CNN and Transfer Learning For Human Activity Recognition

[1]Prof. Kanchan Umanave,[2]Mr.Rohit Vishwakarma, [3]Miss.Roshni Gangan, [4]Miss. Amita Desai

[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering & Technology, Asangaon,Maharashtra,India.

[1]kanchanumavane2020@gmail.com, [2]asrohit98@gmail.com, [3]rvg16898@gmail.com, [4]amita.alekar15@gmail.com

Abstract- With the arrival of the web of Things (IoT), there are significant advancements within the area of human action recognition (HAR) in recent years. HAR is applicable to wider application like elderly care, anomalous behaviour detection and closed-circuit television. Several machine learning algorithms are employed to predict the activities performed by the human in an environment. However, traditional machine learning approaches are outperformed by feature engineering methods which may select an optimal set of features. On the contrary, it's known that deep learning models like Convolutional Neural Networks (CNN) can extract features and reduce the computational cost automatically. during this paper, we use CNN model to predict human activities from Wiezmann Dataset. Specifically, we employ transfer learning to induce deep image features and trained machine learning classifiers. Our experimental results showed the accuracy of 96.95% using VGG16. Our experimental results also confirmed the high performance of VGG-16 as compared to remainder of the applied CNN models.

Keywords-Deep learning, Artificial intelligence, HAR, CNN, ANN.

I. INTRODUCTION

The face and human behavioral pattern play an important role within the face-to-face identification. Visual information might be a key source for such identifications. Surveillance videos provide such visual information which can be viewed as live videos, or it'll be played back for future references. The recent trend of 'automation' has its an impression even within the sector of video analytics. Video analytics could also be used for an outsized kind of applications like motion detection, act prediction, person identification, abnormal activity recognition, vehicle counting, people counting at crowded Places, etc. during this domain, the 2 f factors that are used for personal papers are technically termed face recognition and step recognition sequentially. Between these two techniques, face recognition is more ready for the automatic person by monitoring videos. Face recognition may be accustomed to predict the orientation of a head, which successively will help to prognosticate performance.

Motion recognition with face recognition is notably useful in many applications like the affirmation of an individual, identification of an individual, and detecting the presence or inadequacy of a self at a discrete place and time. Additionally, human interactions like subtitle contact between two individuals, head motion detection, hand gesture recognition, and estimation are accustomed devise a system which will recognize and see suspicious behavior between followers in an examination hall favorably. This method provides a strategy for suspicion human action detection through face recognition.

Computational complexities and time complexities are several key factors while designing a real-time system. The system which uses an algorithm with comparatively

II. AIMS AND OBJECTIVE

a) Aim

The aim is to analyze human action Human activities have an integrated hierarchical structure that intimates the different levels of it, which can be held as a three-level categorization. Front, for the bottom level, there is a microscopic particle and specific action primitives ascertain extra complicated human activities. After the action elemental level, the work/outline emerges as to the second level. Conclusively, the aggregate intercommunications form the top level, which manages human movements that incorporate more than two persons and objects. In this project, we catch this three- level categorization namely action primitives, operations/motions, and cooperations. This three-level categorization varies a little from prior critiques.

b) Objective

Objective is to determine if the person is

1. Inter. with Comp.
2. Riding Horse
3. Running lower time complexity, using fewer hardware resources, and which produces good results is more useful for time-critical applications like bank robbery detection, subject monitoring system, discovering and inscribing questionable movements at the terminal, exam slots, etc.

