Smart Camera System Using Machine Learning

Siddharth Sandbhor[1] , Saurabh Solkar[2]
Department of Computer Engineering, Suman Ramesh Tulsiani Technical Campus FOE, Pune, India
Prof.Bhushan Mahajan
Guide
Department of computer Engineering

Abstract :The current camera system been installed is not smart to detect the anonymous activities and notify it to the authority The old systems were taking to much time to train the module with the dataset and the system used to analyze the data on the server side this was time consuming process. Proposed system will be made smart by using external hardware and software which will detect anonymous activities in the bus to improve security .System will Use raspberry PCB with installed android things/raspbian operating system in it . using Machine Learning the system will detect anonymous activities .In machine learning, a convolutional neural network (CNN, or ConvNet) is a class of deep, feed-forward artificial neural networks, most commonly applied to analyzing visual imagery. In machine learning the pretrained modules (e.g. mobile net ,inception) will be used in this neural networks. Also the dataset (e.g. coco,imagenet) will be used.

Keywords: *Internet of Things, Machine Learning, Convolution Neural Network.*

I. INTRODUCTION

Security is biggest problem of all the time lots of companies making lot's of efforts to built new security system or improve existing system to achieve better and accurate results .

The security camera system that formally we were using that system was just capable of streaming the video. To check the activities every time is always headache it require separate employee to monitor activities sometimes it doesn't give expected output . this system is not proper solution in security point of view.

Formally we were using non smart camera system that camera system not useful for detection of activities over the time technology improved and using machine learning(ML) technology we started detecting activities, detecting objects using ML algorithms. initially this technique faced lots of problems ,problems in terms of processing time ,storage problem, accuracy etc. but over the time because of contributive efforts of lots of developers technology improved and it started giving more accurate outputs. all we know beauty of ML technology is system improves its knowledge. System make itself smarter and accurate by learning the surrounding .

Detection of object is most important task in the whole camera system this can be archive by lots of different ways but every way will not give accurate output .ML Algorithms should be optimized to give accurate output .

Proposed system will use An Tensor flow library which will performing lot's of numerical calculation more efficiently . There are two way one can classify images either way is collect our own data train our own

model but this process is time consuming and not necessary it will give correct output because training model fetching neural network values is time consuming process.

But this process has remedy and that is pre-train model using pre-train one can archive same results or even more effective results with short period of time . in pre-train model model users just need to give inputs and model gives output . proposed system will use mobile net as pre-train model . Mobile net is one of the most accurate model for image processing it has trained over image-net data set which contain over 1000 of different classes .Both of this system together gives impressive outputs. Proposed system will detect object and if system detect any anonymous activity it will notify it to authorized person.

