta^{(D)}, \theta^{(G)})$ while D tries to maximize $V(D)$, $(\theta^{(D)}, \theta^{(G)})$ by updating $\theta^{(D)}$ and $\theta^{(D)}$ respectively. The two players use loss functions that have parameter dependency. For mutual update of parameters Nash equilibrium is used.

$$\text{Min max } V(D,G) = E_{x\sim p_r(x)}[\log D(x)] + E_{z\sim p_g(z)}[\log(1-D(g(z)))] \quad (3)$$

As expressed in Eq. 3, the GAN functionality is provide in terms of min-max optimization. Both the components such as D and G involve in their functionality to achieve optimization in producing new samples.

Reference/ Year	**Techniques**	**Advantages**	**Limitations**	**Dataset**

Wu et al. [4]	GAN and deep learning	Visual privacy and HAR	Privacy leakage risk analysis is to be improved.	A-HMDB 51
Wu et al. [79], 2018	GAN based method	Privacy, HAR	No deep learning is used.	SBU Kinect Interaction
Peng and Schmid [80]	R-CNN with GAN	Visual recognition, privacy	Needs further improvement in training process.	UCF-Sports, J-HMDB and UCF101
Ren and Lee [81], 2018	Multi-task feature learning and GAN	Image synthesis, action recognition, privacy	Privacy is not implemented	Synthetic RGB images
Ryoo et al. [82], 2017	GAN	Privacy and action recognition with low resolution images.	Privacy budget needs to be improved.	YouTube videos
Liu et al. [83], 2017	GAN and deep learning	Face recognition	Privacy is not implemented	MegaFace Challenge 1, YTF and LRW

Table 2: Summary of human action recognition methods based on GAN Architecture

Wu et al. [4] proposed GAN based framework with deep learning for HAR and visual privacy. Their method includes action-attribute correlation in order to achieve better performance and visual privacy. Action distribution and privacy distribution are considered in order to discriminate privacy and preserve it which achieving HAR. Wu et al. [79] proposed a GAN based model for ensuring visual privacy and detect human actions from videos. It makes use of Kinetic Interaction dataset which is publicly available. Peng and Schmid [80] implemented a method based on GAN and R-CNN for privacy and visul recognition. Ren and Lee [81] proposed a multi-task learning approach based on GAN architecture. It uses synthetic images and performs action recognition and preserves privacy. Ryoo et al. [82] proposed GAN bsed method for action recognition and provide privacy to humans. Their research focuses on the low resolution images. Liu et al. [83] on the other hand explores face recognition with deep learning based on GAN architecture.

A. Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

IV. LATEST RESULT FINDINGS

This section presents summary of most recent research findings with respect to AI or deep learning usage for automatic human action recognition

S no	Reference	Authors & Year	Approach	Limitation
1	[84]	Singh et al., 2020	Human action recognition as a remote cloud service	Not suitable for long term actions
2	[85]	Imran et al.,[2020]	Convolutional neural network (CNN), Deep learning	Limited to detection of only violent activities
3	[86]	Kwabena Owusu et al. [2021]	Deep learning	Yet to be improved to deal with large real time datasets.
4	[87]	Aparna Akula et al. [2018]	Deep Learning	It has issue of mis-classifications.
5	[88]	Mina Hashemian et al. [2019]	Privacy-preserving, Activity recognition	Relies on machine learning. Deep learning could improve performance further.

6	[89]	Jixin Liu and Leilei Zhang [2020]	Convolutional Neural Network	Entire network is not used for training.
7	[90]	Huaijun Wang,Jing Zhao et al [2020]	Hybrid Deep Learning	Need to be improved to detect more actions.
8	[91]	Zhongzheng Ren et al. [2018]	GAN based approach	Face anonymization could be improved further.
9	[92]	Zhenyu Wu et al [2020]	Visual privacy, adversarial training	Adversarial training is found inefficient and unstable
10	[93]	Zhenyu Wu et al. [2020]	GAN based	It suffers from privacy leakage risk

Table 3: Latest research findings on human action recognition

As presented in Table 3, various approaches proposed in the most recent works pertaining to automatic human action recognition are provided along with insights.

