

Leveraging Deep Learning for Human Activity Recognition

¹Shereen Uzma ,²Mohammed Emaad Ali,³Mirza Munawar Ali Baig,⁴Mohammed Najeebuddin Quadri

¹Assistant professor, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd

shereenuzmacse13@gmail.com.

^{2,3,4}BE Student, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd.

Emaad.ali7755@gmail.com,mirzamunawaralibaig4@gmail.com,Mnuq2020@gmail.com

Abstract: *With the emergence of the Internet of Things (IoT), progress has been made in recent years in human resource management (HAR). HAR has implications for a huge range of applications, including geriatric care, behavioral trying out, and tracking tools. Many management algorithms had been employed to anticipate how people work around the surroundings. However, conventional device mastering strategies were overtaken with the aid of the features of engineering strategies that can pick the pinnacle layer. Conversely, it's miles recognized that deep learning models that consist of convolution neural networks (CNNs) can cast off capabilities and decrease computational fee routinely. In this paper, we use the CNN version to predict human athletes from the Weizmann dateset. In particular, we name adaptive studying to attain deep image recognition and appropriate gadget studying classifiers. Our test results showed the accuracy of 90-six.Ninety five% of the usageof VGG-16 in comparison to other models the use of CNN.*

Keywords: Activityrecognition,deeplearning,c o n v o l u t i o n neuralnetwork.

I. INTRODUCTION

Human Reputation (HAR) is an critical area of research due to its contributions to aged care, home automation, and surveillance. Much research has been finished on human's popularity for delight within the out of doors world. Some of the prevailing works are both dressed in [1], or no longer

wearing[2] [3]. Wearable as a whole The HAR system makes use of wearable gadgets connected to the human frame. Portable HAR devices are intrusive in nature. Non-portable, as a HAR tool does now not want to have sensors connected to humans or carry a tool for enjoyment functions. Non-wearable structures can be divided into sensor-based [2] and emotion-

and foresight-based totally HAR structures [3]. The entire sensor-based generation uses RF indicators from sensors, including RFID, PIR, and Wi-Fi sensors, to become aware of human activities. Temporal imaginative and prescient uses video, images from depth cameras or IR cameras to discover human activity. Most sensors primarily based on the HAR device have no interference in the discipline but will not have excessive accuracy. Therefore, famous imaginative and prescient-based generation for human entertainment has attracted outstanding hobby at some point of the gift period. Recognizing human hobby from streaming films is difficult.

Video-primarily based human overall performance may be labeled as symbol-based totally questioning and feeling and universal imaginative and prescient based totally at the capability to transport.

[4]. Marker-based method to use motion capture (MoCap) framework primarily based on wearable optical markers. It may be beneficial to capture human motion, but this method gives numerous risks. He desires optical sensors to connect with people and also requires the want for more than one virtual digital camera. However, the imaginative and prescient-primarily based method makes use of RGB or makes use of pictures. Now there is no want for someone to preserve a device or connect

sensors to human beings. Thus, this method has received extra interest these days, for this reason making the HAR framework simple and clean to encompass in many packages.

Most of the comprehensive HAR models primarily based on theory and foreknowledge have been proposed in the literature the usage of conventional gaining knowledge of materials for cognitive pride. However, conventional device learning strategies have these days been changed by the deep learning approach [5]. The most famous kind of deep network is the convolution neural network (CNN). CNN is regularly utilized in areas related to computer vision. It is created through convolution layers of the given picture for processing. In this paper, we use CNN to seize human hobby from Weizmann dataset. We first extract the frames of each sport from the video. In unique, we use rework reputation to gain deep photos and train system mastery classifiers. We construct three one-of-a-type CNN models to classify sports and compare our results with present paintings at the identical statistics. In summary, the primary reason of our paintings is:

The the rest of the paper is as follows: Section II provides a high-level evaluate of video-primarily based HAR models. We offer an overview of the know-how

transition in Section III. Chapter IV defines the studies methodology, data sources, studies method and discusses the experimental results. Conclusions and future snap shots are drawn in bankruptcy V.