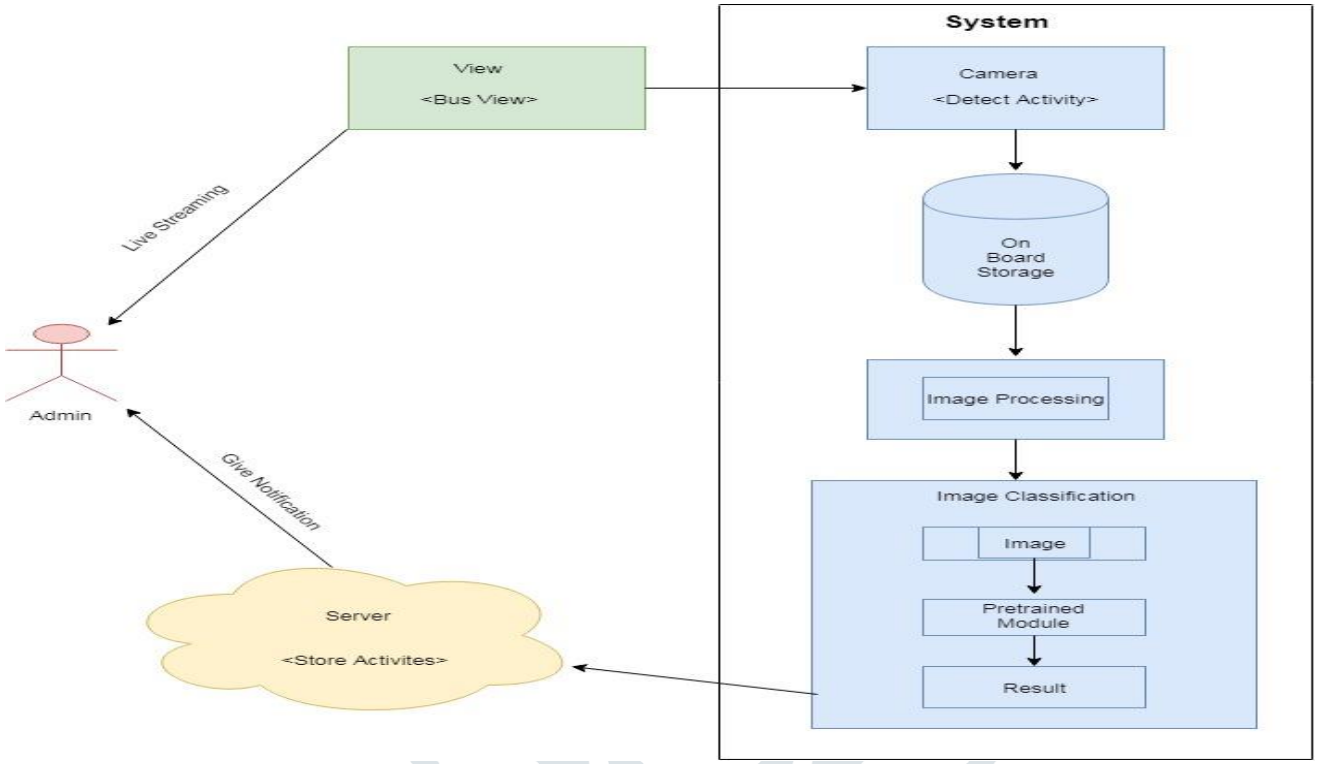

FIGURE 1 : System Architecture

II LITERATURE SURVEY

CHAO YAO 1,2, PENGFEI SUN1 , RUICONG ZHI2 , AND YANFEI SHEN [1]

In this paper, they propose to learn the coexistence features for multi-class object detection. Given an image with multiple class objects, the strong supervision of the region-based annotations are first used as the image label to learn the independent discriminative features for each class. Then, the coexistence relation feature based on the attention mechanism. By combining the independent features and coexistence feature, the classification performance of multi-class object proposals can be consistently improved. Experimental results prove that the proposed end-to-end network outperforms the state-of-the-art object detection approaches, and the learned discriminative features can capture relation to improve classification performance of multi-class objects in the object detection task.

In this paper, they propose to produce a coexistence feature to model contextual relations of multi-class for object detection. Firstly, the discriminative feature for each class in an image are learned based on the attention mechanism. Secondly, a CRN network is utilized to integrate the attention feature of each class into coexistence feature vectors for multi-class object detection. At last, the captured contextual information is connected with the feature vectors from Faster R-CNN, and be exploited to assist the classification of each bounding box proposals. Experimental results prove the efficiency of the designed network, and visualization of learned models also shows

the proposed approach could effectively capture the coexistence relations of multi-class objects. In the future work, we can implement our proposed CRN branch network based on the other state-of-the-arts networks, such as R-FCN [8], SSD [9] and YOLO [10], and ResNet, DenseNet are also able to be utilized as backbone to further verify the performance of our proposed approach in the deeper network and with longer training time.

HANYU WANG1 , PING WANG1 , AND XUEMING QIAN 1,2, (Member, IEEE) [2]

This is a new framework named moving-object proposals generation and prediction framework (MPGP) to reduce the searching space and generate some accurate proposals which can reduce computational cost. In addition they explore the relation of moving regions in feature map of different layers and predict candidates according to the results of previous frames. Last but not least, they utilize spatial-temporal information to strengthen the detection score and further adjust the location of the bounding boxes. Their MPGP framework can be applied to different region-based networks. Experiments on CUHK data set, XJTU data set, and AVSS data set, show that our approach outperforms the state-of-the-art approaches.