III. LITERATURE SURVEY

Paper 1: Online Detection of Unusual Events in Videos via Dynamic Sparse Coding:

The architect recommends enhanced Real- time abnormal event detection in a video rivulet has been a complex hurdle due to the lack of adequate preparation learning, volatilization of the explanations for normalcy and anomaly, time limitations, and mathematical modification of the fitness of any parametric figures. They recommend a fully unsupervised dynamic sparse coding strategy for identifying rare events in videos based on online sparse constructability of inquiry signals from anatomically detected event dictionaries, which sets sparse coding bases. Based on a flash that usual events in a video are more likely to be constructible from an event dictionary, whereas unusual events are not, our algorithm operates a principled bulging optimization formulation that allows both a sparse reconstruction code and an online dictionary to be simultaneously understood and renewed. Our algorithm is entirely unsupervised, making no prior hypotheses of what significant events may look comparable and accordingly the warning the cameras. the very fact that the bottom encyclopedia is renewed in a network fashion because the algorithm senses more data, sidesteps any issues with theory drift. Preliminary results on hours of real-world inspection video and inadequate YouTube videos show that the proposed algorithm could dependably locate the unique events within the video sequence, exceeding the present state-of-the-art processes.

Paper 2: Real- Time Anomaly Detection and Localization in Crowded Scenes:

Introducing a method for real-time irregularity detection and localization in congested displays. Each video is represented as a set of no overlapping cubic bits and is illustrated using two economic and global descriptors. These descriptors arrest the video properties from complex regards. By fusing simple and cost-effective Gaussian classifiers, we can notice typical motions and oddities in videos. The limited and global articles are based on structure similarity connecting nearby patches and the points learned in an unsupervised way, using a sparse autoencoder. Laboratory effects show that our algorithm is analogous to a state-of-the-art mode on UCSD ped2 and UMN benchmarks, but still more extra time- efficient. The researchers verify that our policy can reliably discover and limit aberrations as someday as both appear during a video.

Paper 3: Abnormal Event Detection at 150 FPS in MATLAB:

Fast abnormal event detection satisfies the expanding stomach to process an enormous number of parade videos. Based on the built- in superfluity of video structures, we advance an experienced sparse federation learning framework. That reaches respectable performance in the detection phase without doubting the resulting characteristic. The tiny running background is ensured because the new method efficiently turns the original complex problem into one in which only a rare costless small scale least-square optimization steps are concerned. Our method reaches high detection rates on benchmark datasets at a clip of 140150 frames per second on mediocre when figuring on a plain desktop PC using MATLAB IV. Mahmudul Hasan, Jonghyun Choiy, Jan Neumanny, Amit K. Roy- Chowdhury, Larry S. Davisz, “Learning Temporal Regularity in Video Sequences”. Perceiving meaningful activities in a long video sequence is a challenging problem due to ambiguous definition of ‘meaningfulness’ as well as clutters in the scene. We approach this problem by learning a generative model for regular motion patterns (termed as regularity) using multiple sources with very limited supervision.

IV. EXISTING SYSTEM

Every surviving well-constructed deep FER systems store on two key issues: the insufficiency of abundant diverse training data and the expression of irrelevant variations, like illumination, head pose, and identity. This policy presents relativistic assets and drawbacks of those complex species of techniques with relevancy to 2 open issues (data size requirement and expression-unrelated variations) and extra loci (computation efficiency, performance, and difficulty of network training). Pre- training and fine-tuning became the mainstream in strange FER to settle the matter of insufficient training data overfitting. A sensible procedure that established to be exclusively useful is pre- training and fine-tuning the network in multiple stages using subsidiary data from large-scale discontent or face recognition datasets to small-scale FER datasets.

V. COMPARTIVE STUDY

SR NO.	PAPER TITLE	AUTHOR NAME	TECHNOLO GY	ADVANTAGE	DISADVAN TAGE
1.	Online Detection of Unusual Events in Videos via Dynamic Sparse Coding	Bin Zhou,Li Fei- Fei,Eric P. Xing	CNN.	A major model of neural mechanisms involved in learning in the brain is provided by long-term potentiation	Its time consuming
2.	Real- Time Anomaly Detection and Localization in Crowded Scenes	Mohammad Sabokrou, Mahmood Fathy, MojtabaHoseini, Reinhard Klette	Pre-trained CNN models- Dense Net, ResNet and inception.	Anomaly detection is an important tool for detecting fraud, network intrusion	Nearest neighbour based anomaldetecti on techniques require a distance or similarity measure between two datainstances.
3.	Abnormal Event Detection at 150 FPS in MATLAB".	Cewu Lu. Jianping Shi, Jiaya Jia	ANN.	It achieves decent performance in the detection phase without compromising result quality.	Detection phase wihich is little bit time consuming

