Deep Learning for Human Pose Classification using Multi View Dataset

B. Gnana Priya, M. Arulselvi

Abstract--- Human pose classification isvery challenging area of work in research in modern times. It widely supports in understanding a human poses and its further sequence of actions. Many standard human pose datasets were created and a wide research is taking place.Our main target is to create a multiview dataset containing novel actions which are different from normal poses. Actions from Karate martial arts and Bharathanatyam dance poses are captured. We use Deep Convolutional neural networks to classify the poses without any feature extraction.

Keywords--- ActionClassification, Deep Learning, Convolutional Neural Network, Karate andBharathanatyam Dataset

I. INTRODUCTION

Machine Learning is the fast emerging technology which is poised to dominate in almost every walks of today's life. Machine Learning not only analyses data, but also predicts future responses/actions aimed towards greater results. Artificial Intelligence will be the prime support system for human resources, human actions and performances in all areas of application. Human poses are very essential for a broad range of applications likeassisted living, human computer interface, surveillance,image indexing, activity recognition,image retrieval and so on. Human pose classification is done to recognize the actions that humans are performing.The need to understand human appearance and its related aspects is to analyze humans and to find the human interaction with surroundings that are very essential for industrial applications nowadays. Key Information like posture, outlook,gesture, etc are in great demandand are important for business.

The Multiview dataset which is employed in this project contains actions from Karate, anunarmed martial-arts discipline contains defensive and counter attacking body movements and Bharathanatyam, a classical Indian dance form. Karate employs kicking, striking, and defensive blocking with arms and legs. KATA in Karate is the formalized flow of movements for various offensive and defensive postures. Bharatanatyam has sixty four principles of hand, foot, face and body movements in coordination which are performed to accompaniment of dance syllables. Various dance steps are based on the dancer's balanced body weight distribution, firm position of lower limbs, pretty hand movements that flow around their body.The poses are very different from ournormal day to day actions.These poses are captured and are used to train our classifier.

Manuscript received June 10, 2019.

B. GnanaPriya, Assistant Professor, Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamil Nadu, India. (e-mail: priyamvatha.joey@gmail.com)

Dr.M. Arulselvi, Assistant Professor, Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamil Nadu, India. (e-mail: marul@gmail.com)

Using Convolutional Neural Networks (CNN) for deep learning is becoming more popular nowadays due to ability of CNNs to learn feature extraction directly. In CNN we can use a already trained network as a base model and build a new network on top of it for new recognition tasks. This enables the network to produce easy, fast and effective results.The inputs are separated into classes for classification depending on the objective function. The layers extract complex features by stacking and down sampling from different parts of the input. Convolutional Layer**,** Pooling Layerand Fully-Connected Layer are the three varieties of layers used here. Convolutional layer uses kernels to detect features all over the image. The Kernelscarries out a convolution operationin which element wise product is taken first and sum of the matrices are taken. Generally, large amount of computation is needed. This can be reduced by inserting Pooling layers between convolutional layers.This also prevents overfitting and reduce the parameters.

II. RELATED WORKS

Image classification using convolutional neural network became standard since Alex Krizhevsky, Geoff Hinton and IlyaSutskevar won ImageNet [1]. [2] uses network for large scale image recognition and[3] reviews the various methods for image classification. TianmeiGuo [4] explained that Convolutional Neural Networks are used in deep learning and has achieved great performance on image classification.EmineCengil[5] used The Caffe library, for cats and dogs classification, a kaggle dataset using Convolutional Neural Networks running on GPU. Travis Williams [6] developed a Convolutional Neural Networks (CNN) to classify handwritten digits. He converted Data into wavelet domain,lower dimension to achieve faster processing time. Er-Xin Shang [7] uses CNN network for spam image classification. Support vector machine technique employed with minimum a margin-based loss to obtain a lower level feature representation and better performance is achieved in distinguishing different spam images. Qing Li [8] uses CNN for medical image classification. They use the network to classify different categories of lung disease patterns. JunhoYim[9] proposes a method in which features from multiple layers are combined to form the target model. CNN models which are already trained and learned features are reused from multiple layer to extract features.

Yu Kong[10] presented a survey of various techniques in action recognition and prediction.