Experiments show that the moving-object detection proposals generation and confidence-based proposals prediction are complementary and all contribute to performance improvements. Only use of one of the two parts cannot play the greatest role. In addition, propose a proposals adjustment method, which is also effective to improve the detection results. We refine the results which contributes to the location precision of true positive and the suppression of false positives. Compared to traditional methods and other deep networks, our method shows superior performance in average-precision.

BERNARDO AUGUSTO GODINHO DE OLIVEIRA , (Student Member, IEEE), FLÁVIA MAGALHÃES FREITAS FERREIRA , (Member, IEEE), AND CARLOS AUGUSTO PAIVA DA SILVA MARTINS , (Senior Member, IEEE) [3]

The intrinsic ability of humans to rapidly detect, differentiate, and classify objects allows us to make quick decisions in regards to what we see. Several appliances can make use of fast and lightweight automated object detection for images or videos. Throughout the last five years, the technology industry has constantly introduced computational and hardware solutions, such as devices with impressive processing and storage capabilities. However, object detection methods usually require either high processing power or large storage availability, making it hard for resource constrained devices to perform the detection in real-time without a connection to a powerful server. The model presented in this paper requires only 95 megabytes of storage and took 113 ms in average per image running on a laptop CPU, making it suitable for standalone devices that can be used on the go.

The FLODNet is capable of quickly identifying the objects present in the image by evaluating each region of the image just once. Since it only requires a single pass throughout the image, the proposed network reduces the computational cost of the evaluation, which reduces the hardware requirement, energy consumption and execution time. Some trade-offs on the proposed architecture are expected, since it reduces resources usage. For FLODNet, the drawbacks are mostly related to the aspect ratio and size of the proposed bounding boxes which might not accurately represent the format of the object. The exactness of the positioning of the proposed bounding boxes are weakened in favor of a faster detection, without impairing the object recognition The training techniques presented in this paper, especially when combined, improved the IoU of the bounding boxes proposed by the network. Pre-training the network on a more diverse and bigger dataset allowed the convolution layers to learn more generalized filters, reducing the overfit and improving the IoU while allowing a faster training for new datasets. Although the usage of D.A. increased the training time by adding new samples to the the dataset, the IoU improved greatly. Since the FLODNet uses a bounding-box approach to represent the detected objects, it is limited to objects that can be easily surrounded by a squared polygon. For objects that requires segmentation to obtain an accurate localization, like wires and snakes, the proposed architecture might not work properly [57]. Several appliances can benefit from the FLODNet without being heavily impacted by the positioning of the bounding boxes. Among others, food detection, license plate detection and road sign detection are examples of tasks that do not require high exactness of the bounding box positioning but can greatly make use of the quick detection. For those tasks it is important to achieve real-time detection on low specification devices, justifying the usage of the Eq. 4. Some of those devices may rely on batteries or solar panels, highlighting the importance for low energy consumption. Since the bounding box positioning is the greatest drawback of the FLODNet, a future work can improve the positioning without impacting too much the performance of the detection. It would make the architecture suitable for even more applications, since most of them can make use of a faster detection.

Thomas Blaschke, Bakhtiar Feizizadeh, and Daniel Holbling [4]

The main objective of this research was to establish a semiautomated object-based image analysis (OBIA) methodology for locating landslides. We have detected and delineated landslides within a study area in north-western Iran using normalized difference vegetation index (NDVI), brightness, and textural features derived from satellite imagery (IRS-ID and SPOT-5) in combination with slope and flow direction derivatives from a digital elevation model (DEM) and topographically oriented gray-level cooccurrence matrices (GLCMs). We utilized particular combinations of these information layers to generate objects by applying multiresolution segmentation in a sequence of feature selection and object classification steps. The results were validated by using a landslide inventory database including 109 landslide events. In this study, a combination of these parameters led to a high accuracy of landslide delineation yielding an overall accuracy of 93.07%. Our results confirm the potential of OBIA for accurate delineation of landslides from satellite imagery and, in particular, the ability of OBIA to incorporate heterogeneous parameters such as DEM derivatives and surface texture measures directly in a classification process. The study contributes to the establishment of geographic object-based image analysis (GEOBIA) as a paradigm in remote sensing and geographic information science

The availability of new remote sensing technologies for the detection and mapping of landslides may facilitate the production of landslide maps, as well as the definition of suitable criteria for evaluating the quality of such maps. OBIA offers comprehensive and flexible methods for landslide detection and mapping as it allows the integration of data from different sources, taking into account the most appropriate spectral, spatial, contextual, or textural properties while at the same time reducing the influence of single pixel reflectance.

