INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

IJARIIT

ISSN: 2454-132X

Impact factor: 4.295

(Volume 5, Issue 1)

Available online at: www.ijariit.com

Theft detection using computer vision

Siddesh Patil
siddesh.patil15@vit.edu
Vishwakarma Institute of Technology, Pune, Maharashtra

Mrunal Shidore
mrunal.shidore@vit.edu
Vishwakarma Institute of Technology, Pune, Maharashtra

Tejas Prabhu
tejas.prabhu15@vit.edu
Vishwakarma Institute of Technology, Pune, Maharashtra

Sanket Yenare
sanket.yenare15@vit.edu
Vishwakarma Institute of Technology, Pune, Maharashtra

Vaibhav Somkuwar
vaibhav.somkuwar15@vit.edu
Vishwakarma Institute of Technology, Pune, Maharashtra

ABSTRACT

Theft is one of the most common criminal behaviors and it is increasing day by day. It has become one of the never-ending problems of the world. In the USA alone, about 7000 cases of baggage theft are reported to Transportation Security Administration (TSA) for theft on airports. Most of the screening takes place before the passengers and their luggage can get on the plane. In order to stop these increasing theft on the airports, a robust system is required which would detect if there is any robbery taking place and would inform the respective authorities. So to solve this problem, we will be creating a system which would accurately track the people and their baggage's and would detect if there is any theft and would inform the authorities.

Keywords— *Theft, Computer vision, CNN*

1. INTRODUCTION

Loss of luggage can happen with anyone irrespective of the circumstances and conditions. Luggage and bags with important documents or precious things and can be lost or theft with which people can lose their important material. Generally, it is seen that people get robbed in public areas like railway stations, bus stands, and other public and private areas. Also, people can even forget their luggage and bags which can have important and necessary things. So it is very necessary to track down the bags in case of loss and theft. A lot of other tracking systems and devices are already present like car tracking system.

Our solution aims at identifying unattended baggage in public areas like railway stations, airports and so on and then triggering an alarm. We will be training our models to detect and track the people and their bags at any particular place using *keras* and *tensor flow*.

2. LITERATURE SURVEY

Image processing technique plays an important role in the detection of theft in sensitive areas. The merits of the proposed scheme for theft detection lie in the segmentation of the humans and their baggage's by FCM clustering and determination of the MO from the minima in average pixel intensity plot of the mouth region. An important aspect of this paper is the design of an emotion control scheme. The accuracy of the control scheme ensures convergence of the control algorithm with a zero error, and repeatability ensures the right selection of audio-visual stimulus. The interpersonal communication between a human being and a computer can be increased rapidly by combining all the pros.

3. CONVOLUTION NEURAL NETWORK

Convolution can be used to achieve the blurring, sharpening, edge detection, noise reduction, which are not easily achieved by other methods. It is represented in mathematical form as follows-

$$g(x,y) = f(x,y) * h(x,y) \tag{1}$$

Where,
$g(x,y)$ = *output image*
$f(x,y)$ = *input image*
$h(x,y)$ = *filter/kernel/mask*

In the convolution method center of the kernel is kept at each element of the image input and the corresponding elements are multiplied and added together.

CNN is a type of neural networks, is made up of a large number of neurons with weights and biases which defines the relations between them. Each neuron receives several inputs, takes a weighted sum over them, pass it through an activation function and responds with an output.

As the name suggests convolution process lies at the heart of this neural network. It is comprised of three major components, namely Convolution layer, pooling layer, and Fully-connected layer.

