INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 3 - V7I3-1445)
Available online at: https://www.ijariit.com

Human activity recognition system

Akash Kumar
akash2508.k@gmail.com
New Horizon College of Engineering, Bengaluru, Karnataka

Varshini Shenoy
varshasn99@gmail.com
New Horizon College of Engineering, Bengaluru, Karnataka

Puneet Tiwari
puneettiwari1999@gmail.com
New Horizon College of Engineering, Bengaluru, Karnataka

ABSTRACT

Understanding the activities of human from videos is demanding task in Computer Vision. Identifying the actions being accomplished by the human in the video sequence automatically and tagging their actions is the prime functionality of intelligent video systems. The goal of activity recognition is to identify the actions and objectives of one or more objects from a series of examination on the action of object and their environmental condition. The major applications of Human Activity Recognition vary from Content-based Video Analytics, Robotics, Human-Computer Interaction, Human fall detection, Ambient Intelligence, Visual Surveillance, Video Indexing etc. The Experimental Evaluation of various papers are observed efficiently with the various performance metrics like Precision, Recall, and Accuracy.

Keywords— Activity Recognition, Deep Learning, Training dataset, ResNet, OpenCV.

1. INTRODUCTION

1.1 Domain Introduction

Deep Learning is an Artificial Intelligence function that almost works as if a human brain would and processes data while creating patterns that can be used for decision making. Deep learning has constantly been evolving with the digital era, which has resulted in an explosion of data in all forms. This data is drawn from sources like social media, ecommerce platforms, and online cinemas. This huge collection of data is readily accessible and can be shared with the help of cloud computing.

However, this data is so vast that it could take years or maybe decades for humans to understand it and extract necessary information. With deep learning, a lot of new applications of computer vision techniques have been introduced and are now becoming parts of our everyday lives. These include face recognition, photo stylization etc. which are based on image recognition. Image recognition means to be able to recognize certain objects through the use of algorithms and machine learning concepts.

1.2 Problem Definition

This project focuses on building a Human Activity Recognition System with OpenCV, TensorFlow and Deep Learning. The goal is to train a custom Deep Learning model to detect human activity being performed on the input stream like a pre-recorded video or a live video input. One of the primary workings of this system is to identify the actions that humans perform and hence tag them. The model trained is capable of identifying over 400 activities like cycling, eating, jogging, etc. with accuracy of 70% to 95%.

1.3 Objective

The primary objective of this project is to solve human centred problems from healthcare to security by inferring several simple human activities. Through this project we will be able to identify the actions of several objects and their environmental behaviour.

Steps to be followed are:
- Take video stream as input
- Process video and extract individual frames
- Pass the frame through trained human activity recognition model
- Compare the frames and predict the type of activity
- Classify the activity and caption it back on the individual frames
- Stream the video to the output

1.4 Scope of the project

The major applications of this system can be seen in robotics, fall detection for humans, AI, video surveillance and many more. This field has gained popularity due to the increase in the websites that share the videos. Hence there is a dire need to develop an efficient way to store the videos. The applications that are required for designing these interfaces by establishing a context-aware computing and so on. The detection of abnormal activities in the video surveillance is not limited to detecting unauthorized people entry and abnormal crowd behaviour.

2. LITERATURE SURVEY

2.1 Technology

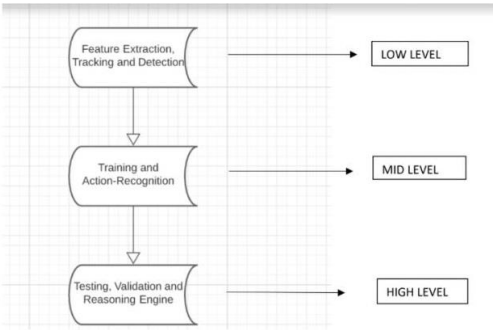
Residual Networks (ResNet) – Deep Learning:

After the first CNN-based architecture (AlexNet) that win the ImageNet 2012 competition, every subsequent winning architecture uses several deep neural networks to reduce the error rate. It can work for a few layers, but when we increase the number of layers, there is a common problem in deep learning associated with that called Vanishing/Exploding gradient. This causes the gradient to become 0 or too large. Hence when we increase the number of layers, the training and test error rate also increases.