Bin Xiong, Xiaoqing Ding [5]

A method was developed to detect generic objects using a single query image. The query image could be a typical real image, a virtual image, or even a hand-drawn sketch of the object. Without a training process, the key problem is how to describe the object class from only one query image with no pre-segmentation or other pre-processing procedures. The method introduces densely computed Scale-Invariant Feature Transform (SIFT) as the descriptor to extract “gradient distribution” features of the image. The descriptor emphasizes the edge parts and their distribution structures, which are very representative of the object class, so it is very robust and can deal with virtual images or hand-drawn sketches. Tests on car detection, face detection, and generic object detection demonstrate that the method is effective, robust, and widely applicable. The results using queries of real images compare well with other training-free methods and state-of-the-art training-based methods.

This paper describes a generic object detection approach using a single query image without training. The query image could be a typical real image, a virtual image, or even a hand-drawn sketch of the target object class. The detection process is very similar to “template matching” with the DSIFT used as the descriptor to extract the “gradient distribution” features of the image. Dimensionality reduction is then used to determine the salient characteristics. The Euclidean distance is used as the similarity measurement with a two-step decision approach with a non-maximum suppression algorithm to get the final result. The system was tested for car detection, face detection, and generic object detection

III PRETRAINED MODULE

MOBILENET

- Mobilenet are small, low-latency, low-power models parameterized to meet the resource constraints of a variety of use cases. Mobilenet is an architecture which is more suitable for mobile and embedded based vision applications where there is lack of compute power. This architecture uses depthwise separable convolutions which significantly reduces the number of parameters when compared to the network with normal convolutions with the same depth in the networks. This results in light weight deep neural networks. This architecture was proposed by Google. They can be built upon for classification, detection, embeddings and segmentation similar to how other popular large scale models, such as Inception, are used. MobileNets can be run efficiently on mobile devices with TENSOR FLOW MOBILE.It’s top1 accuracy is 70% and top 5 accuracy is 89%.Also it comes with several versions.

(b) Depthwise Convolutional Filters

DATASET

COCO is a large-scale object detection, segmentation, and captioning dataset. COCO has several features:

- Object segmentation
- Recognition in context
- Superpixel stuff segmentation
- 330K images (>200K labeled)
- 1.5 million object instances
- 80 object categories
- 91 stuff categories
- 5 captions per image
- 250,000 people with keypoints