II. LITERATURE SURVEY

Human choice strategies start from hand-crafted feature-based techniques to the intensity of acquisition of AI-primarily based techniques. In a associated research, the authors (Vrigkas, Nikou, and Kakadiaris 2015) have studied the interest of the humans by using dividing its assets into gathering samples and their services; In the same sub classes, the division is based totally on the release of improvement and numerous HAR games. In the primary categorization, the authors review uni modal and multi modal HAR strategies. In the uni modal class, region-time, stochastic, rule-based, and exceptional-based modes are grouped together. Simultaneously, multi modal units the possibilities, behaviors and relations of human sports.

Reining et al. (Reining et al. 2019) performed a literature assessment on HARs for manufacturing and distribution. This research presents an in-depth evaluate of modern-day HAR strategies, which includes widely wide-spread models and

deep fashions. This observe may be very beneficial for business. Beddiar et al. (Beddiar et al. 2020) studied the human experience of following mind and perceptions and divided all the subjects into the subsequent: the comply with-up gaining knowledge of features and objects guide, in which the authors analyzed exceptional strategies, together with their reference materials. The authors also discuss with relevant literature based on all sorts of human pursuits - simple human conduct, orientation, conduct, interaction, group motion and sports, which educate HAR principles at the lowest degree. Similarly, Zhu et al. (Zhu et al. 2016) also examined the craft and ability to determine the goods. Unlike (Beddiar et al. 2020), the authors first analyzed the problems of guide strategies and then confirmed the development of the deep observe of HAR in a brief time frame , till 2016. A assessment by using Zhang et al. (S. Zhang et al. 2017) targeted on the development of contemporary enterprise recognition techniques in phrases of enterprise representation and HAR approach distribution. This investigation shows the activities represented in keeping with international, nearby, and improvement efforts, at the same time as the categorization of the distribution of ideas is specifically based totally on the standard,

hate differences, and designs. The brief-term model with the HAR dataset indicates a great typical performance within the experimental effects. All HAR studies protected in this summary are from 2017.

Another take a look at become performed with the assist of Herath et al. (Herath, Harandi, and Porikli 2017) from the equal yr offered a comparable look at, starting with the director of the HAR technique - a method primarily based at the handcraft to achieve in-intensity expertise of process-based totally structures. This have a look at is the primary to present an in-depth understanding, together with HAR records, which has been missing in preceding research. But that consists of the textual content up to 2016; So, circulate ahead after that.

I want to make a present for scientists. The course to future set with the help of those authors is properly defined, it's miles a brilliant motivation to be used in the academic community.

A look at performed by way of Koohzadi et al. (Koohzadi and Charkari 2017) studied the function of deep gaining knowledge of in images and video for HAR. The strategies are commonly divided into 5 kinds: deep supervised generative, supervised deep discriminating, unsupervised deep, semi-supervised deep and hybrid. An vital function on this

evaluation is the effects, guidelines, and hints related to the choice of master intensity for HAR within the 5 categories cited above. The writer also mentioned the techniques for understanding space-time illustration extensive in terms of down load time, a zero.33 measure of traditional 2D video processing. Neeb et al. (Nweke et al. 2018) supplied a comprehensive assessment of deep gaining knowledge of techniques for HAR primarily based on mobile and wearable gadgets. The categorization of the method is generative, discriminating and hybrid explaining the best and the bad. This take a look at measures the depth of the purchase.

III. EXISTING SYSTEM

HAR system type

There are several important equipment-based HAR system features:

Vision-based HAR

Cameras appear to be attached in various locations for surveillance, video recording and storage on servers. Digital camera feeds or video recorders are used for surveillance work. For example, (Htike et al. 2014) studied human body image recognition for video surveillance using a static digital camera. This type of HAR is used for road protection, public safety, visitor control, crowd control, etc. Figure 1 shows the overall thinking and feeling

process of people who like to experiment with the gadget.

