IJCSE International Journal of Computer Sciences and Engineering Open Access

Survey Paper Vol.-6, Issue-10, Oct 2018 E-ISSN: 2347-2693

Deep Learning for Human Action Recognition – Survey

K. Kiruba[1], D. Shiloah Elizabeth[2*], C Sunil Retmin Raj[3]

[1] Dept. of Computer Science and Engineering, Anna University, CEG Campus, Chennai, India
Dept. of Computer Science and Engineering, Anna University, CEG Campus, India
[3]Dept. of Information Technology, Anna University, MIT Campus, India

[*]Corresponding Author: shiloah@annauniv.edu

Available online at: www.ijcseonline.org

Accepted: 15/Oct/2018, Published: 31/Oct/2018

***Abstract*—** Human action recognition (HAR) in visual data has become one of the attractive research area in the field of computer vision including object detection, recognition, retrieval, domain adaptation, transfer learning, segmentation etc. Over the last decade, HAR evolved from heuristic hand crafted feature to systematic feature learning namely deep feature learning. Deep feature learning can automatically learn feature from the raw inputs. Deep learning algorithms, especially Convolutional Neural Network (CNN), have rapidly become a methodology of choice for analysing recognition of videos. In this paper, details of recent trends and approaches of deep learning including CNN, Recursive Neural Network (RNN), Long Short term Memory (LSTM) and Autoencoders which are used in HAR are discussed. The challenges are identified to motivate the researchers for future works.

***Keywords*—**HAR, CNN, LSTM, Deep Learning model.

I. INTRODUCTION

In real world environment, recognizing human action plays a vital role in a variety of domains including intelligent video surveillance, shopping behavior analysis, abnormal activity recognition, human interaction analysis, crowd analysis etc. The challenges in accurate recognition of action is cluttered background, occlusion, view point variations, multi-scale chances, camera jitter, illumination changes, and clothing and appearance variations [1], [2], [3]. Most of the methods perform the HAR in two steps. In the first step, handcrafted features are computed from the raw video frames. In the second step, classifier is built based on the obtained features to recognize the action. Figure 1(a) shows the representation of machine learning approaches for Human Action Recognition (HAR). The choice of feature extraction is problem dependent and it is hard to choose appropriate features in real time scenarios. Especially for HAR, different action classes may appear differently in terms of their appearances, viewing angles and motion patterns. Hence, learning the discriminative features from raw video frames using deep learning has attracted the researchers in recent years as it can discover effective patterns in large scale and complex datasets. Figure 1 (b), (c), (d) shows the deep learning approaches for HAR.

Deep learning in computer vision was introduced for image classification. The effective results of deep learning attracts the attention to the researchers in computer vision applications including human imperfections, pedestrian detection, body pose estimation, glance classification, emotion recognition, real time cognitive load estimation, human- centered vision for autonomous vehicles, etc. In HAR, Convolutional Neural Network (CNN) works directly on the raw inputs by extracting spatial and temporal features. It can learn a hierarchy of features by extracting high level features from low level ones. CNN have been invariant in terms of pose, clutter and lighting. It can capture the motion information encoded in multiple adjacent frames by performing 3D convolutions. It can be trained using supervised or unsupervised approach. CNN yields better performance using training with proper regularization [4].

In this survey, the main focus is on deep learning for HAR. The rest of the paper is organized as follows. Section II details the Convolution neural network including 2D CNN and 3D CNN. Section III details the CNN + LSTM combination for HAR. Section IV provides the details of auto encoder for HAR. Section V provides some challenges and suggest some future works to researchers. Finally, Section VI concludes the deep learning for HAR.

II. CONVOLUTIONAL NEURAL NETWORK (CNN)

CNN is categorized as 2D and 3D CNN model. In 2D CNN, video frames are treated as still images, the CNN convolution operation is performed over the individual frames to recognize the action. Hence, it does not consider the motion information encoded in multiple contiguous frames. In order to overcome this limitation, 3D CNN has been introduced which captures the discriminative features along both the spatial and the temporal information.