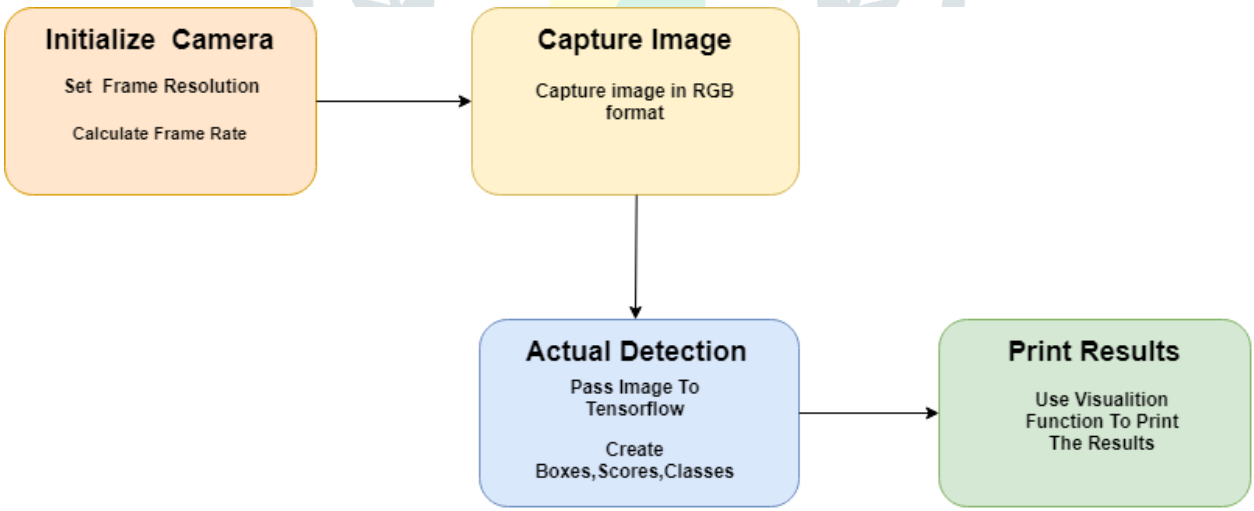

Fig: Process Flow

Initialize Camera:- First we have to initialize the camera as picamera. Then we have to set the frame resolution in height and width the lesser the resolution more is the frame rate per seconds(fps). After setting the frame resolution we have to calculate the frame rate this is the optional part of the code it is not mandatory.

Capture Image:- After initializing the camera we have to capture the actual image from the surrounding. We will capture the image in continuous form as the video. We will capture the image in the Red Green Blue(RGB) format and we will resize it to increase the efficiency of the model. After Capturing the image we will truncate it later on.

Actual Detection:- In this stage we will perform the actual detection part. After capturing the image we will pass it to the tensorflow model for mathematical calculations and computations. After that we will send that data to the pretrained module named mobile net. The pretrained module will perform the classification of the image and then detect the objects and draw the boxes around the objects. Then probability of that object will be displayed with the label name with it. After detection if there are anonymous objects found then the mail with the object name will be sent to the authorized person.

Print Result:-

After image processing and classification the results are printed in new tab with the boxes around the objects and probability with the label of that object

IV RESULT

Condition	Object	Times of Detection	Time Taken
Normal Lights	Fork	9/10	0.90 ms
	Scissors	8/10	0.82 ms
Dark Lights	Fork	5/10	1.02 ms
	Scissors	4/10	1.22 ms
Unstable Condition	Fork	6/10	0.92 ms
	Scissors	5/10	0.88 ms
Confusing Surrounding	Fork	8/10	0.96 ms
	Scissors	7/10	0.89 ms
Multiple Same Objects	Fork	9/10	0.98 ms
	Scissors	9/10	1.08 ms

V CONCLUSION

Proposed system is able to detect and notify in case any anonymous activity happen . we will be able to add one more layer in security using this system.The authorized person will get a mail when the particular activity or object is detected by the system

VI References

[1] Thomas Blaschke, Bakhtiar Feizizadeh, and Daniel H¨olbling "Object-Based Image Analysis and Digital Terrain Analysis for Locating Landslides in the Urmia Lake Basin, Iran" DECEMBER 2014

[2] Bin Xiong, Xiaoqing Ding "A Generic Object Detection Using a Single Query Image Without Training*"April 2012

[3] BERNARDO AUGUSTO GODINHO DE OLIVEIRA , (Student Member, IEEE), FLÁVIA MAGALHÃES FREITAS FERREIRA , (Member, IEEE), AND CARLOS AUGUSTO PAIVA DA SILVA MARTINS , (Senior Member, IEEE)"Fast and Lightweight Object Detection Network: Detection and Recognition on Resource Constrained Devices "February 5, 2018,.

[4] CHAO YAO 1,2, PENGFEI SUN1, RUICONG ZHI2, AND YANFEI SHEN3,4 "Learning Coexistence Discriminative Features for Multi-Class Object Detection" July 4, 2018

[5] HANYU WANG1, PING WANG1, AND XUEMING QIAN 1,2, (Member, IEEE)"MPNET: An End-to-End Deep Neural Network for Object Detection in Surveillance Video" May 24, 2018,