2.2 Existing System

HAR system gained its significance as a result of increase in the number of surveillance cameras. The task can be difficult due to changes in action performance and inter-personal differences. To understand the action in the video, there are two things to be noted- Action recognition and action localization. One is 'what action' is performed in the video and the other is 'where the action' of interest is taken place. The problem of assigning videos into several predefined action classes is known as action recognition and action localization. Training includes both positive and negative samples. After the training phase we can test the videos. The HAR system follows a step-wise process.

FLOW OF PROGRAM



2.3 Proposed System

In this study, the proposed concept is advanced primarily based totally on Python 3, Keras, OpenCV, ResNet, and TensorFlow. The foremost cause of this device is to system the enter video flow for human detection and similarly processing the character frames of the enter video to expect which hobby is being carried out with the aid of using the human. After the prediction made, that's correct as much as 94%, the frames are captioned and the end result is given to the output. One of the most important part of this study is the classification of activity being performed by the human and this feature is dependent on object detection framework.

3. REQUIREMENT ANALYSIS

3.1 Functional Requirements

- System must be able to read video sequence as input
- System should be able to extract each frame from the video input for processing
- System must be able to pre-process the frames extracted from the input and resize or crop it to the required threshold size.
- System should be able to compare the frames with the trained weights.
- After comparing, the system should be able to categorize the input sequence into various classes with acceptable accuracy.

3.2 Non-Functional Requirements

Security

No outside entity shall be allowed to modify content of code without proper authorization.

Usability

- Self-learning support must be available.
- System must be intelligent enough to suggest through proper steps as you continue using the system.
- System should be able to recognize all kinds of activities a human can perform.
- There should not be any restriction on the type of input video stream the system can process.

Reliability

- The system should be able to recover in time.
- System should be able to handle any exceptions properly

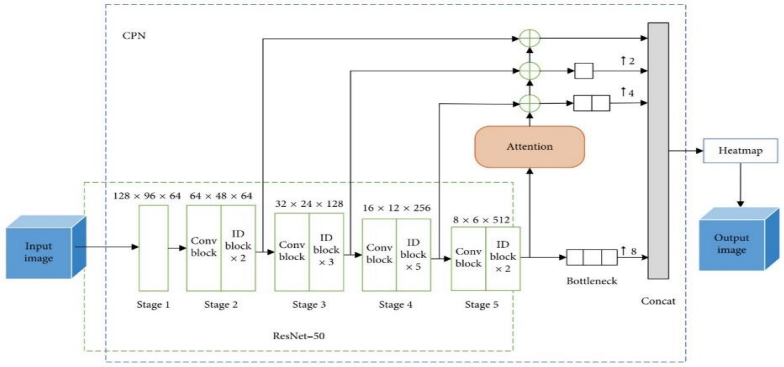
4. DESIGN

4.1 Design Goals

INPUT: Take the pre-recorded video as input from the user through command line backend.
OUTPUT: The output should caption the frames of input video based on the prediction made by trained activity recognition model.
EFFICIENCY: The system should be able to identify human from any other similar objects in the input video stream and appropriately classify the activity with acceptable accuracy.