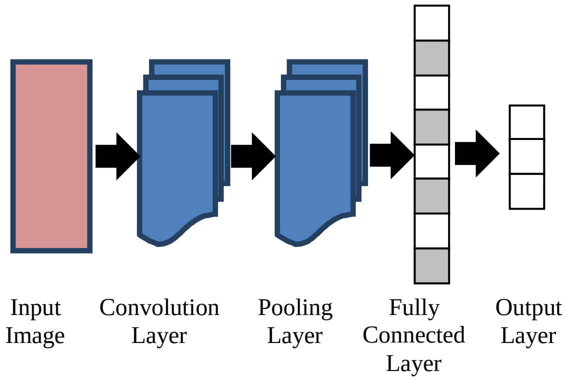

Fig. 1: Typical Convolution Neural Network

3.1 Convolution Layer
This is the main part of the entire structure of CNN. The convolution layer comprises of a set of independent filters. In this layer, the convolution operation, which is an element-wise product and sum, is carried out between the input image layer and kernel. If the required feature is present in that image then convolution operation results in a high valued real number. Otherwise, the operation yields low value. This layer is used to extract the features of the input images. The feature map obtained after the convolution operation is then passed to the activation function. The obtained features maps depend on which kernel we have used. Also, the final results on convolution layers depend on which activation function we are using. Several convolution layers can be connected to improve the functioning of the algorithm.

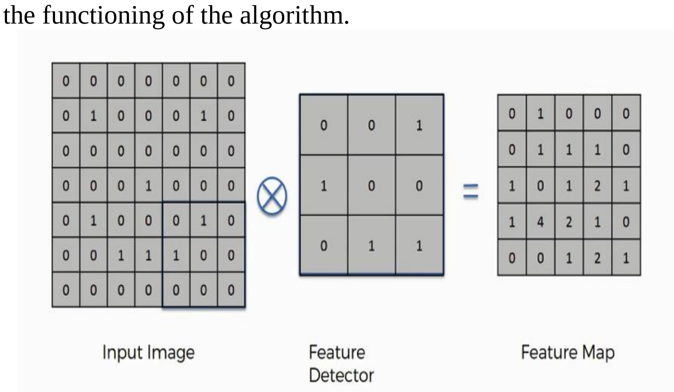

Fig. 2: Convolution operation

3.2 Pooling Layer
Due to the cascading of the convolution layers, after several layers, the size of the feature map keeps on reducing. The reduction in the representation's size is known as downsampling. To reduce the amount of memory consumed by the network and speed up the training process, we try to reduce the redundancy present in the input feature. There are a couple of ways for downsampling an image namely max pooling, average pooling, etc. In max pooling, a window passes over an image according to a set stride (how many units to move on each pass). At each step, the maximum value within the window is pooled into an output matrix, hence the name max pooling. In the case of average pooling, the average value is pooled into the output matrix. The output size of the max pooling operation can be calculated using the following equation: (2)

$$n_{out} = floor\left(\frac{n_{in} - f}{s}\right) + 1$$

Where n_{in} denotes the dimension of the input image, f denotes the window size, and s denotes the stride. Pooling layers are stacked over convolution layers.

3.3 Fully Connected layer
The CNN system has fully connected layer as the last component. At the end of this layer, the output layer is connected. The output of the pooling layer is flattened into a single feature layer and passed through a network of neurons to predict the output probabilities. Generally, the softmax function is used as an activation function at the end of a fully connected layer. The rows are concatenated to form a long feature vector. If multiple input layers are present, its rows are also concatenated to form an even longer feature vector. At each dense layer, the feature vector is multiplied by the layer's weights, summed with its biases, and passed through a non-linear function such as softmax.

