

Human Activity Recognition using Transfer Learning

Neelam Chandolika¹, Rutuja Taywade², Abhishek Tiwari³, Ayushi Roy⁴ and Rohit Sharma⁵

¹⁻⁵Vishwakarma Institute of Technology, Pune

Email: neelam.chandolika@vit.edu, rutuja.taywade21@vit.edu, abhishek.tiwari21@vit.edu, ayushi.roy21@vit.edu, karmchand.rohit@vit.edu

Abstract— In recent years, there have been substantial developments in the field of human activity recognition (HAR). Many machine learning techniques have been used to anticipate human activity. Contrarily, deep learning models like Convolutional Neural Networks (CNN) are known to automatically extract features and lower processing costs. In this study, we predict human behavior's using two dataset –training, set, testing, set from Kaggle Dataset using the CNN model and data analysis using python. In proposed system applied transfer learning particularly to collect significant characteristics of image and trained machine learning classifiers, predict given image activity.

Index Terms— Human activity recognition, Deep CNN.

I. INTRODUCTION

Human activity recognition (HAR) is a research topic that is currently being explored because it has applications in aged care, automated homes, and surveillance systems. Transfer learning is useful for visual categorization of problems. Transfer learning if applied for Human activity recognition, gives wonderful results. It improves training accuracy. On the recognition of human activities, numerous research has been conducted in the past. [1] It can be difficult to identify human activity in the streaming video. It does not call for the user to wear any sensors or carry any gadgets. As a result, this methodology is currently receiving increasing attention, making the HAR framework simple and convenient to implement in various applications. For activity recognition, the majority of the vision based HAR systems that were proposed in the literature used conventional machine learning methods. Deep learning techniques, however, have recently outpaced conventional machine learning techniques [2]. Convolutional neural networks are the most prevalent kind of deep learning technique (CNN). CNN is frequently used in computer vision-related fields. It is made up of a number of convolutional layers through which images are processed.

II. MOTIVATION

Any automated system which identifies human activities are beneficial for various application areas. This research provides a systematic method for automation of Human Activity Recognition Using Transfer Learning.

III. PROBLEM DOMAIN

This research's domain is application of machine learning and related algorithm core area belongs to computer science and application.

IV. PROBLEM STATEMENT

The goal of this research is to develop a model that can recognize fundamental human behaviors such as sitting, sleeping, running, clapping, and eating. The model will be given a collection of images showing individuals performing various behaviors. The label for an image will be whatever action that is being performed in that image. Before the model is able to predict the label of an input (image) that it has never seen before, it must first learn this correlation. The model would technically need to be able to discriminate between various human activities.

Because of advancements in computer vision, computers are growing better at handling some extremely complicated tasks (like comprehending an image). There are models being created that, when given an image, can either predict what the image is about or determine whether or not a specific object is included in the image. The structure and function of the human brain were taken into consideration while creating these models, which are frequently referred to as "neural networks" or "artificial neural networks". Deep learning, a subset of machine learning, is the study of these neural networks. Various versions of these networks have been used to several different problems over time.

V. INNOVATIVE CONTENT

Various researchers worked on human activity recognition; this paper showcase an automated system in which transfer learning is combined with VGG-16 with 5 epochs. By applying this a higher accuracy can be achieved for human activity recognition.

VI. RELATED WORK OR LITERATURE STUDIES

Deep et.al [1] Described a technique for categorizing behaviors of people using RGB-D data and deep convolutional neural networks. The experimental data on the UTD-MHAD dataset shows that combining different modalities can result in performance that is superior to using each modality alone. By fine-tuning a pre-trained model like VGG-16, cutting-edge results can be obtained even on a tiny dataset. Yu, Xiangchun et.al [2] In this study, it suggests using deep ensemble learning techniques to recognize human motion in still photos automatically. Due to the nature of human motions like "Phoning," "Riding a Horse," and "Running," static cue-based methods are necessary. In addition to video-based techniques, still photographs can also be used to recognize human actions. Consequently, an essential question is how to effectively reduce overfitting when training our model. Latha et.al [3] In this paper, an innovative technique (ReHAR) for detecting human activity in videos is proposed. The proposed model operates existing innovative approaches in terms of accuracy on datasets that include individual activity and group activity, and it is entirely trainable. Shi et.al [4] presented deep learning model for recognition of human activity.