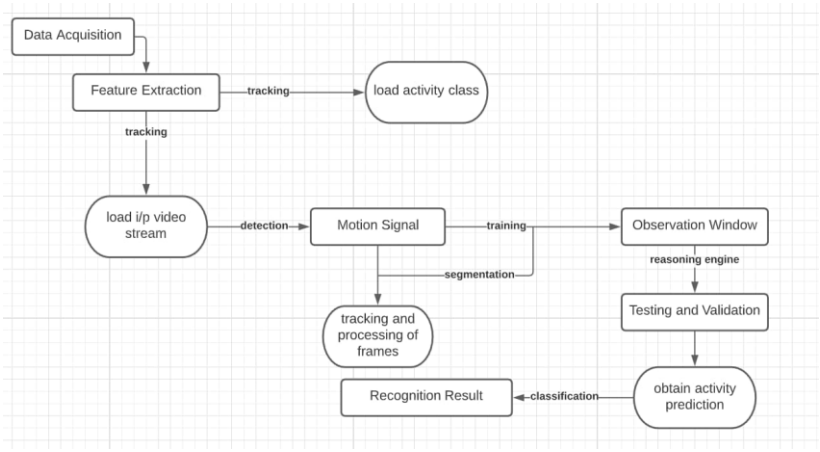
4.2 Architecture

Since ResNets can have variable sizes, depending on how big each of the layers of the model are, and how many layers it has, we are going to follow the one described by the authors in the paper [1] — ResNet 34 — in order to explain the structure after these networks. In here we can see that the ResNet (the one on the right) consists on one convolution and pooling step (on orange) followed by 4 layers of similar behaviour.The dotted line is there, precisely because there has been a change in the dimension of the input volume (of course a reduction because of the convolution). Note that this reduction between layers is achieved by an increase on the stride, from 1 to 2, at the first convolution of each layer; instead of by a pooling operation, which we are used to see as down samplers.

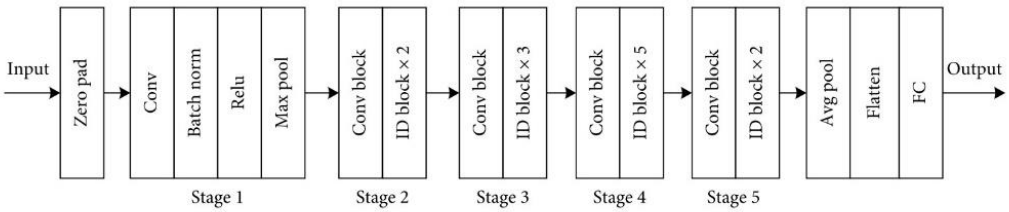


CPN network with attention model introduced using ResNet

4.2.1 Data Flow Diagram



4.2.2 Network Structure of ResNet-50



Basic network structure of ResNet-50

5. APPLICATIONS

Advances in today's modern technologies have blessed us with innovative ways to improve the standard of living of elderly and disabled people. Assisted and active living systems use HAR techniques and analysis to monitor as well as assist residents to secure their safety. A smart home is an environment equipped with sensors that enhance the safety and well-being of residents whilst monitoring their health conditions.

Hence, such smart homes integrated with the HAR system, help improve the level of independency as well as the quality of life of people who need support in terms of physical and cognitive functions. Basically, in a smart home, the behaviour of residents and their interaction with the environment is monitored by performing analysis on the data collected by the sensors.

6. FUTURE WORK

A greater effective software may be advanced to Understanding and detection of day by day existence sports in long-time period videos. Although numerous full-size assessment papers have already been posted withinside the widespread HAR topics, the developing technology withinside the subject in addition to the multi-disciplinary nature of HAR activate the want for steady updates withinside the subject. Indeed, maximum laptop imaginative and prescient packages which include human laptop interaction, digital reality, security, video surveillance and domestic tracking are particularly correlated to HAR tasks. This establishes new fashion and milestone withinside the improvement cycle of HAR structures. Here we gift unique insights on present works and the methodologies utilized by researchers for spotting the human sports.

Comparison amongst special human sports with the aid of using similarity structures is mainly hard because of the terrific form of strategies carried out to symbolize likeness and the dependence that the outcomes gift of the used picture dataset. This could be beneficial to the researchers for his or her destiny studies path on this area.

7. REFERENCES

[1] T. Lan, Y. Wang, and G. Mori, "Discriminative figure-centric models for joint action localization and recognition," in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011, pp. 2003–210.

[2] M.Raptis, I.Kokkinos and S.Soatto," Discovering discriminative action parts from midlevel video representations" ,in Proc ,IEEE Conf.Comput.Vis,.Pattern Recog.,Jun 2012,pp.1242-1249.

[3] M.Jain,J.van Genmert,H.Jegou ,P.Bouthemy and C.Snoek ,"Action localization with tubelets from motion" ,in Proc IEEE Conf, Comput.Vis.Pattern Recog. Jun 2014 pp 740747

[4] H. Zhang and O. Yoshie, "Improving human activity recognition using subspace clustering," in Machine Learning and Cybernetics (ICMLC), 2012 International Conference on, vol. 3, July 2012, pp. 1058–1063.

[5] N. Robertson and I. Reid, "A general method for human activity recognition in video,"Computer Vision and Image Understanding, vol. 104, no. 2, pp. 232–248, 2006.