VI. PROBLEM STATEMENT

Design a project using transfer learning identifying various humans using CNN, like alteration learning as an example, CNN has been universally implemented in individual domains of research. which can be more useful for time-critical reinforcements like bank robbery exposure, patient monitoring system, detecting and reporting irregular actions at the train depot, exam holes, etc.

VII. PROPOSED SYSTEM

Customary bank checks, bank credits, credit cards, and different authoritative reports are an important piece of the innovative economy. They are gone altogether the essential mediums by which people what's more, associations exchange cash and pay bills. Indeed, even today all of these exchanges particularly money-related require the signature to be verified. The unavoidable symptom of signature is that they will be misused to feign records genuineness. Consequently, the necessity for investigating in sufficient mechanized answers for point acknowledgment and confirmation has grown in late years to avoid living helpless upon misguidance Proposed System evaluated signature attestation system using real signatures through aforementioned approach System will halt valid and illogical signature from the user then preprocessing and have descent operation thereon signatures for this centroid x, centroid y, solidity, oddity and skew x, skew y are used. The order will show the correctness of the signature and for affirmation, it'll show the result whether the signature is natural or produced. An accidental forest classifier and neural network are accepted in the proposed system.

VIII. ALGORITHM

The general idea of working of proposed system algorithm is given as follow:

Step.1: Start

Step.2: Login

Step.3: Authentication.

Step.4: Input image

Step.5: Partition the data into training and testing splits using 80% of

the data for training and the remaining 20% for testing

Step.6: Construct model A model.add(layers.Dense(256, activation='relu', input_dim=4* 4 * 512)) model.add(Dropout(0.5)) model.add(layers.Dense(128, activation='relu')) model.add(Dropout(0.5)) model.add(layers.Dense(64, activation='relu')) model.add(layers.Dense(7, activation='sigmoid'))

Step.7: Construct model B model6 = models.Sequential()

model6.add (layers.Dense(256, activation='relu', input_dim=4 * 4 * 512)) model6.add(Dropout(0.5)) model6.add(layers.Dense(128, activation='relu')) model6.add(Dropout(0.3)) model6.add(layers.Dense(64, activation='relu')) model6.add(layers.Dense(7, activation='sigmoid'))

Step.8: Construct model C model7 = models.Sequential()

model7.add (layers.Dense(256, activation='relu', input_dim=4 * 4 * 512)) model7.add(Dropout(0.5)) model7.add(layers.Dense(128, activation='relu')) model7.add(Dropout(0.5)) model7.add(layers.Dense(64, activation='relu')) model7.add(Dropout(0.5)) model7.add(layers.Dense(32, activation='relu')) model7.add(layers.Dense(7, activation='sigmoid'))

Step.9: Test Accuracy of Model A **Step.10:** Optimize Model A **Step.11:** Test Accuracy of Model B **Step.12:** Optimize Model B **Step.13:** Test Accuracy of Model C **Step.14**: Optimize Model C **Step.15:** Test

IX. MATHEMATICAL MODEL

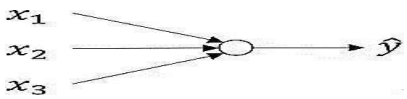

Fig 9.1 model diagram

$$L(X_i, Y_i) = -\sum_{j=1}^{C} y_{ij} * \log(p_{ij}) \quad (3)$$

$$y_{ij} = \begin{cases} 1, & \text{if } i_{th} \text{ element is in class } j \\ 0, & \text{if } i_{th} \text{ element is not in class } j \end{cases}$$

Fig.9.1. Single Neuron network

Let's see the simple neural network in Fig.7.2.1, let output neurons be *ŷ*, it then receives input from activation function neuron *a*. And respectively *a* get the input from three other activation *x1*, *x2* and *x3*. Which symbolize by *X1*, *X2*, and *X3* for their neurons name.