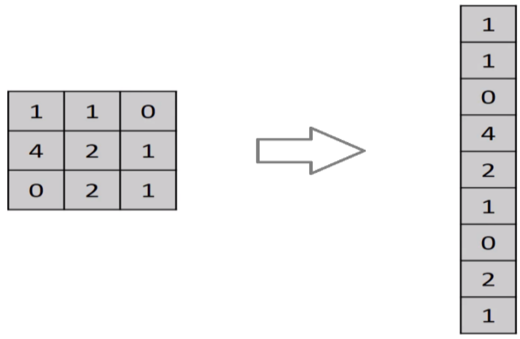

Fig. 3: Fully connected layer

4. THE ARCHITECTURE OF THE MODEL

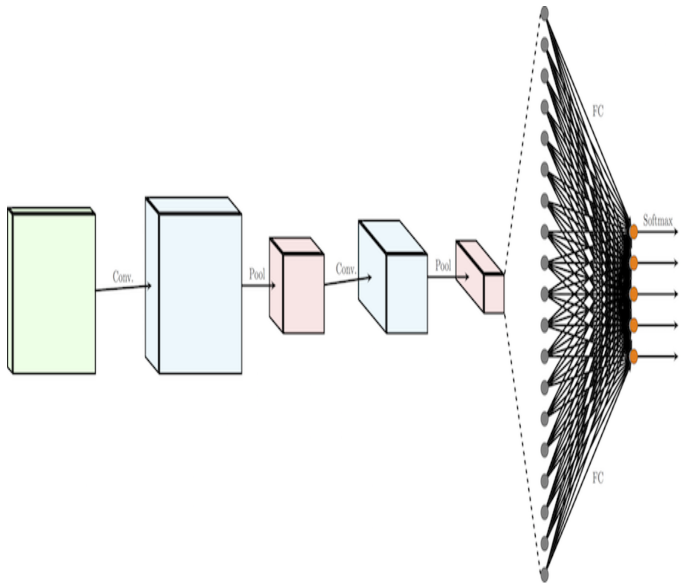

Fig. 4: General architecture of a CNN

The proposed model consists of three convolution layers, one max pooling layer, two average pooling layers and at last three dense layers of fully connected layers. The first convolution layer has input values of dimensions 48 x 48 x 1 and 64 kernels of size 5 x 5. Relu activation function is used here. Max pooling is done further with kernel size 5 x 5 with a stride of 2. The second convolution layer uses 64 kernels of size 3 x 3 with the activation function of Relu. Convolution is done 2 times here and average pooling is used with a kernel of 3 x 3 size with stride 2. In the third convolution layer, convolution is done two times with 128 kernels of 3 x 3 size and again average pooling is done with the same characteristics as that of the previous layer. After that at the end feature map is flatten into a dense layer of size 1024. To overcome the regularization problem, dropout method is used with a dropout value of 0.2 after each dense layer. At end of the CNN, the last layer is a dense layer with a size of 7 and the activation function used is softmax. The algorithm is optimized using Adam optimizer.

The model works best when it is training until 6 epochs with a batch size of 256.

5. EXPERIMENTAL SETUP

The processor used for this purpose is Intel(R) Core i7-8750H CPU @ 2.20GHz. The system has an 8 GB of memory installed and it has x64 based processor.

We have used fer2013 dataset for facial expression detection. It has almost 35888 total samples. It means it has 35888 rows or observations. It has 3 columns namely emotions, pixels and usage. Each image is of size 48 x 48 in grayscale. The entire dataset is split into two parts. Training part has 28710 samples and testing part has 7177 samples.

From the analysis, the model is successful in identifying emotions with a probability of up to 0.5. The graph gives us the probabilities of all the seven emotions present in the image.

In the above-mentioned table, the rows give the actual emotions and columns give predicted emotions. The diagonal values give us the true positive outcomes means the number of correctly predicted emotions. All the other values are false predictions.

6. RESULT

In Figure 5 we can see that the person and bag are detected by the firmware and allotted the same number to both of them. In this way, the firmware understands that the green bag belongs to that particular person and hence a link is drawn on the screen. In Figure 6 when the person moves away from the bag the link continues to follow the person. In this way, on the screen, we can see the person moving away and the link following it.