Li et al. have presented an image classification scheme to action recognition with 3D skeleton videos [5]. In this work they have proposed three different approaches. First, they have introduced a video domain translation-scale invariant image mapping. It transforms the 3D skeleton videos to color images. They are called as skeleton images. Second, a multi-scale dilated CNN is designed for the classification of skeleton images. It improves the frequency adaptiveness and find the discriminative temporal spatial cues for the skeleton images. In order to improve the generalization and robustness of the proposed method, they have utilized different kinds of data augmentation strategies. They have experimented their work on NTU RGB+D, UTD-MHAD, MSRC-12 and G3D dataset and achieved accuracy of 80.2%, 85.0%, 87.7%, 99.4% and 93.1% respectively.

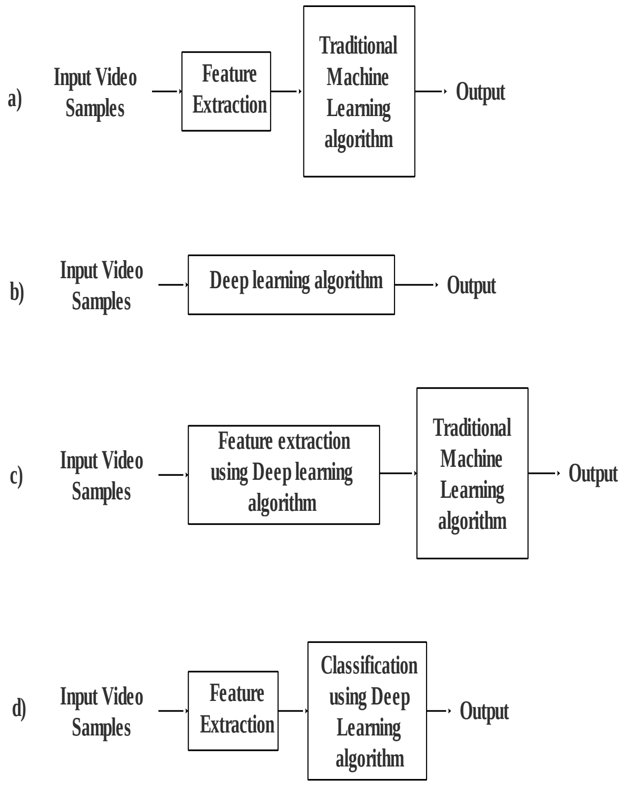
a)
Input Video Samples
Feature Extraction
Traditional Machine Learning algorithm
Output
b)
Input Video Samples
Deep learning algorithm
Output
c)
Input Video Samples
Feature extraction using Deep learning algorithm
Traditional Machine Learning algorithm
Output
d)
Input Video Samples
Feature Extraction
Classification using Deep Learning algorithm
Output

Figure 1. a) Representation of traditional Machine learning flow for HAR. b) Representation of Deep Learning flow for HAR. c) Representation of combined machine learning and deep learning for HAR. d) Representation of combined image processing and deep learning classification for HAR.

Ding et al. have reported on the modelling and analysis of time series using a linear dynamical system (LDS) [6]. In this work, they have used tensor based LDS (tLDS) for modelling tensor observations in time series. It used in Tuker decomposition which estimates the parameters of the LDS model as action descriptors. These actions can be expressed as a subspace correspondingly to a point on a grassmann manifold. It is classified using dictionary learning and sparse coding technique. They have experimented their work on MSRAction3D, UCFKinect and Northwestern-UCIA multi-view action dataset and achieved the recognition rate of 94.85, 96.48 and 92.99 respectively.

Ke et al. have presented a clip representation for skeleton based 3D action recognition [7]. They have proposed to transform each channel of the 3D coordinates of a skeleton sequence into a clip with the same length. Each generated clip is given to multi-task CNN. Multi task CNN consists of several convolution layers, fully connected layer and softmax which is used to learn features from the generated clips. The time step wise concatenation layer concatenates the CNN representations of the three clips at the same time step. It produced K feature vectors. These K feature vectors are then processed in parallel and K sets of class scores are produced. Human action classes are classified based on the scores. They have experimented their work on six challenging datasets and its performance on NTU RGB-D dataset is 87.3%, Northwestern-UCLA multi view action 3D dataset is 86.82%, SBU Kinect interaction dataset is 94.17%, CMU dataset is 94.28%, UTKinect action3D dataset is 99% and Berley MHAD dataset is 100%.

