



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 7, July 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

PoseSceneNet: Integrating Human Pose Estimation with Scene Understanding using CNN

Asma Noorain

Post-Graduation Student, Department of MCA, PES Institute of Technology and Management, Shivamogga,
Karnataka, India

ABSTRACT: The science of computer vision is expanding rapidly each year, and human action recognition is one of its applications in numerous industries, including the video surveillance, sports performance analysis, and healthcare (for example, tracking patients' workouts). In this proposed paper, tensor flow and Python 3 are used to find the predicted scene and predicted pose in images. Using a dataset of about 1500 images, MobileNet, a CNN trained on the ImageNet Large Visual Classification dataset was retrained for binary pose recognition (sitting or upright). To categorize situations around the individual in question, We use the Stanford 40 dataset, which consists of 40 different daily activities, to retrain the Inception-v3 architecture. The project employs an implementation for pose recognition and evaluation called OpenPose, which is based on TensorFlow.

KEYWORDS: Pose Estimation, Scene Recognition, Inception v3, Mobilenet v2.

I. INTRODUCTION

The process of using deep neural networks to analyse human motion is known as "human activity recognition". Human motion can be defined as actions gestures or behaviour captured. The movement data is transformed into action commands that computers may use to execute and analyse the human activity recognition code. The Human Action Recognition project aims to create a model that can process image input from many sources such as security cameras. Identify the activities that the individuals are performing which is also known as human activity recognition. Currently, many public venues including as hospitals, schools, universities, businesses, libraries, parks, metro stations and highways, use video surveillance to monitor traffic flow.

The class Mobilenet Network derives from network_base. In order to manage the convolutional layer widths and trainability Base Network is initialized using input tensors and parameters. The network design is described in the setup method. It consists of two primary components the OpenPose component which refines feature maps to create posture estimate outputs and MobileNet v2 which uses separable convolutions for feature extraction. The network consists of many stages containing layers for convolution, upsampling and concatenation. It offers ways to obtain restorable variables, refinement stages and final output tensors and it has loss functions for keypoint and limb identification tasks. It enables efficient PoseEstimation Evaluation in the project. TensorFlow is used to create and load neural network models for posture estimation. It has functions like get_network(), which builds and returns a neural network model depending on the supplied type such as MobileNet Network, MobileNet Network Thin, Cmu Network, or Person Lab Network, and _get_base_path(), which gets the base path for model files.

- To identify and calculate the locations of human body joints in pictures and video frames using the openpose algorithm.
- Create a reliable model that can categorize various human activities according to estimated postures and temporal dynamics.

The proposed approach can process a single input image and provide annotated results for human posture estimation action categorization and scene interpretation. To detect human key points, it starts by importing an image and initializing a pretrained posture estimation model. The skeletal structure is separated by drawing the detected postures onto a white, blank picture. After that, the original image is examined for scene context, and this skeleton image is

utilized for action categorization. The system shows the captioned image records pertinent processing stages for troubleshooting, and recognizes the initial image with the expected action and scene descriptors.

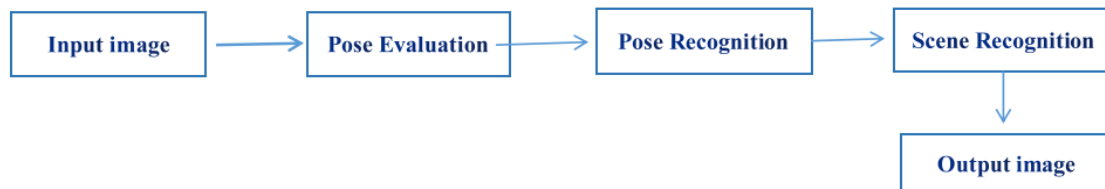


Figure 1.1: Proposed Methodology

The block design for suggested module is shown in Figure 1.1. The first step is to enter an image into the project and add the image's name to the code using .jpg extension. The Pose Evaluation, the process of determining where different bodily parts are located in a picture is conducted using Convolutional Neural networks. The next Step where Pose Recognition is Estimated by MobileNet v2 Architecture due to its Simplified design helps to keep the accuracy and computing efficiency under check. The next task is accurately identifying the scene in the image using Inception v3 architecture. The image's output is then created with captions that include "Predicted Scene" and "Predicted Pose" along with accuracy rate and draws pedestrian's skeleton structure.