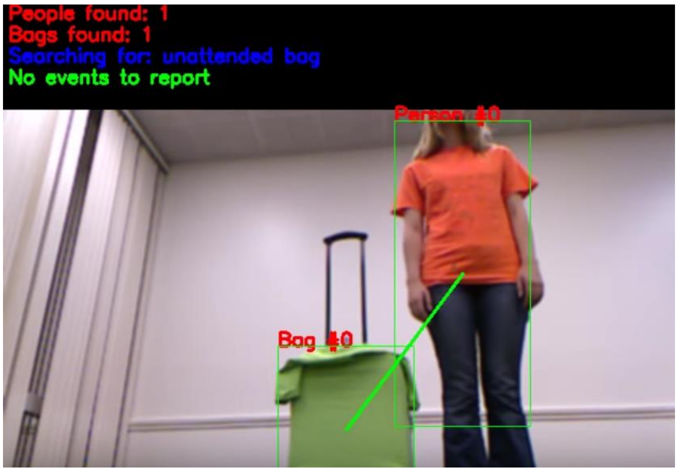

Fig. 5: Person with the bag

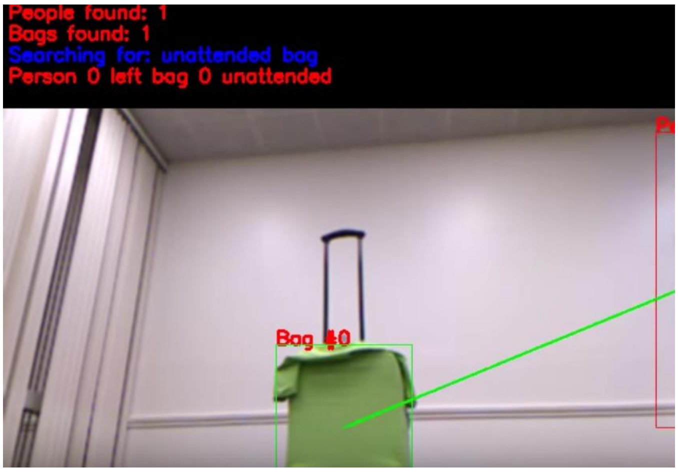

Fig. 6: Link between the person and the bag

7. CONCLUSION

Using highly advanced algorithm like CNN gives you an edge over other traditional algorithms like RNN, SVM etc. This algorithm successfully classifies the facial emotions with accuracy up to 94 percentage. Keras toolkit is used for this purpose. After trying out different layer combinations and number of iterations, we came up with the aforementioned improvised model.

8. REFERENCES

[1] Martin Spengler and Bernt Schiele "Automatic Detection and Tracking of Abandoned Objects", Perceptual Computing and Computer Vision Computer Science Department ETH Zurich, Switzerland

[2] Fahian Shahriar Mahin1, Md. Nazmul Islam1, Gerald Schaefer2, and Md. Atiqur Rahman Ahad1 "A Simple Approach for Abandoned Object Detection", 1 Department of Electrical and Electronic Engineering, University of Dhaka, Bangladesh 2 Department of Computer Science, Loughborough University, U.K.

[3] Dima Damen, David Hogg "BICYCLE THEFT DETECTION" School of Computing, University of Leeds, LS2 9JT, Leeds

[4] Joseph Redmon∗, Santosh Divvala∗†, Ross Girshick¶, Ali Farhadi∗†, " You Only Look Once: Unified, Real-Time Object Detection ", University of Washington∗, Allen Institute for AI†, Facebook AI Research¶

[5] Alexandros Andre Chaaraoui 1,*, Jose Ram ´ on Padilla-L ´ opez ´ 1 , Francisco Javier Ferrandez-Pastor ´ 1 , Mario Nieto-Hidalgo 1 and Francisco Florez-Revuelta, " A Vision-Based System for Intelligent Monitoring: Human Behaviour Analysis and Privacy by Context ", Sensors 2014, 14, 8895-8925; doi:10.3390/s140508895

[6] Rajesh Kumar Tripathi1 · Anand Singh Jalal1 · Subhash Chand Agrawal " Suspicious human activity recognition: a review", I © Springer Science+Business Media Dordrecht 2017.