Ijjina and Mohan have proposed an approach for HAR using genetic algorithms (GA) and Deep Convolutional Neural Network [32]. In their work, they have initialized the GA population and evaluated the fitness of GA chromosomes. It is used to initialize a CNN classifier. This GA framework has been executed for multiple cycles and the final GA population has been reached which is initialized the CNN classifiers. Finally, they have combined the classification evidences of CNN classifiers generated using GA. They have experimented their work on UCF50 dataset and achieved 99.98% of recognition accuracy.

Ijjina and Mohan have proposed a hybrid deep neural network model for HAR in videos [33]. They have fused the evidences generated by homogeneous models arranged in a parallel topology. They have built the ensemble of classifiers to recognize human actions from action bank features using CNN classifier. They have experimented their work on UCF50 dataset and achieved 99.68% of recognition accuracy.

Xu et al. have proposed a fully coupled two stream spatio temporal architecture for extremely low resolution action recognition videos [10]. They have coupled the C3D and

RNN (Recurrent Neural Network) architectures. Gated Recurrent Unit (GRU) has been used to extract the features from C3D and RNN. They have extracted efficient spatial and temporal features using CNN on video frames and optical flows. Finally, they have aggregated the features into a robust feature representation to recognize the actions in the entire video. The final fusion of features evaluated among different fusion techniques such as sum fusion, max fusion, concatenated fusion and convolution fusion. They have achieved better results using sum fusion approach. They have experimented their work on HMDB51 and Dogcentric dataset and achieved the recognition accuracy of 44.96% on HMDB51 dataset and 73.19% on Dogcentric datasets.

Ignatov et al. have proposed an architecture which combines a shallow CNN for unsupervised local feature extraction together with statistical features that encode global characteristics of the time series [11]. The CNN architecture consists of convolutional layer, Nonlinearity, pooling layer, fully connected layer and softmax layer. In their work they achieved better performance in the combination of CNN + statistical features + data centering. They have used accelemeter data such as WISDM and UCI datasets and achieved 93.32% on WISDM dataset and 97.63% on UCI dataset.

Cao et al. have introduced 3DCNNs for HAR [12]. It simultaneously learn the spatial and temporal features. They have not directly use the activations of fully connected layers of 3D-CNN. They have chosen the selective convolutional layer activations to form a discriminative descriptor for video. They have proposed a novel JDD (3-D Joins pooled Deep convolution Descriptor) C3D (Convolutional 3D network) for action recognition. They have proposed two-stream bilinear C3D model. It can learn the guidance from the body joints and capture the spatio-temporal features simultaneously.

Wang et al. have proposed an end to end pipeline called as two-stream 3-D ConvNets fusion for human action recognition in videos [13]. Existing 3-D ConvNets requires a fixed sized input video and fixed length of the frames which may reduce the quality of video analysis. To overcome this issue, in this proposed work require only arbitrary size and length to recognize the human actions in videos. This deep model consists of Spatial Temporal Pyramid Pooling (STPP) ConvNet and Long Term Temporal Modelling. RGB and Optical Flow features are extracted using STPP ConvNet. Finally a set of spatial and motion features are concatenated. Long Short-Term Memory (LSTM) or CNN is applied on this concatenated features for action classification. They have experimented their work in UCF101 dataset, HMDB51 and ACT datasets. They have achieved 91.6% of average accuracy in UCF101, 69.0% in HMDB51, 81.7% in ACT dataset using STPP+LSTM fusion.

Cao et al. have proposed a selective connected layers activations instead of using fully connected layers to form a discriminative description for video [12]. In their work preliminary work focuses on Joints pooled 3-D deep convolutional descriptor (JDD). The pooling activation on 3D feature maps is computed as follows. First, Video clips are split and fed into a 3D CNN for convolution computation. The estimated body joints in the video frames have been used to localize points in 3D feature maps. From every channel, activations performed on each corresponding point of body joint are pooled. Finally aggregation of pooled activations of all the video clips is named as JDD. Additionally they have proposed a two stream bilinear model to learn the guidance from the body joints and extracts the spatio-temporal features simultaneously. Two stream bilinear C3D contains feature stream and attention stream. The feature stream inherits the Conv structure of original C3D which extracts spatio-temporal features. In the attention stream is pretrained which locates the key points in 3D Conv feature maps. The two stream are formulated and pooling together to recognize the actions. They have achieved 83% of accuracy using fusion layers of JDD (conv5b +conv4b) on subJHMDB dataset. Additionally, 82% of recognition accuracy has been achieved on the same dataset using two stream bilinear C3D.