V. HUMAN ACTION RECOGNITION PUBLIC DATASETS

Research on human action recognition encapsulate the usage of different datasets. In fact, it is indispensable to have such research activity with diversified datasets available. Fortunately, number of datasets are available for this kind of research. Table 1 shows the list of publicly available datasets and their details.

SL. NO.	Dataset Name	Videos	Classes	Dataset URL	References
1	HMDB	6849	51	https://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database	[36], [37], [38], [39], [40]
2	UCF101	13320	101	http://crcv.ucf.edu/data/UCF101.php	[41], [42], [43], [44],[45]
3	Hollywood extended	937	16	https://www.di.ens.fr/willow/research/actionordering/	[46], [47], [48], [49], [50]
4	Breakfast dataset	1989	10	https://serre-lab.clps.brown.edu/resource/breakfast-actions-dataset/	[51], [52], [53], [54], [55]
5	JHMDB	928	21	http://jhmdb.is.tue.mpg.de/	[56], [57], [58], [59], [60]
6	Charades	9848	157	https://prior.allenai.org/projects/charades	[61], [62], [63], [64], [65]
7	AVA	57600	80	https://research.google.com/ava/	[66], [67], [68], [69]
8	Kinetics	300000	400	https://www.deepmind.com/research/open-source/open-source-datasets/kinetics	[70], [71], [72], [73], [74]
9	Epic-Kitchens	432	149	https://epic-kitchens.github.io/2018	[75], [70], [61],
10	Something-Something	220847	174	https://20bn.com/datasets/something-something	[76], [77], [70],
11	Moments	1	339	http://moments.csail.mit.edu/	NA

Table 4: Details of publicly available datasets on human actions

Each dataset is made up of many video samples and classes (actions) suitable for the research. And each dataset is widely used in the research. The references of the usage of datasets are provided. These datasets help in designing novel frameworks using ML and ML for improving the state of the art.

VI. RESEARCH GAPS

Important insights found in the research are given here. First, human action recognition is one of the computer vision applications that has become more useful in the contemporary era. Second, automatic recognition of human actions and notifying authorities concerned saves time and effort besides helping in evidence recovery in investigations and digital forensics. Third, it is important to have privacy preserving approach to human action recognition to ensure the privacy of people is not lost in certain applications. GAN based approaches became more attractive due to their power to have data augmentation and improve quality in the training process. Here are the research gaps identified. First, there is need for exploiting emerging technologies in AI for more efficient means in detection of human actions automatically from running or streaming videos. Second, there is scope for further investigation and enhancement of GAN based approaches as they do have potential to improve quality in training process. Third, a holistic or comprehensive framework based on AI is essential for leveraging the performance of video analytics. Fourth, there is need for provision of privacy preserving approaches as it can ensure the privacy of citizens is not lost in certain applications.

VII. CONCLUSION AND FUTURE WORK

This paper throws light on different methods of human action recognition using machine learning (ML) and deep learning (DL) that come under Artificial Intelligence (AI). It also reviews methods on privacy preserving action recognition and Generative Adversarial Networks (GANs). This paper also provides different datasets being used for human action recognition research besides giving an account of research gaps that help in pursuing further research in the area of human action recognition. Important insights found in the research are given here. First, human action recognition is one of the computer vision applications that has become more useful in the contemporary era. Second, automatic recognition of human actions and notifying authorities concerned saves time and effort besides helping in evidence recovery in investigations and digital forensics. Third, it is important to have privacy preserving approach to human action recognition to ensure the privacy of people is not lost in certain applications. GAN based approaches became more attractive due to their power to have data augmentation and improve quality in the training process. In future, we intend to propose a comprehensive framework for privacy preserving action recognition using AI enabled approaches to fill the research gaps identified.

VIII. DECLARATIONS

A. *Funding*

There was no financial support received by the authors in this research.

B. *Conflicts of Interest*

The authors declare no conflicts of interest.

REFERENCES

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [1]. Where appropriate, include the name(s) of editors of referenced books. The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first . . ."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] Rajput, Amitesh Singh; Raman, Balasubramanian; Imran, Javed (2020). Privacy-preserving human action recognition as a remote cloud service using RGB-D sensors and deep CNN. Expert Systems with Applications, 152, p1-15.