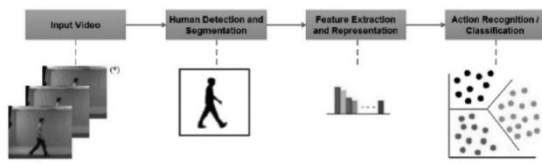


Figure 1. A typical human activity recognition system. (Image sequence (*) from Southampton database <http://www.gait.ecs.soton.ac.uk/>. Accessed: 2016-10-12).

As all HAR sensors

Smartphones have come to be widely wide-spread equipment for language trade and, greater currently, technology for human gaining knowledge of. Built-in sensors in smartphones can capture non-stop statistics on human activity. Wan et al. (S. Wan et al. 2020) analyzed human hobby using telephone sensors. In this way, the statistics is furnished by the accelerometer and gyroscope sensors incorporated into the cell smartphone, and then the information acquired by way of the system concerning the input received used to apprehend human activity. This form of HAR is useful for tracking sufferers, monitoring the sports of sports activities individuals, etc., however can't be used in large-scale software program-based totally human packages. For home/public space protection, surveillance, and plenty of others.

Motivation

Many factors influenced us to finish this research. However, a few essential points are listed below:

●In the ImageNet 2012 Competition, we determined that a convolution neural community (CNN) named AlexNet (Krizhevsky, Sutskever, and Hinton 2012) regarded to be nonetheless gaining knowledge of. Utilized in making snapshots and has validated to be a splendid assist for making movies. The usage of CNN. Researchers are now specializing in in-depth understanding of HARs based on real-time video.

"Film scientists were working to increase deep networks at the same time as the use of GPU mining with massive subjects. Therefore, many deep studying techniques were evolved for HAR, which ought to turned into offered among st future students.

"There are many benefits of HAR video datasets additionally result in the community of wondering and visualization to interpret the opposition process. Therefore, we are invited to introduce the readers to the set modern-day maximum superior deep learning method for HAR and their assessment on trendy HAR video datasets.

Assistance

The contemporary studies targets to provide an evaluation of imaginative and prescient-based totally human interest records, primarily based on deep getting to know techniques and contemporary getting

to know of statistics sets. Picture video. These video documents contain movies recorded by static cameras built from unique locations. This article provides an in-situ evaluation - first introducing the video facts after which describing the HAR approach inside the kingdom. Introducing deep information is the primary aim. At the equal time, we additionally briefly discuss the follow-up characteristic and the mixture technique to reveal the evolution of the HAR method. In preceding literature research, researchers have recommended the continuation of taxonomy for classes of HAR techniques. We introduce the new part of our textual content by way of introducing a taxonomy of schooling based on HAR thoughts, primarily based on all their social models, and classifying each concept in those lessons. In addition, the true power of our evaluation, which makes it precise to exceptional studies; are the following:

- We have multiplied almost all of the most superior deep getting to know fashions supplied in the literature thus far, past the human hobby assignment in reputation.

⊙◆ We present the performance of each deep mastering version with their benchmark statistics.

"We show the evolution of strength reputation from traditional ideas to the exceptional ideas.

⊙◆ This article is supposed to help researchers in this discipline to reap particular HAR records with thoughts and statistics.

In addition, the motive behind selecting the data evaluation length between January 2011 and May 2021 for this have a look at is that many in-depth studies have been developed at some stage in this era. We consist of some articles that feature the nice handmade HAR strategies for history, which can be out of this variety. We decided on a huge a part of the literature published in this era and considered only the articles published simplest in the peer-reviewed journals Scopes and WoS.

The entire article is organized as follows: Section 2 gives the literature evaluate in this area, and Section three lists.

IV. IMPLEMENTATION

A. DATA SET

Hobby. Table I suggests the whole quantity of frames per hobby based totally at the extracting frames for all nine human beings. The entire dateset is divided into Training (70%), Validation(10%), and Testing (20%).