III. Long Short Time Memory (LSTM)

In machine learning community Recurrent Neural Network (RNN) [13], [20] has got lot of attention. RNN is effective in short term dependencies. LSTM plays a major role in long term dependencies. LSTM can selectively remember or forgot memories. LSTM network is composed of different memory blocks called cells. Cell states and hidden state are responsible for transferring information to the next cell. The memory blocks are capable of remembering and manipulating things through forget gates, input gate and output gate. LSTM based human action recognition method became trend [14-25].

Ullah et al. have proposed a novel action recognition method by processing the video sequences data using CNN and DB_LSTM (Deep-Bidirectional) Network [9]. In their work, they have extracted the deep features from every sixth frame to reduce the redundancy and complexity. Then, the sequential information is learnt using DB-LSTM network. DB-LSTM network consist of multiple forward pass and backward pass of DB-LSTM to increase its depth. This method is able to learn the long term sequences and analyze video for a certain time interval. They have experimented their work on UCF-101, Youtube 11 actions, HMDB51 datasets. They have achieved 92.84%, 87.64% and 91.21% of average recognition score using the proposed DB-LSTM for action recognition over Youtube, HMDB51, UCF101 dataset.

Das et al. have proposed an effective method for daily living action recognition. In the proposed work they have fused the appearance based CNNs and temporal evaluation of skeleton with LSTM [26]. They have used two different inputs namely skeleton and RGB frames. The skeleton features processed with LSTM layers and RGB frames have been modelled using Part based CNN (P-CNN) features. The two input processes are performed independently and separate SVM classifiers are used. The fusion of the classifier scores are used to recognize the human actions. They have experimented their work on CAD-60 and MSRDailyActivity3D dataset. They have achieved 97.06% of accuracy and 96.25% of accuracy on CAD-60 and MSRDailyActivity 3D dataset using P-CNN + (Body Coordinates + LSTM) + SVM (fusion).

Nunez et al. have proposed a deep learning based approach for temporal 3D pose recognition problems based on the combination of CNN and LSTM recurrent network [27]. In this work, they have used two stage of training strategy. First, it focuses on CNN training and secondly, adjusts the full method (CNN + LSTM) . They have experimented their work on six different datasets including MSR Action 3D dataset, MARDailyActivity 3D dataset, UTKinect action 3D dataset, NTURGB+D dataset, Montalbano V2 dataset and Dynamic Hand Gesture dataset. They have achieved 96.0% of accuracy on MSRAction3D dataset, 63.1% of accuracy MSRDailyActivity 3D, 99.0% on UTKinect Action3D dataset, 76.21% on Montalbano V2 and 81.1% of accuracy in Dynamic Hand Gesture Dataset.

Li et al. have presented a videoLSTM which starts from the soft attention LSTM model (ALSTM) [8]. Additionally they have introduced two novel modules including convolutional ALSTM and motion based attention networks. VideoLSTM makes three novel contributions including a convolutions to exploit the spatial corrections in images, CNN to allow for motion information which generates motion based attention maps and attention maps to localize the action spatio-temporally. They have experimented their work on UCF101, HMDB51, Thumos S13 localization datasets.

Sharma et al. have proposed a soft attention based model for HAR in videos [28]. They have extracted deep features that include both spatial and temporal information using multi-layered RNNs with LSTM units. In the soft attention mechanism, video input videos are fed to CNN which produces the feature cube. The feature cube model as an average of the feature slices weighted according to the location softmax. The feature slices are taken as input to the recurrent model at each time step. The proposed recurrent model consists of three LSTM layers and it predicts the new location probabilities and class label. They have evaluated their model on UCF-11, HMDB51 and Hollywood2 datasets. They have achieved 41.3% of recognition accuracy on HMDB51 and 43.9% of mean average precision on Hollywood2 dataset. They have achieved best result in UCF-11 dataset and experimented with various parameters with soft attention based model such as average pooled LSTM (82.56%), max pooled LSTM (81.60%) and soft attention model (@30fps, lemmeda =0, 1 and 10).