[2] Wang, Zihao W.; Vineet, Vibhav; Pittaluga, Francesco; Sinha, Sudipta N.; Cossairt, Oliver; Kang, Sing Bing (2019). [IEEE 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) - Long Beach, CA, USA (2019.6.16-2019.6.17)] 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) - Privacy-Preserving Action Recognition Using Coded Aperture Videos. , p1–10.

[3] Yan, Jiawei; Angelini, Federico; Naqvi, Syed Mohsen (2020). [IEEE ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) - Barcelona, Spain (2020.5.4-2020.5.8)] ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) - Image Segmentation Based Privacy-Preserving Human Action Recognition for Anomaly Detection. , p8931–8935.

[4] Wu, Zhenyu; Wang, Haotao; Wang, Zhaowen; Jin, Hailin; Wang, Zhangyang (2020). Privacy-Preserving Deep Action Recognition: An Adversarial Learning Framework and A New Dataset. IEEE Transactions on Pattern Analysis and Machine Intelligence, p1–14.

[5] Dai, Ji; Wu, Jonathan; Saghafi, Behrouz; Konrad, Janusz; Ishwar, Prakash (2015). [IEEE 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) - Boston, MA, USA (2015.6.7-2015.6.12)] 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) - Towards privacy-preserving activity recognition using extremely low temporal and spatial resolution cameras. , p68–76.

[6] Chaudhary, Sachin; Dudhane, Akshay; Patil, Prashant; Murala, Subrahmanyam (2019). [IEEE 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) - Taipei, Taiwan (2019.9.18-2019.9.21)] 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) - Pose Guided Dynamic Image Network for Human Action Recognition in Person Centric Videos. , p1–8.

[7] Akula, Aparna; Shah, Anuj K.; Ghosh, Ripul (2018). Deep Learning Approach for Human Action Recognition in Infrared Images. Cognitive Systems Research, p1-19.

[8] Imran, J., Raman, B., & Rajput, A. S. (2020). Robust, efficient and privacy-preserving violent activity recognition in videos. Proceedings of the 35th Annual ACM Symposium on Applied Computing. P1-8.

[9] Angelini, Federico; Yan, Jiawei; Naqvi, Syed Mohsen (2019). [IEEE ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) - Brighton, United Kingdom (2019.5.12-2019.5.17)] ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) - Privacy-preserving Online Human Behaviour Anomaly Detection Based on Body Movements and Objects Positions. , p8444–8448.

[10] Amsaleg, Laurent; Guðmundsson, Gylfi Þór; Gurrin, Cathal; Jónsson, Björn Þór; Satoh, Shin'ichi (2017). [Lecture Notes in Computer Science] MultiMedia Modeling Volume 10132 ||

Spatio-Temporal VLAD Encoding for Human Action Recognition in Videos. , 10.1007/978-3-319-51811-4(Chapter 30), p365–378.

[11] Jixin Liu and Leilei Zhang. (2020). Indoor Privacy-preserving Action Recognition via Partially Coupled Convolutional Neural Network. International Conference on Artificial Intelligence and Computer Engineering (ICAICE), p292-295.

[12] Hao, Y., Shi, Z., & Liu, Y. (2020). A Wireless-Vision Dataset for Privacy Preserving Human Activity Recognition. 2020 Fourth International Conference on Multimedia Computing, Networking and Applications (MCNA). P1-9.

[13] Banerjee, Dighanchal; Rani, Smriti; George, Arun M.; Chowdhury, Arijit; Dey, Sounak; Mukherjee, Arijit; Chakravarty, Tapas; Pal, Arpan (2020). [IEEE 2020 International Joint Conference on Neural Networks (IJCNN) - Glasgow, United Kingdom (2020.7.19-2020.7.24)] 2020 International Joint Conference on Neural Networks (IJCNN) - Application of Spiking Neural Networks for Action Recognition from Radar Data. , p1–10.

[14] Gutoski, Matheus; Lazzaretti, AndrÃ© EugÃªnio; Lopes, Heitor SilvÃ©rio (2020). Deep metric learning for open-set human action recognition in videos. Neural Computing and Applications, p1-14.