II. LITERATURE REVIEW

The study is carried out the experimentation on Robust Human Activity Recognition Approach Using OpenPose, Motion Features and Deep Recurrent Neural Network where the The detection of human activity has become simpler because to developments in computer vision and pattern recognition. This study offers a reliable method for identifying anatomically significant spots from RGB pictures in human movement recognition using the open-source library OpenPose. We use these critical locations to generate strong motion characteristics based on their motions over consecutive frames. A Recurrent Neural Network containing Long Short-term Memory cells is used to recognise actions related to these properties. The strategy is person-independent, utilising several camera angles on different subjects. This study offers a reliable method for identifying anatomically significant spots from RGB pictures in human movement recognition using the open-source library OpenPose. The proposed activity identification method holds promise for both practical applications in elderly caring, enabling more independence and research in image processing and computer vision[1].

This study is carried out the research on Human activity recognition by using convolutional neural network which states as Recently, the Human Activity Recognition system has been the subject of research. HAR studies peripheral support life using automated monitoring systems and smart home sensors. The previous method that was looked at for a few activities was less precise and relied on stationary motion. This study looks at how thermal imaging cameras could improve the precision of motion detection in intelligent surveillance systems. For picture recognition, we use Convolutional Neural Networks a well-liked deep learning architecture. they propose a system that makes use of CNN and thermal cameras to improve accuracy and offer more functionality[2].

This research is based on Human Activity Recognition Using Pose Estimation and Machine Learning Algorithm proposes a method for recognizing and classifying human activity based on a person's stance in photographs and is divided into two sections the estimation of an individual's posture and the use of position to categorize activities. Pose estimation involves recognizing 18 body key points and joint positions. We utilized the OpenPose package for pose estimation tasks. The activity categorization task utilises multiple logistic regression. Using dataset, they evaluated the accuracy of several regression and classification techniques and dataset was divided into two sections one used to train the model and other to assess its functionality[3].

III. METHODOLOGY

Human Activity identification (HAR) system that makes use of pose classification and scene identification, two parallel processes. The ImageNet datasets is used to train the MobileNet v2 model, which is employed in the posture classification process, to recognize various human positions, including sitting and standing. In the meanwhile, as part of the scene identification process, the Inception v3 model is trained on the Stanford datasets to recognize the action in the picture. Using the complementing information that the human position and the surrounding scene supply, the outputs from the two systems are then integrated to identify the overall activity that is taking place in the picture.