Moreover, the weight connected from *X1*, *X2*, and

X3 to activation function *a* are *w1*, *w2*, and

w3. So, the calculation for output can be denoted by (1).

$\hat{y} = a = w1x1 + w1x1 + w1x1$

After that, we can calculate the loss function network above. The loss function is a measure of the difference between the prediction of ŷ, and the true value (ground truth), in other words, it is an error calculation for one stage of training. This function can be seen (2).

In this research, we use categorical cross-entropy as loss function, because we want to classify each person by his/her face. This function will compare the distribution of predicted face, by true and false which set to 1 for true and 0 if false. The true class of a person's face represents as a one-hot encoded vector, which is we get the lower loss if the model output ve Class symbolize by

C, where *Xi* is the input vector for one-hot encoded target vector *Yi*, and

pij is probability that *Ith* element in class *j*.

Class symbolizes by *C*, where *Xi* is the input vector for one-hot encoded target vector *Yi*, and

pij is probability that *Ith* element in class *j*.

X. SYSTEM ARCHITECTURE

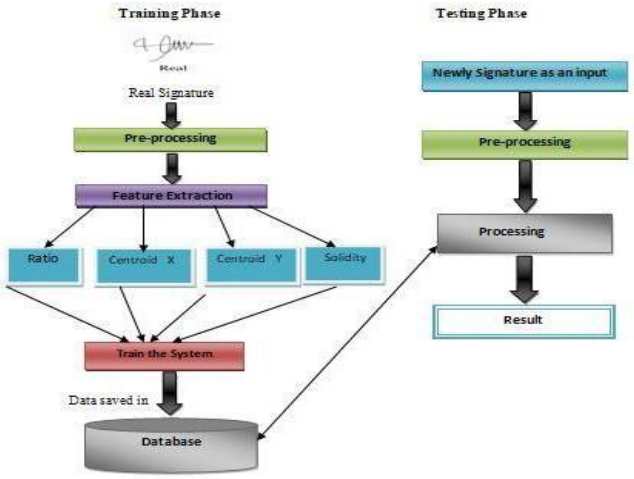

Fig no.1 System Architecture

Description: There are 2 types of phases: 1. Training phase 2. Testing phase

Training phase: The system takes the signature five confirmed and five unreasonable signatures of the user from that three signatures are used for the practice phase and two used for the examination phase subsequent that remarkable preprocessing operation function on such signature then emphasize extraction organization implemented on that subscription for that it uses ratio, centroid x, centroid y, and solidity extraction function.

In the testing phase, it supports the signature of the new user. The system will take five valid or unreasonable signatures of the new user and it commands to show the result whether the signature is natural or manufactured.

XI. ADVANATGES

That method may be employed in video examination systems, human-computer intercommunication, and robotics for civilized behaviour characterization, claim multiple movement recognition operations.

Unit of mysterious learning's main benefit over other machine learning algorithms is its potential to execute ultimate engineering beyond on its own. A long learning algorithm will consider the knowledge to look for features that compare and mix them to enable faster learning externally being explicitly mentioned to try and do so.