Various existing models, action databases, algorithms, their evaluation and problems were discussed.Diogo[11] proposed a multitask framework for both two and three dimensional pose estimation from still images as well as action recognition from videos. Arjun[12] uses multilayer CNNs and modified learning techniques to learn features. They used a bottom up and weak spatial model that produced best results**.[13]**Hourglass Network [14] proposed a stateof-the-art architecture for bottom-up and top-down inference with residual blocks. Chou[15] employed two stacked hourglass networks, used one as generator and other as discriminator**.**[16]addresses multicontext view for estimating poses. Tompson[17] used multiple links of convolutional networks to mix the features from an image dataset, and also used Markov Random Field for post-processing. Chen and Yuille [18] introduced the ConvNet to learn both the unary and the pair wise term of a tree structured graphical model. Many of the past works normally uses a manual designed multi context representations, e.g., multiple bounding boxes [19] or multiple image crops [20], and lack of flexibility and diversity for modelling the multi-context representations**.** [21, 22] uses advanced GAN's(Generative Adversial Networks) for learning.

Existing benchmarks datasets includes aspects of the human pose estimation such as sport scenes,frontal-facing people, and people interacting with objects, pose estimation in group photos and pose estimation of people performing synchronized activities. Few of the challenges in predicting human pose coordinates includes the foreshortening of limbs, occlusion of limbs, rotation and orientation of the figure, and overlap of multiple subjects. Although CNN gives the highest classification scores, it requires many number of parameters to be trained, a bulge amount of memory and abundant still images for training.

III. PROPOSED WORK& RESULTS

Computer vision and Machine Learning tasks heavily rely on Convolutional neural networks due to their simplicity and reduced number of parameters. We consider the action classification of various poses as a multi-class classification problem.Sample of the multiview dataset containing images from karate and Bharathanatyam dataset are presented in Fig (1). The network architecture is shown in Fig (2).Table (1) shows the number of images considered for each pose. The karate images are captured afresh in an open environment and Bharathanatyam poses in studio both with cluttered background. The images are captured using three cameras placed around the capture space. Twenty different persons performed the same actions repeatedly and are captured. The poses are from different parts of the action sequence while doing karate Kata and different parts of dance sequence for Bharathanatyam.Original captured image which are clear enough for processing are only taken. The images are taken

Fig. 1: Sample images from multiview dataset (Karate and Bharathanatyam

Table 1: Total number of images in multiview dataset

Action Category	No. of Actions	No. of Images
KARATE	1. Karate Action 1	100
	2. Karate Action 2	100
	3. Karate Action 3	110
	4. Karate Action 4	100
	5. Karate Action 5	98
	6. Karate Action 6	97
	7. Karate Action 7	95
	8. Karate Action 8	98
	9. Karate Action 9	102
	10. Karate Action 10	110
	TOTAL	**1010**
BHARATHANATYAM	1. Action 1	110
	2. Action 2	105
	3. Action 3	105
	4. Action 4	110
	5. Action 5	110
	6. Action 6	105
	7. Action 7	105
	8. Action 8	100
	9. Action 9	100
	10. Action 10	100
	TOTAL	**1050**

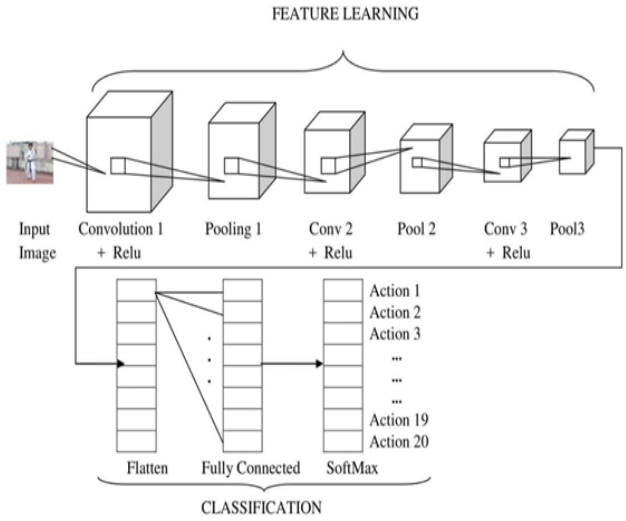

Fig. 2: Feature extraction and classification flow graph

such that our system can recognize foreshortening and occlusion of limbs, orientation and rotation of the same image. The CNNs need a large amount of training data, so we perform Data augmentation.The captured images are rotated to different angles from the original angle. We got around 3500 images after augmentation. Each of our images contains only single subject and are centred in the image. The proposed CNN takes as input a greyscale image of size 200 X 200 pixels and outputs a vector of numbers representing the probabilities of each of the activity labels corresponding to the 20 categories.