El-Ghaish et.al [5] resented a hybrid deep learning architecture by combining CNN and LSTM. Transfer Learning from the VGG16 to NTU RGB+D then from NTU RGB+D to any small dataset. The MCLP model presented in this research blends CNN, LSTM, and pre-trained VGG16 deep architectures to capture the spatiotemporal features—body position, part form, and body motion—in human activities. The recommended MCLP combines the output scores of CL1D, CL2D, and CMHI utilizing the Late Multiply Fusion module to calculate the combined probability of the action done and arrive at the final decision. Abraham et.al in their book [6] presented numerous methods which can be applied to various domain. Li et.al [7] in their study used LSTM to predict the human activities based on the generated representations. UCFSports Action Dataset and NCAA Basketball Dataset are used in their study. Li et.al [8] Author presented a novel method to push the classification performances. For this a hybrid fusion scheme is proposed which combine soft and hard fusion. Their framework is based on Bi-LSTM.

Allah Bux Sargano et.al [9] presented CNN as source and SVM-KNN for action classification. Ann et.al [10] in their study presented a review of various methods for human activity recognition. Srikanth Tammina [11] in this research work proposed transfer learning for classifying images. Author in the thesis applied models ResNet, VGG16 and Inception V3 for human activity recognition.

VII. METHODOLOGY

For human activity recognition; proposed system at first extracts features from data, then it loads image to detect motion signal. Further tracking and processing of each frame is performed. In this study, we employ the Kaggle Dataset and CNN to identify human activity. From the photos, we first extracted the frames for each activity. Transfer learning is specifically used to create trained machine learning classifiers and deep picture features. To

categories activities, we used three different CNN models, and we compared the findings to another research on the same dataset.

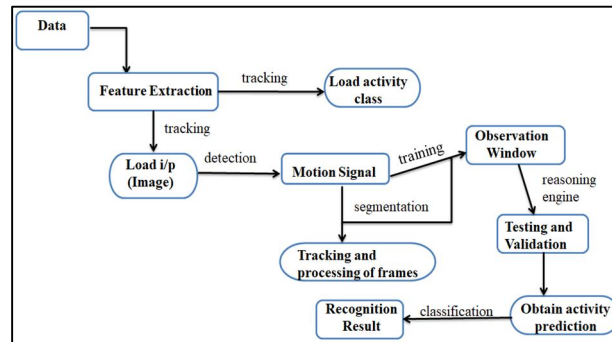


Fig 1. Proposed Diagram of system.

VGG-16CNN model is applied. One of the greatest computer vision models available today is called the VGG model, which stands for Visual Geometry Group. It is an algorithm for object detection and classification that has a classification accuracy of 92.7% when classifying 1000 images into 1000 different categories. Six models with various numbers of trainable layers are used in the experiment. A total of 21 layers—13 Convolutional layers (feature extractors), 5 Max Pooling layers (spatial compression), and 3 Dense layers—have been added, but only 16 of those layers—the learnable parameters layer—are weighted layers. 224, 244 tensors with 3 RGB channels are the size of the input tensor

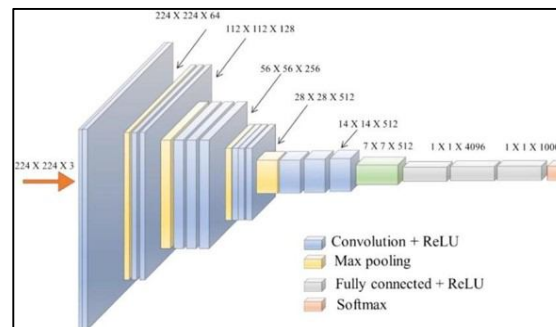


Fig 2. VGG-16-CNN model

Following steps give details of experiments done in this research work

Step 1: There were 15 labels [using laptop, hugging, drinking, dancing, cycling, calling, laughing fighting, listening to music, texting, Sitting, Sleeping, Clapping, Eating, Running] in fig 4. We are working on 5 labels [Sitting, Sleeping, Clapping, Eating, Running]. We are predicting the human activity from the image through this label, shown in below Fig 3.



Fig 3. Label dataset image

```
In [15]: df = df.loc[(df['label'] == "sitting") | (df['label'] == "sleeping") | (df['label'] == "clapping") | (df['label'] == "running")] (
empty= pd.DataFrame()
pd.concat([empty,df], ignore_index=True)
empty.head()

Out[15]: ---

In [16]: df
Out[16]:
```

	filename	label
0	image_1.jpg	sitting
3	image_4.jpg	sleeping
5	image_6.jpg	sleeping
8	image_9.jpg	clapping
12	image_13.jpg	clapping
---	---	---
12582	image_12583.jpg	clapping
12583	image_12584.jpg	sleeping
12585	image_12586.jpg	sitting

Step 2: The dataset of 12000 images. In the target variable there are 15 labels out of which 5 labels counts are shown below from the total dataset. Each label has equal number of counts.

```
In [19]: # Checking Label counts

counts = df['label'].value_counts()
counts

Out[19]: sitting      840
sleeping      840
clapping      840
eating        840
running       840
Name: label, dtype: int64
```

Fig 5. Labels

Step 3: Visualizing label data using library plotly. Display data in percentage in pie chart.