[15] Honghui Xu, Zhipeng Cai,Daniel Takabi, and Wei Li. (2021). Audio-Visual Autoencoding for Privacy-Preserving Video Streaming. IEEE INTERNET OF THINGS JOURNAL. 20 (20), p1-13.

[16] Chen, Chen; Liu, Kui; Kehtarnavaz, Nasser (2016). Real-time human action recognition based on depth motion maps. Journal of Real-Time Image Processing, 12(1), p155–163.

[17] Liu, Jixin; Xia, Yinyun; Tang, Zheng (2020). Privacy-preserving video fall detection using visual shielding information. The Visual Computer, p1-12.

[18] Ronald Poppe (2010). A survey on vision-based human action recognition. , 28(6), p976–990.

[19] Farrajota, M.; Rodrigues, João M. F.; du Buf, J. M. H. (2018). Human action recognition in videos with articulated pose information by deep networks. Pattern Analysis and Applications,p1-12.

[20] Perera, Asanka G.; Law, Yee Wei; Ogunwa, Titilayo T.; Chahl, Javaan (2020). A Multiviewpoint Outdoor Dataset for Human Action Recognition. IEEE Transactions on Human-Machine Systems, p1–9.

[21] Latha, BM; Manjula, BK; Venkata Sumana, CH; Hemalatha, KL (2020). Human Action recognition using STIP Evaluation techniques. IOP Conference Series: Materials Science and Engineering, 925, p1-12.

[22] Berlin, S. Jeba; John, Mala (2020). Particle swarm optimization with deep learning for human action recognition. Multimedia Tools and Applications, p1-23.

[23] Ren, Ziliang; Zhang, Qieshi; Gao, Xiangyang; Hao, Pengyi; Cheng, Jun (2020). Multi-modality learning for human action recognition. Multimedia Tools and Applications, p1-19.

[24] Wang, Hao; Yang, Yanhua; Yang, Erkun; Deng, Cheng (2017). Exploring hybrid spatio-temporal convolutional networks for human action recognition. Multimedia Tools and Applications, 76(13), p15065–15081.

[25] Xu, Lina; Pombo, Nuno (2019). [IEEE 2019 IEEE 5th World Forum on Internet of Things (WF-IoT'19) - Limerick, Ireland (2019.4.15-2019.4.18)] 2019 IEEE 5th World Forum on Internet of Things (WF-IoT) - Human Behavior Prediction Though Noninvasive and Privacy-Preserving Internet of Things (IoT) Assisted Monitoring. , p773–777.

[26] Sun, Mingxuan; Wang, Qing; Liu, Zicheng (2020). [IEEE 2020 IEEE International Conference on Multimedia and Expo (ICME) - London, United Kingdom (2020.7.6-2020.7.10)] 2020 IEEE International Conference on Multimedia and Expo (ICME) - Human Action Image Generation with Differential Privacy. , p1–6.

[27] Demir, U., Rawat, Y. S., & Shah, M. (2021). TinyVIRAT: Low-resolution Video Action Recognition. 2020 25th International Conference on Pattern Recognition (ICPR).p1-8.

[28] Rashid Minhas; Aryaz Baradarani; Sepideh Seifzadeh; Q.M. Jonathan Wu (2010). Human action recognition using extreme learning machine based on visual vocabularies. , 73(10-12), p1906–1917.

[29] Chakraborty, S., Mondal, R., Singh, P. K., Sarkar, R., & Bhattacharjee, D. (2021). Transfer learning with fine tuning for human action recognition from still images. Multimedia Tools and Applications, 80(13), p20547–20578.

[30] Wang, Q., & Chen, K. (2017). Alternative Semantic Representations for Zero-Shot Human Action Recognition. Lecture Notes in Computer Science, p87–102.

[31] Liu, Li; Shao, Ling; Rockett, Peter (2013). Boosted key-frame selection and correlated pyramidal motion-feature representation for human action recognition. Pattern Recognition, 46(7), p1810–1818.

[32] Patel, Chirag I; Garg, Sanjay; Zaveri, Tanish; Banerjee, Asim; Patel, Ripal (2016). Human action recognition using fusion of features for unconstrained video sequences. Computers & Electrical Engineering, p1-18.