Grushin et al. have proposed robust human action recognition via LSTM [14]. They have experimented their work on KTH dataset with HOF descriptors as input features with LSTM without recurrent cell connections and achieved 90.7% of accuracy. Srivatava et al. have used the LSTM encoder-decoder framework to learn video representations [25]. They have constructed the representation from the sequence of frames using encoder LSTM. It is decoded through another LSTM to produce the output sequence. They have experimented their work on UCF-101 and HMDB-51 datasets and achieved 75.8% and 44.0% of recognition accuracy, respectively using composite model with conditional future predictor.

IV. AUTOENCODERS

Autoencoders are one of the type in neural network [29]. It consists of encoder and decoder. It aims to compress their inputs x into a latent space representation, h=f(x) where h is dimensional reduction of x. It reconstructs the output from this latent space representation, r=g(h). The applications of autoencoders are data denoising and dimensionality reduction for data visualization. It performs two ways of learning namely, under complete and over complete representations. Types of autoencoders [30], [31] are listed in Figure 2.

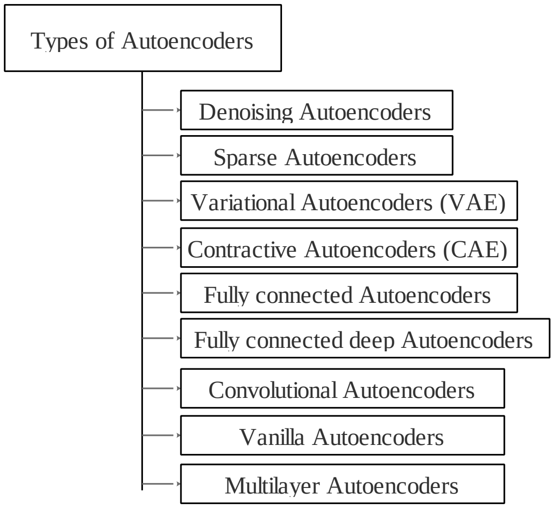
Types of Autoencoders
Denoising Autoencoders
Sparse Autoencoders
Variational Autoencoders (VAE)
Contractive Autoencoders (CAE)
Fully connected Autoencoders
Fully connected deep Autoencoders
Convolutional Autoencoders
Vanilla Autoencoders
Multilayer Autoencoders

Figure 2. Types of Autoencoders.

V. DISCUSSION

Some limitations of Deep learning are provided below,

1. Enormous training data: Deep learning requires very large amount of training data to reduce better results. The large scale data requires various number of parameters.
2. Slow learner: Deep learning is a slow learner and extremely computationally expensive to train. It can be improved by GPU training speed. Even though the learning speed may be adjusted by learning rate and the adjustment in learning rate may affects the reliability of the deep neural network.
3. Over fitting is the most common problem in neural networks. It cannot perform well if any change in camera and illumination.
4. Deep learning is sometime considered as black box operation.
5. Determining the topology, training method and hyper parameters.

VI. Conclusion

Learning the deep features directly from raw data is an active research area in computer vision. Deep learning is the most effective solution for complex data and temporal information extraction especially in video analytics. In this survey, the deep learning techniques including CNN, LSTM, RNN, combination of CNN+LSTM and Autoencoders are presented in the field of HAR. Finally, the limitations and challenges of deep learning have been discussed.

Acknowledgment

The author expresses their sincere and reverential gratitude to University Grant Commission (UGC), India, for the obtained financial support during this research.

References

[1] D.D. Dawn, S.H. Shaikh, "A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector", The Visual Computer, Vol. 32, Issue. 3, pp. 289-306, March 2016.

[2] S. Herath, M. Harandi, F. Porikli, "Going deeper into action recognition: A survey", Image and vision computing, Vol. 60, pp. 4-21, April 2017.