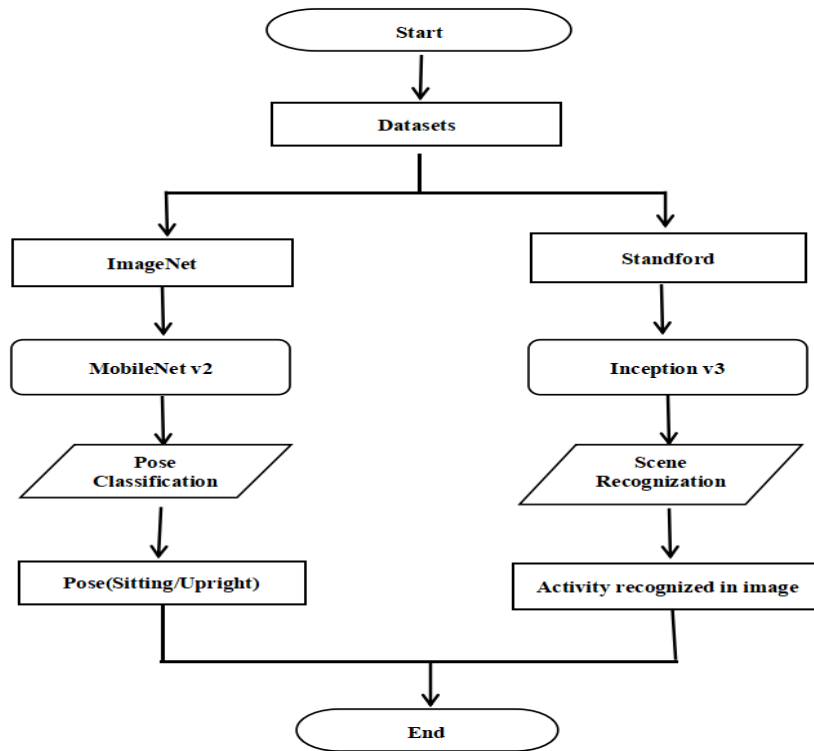


Figure 3.1 : Data flow Diagram

IV. ALGORITHM

1. Convolutional Neural Network

They are used to learn hierarchical features, record spatial hierarchies in pictures and give translation invariance, they are extensively utilised for posture and scene identification. CNNs are capable of recognising intricate patterns and spatial correlations between body components, which is necessary for precisely identifying human postures in pose identification. CNNs are particularly good at extracting and separating significant characteristics from complicated patterns and various objects inside scenes when it comes to scene identification.

Architecture build on CNN

1. MobileNet v2 Architecture

Specifically for mobile and embedded devices the MobileNet V2 architecture is built for effective and high-performance image processing. Let us understand the working of MobileNet v2. First, 224x224 x3 224x224x3 standard input picture is used, and then the data is processed using an initial convolution layer. Multiple inverted residual blocks, each having a 1x1 convolution with ReLU6 activation (expansion phase), a depthwise convolution, and a 1x1 linear convolution (projection phase), make up the foundation of MobileNet V2. Across the network, these blocks configurations including stride and expansion factors vary. The design is further optimized by bottleneck layers, which are 1x1 linear convolutions devoid of activation functions. The network finishes with a

fully connected layer that links to the last classification layer, then a global average pooling layer that averages every feature map to a single spatial dimension. A balance between computing efficiency and model correctness is ensured by this approach.

2. Inception v3 Architecture

This Model is constructed step by step starting with an input layer that handles grayscale photos with defined dimensions. The model includes numerous convolutional and max pooling layers with core Inception modules which employ parallel convolutions with varying filter sizes and pooling layers. To reduce overfitting, dropout layers and local response normalization are used. The last layers include a regression layer that employs the momentum optimizer and categorical cross-entropy loss and fully connected layer for multi-class classification using a softmax activation function. The function returns a deep neural network model that is ready for training and prediction. This advanced design is appropriate for challenging jobs including scene categorization

V. IMPLEMENTATION

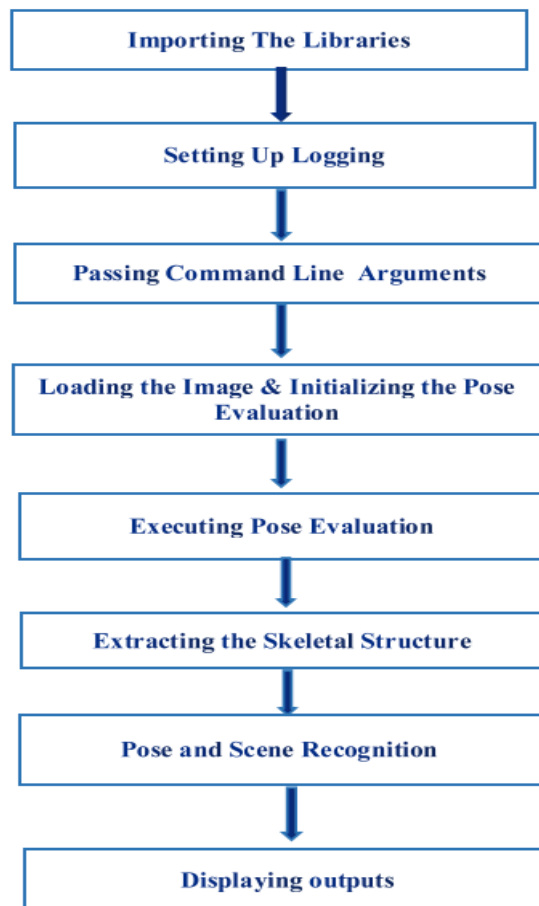


Figure 5.1 : Program flow

Figure:5.1 explains the Flow of the Program. Importing the required libraries and configuring logs for debugging are the first steps. The command-line parameters for the input picture and a possible debug flag are parsed by the main execution block. It scans the input picture, records its dimensions, and uses the 'mobilenet_thin' network to initialize a pose estimate model. Next, to extract the skeletal structure, the model infers human positions from the image and draws them on either the original or an unfilled white image. Separate pre-trained algorithms are given these photos to classify

scenes and actions. The script saves the original image to a file, annotates it with the scene labels and anticipated action, then opens the file in a window.