TABLE I
DATASET STATISTICS IN TERMS OF NUMBER OF FRAMES PER ACTIVITY

Activity	Number of Frames
Bend	639
Jack	729
Jump	538
Run	346
Side	444
Skip	378
Walk	566
Wave1	653
Wave2	624
Total	4917

B. Discussion and effects

To discover sports activities, we experimented with three special convolution neural networks (CNN) to understand interest, in particular VEG-sixteen, VGG-19 and Google's InceptionNet-v3. We used trans formative knowledge acquisition to help expertise gained from massive databases consisting of Image Net. The alternate acquires understanding about shifting records from preceding education to introduce new regions into the neural network. We completed checking out at the Weizmann dateset using information received from previous electricity schooling on Image Net. The functions are extracted from the modern-day set of CNNs. We carried out transfer learning on the VGG-16th version of CNN and carried out an accuracy of ninety-six, ninety-5 percent.

For VEG-capture, the picture of duration 224×224 is given as enter and the characteristic of layer fc1 is extracted, which offers

4096-dimensional vector for each photograph.

We extensively utilized the transition to different CNN models including VGG-19 and Google's InceptionNet-v3 to study the implementation of different CNN fashions. Google's VGG-19 and InceptionNet-v3 scored 90-six.54% and 95.63%, respectively. The experimental effects verify that VGG-16 performs higher than the rest of the CNN fashions in spite of everything models have been used to modify the mastering. Table II suggests the accuracy price, precision, misses and fl score of the CNN model. The confusion matrix of our one-of-a-kind CNN version is shown in Figures 2, three and 4.

We compared the results received by several unique techniques that don't use the Weizmann dateset transformation strategy with our approach. Test outcomes showed that the proposed exchange to understand facts stability done higher in recognition rating. The fact is the fact is advanced using 1-6% by soliciting for instructional adjustments. An assessment of the outcomes of the usage of a modified model of the VGG-16

The version and other techniques are shown in Table III. The difference with the brand new approach lies in exploring the efficiency of understanding conversion

whilst striving with CNN mode to enhance the recognition ranking.

TABLE II
RESULTS ON ACTIVITY RECOGNITION BASED ON DIFFERENT CNN MODELS IN TERMS OF ACCURACY SCORE, PRECISION, RECALL, AND F1-SCORE

Model	Accuracy (in %)	Precision (in %)	Recall (in %)	F1-score (in %)
VGG-16	96.95	97.00	97.00	97.00
VGG-19	96.54	97.00	97.00	96.00
Inception-v3	95.63	96.00	96.00	96.00

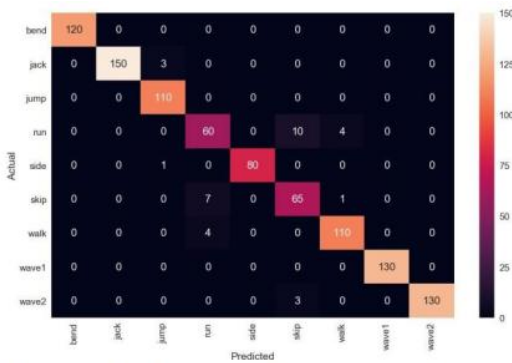


Fig. 2. Confusion Matrix for recognizing 9 activities on Weizmann Dataset using VGG-16 Convolution Neural Network

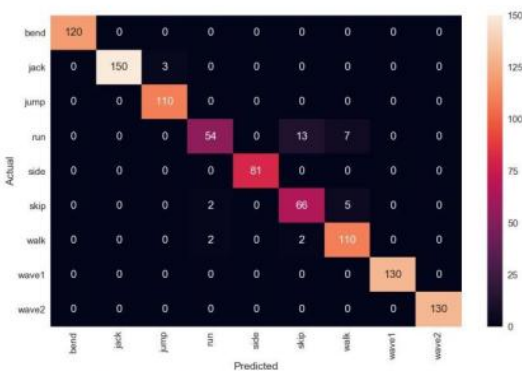


Fig. 3. Confusion Matrix for recognizing 9 activities on Weizmann Dataset using VGG-19 Convolution Neural Network