[33] Kishore K. Reddy, Mubarak Shah (2013). Recognizing 50 human action categories of web videos. , 24(5), p971–981.

[34] Liu, Jixin; Zhang, Ruxue; Han, Guang; Sun, Ning; Kwong, Sam (2020). Video action recognition with visual privacy protection based on compressed sensing. Journal of Systems Architecture, p1-14.

[35] Yi, Yang; Lin, Maoqing (2015). Human Action Recognition with Graph-Based Multiple-Instance Learning. Pattern Recognition, p1-56.

[36] Lei Wang, Piotr Koniusz and Du Q. Huynh3. (219). Hallucinating IDT Descriptors and I3D Optical Flow Features for Action Recognition with CNNs, p1-12.

[37] Joao Carreira and Andrew Zisserman†. (2017). Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. IEEE Conference on Computer Vision and Pattern Recognition, p4724-4734

[38] Vasileios Choutas1, Philippe Weinzaepfel , Jer ome Revaud and Cordelia Schmid Inria. (2018). PoTion: Pose MoTion Representation for Action Recognition. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, p7024-7034.

[39] AJ Piergiovanni and Michael S. Ryoo. (2019). Representation Flow for Action Recognition. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR , 9937-9945.

[40] Jue Wang , Anoop Cherian , Fatih Porikli , Stephen Gould. (2018). Video Representation Learning Using Discriminative Pooling. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition , 1149-1158.

[41] Vasileios Choutas , Philippe Weinzaepfel , Jer ome Revaud , Cordelia Schmid Inria. (2018). PoTion: Pose MoTion Representation for Action Recognition. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7024-7033.

[42] Joao Carreira , Andrew Zisserman. (2017). Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 4724-4733.

[43] Jiagang Zhu , Zheng Zhu , Wei Zou. (2018). End-to-end Video-level Representation Learning for Action Recognition. 2018 24th International Conference on Pattern Recognition (ICPR, 645-650.

[44] Nieves Crasto, Philippe Weinzaepfel, Karteek Alahari, Cordelia Schmid. (2019). Motion-Augmented RGB Stream for Action Recognition, 1-11.

[45] Ali Diba , Mohsen Fayyaz, V ivek Sharma , M.Mahdi Arzani , Rahman Yousefzadeh, Juergen Gall , Luc Van Gool. (2018). Spatio-Temporal Channel Correlation Networks for Action Classification. 1-16.

[46] Jun Li, Peng Lei, Sinisa Todorovic. (2019). Weakly Supervised Energy-Based Learning for Action Segmentation. IEEE Xplore, 6243-6251.

[47] Yaser Souri, Mohsen Fayyaz, Luca Minciullo, Gianpiero Francesca, and Juergen Gall. (2015). Fast Weakly Supervised Action Segmentation Using Mutual Consistency. JOURNAL OF LATEX CLASS FILES. 14, 1-13.

[48] Chien-Yi Chang, De-AnHuang, Yanan Sui, Li Fei-Fei, Juan Carlos Niebles. (2019). D3TW: Discriminative Differentiable Dynamic Time Warping for Weakly Supervised Action Alignment and Segmentation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3541-3550.

[49] Hilde Kuehne, Alexander Richard, Juergen Gall. (2017). Weakly supervised learning of actions from transcripts. University of Bonn, Institute of Computer Science III, 1-12.

[50] Hilde Kuehne, Alexander Richard, and Juergen Gall. (2018). A Hybrid RNN-HMM Approach for Weakly Supervised Temporal Action Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1-14.

[51] Jun Li, Peng Lei. (2019). Weakly Supervised Energy-Based Learning for Action Segmentation. IEEE Xplore, 6243-6251.

[52] Hilde Kuehne, Juergen Gall, Thomas Serre. (2016). An end-to-end generative framework for video segmentation and recognition. arXiv:1509.01947v2 [cs.CV], 1-8.

[53] Hilde Kuehne, Ali Arslan, Thomas Serre. (2014). Recovering the Syntax and Semantics of Goal-Directed Human Activities, 1-8.