VI. RESULTS

Results are accessed through enter images across the web and image name is added in the program file. The program execution starts with tensor Flow log messages. Next, the system logs details about the CPU capabilities loads the mobilenet_thin model from a specified directory, initializes it and notifies the user about deprecated TensorFlow methods. In order to detect human poses, the system performs image processing and postprocessing on an input camera image with 996 x 664 pixels. It analyzes the image during classification, recording an assessment time in seconds for the entire process and time in seconds for the initial step. The system is able to identify a variety of activities; it logs limited confidence for certain activities and the highest confidence level (99.992%) for an activity that is shown in a picture. When concluded, the assessment time for processing a single image came in seconds. This log sequence shows how the system can process photos and recognize activities in almost real-time with accuracy.

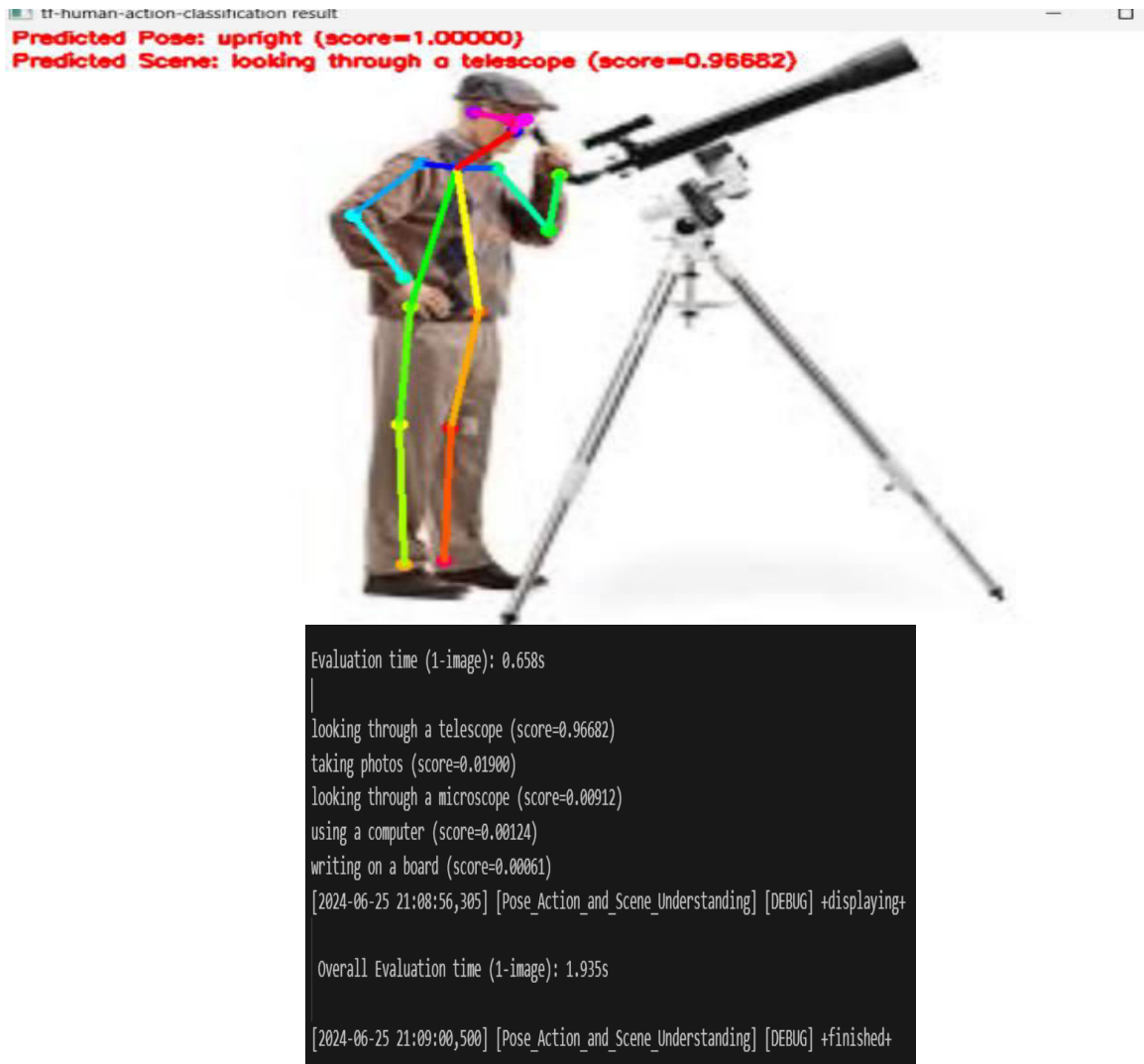


Figure 6.1: Pose and Scene Recognition.

Figure 6.1: depicts that an input picture with dimensions of 798x585 is processed by the network through image processing post processing and classification stages. Efficient model performance is demonstrated by the evaluation

times of 0.220 seconds and 0.675 seconds for processing a single picture in two separate cases. With a score of 0.96682, "looking through a telescope" has the most confidence among the classified activities. Other activities, such as "taking photos" and "using a computer," have lower ratings. The overall time taken for evaluating the image is 1.898 seconds, concluding with a display step.

VII. CONCLUSION

The proposed project with an average inference time of 1.35 seconds per image, the module successfully recognizes and identifies human postures and actions in static photos. nevertheless, improvement is required for real-time performance. It emphasizes that although the module can manage tasks involving people in close proximity, it has trouble with tasks needing a more comprehensive semantic comprehension of the surroundings. The project highlights the possibility for future advancements, such as real-time performance optimization and managing video streams for temporal action detection, and recommends that the current solution serves as a foundation for more general activity recognition modules.