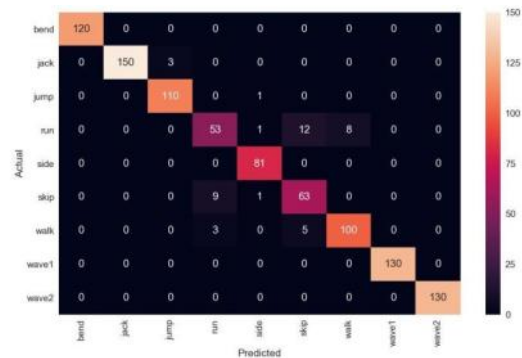


Fig. 4. Confusion Matrix for recognizing 9 activities on Weizmann Dataset using Inception-v3 Convolution Neural Network

Figures 2, 3 and 4 show the confusion matrix of 3 different constitutional neural networks (CNN) after applying the knowledge transfer, which was used to classify frames from different sports using VGG -16, VGG-19 and Google's Inception Net. - v3 one. It is clear from figures 2, 3 and 4 that VGG-16 does not understand its perception when predicting running as crossing, VGG-19 does not understand its perception when predicting run as a jump and jump as a walk, and Google's Inception Net-. v3 has a typo in prediction functions such as jumps, which can be compared to the trust view. The work of the research transferred to CNN fashion has made sure of the popularity. However, the transformation learning method used in our work with the reference data from the weight before the learning of the Image net will be affected, because the Image net includes images from many special class.

V. CONCLUSION

We use a CNN model to expect human sports from the Weizmann data set. We experimented with three unique constitutional neural networks (CNNs) for cognitive processing. We implemented transfer gaining knowledge of to attain deep image and gadget getting to know classifiers. Our check consequences

confirmed ninety six.95% accuracy using VEG-sixteen with the use of adaptive learning. Our experiments showed that VGG-16 outperforms other CNN fashions in characteristic extraction. Our experimental effects with the modified method also confirmed the overall performance of VEG-sixteen compared to the state-of-the-art approach.

In the future, we plan to preserve this research with the aid of developing information-based and content material-aware paintings for people. We will even keep our work to apprehend humans' hard sports together with cooking, reading books and watching TV.

REFERENCES

1. L. Yao, Q. Z. Sheng, X. Li, T. Gu, M. Tan, X. Wang, S. Wang, and W. Ruan, "Compressive representation for device-free activity recognition with passive rfid signal strength," *IEEE Transactions on Mobile Computing*, vol. 17, no. 2, pp. 293–306, 2018.
2. I. Lillo, J. C. Niebles, and A. Soto, "Sparse composition of body poses and atomic actions for human activity recognition in rgb-d videos," *Image and Vision Computing*, vol. 59, pp. 63–75, 2017.

3. W. Zhu, C. Lan, J. Xing, W. Zeng, Y. Li, L. Shen, and X. Xie, "Co-occurrence feature learning for skeleton based action recognition using regularized deep lstm networks," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
4. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
5. J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 248–255.
6. Prasadu Peddi (2016), Comparative study on cloud optimized resource and prediction using machine learning algorithm, ISSN: 2455-6300, volume 1, issue 3, pp: 88-94.
7. K. Simonyan and A. Zimmerman, "Two-stream convolutional networks for action recognition in videos," in *Advances in neural information processing systems*, 2014, pp. 568–576.

8. G. Gkioxari, R. Girshick, and J. Malik, "Contextual action recognition with r* cnn," in *Proceedings of the IEEE international conference on computervision*, 2015, pp. 1080–1088.
9. Prasadu Peddi (2017) "Design of Simulators for Job Group Resource Allocation Scheduling In Grid and Cloud Computing Environments", ISSN: 2319- 8753 volume 6 issue 8 pp: 17805-17811.
10. Prasadu Peddi and Dr. Akash Saxena (2014), "EXPLORING THE IMPACT OF DATA MINING AND

MACHINE LEARNING ON STUDENT PERFORMANCE", *International Journal of Emerging Technologies and Innovative Research* (www.jetir.org), ISSN:2349-5162, Vol.1, Issue 6, page no.314-318, November-2014, Available: <http://www.jetir.org/papers/JETIR1701B47.pdf>