[54] Hilde Kuehne, Alexander Richard, and Juergen Gall. (2018). A Hybrid RNN-HMM Approach for Weakly Supervised Temporal Action Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence , 1-14.

[55] Qian Wang, Ke Chen. (2019). Multi-Label Zero-Shot Human Action Recognition via Joint Latent Ranking Embedding. The University of Manchester, UK, 1-27.

[56] An Yan, Yali Wang, Zhifeng Li, Yu Qiao. (2019). PA3D: Pose-Action 3D Machine for Video Recognition. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR, 7914-7923.

[57] Choutas V, Weinzaepfel P, Revaud J & Schmid C. (2018). PoTion: Pose MoTion Representation for Action Recognition. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7024-7033.

[58] Mohammadreza Zolfaghari , Gabriel L. Oliveira, Nima Sedaghat, and Thomas Brox. (2017). Chained Multi-stream Networks Exploiting Pose, Motion, and Appearance for Action Classification and Detection. @cs.uni-freiburg.de, 2904-2913.

[59] Cherian A ,Koniusz P & Gould S.. (2017). Higher-order Pooling of CNN Features via Kernel Linearization for Action Recognition. 2017 IEEE Winter Conference on Applications of Computer Vision, 130-138.

[60] Xiaojiang Peng, Changqing Zou, Yu Qiao and Qiang Peng. (2014). Action Recognition with Stacked Fisher Vectors. Lecture Notes in Computer Science, 581-595.

[61] Chao-Yuan Wu, Christoph Feichtenhofer, Haoqi Fan. (2019). Long-Term Feature Banks for Detailed Video Understanding. The University of Texas at Austin,1-12.

[62] Lei Wang, Piotr Koniusz, Du Q. Huynh. (2019). Hallucinating IDT Descriptors and I3D Optical Flow Features for Action Recognition with CNNs. Australian National University, 1-12.

[63] Noureldien Hussein, Efstratios Gavves, Arnold W.M. Smeulders. (2019). Timeception for Complex Action Recognition. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 254-263.

[64] AJ Piergiovanni and Michael S. Ryoo. (2018). Learning Latent Super-Events to Detect Multiple Activities in Videos. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5304-5313.

[65] Jue Wang1,3 Anoop Cherian , Fatih Porikli , Stephen Gould. (2018). Video Representation Learning Using Discriminative Pooling. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1149-1158.

[66] Rohit Girdhar, Joao Carreira, Carl Doersch, Andrew Zisserman. (2019). Video Action Transformer Network. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 244-253.

[67] Christoph Feichtenhofer, Haoqi Fan Jitendra, Malik Kaiming He. (2019). SlowFast Networks for Video Recognition. 2019 IEEE/CVF Conference on Computer vision, 6201-6210.

[68] Sovan Biswas, Yaser Souri, Juergen Gall. (2019). Hierarchical Graph-Rnns for Action Detection of Multiple Activities. 2019 IEEE International Conference on Image Processing (ICIP), 3686-3690.

[69] Gu C , Sun C , Ross D. A , Vondrick C , Pantofaru C , Li Y Malik J. (2018). AVA: A Video Dataset of Spatio-temporally Localized Atomic Visual Actions. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6047-6056.

[70] Deepti Ghadiyaram, Du Tran, Dhruv Mahajan. (2019). Large-scale weakly-supervised pre-training for video action recognition. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 12038-12047.

[71] Du Tran, Heng Wang , Lorenzo Torresani, Matt Feiszli. (2019). Video Classification with Channel-Separated Convolutional Networks, 1-11.

[72] Xiang Long , Chuang Gan , Gerard de Melo , Jiajun Wu , Xiao Liu , Shilei Wen. (2017). Attention Clusters: Purely Attention Based Local Feature Integration for Video Classification. Massachusetts Institute of Technology, 1-11.

[73] AJ Piergiovanni and Michael S. Ryoo. (2019). Representation Flow for Action Recognition. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 9937-9945.

[74] AJ Piergiovanni, Anelia Angelova, Alexander Toshev, Michael S. Ryoo. (2019). Evolving Space-Time Neural Architectures for Videos. {ajpiergi,anelia,toshev,mryoo}@google.com, 1-18.