REFERENCES

1. Gupta, Abhay, Kuldeep Gupta, K. N. Mahalakshmi Gupta and Kapil O Gupta, human activity recognition using pose estimation and machine learning algorithm, International Symposium on Intelligent Control (2021).
2. Gulzar, Zameer & Leema, Anny (2019), Human Activity Recognition using Machine Learning Classification Techniques, International journal of Innovative Technology and Exploring Engineering, 9. 10.35940/ijitee.B7381.129219.
3. Fernandes, B., Alag, G., & Kumar, S. (2019), HARM - The Human Action Recognition Module, In Lecture notes in computational vision and biomechanics (pp. 419–433). https://doi.org/10.1007/978-3-030-32040-9_43.
4. Chin Ann, Ong & Lau, Bee. (2015). Human activity recognition: A review, Proceedings - 4th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2014. 389-393. 10.1109/ICCSCE.2014.7072750.
5. Nguyen, Phat & Thi, Nguyen Ngoc & Ngoc Thien. (2021), Proposing posture recognition system combining mobilenetv2 and lstm for medical surveillance, IEEE Access, PP. 1-1. 10.1109/ACCESS.2021.3138778.
6. C. Shiranthika, N. Premakumara, H. -L. Chiu, H. Samani, C. Shyalika and C. -Y. Yang, "Human Activity Recognition Using CNN & LSTM," 2020 5th International Conference on Information Technology Research (ICITR), Moratuwa, Sri Lanka, 2020, pp. 1-6, doi: 10.1109/ICITR51448.2020.9310792.
7. K. Xia, J. Huang and H. Wang, "LSTM-CNN Architecture for Human Activity Recognition," in IEEE Access, vol. 8, pp. 56855-56866, 2020, doi: 10.1109/ACCESS.2020.2982225.
8. Yesha Sarwat¹, Dr. K Vaidehi² and M. Sowmya³ 1PG Scholar, 'Human Activity Recognition Using CNN and LSTM Methods' Department of Computer Science & Engineering, 2Professor, Department of Artificial Intelligence & Data Science, 3Research scholar, Department of CSE, Annamalai University Stanley College of Engineering and Technology for Women, India.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details