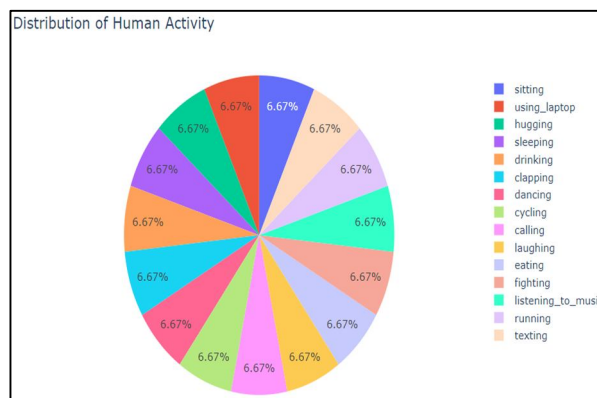


Fig 6. Visual display of labels.

Step 4: In dataset, there are 15 label and we have taken 5 label are build in binary format stored in array format using numpy library, if label has positive; it is evaluated to '1'. If label isn't there in the dataset taken, it is evaluated to '0'.

```
y_train = to_categorical(np.asarray(train_csv['label']).factorize()[0])
print(y_train[0])

[1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
```

Fig 7. Binary Format.

Step 5: Summary of model with different layer used vgg16, flatten, dense, dense_1 shown in Fig 8.

```
In [28]: model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
model.summary()
```

Model: "sequential"		
Layer (type)	Output Shape	Param #
vgg16 (Functional)	(None, 512)	14714688
flatten (Flatten)	(None, 512)	0
dense (Dense)	(None, 512)	262656
dense_1 (Dense)	(None, 5)	2565
Total params: 14,979,909		
Trainable params: 265,221		
Non-trainable params: 14,714,688		

Fig 8. Summary of the Layers.

We have taken the epoch of 5 to fit our model. At first epoch=1 we were getting the accuracy of 0.64. At each epoch the accuracy goes increasing. At the end we are getting accuracy of 0.9980.

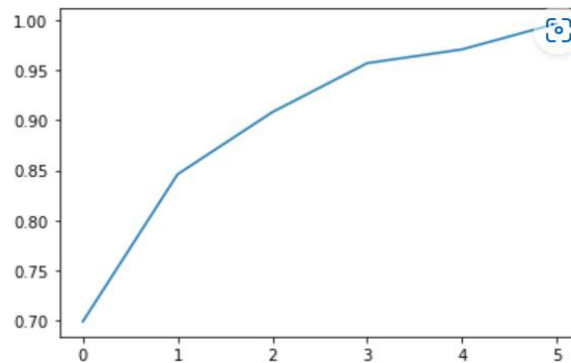


Fig 9. Accuracy graph.

At the time of epoch=1 we were getting the loss of 1.82. At the end means at epoch =5 we are getting the loss of 0.001.

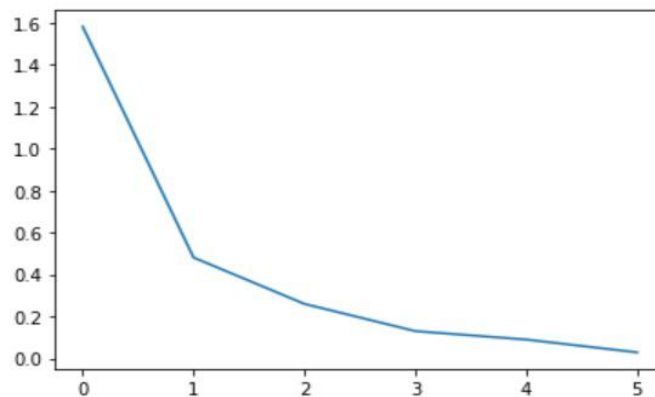


Fig 10. Loss graph

VIII. RESULT

In the prediction of test image, we are getting the probability of 99.9 %. The prediction class is also matching to the given image provided for test for example, it predicts the correct output depends on user input show in below image Fig 11,12 output is correctly predict that is running.



Fig 11. Result (sitting)



Fig 12. Result (eating)

In above, results are correctly predicted, i.e. sitting, eating with probability of 100 % and predicted class. Overall accuracy of proposed system is 99.9%.