XII. DESIGN DETAILS

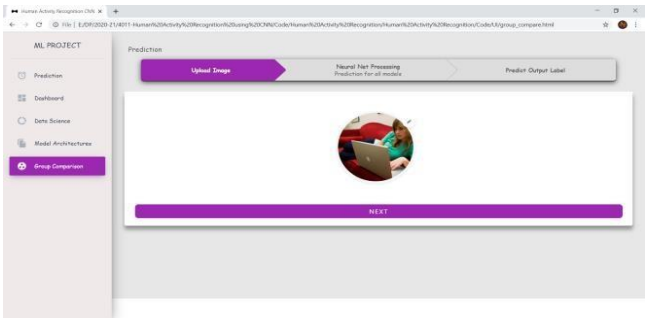

Fig no.3 Interacting with Computer

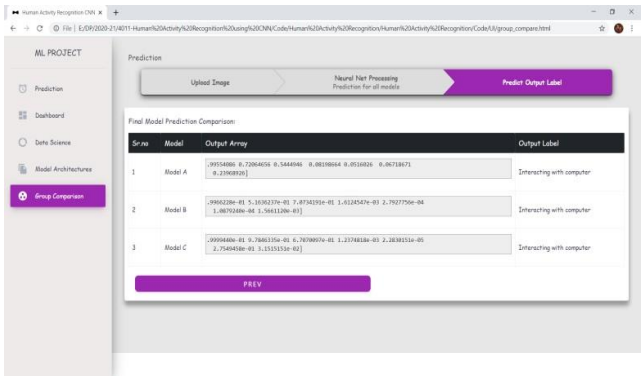

Fig no.3 Horse Riding

XIII. CONCLUSION

Thus, we have tried to implement the paper "Bin Zhou, Li Fei-Fei, Eric P. Xing Mohammad Sabokrou, Mahmood Fathy, MojtabaHoseini, Reinhard Klette Cewu Lu. Jianping Shi, Jaya Jia ", "*CNN And Transfer Learning For Human Activity Recognition",* IEEE 2020 and according to the implementation the conclusion is as follows Project used CNN models to predict the human activities. We are going to experiment with different Convolutional Neural Networks (CNN) for activity recognition. We will employ transfer learning to urge the deep image features and trained machine learning classifiers.

REFERENCE

[1] B. Bhandari, J. Lu, X. Zheng, S. Rajasegarar, and C. Karmakar, "Noninvasive Sensor Based Automated Smoking Activity Detection," in Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017, pp. 845–848.

[2]L. Yao, Q. 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.

[3] I. Lillo, J.C.Niebles, and A.Soto, "Sparse composition of body poses and atomic actions for human action recognition in rgb-d videos," Image and Vision Computing, vol. 59, pp. 63–75,2017.

[4] W. Zhu, C. Ln, 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 computing, 2016.

[5] C. Szegedy, W. Liu, Y. J, P. Sermanet, S. Reed, Anguelov, 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.

[6] 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.

[7] A. Jalal, N.Sarif, J. T. Kim, and T.S. Kim, "Human Activity Recognition Via Recognized Body Parts Of Human Depth Silhouettes For Residents Monitoring Services At Smart Home," Indoor and built environment, vol.22, no.1pp.271–279,2013.

[8] K. Simonyan and A. Zisserman, "Two-Stream Convolutional Networks for Action Recognition in Videos," in Advances in neural information science systems, 2014, pp. 568–576.

[9] G. Gkioxari, R. Girshick, and J. Malik, "Contextual Action Recognition With R* Cnn," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1080–1088.

[10] L. Wang, Y. Xiong, Z. Wang, AND Y. Qiao, "Towards Good Practices for Very Deep Two- Stream Convnets," arXiv preprint arXiv: 1507.02159, 2015.

[11] D. Tran, L. Bourdev, R. Fergus, L. Torresani, AND M. Paluri, "Deep End2end Voxel2voxel Prediction," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2016, pp. 17–24.

[12] P. Wang, W. LI, J. Wan, P. Ogunbona, AND X. Liu, "Cooperative Training of Deep Aggregation Networks for Rgb-D Action Recognition," in Thirty Second AAAI Conference on computer science,2018.

[13] S. Ji, W.Xu, M.Yang, AND K.Yu, "3d Convolutional Neural Networks for Act Recognition," IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 1, pp. 221–231,2013.