[75] Evangelos Kazakos , Arsha Nagrani , Andrew Zisserman , Dima Damen. (2019). EPIC-Fusion: Audio-Visual Temporal Binding for Egocentric Action Recognition. The final published version of the proceedings is available on IEEE Xlore, 5492-5501.

[76] Evangelos Kazakos , Arsha Nagrani , Andrew Zisserman , Dima Damen. (2019). EPIC-Fusion: Audio-Visual Temporal Binding for Egocentric Action Recognition, 5492-5501.

[77] Nieves Crasto, Philippe Weinzaepfel, Karteek Alahari, Cordelia Schmid. (2019). MARS: Motion-Augmented RGB Stream for Action Recognition, 1-11.

[78] Goyal R , Kahou S. E , Michalski V , Materzynska, J Westphal, S Kim H , Memisevic R.. (2017). The "something something" video database for learning and evaluating visual common sense. 2017 IEEE International Conference on Computer Vision, 5843-5851.

[79] Zhenyu Wu, Zhangyang Wang, Zhaowen Wang and Hailin Jin. (2018). Towards Privacy-Preserving Visual Recognition via Adversarial Training: A Pilot Study. Springer, P1-19.

[80] Peng, X., Schmid, C.: Multi-region two-stream r-cnn for action detection. In: ECCV (2016)

[81] Ren, Z., Lee, Y.J.: Cross-domain self-supervised multi-task feature learning using synthetic imagery. In: CVPR (2018).

[82] Ryoo, M.S., Rothrock, B., Fleming, C.: Privacy-preserving egocentric activity recognition from extreme low resolution. AAAI (2017).

[83] Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., Song, L.: Sphereface: Deep hypersphere embedding for face recognition. In: CVPR (2017)

[84] Rajput, Amitesh Singh; Raman, Balasubramanian; Imran, Javed (2020). Privacy-preserving human action recognition as a remote cloud service using RGB-D sensors and deep CNN. Expert Systems with Applications, 152, p1-15.

[85] Javed Imran;Balasubramanian Raman;Amitesh Singh Rajput; (2020). Robust, efficient and privacy-preserving violent activity recognition in videos . Proceedings of the 35th Annual ACM Symposium on Applied Computing, p1-8.

[86] Kwabena Owusu-Agyemeng;Zhen Qin;Hu Xiong;Yao Liu;Tianming Zhuang;Zhiguang Qin; (2021). MSDP: multi-

scheme privacy-preserving deep learning via differential privacy . Personal and Ubiquitous Computing, p1-13.

[87] Akula, Aparna; Shah, Anuj K.; Ghosh, Ripul (2018). Deep Learning Approach for Human Action Recognition in Infrared Images. Cognitive Systems Research, p1-19.

[88] Hashemian, Mina; Razzazi, Farbod; Zarrabi, Houman; Moin, Mohammad Shahram (2019). A privacy-preserving distributed transfer learning in activity recognition. Telecommunication Systems, p1-11.

[89] Jixin Liu;Leilei Zhang; (2020). Indoor Privacy-preserving Action Recognition via Partially Coupled Convolutional Neural Network . 2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE), p1-4.

[90] Wang, Huaijun; Zhao, Jing; Li, Junhuai; Tian, Ling; Tu, Pengjia; Cao, Ting; An, Yang; Wang, Kan; Li, Shancang (2020). Wearable Sensor-Based Human Activity Recognition Using Hybrid Deep Learning Techniques. Security and Communication Networks, 2020, p1–12.

[91] Zhongzheng Ren, Yong Jae Lee and Michael S. Ryoo. (2018). Learning to Anonymize Faces for Privacy Preserving Action Detection, pp.1-17.

[92] Zhenyu Wu, Zhangyang Wang, Zhaowen Wang and Hailin Jin. (2020). Towards Privacy-Preserving Visual Recognition via Adversarial Training: A Pilot Study, pp.1-27.

[93] Wu, Zhenyu; Wang, Haotao; Wang, Zhaowen; Jin, Hailin; Wang, Zhangyang (2020). Privacy-Preserving Deep Action Recognition: An Adversarial Learning Framework and A New Dataset. IEEE Transactions on Pattern Analysis and Machine Intelligence, p1–14