IV. NETWORK ARCHITECTURE

We use Keras API which allows us to built networks easily, extend them and add new modules in a simple manner. This model is inspired from VGGNet and uses a 3x3 CONV filters. A Sequential model in which the desired layer can be added one above the other is used for building our network. The Fully connected layer wherein all the neurons in any one layer are all connected to the neurons in the previous layer is used to build a feed forward network. We use ReLU activation function that makes the network to learn non linear decision boundaries. SoftMax layer is used as the final layer for multiclass classification.

Our network takes advantage of batch normalization and dropout layer. Batch normalization normalizes the input as it passes to the next layer, thus reducing the number of epochs for training our network. Dropout layer used to prevent overfitting and ensure that the parameters of the network are not getting biased towards training data. It will drop random connections during our training and dependency of training set may get reduced. Stochastic Gradient Descent optimizer used to configure the network. For multiclass classification the loss type used is categorical cross entropy. The accuracy and loss are the metrics we are tracking during the training process. Batch size, Number of epochs and the learning rate are the hyperparameter we fine tune to get our results. Fit() function in Keras used for training the network. The network is trained for 300 epochs. We use scikit-learns classification_report to view our final results.

V. EVALUATION OF NETWORK

A directory structure is created to arrange our training and validation datasets for processing. Separate sub-directory for each class is created and images are stored in the training and validation directories. We are going to use the VGG16 model with weights pre-trained on ImageNet. To save the bottleneck features from VGG16 model we include a function save_bottlebeck_features() .We take 3200 images for training and 300 images for validation. Initially we obtain a accuracy of 54% since our dataset is trained for the first time. On fine tuning various parameters we finally got a accuracy of 62%. Since we have only small number of images it is not sufficient to train our network. The future work will be to improve the size of our dataset and images need to be augmented to increase the size.

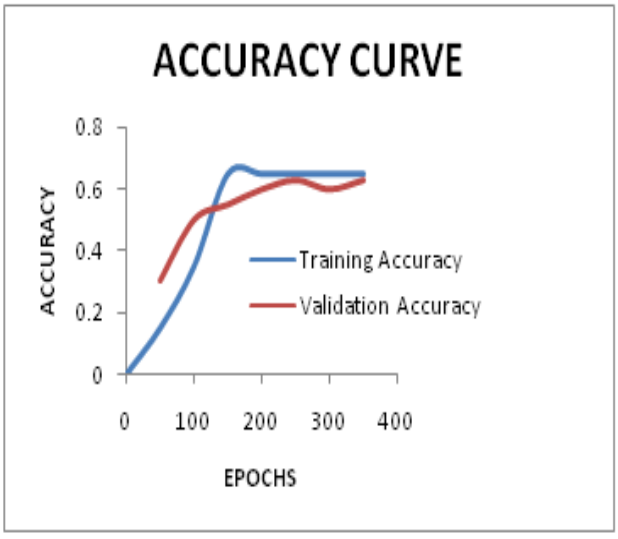

Fig. 3: Training and Validation accuracy for 300 epochs

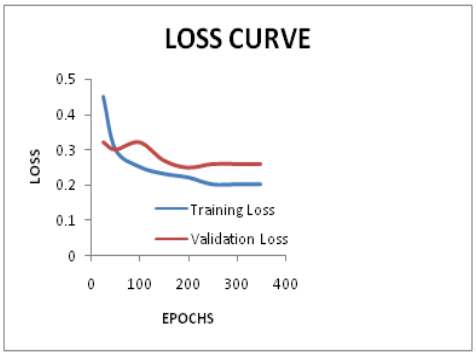

Fig. 4: Training and Validation loss for 300 epochs

The network recognizes some of the poses accurately. But, some of the poses are poorly recognised. We need to adjust the various network parametersinorder to obtain this accuracy. The karate and Bharathanatyam data were separately taken and we acquired 68% and 62% classification accuracy respectively. We combined both the data and we acquired an overall accuracy of 62% for the multiviewdataset. The accuracy and loss for training and validation data are plot and are shown in Fig (3) and Fig (4). The confusion matrix is also plot to find the individual accuracy of each pose.

VI. CONCLUSION

In this work few Karate and Bharathanatyam moves that are original and naive images captured were taken and action classification is done. The classification is carried out based on deep learning algorithm using Keras library running in top of Tensorflow. The proposed work classifies the poses with an accuracy of 62%. The aim is to build a multi-view dataset including actions from martial, sports and dance for various poses. In future the work can be extended for classifying many diverse poses. The classification of certain poses were poor and further tuning of parameters or applying other machine learning techniques will yield better results.