[3] C. Indhumathi, V. Murugan, "A Survey on Neural Network based Approaches and Datasets in Human Action Recognition", International Journal of Computer Sciences and Engineering, Vol. 6, Issue 6, June 2018.

[4] N.S. Lele, "Image Classification using Convolutional Neural Network" Vol. 6, Issue 3, PP.22-26, June 2018.

[5] B. Li, M. He, Y. Dai, X. Cheng,Y. Chen, "3D skeleton based action recognition by video-domain translation-scale invariant mapping and multi-scale dilated CNN" Multimedia Tools and Application, pp. 1-21, January 2018.

[6] W. Ding, K. Li, E. Belyaev, F. Cheng, "Tensor-based linear dynamical systems for action recognition from 3D skeletons" Pattern Recognition, Vol. 77, pp. 75-86, 2018.

[7] Q. Ke, M. Bennamoun, S. An, F. Sohel, F. Boussaid, "Learning Clip Representations for Skeleton-Based 3D Action Recognition", IEEE Transactions on Image Processing, Vol. 27, No. 6, June 2018.

[8] Z. Li, K.G.E. Gavves, M. Jain, C.G.M. Snoek, "VideoLSTM convolves, attends and flows for action recognition", Computer Vision and Image Understanding, Vol. 166, pp.41–50, 2018.

[9] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, S.W. Baik, "Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features", Special Section on Visual Surveillance and Biometrics: Practices, Challenges, And Possibilities, Vol. 6, 2018.

[10] Y. Xu, L. Wang, J. Cheng, H. Xia, J. Yin, "DTA: Double LSTM with temporal-wise attention network for action recognition" IEEE International Conference on Computer and Communications (ICCC), pp. 1676-1680, December, 2017.

[11] A. Ignatov, "Real-time human activity recognition from accelerometer data using Convolutional Neural Networks" Applied Soft Computing. Vol.62, pp. 915–922, 2018.

[12] C. Cao, Y. Zhang , C. Zhang and H. Lu, "Body Joint Guided 3-D Deep Convolutional Descriptors for Action Recognition" IEEE Transactions on Cybernetics, Vol. 48, No. 3, March 2018.

[13] X. Wang, L. Gao, P. Wang, X. Sun and X. Liu, "Two-Stream 3-D convNet Fusion for Action Recognition in Videos With Arbitrary Size and Length", IEEE Transactions On Multimedia, Vol. 20, No. 3, March 2018.

[14] A. Grushin, D.D. Monner, J.A. Reggia, A. Mishra,"Robust Human Action Recognition via Long Short-Term Memory", International Joint Conference on Neural Networks (IJCNN), 2013.

[15] S. Nitish, M. Elman, S. Ruslan, "Unsupervised Learning of Video Representations using LSTMs", Proceedings of the 32nd International Conference on Machine Learning, France, 2015.

[16] H. Gammulle, S. Denman, S. Sridharan, C. Fookes, "Two stream lstm: A deep fusion framework for human action recognition", IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 177-186, March 2017.

[17] X. Wang, L. Gao, Song, J. and Shen, H., "Two stream lstm CNN: saliency-aware 3-D CNN with LSTM for video action recognition". IEEE Signal Processing Letters, Volume 24, Issue 4, pp.510-514, 2017.

[18] J. Liu, G. Wang, L.Y. Duan, K. Abdiyeva, and A.C. Kot, "Skeleton-based human action recognition with global context-aware attention LSTM networks", IEEE Transactions on Image Processing, Volume 27, Issue 4, pp.1586-1599, 2018.

[19] I. Lee, D. Kim, S. Kang, S. Lee, "Ensemble deep learning for skeleton-based action recognition using temporal sliding lstm networks" IEEE International Conference on Computer Vision (ICCV), pp. 1012-1020 , October, 2017.

[20] J. Liu, G. Wang, P. Hu, L.Y. Duan, A.C. Kot, "Global context-aware attention lstm networks for 3d action recognition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR) ,Vol. 7, pp. 43, July 2017.

[21] L. Wang, X. Zhao, Y. Liu, "Skeleton Feature Fusion based on Multi-Stream LSTM for Action Recognition", IEEE Access, September, 2018.