Result of proposed system is compared with other existing work, Haobo Li, Aman Sheathe, Hadi Heidari, Julien Le Kernec and Francesco Fioranelli [8], The suggested bi-LSTM performs two strong hard-fusion methods utilizing the matrix of confusion of both sensors, as well as soft feature fusion between wearable sensors and radar data. The classification performances are further enhanced to 96% accuracy in differentiating regular activity. In another existing work [5] Convolution neural networks (CNNs), long short-term memory (LSTM), and a _netuned pre-trained architecture are all combined into one hybrid architecture in the deployed architecture MCLP:CNN+LSTM+VGG16, , the suggested approach MCLP is better 84.3% in CS and 90.4% in CV. Another existing work in this proposed system CNN algorithm is used with method of Depth+RGB (product Rule) the accuracy is 91.2 % which is best method ,are compared with our proposed system where our accuracy with of 99.9% for 5 different activities of human.

IX. CONCLUSION

In this paper, CNN models are used to predict human behaviour from the Kaggle Dataset. Convolutional neural networks (CNN) were used in experiments to identify activities. To get the deep image features and put transfer learning to use in training the machine learning classifiers. Results indicated that when transfer learning had been used combined with VGG-16 in the proposed system with 5 epochs, accuracy is 0.9980, has increasing epochs can get better accuracy. The outcomes of an experiment utilizing the transfer learning approach also showed how well the VGG-16 performed.

FUTURE WORK

Human activity recognition system is the trending topic these days and these industries will be leading the advancement of facial recognition in 2022 and beyond. In this proposed system, we have taken into account the images, but in the future, we may also take into account the real-time videos to recognize different activities in it with other deep learning algorithm and recognition technique. In future it can be extended to recognize multiple activities performed in single frame and also can be used for blind people by giving feature of sound so they can hear the human activity.

REFERENCES

- [1] Deep, Samundra, and Xi Zheng. "Leveraging CNN and transfer learning for vision-based human activity recognition." In 2019 29th International Telecommunication Networks and Applications Conference (ITNAC), pp. 1-4. IEEE, 2019.
- [2] Yu, Xiangchun, Zhe Zhang, Lei Wu, Wei Pang, Hechang Chen, Zhezhou Yu, and Bin Li. "Deep ensemble learning for human action recognition in still images." *Complexity* 2020 (2020): 1-23.
- [3] Latha, K., and Mohamed Musammil. "Human Action Recognition Using Deep Learning Methods (CNN-LSTM) Without Sensors." In 2022 1st International Conference on Computational Science and Technology (ICCST), pp. 447-451. IEEE, 2022.
- [4] Shi, Xiaoran, Yaxin Li, Feng Zhou, and Lei Liu. "Human activity recognition based on deep learning method." In 2018 International Conference on Radar (RADAR), pp. 1-5. IEEE, 2018.
- [5] El-Ghaish, Hany, Mohamed E. Hussien, Amin Shoukry, and Rikio Onai. "Human action recognition based on integrating body pose, part shape, and motion." *IEEE Access* 6 (2018): 49040-49055.

- [6] Abraham, Ajith, Andries Engelbrecht, Fabio Scotti, Niketa Gandhi, Pooja Manghirmalani Mishra, Giancarlo Fortino, Virgilijus Sakalauskas, and Sabri Pllana, eds. *Proceedings of the 13th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2021)*. Vol. 417. Springer Nature, 2022.
- [7] Li, Xin, and Mooi Choo Chuah. "Rehar: Robust and efficient human activity recognition." In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 362-371. IEEE, 2018.
- [8] Li, Haobo, Aman Shrestha, Hadi Heidari, Julien Le Kernec, and Francesco Fioranelli. "Bi-LSTM network for multimodal continuous human activity recognition and fall detection." *IEEE Sensors Journal* 20, no. 3 (2019): 1191-1201.
- [9] Sargano, Allah Bux, Xiaofeng Wang, Plamen Angelov, and Zulfiqar Habib. "Human action recognition using transfer learning with deep representations." In *2017 International joint conference on neural networks (IJCNN)*, pp. 463-469. IEEE, 2017.
- [10] Ann, Ong Chin, and Lau Bee Theng. "Human activity recognition: A review." In *2014 IEEE international conference on control system, computing and engineering (ICCSCE 2014)*, pp. 389-393. IEEE, 2014.
- [11] Tammina, Srikanth. "Transfer learning using vgg-16 with deep convolutional neural network for classifying images." *International Journal of Scientific and Research Publications (IJSRP)* 9, no. 10 (2019): 143-150.
- [12] Albert Florea, George, and Filip Weild. "Deep Learning Models for Human Activity Recognition." (2019).