ACKNOWLEDGMENT

We express our gratitude and sincere thanks toShihanDr.A.R.Sundar - 7th Dan, his juniors and all students for demonstratingKarate poses and KalaimamaniParvathi Ravi Ghantasala and her disciples for Bharathanayam poses.

REFERENCES

1. A. Krizhevsky, I. Sutskever, and G. E. Hinton. “Imagenet classification with deep convolutional neural networks”, In NIPS, 2012.
2. Karen Simonyan and Andrew Zisserman, “Very Deep Convolutional Networks for Large-scale Image Recognition” ICLR 2015
3. WaseemRawat, Zenghui Wang, “Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review”, Neural Computation © 2017 Massachusetts Institute of Technology, Volume 29, Issue 9, September 2017.
4. TianmeiGuo, Jiwen Dong ,HenjianLi'YunxingGao, “Simple Convolutional Neural Network on Image Classification” , IEEE 2nd International Conference on Big Data Analytics at Beijing, China on 10-12 March 2017
5. Emine CENGIL , Ahmet ÇINAR , Zafer GÜLER, “A GPU-Based Convolutional Neural Network Approach for Image Classification”, International Conference on Artificial Intelligence and Data Processing Symposium at Malatya, Turkey on 16- 17 September 2017.
6. Travis Williams, Robert Li, “Advanced Image Classification using Wavelets and Convolutional Neural Networks”, 15th International Conference on Machine Learning and Applications at Anaheim, CA, USA on 18-20 December 2016.
7. Er-xinshang, Hong-gang zhang, “Image Spam Classification Based On Convolutional Neural Network”, International Conference on Machine Learning and Cybernetics at Jeju, South Korea on 10-13 July 2016
8. Qing Li, WeidongCai, Xiaogang Wang, Yun Zhou, David Dagan Feng and Mei Chen, “Medical Image Classification with Convolutional Neural Network”, 13th International Conference on Control Automation Robotics and Vision at Singapore on 10-12 December 2014
9. JunhoYim, JeongwooJu, Heechul Jung, and JunmoKimt “ Image Classification Using Convolutional Neural Networks With Multi-stage Feature” ,Robot Intelligence Technology and Applications
10. Yu Kong, Yun Fu “Human action recognition and Prediction: A Survey”, Journal of Latex class files, Vol.13, Sep 2018
11. DiogoC.Luvizon, David, Hedi “2D/3D Pose Estimation and Action Recognition using Multitask Deep Learning”, arXiv:1802.09232v2, Mar 2018
12. Arjun Jain, Jonathan, Andriluka “Learning Human pose Estimation Features with Convolutinal Networks” , arXiv 1312.7302v6, Apr 2014
13. Bulat and Tzimiropoulos. “Human pose estimation via Convolutional Part Heat map Regression” In European Conference on Computer Vision (ECCV), pages 717–732, 2016.
14. A. Newell, K. Yang, and J. Deng. “Stacked hourglass networks for human pose estimation”, In Proc. Eur. Conf. Comp. Vis., pages 483–499, 2016.
15. Chia-jungchou, Jui-Ting “Self Adversarial training for Human pose Estimation”, arXiv:1707.02439v2, Aug 2017
16. Xiao Chu, Wei Yang, Wanli “ Multi-context attention for Human pose Estimation” arXiv:1702.07432v1, Feb 2017
17. J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler “ Efficient object localization using convolutional networks” ,In Proc. IEEE Conf. Comp. Vis. Patt. Recogn., pages 648–656, 2015.
18. X. Chen and A. L. Yuille“ Articulated pose estimation by a graphical model with image dependent pair wise relations”, In NIPS, 2014.
19. V. Ramakrishna, D. Munoz, M. Hebert, J. A. Bagnell, and Y. Sheikh “Pose machines: Articulated pose estimation via inference machines”, In ECCV. 2014.
20. Wei, S.E.Ramakrishna, V.Kanade, T.Sheikh, “Convolutional pose machines”, CVPR, 2016
21. A.Raford, L.Metz and S.Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks”, arXiv, 2015
22. T.Salimans, I.J.goodfellow, W.Zaremba, V.Cheung, “Improved techniques for training GAN's”, In Proc. Advances in neural Inf. process systems, 2016.