[22] C. Li, P. Wang, S. Wang, Y. Hou, W. Li, "Skeleton-based action recognition using LSTM and CNN" IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 585-590, July 2017.

[23] S. Song, C. Lan, J. Xing, W. Zeng, J. Liu, "Spatio-Temporal Attention-Based LSTM Networks for 3D Action Recognition and Detection" IEEE Transactions on Image Processing, Volume 27, Issue 7, pp. 3459-3471, July, 2018.

[24] Y. Yuan, X. Liang, X. Wang, D.Y. Yeung, A. Gupta," Temporal Dynamic Graph LSTM for Action-Driven Video Object Detection" International Conference on Computer Vision, pp. 1819-1828, October, 2017.

[25] N. Srivastava, E. Mansimov, R. Salakhudinov, "Unsupervised learning of video representations using LSTMs", International conference on machine learning, pp. 843-852, Junuary 2015.

[26] S. Das, M. Koperski, F. Bremond, G. Francesca, "A Fusion of Appearance based CNNs and Temporal evolution of Skeleton with LSTM for Daily Living Action Recognition" arXiv preprint arXiv:1802.00421, 2018.

[27] J.C. Núñez, R. Cabido, J. Pantrigo, A. Montemayor, J. Vélez, "Convolutional Neural Networks and Long Short-Term Memory for skeleton-based human activity and hand gesture recognition" Pattern Recognition Vol. 76, pp.80–94, 2018.

[28] S. Sharma, R. Kiros, R. Salakhutdinov, "Action Recognition using Visual Attention", arXiv preprint arXiv:1511.04119, 2015.

[29] X. Qinkun, S. Yang, "Human Action Recognition Using Autoencoder", IEEE International Conference on Computer and Communications, 2017.

[30] G. Ian, B. Yoshua, C. Aaron," Deep Learning" MIT Press, 2016.

[31] P. Josh, G. Adam, "Deep Learning: A Practitioner's Approach", O'reilly, 2017.

[32] E.P. Ijjina and C.K. Mohan, "Human action recognition using genetic algorithms and convolutional neural networks", Pattern Recognition, 2016.

[33] E.P. Ijjina and C.K. Mohan, "Hybrid deep neural network model for human action recognition" Applied soft computing, 2015.

[34] J. Donahue, L.A. Hendricks, S. Guadarrama and M. Rohrbach, "Long-term Recurrent Convolutional Networks for Visual Recognition and Description" Conference on Computer Vision and Pattern Recognition (CVPR 2015), 2015.

[35] M. Xu, A. Sharghi, X. Chen, D.J. Crandall, " Fully-Coupled Two-Stream Spatiotemporal Networks for Extremely Low Resolution Action Recognition", IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1607-1615, March 2018.

[36] L.N. Pondhu, G. Pondhu, "Tuning Convolution Neural networks for Hand Written Digit Recognition", International Journal of Computer Sciences and Engineering,Vol. 6, Issue 8, August 2018.

Authors Profile

K. Kiruba received B.E. degree in Computer Science and Engineering from Anna University, India in 2012 and M.E. degree in Computer Science and Engineering from Madras Institute of Technology, Anna University, Chennai, India in 2015. She is a research scholar in the Department of Computer Science and Engineering, College of Engineering, Guindy, Anna University, Chennai, India and pursuing her research in Visual Object Recognition and Retrieval.

D. Shiloah Elizabeth received B.E. degree in Electronics and Communication Engineering from Manonmaniam Sundaranar University, India in 2001, M.E. degree in Computer Science and Engineering from Manonmaniam Sundaranar University in 2002 and Ph.D. degree from College of Engineering Guindy Campus, Anna University in 2010. She is an Assistant Professor in the Department of Computer Science and Engineering of College of Engineering Guindy, Anna University. Her current research interests include image processing and Machine learning.

Sunil Retmin Raj C received B.E. degree in Electronics and Communication Engineering from Bharathiyar University, India in 1999, M.E. degree in Applied Electronics from Madurai Kamarajar University, India in 2000 and Ph.D. degree from College of Engineering Guindy Campus, Anna University in 2014. His current research interests include image processing and soft computing.