iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.43989

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

People Counting in Crowd: Faster R-CNN

Saravana Kumar[1], Kranthi Kumar[2], M. Vishnu Menon[3], K. Anirudh Reddy[4], B. Mahender Yadav[5], B. M. N. Sai Pavan[6]
[1,2]*Associate Professor,* [3,4,5,6] *Students, Department of Information Technology,*
Sreenidhi institute of Science and Technology, India

Abstract-Because of its vast range of operations, people counting in crowds is a significant challenge in the field of computer vision. To achieve further dependable results of crowd counting, head discovery grounded ways are used rather than viscosity chart grounded crowd counting ways. This is because, in case of viscosity charts, it isn't always the correct position which contributes to final crowd count. This leads to unreliable results especially in case of false positives. This makes the entire task of head discovery in crowded scenes a grueling one to be solved. While face discovery has reached maturity, the more general task of chancing people in images and videotape still remains to be veritably challenging. Count of people may also be demanded for the statistical purposes which help to concoct marketing strategies or it may be used for crowd control in colorful situations image processing is a fashion for applying operations on an image in order to ameliorate it or prize precious information from it. In our design, input to the system is an image/ videotape of the surveillance system which further divided into image frames. Our proposed system gives the count of people in the scene using the Faster R- CNN object discovery algorithm.

I. INTRODUCTION

A. Motivation

Concerts, political speeches, rallies, marathons, and stadiums are all examples of circumstances where crowds occur. Crowd counting, also known as density estimate, assists in crowd management for safety and surveillance purposes, such as law enforcement officer deployment and the detection of unusual behavior. It may also be used to determine the number of commuters, which is crucial for the construction of public transportation infrastructure. It may also be used to assess the political significance of demonstrations or protests, as different estimations are frequently given for the same event. And, because counting through turnstiles or by humans is not always practicable or practical, To estimate the number of individuals in dense crowds, use computer vision-based techniques.

B. Problem Statement

In a congested atmosphere, counting individuals has become a need for their safety. This count is essential in retail since it helps you to see when the business is busiest and plan accordingly. When a firm or government wants to know how crowded an area is at any given time, crowd counts is essential for infrastructure planning. It's also used to avoid congestion and create an escape route. With malls and subways 4 becoming increasingly popular, it's vital to estimate how many people will visit. There are a number of automatic detection techniques available. Although methods have been found to address this problem, achieving sufficient accuracy and counting efficiently remains a difficulty.

C. Objectives

The study's main goal is to use head detection to count the number of persons present. It aids in the correct evaluation of crowd count statistics . The system is linked to a simple graphical interface that allows users to complete jobs, keep track of persons by taking action against overcrowding, and aid with evacuation routes. The project's purpose is to help users arrange their resources by giving accurate counting data. This graphic may be used to keep track of how many people are attending an event. It might also be used to keep track of traffic and make marketing decisions.

II. LITERATURE SURVEY

A. Existing System

To determine the number of persons in a group, the Jacob method was utilized. By counting the number of individuals in the grid squares and increasing the number of squares, they attempted to estimate the extent of the demonstrations. It only works when you have a grid pattern, and if the grid is not convenient, it is sufficient to estimate the total area and density of your collection. It is difficult to estimate a large number.

For counting individuals, there are two sorts of deep learning approaches: regression-based methods and detection-based methods Individuals are tallied in regression-based systems by building a relationship between crowd size and picture attributes.

In detection-based systems, a density map is generated from images and density maps are combined to determine the number of individuals.

Zhang proposed a multi column convolutional neural network (MCNN). Shunqiang Liu used a CNN based feature to count crowds. .The first step is to develop a study template that can be used to create a density map for each image. We use the convolution with the help of Gaussian nuclei to create a density map and then model it to get the number of objects when combined The next step is to drag the entire convolution line to turn the image into a density map and then combine them to determine the number of objects.

U-Net and Full Convolutional Regression Network (FCRN) were taken into account when designing FCN. U-Net is a widely used FCN for segmenting biological data images. The embedded image is processed by the spastic layer block followed by a joint layer (lower modeling). The same procedure is repeated numerous times, putting the following block at risk. On the network, significant parts of the input picture are compressed (and encoded). The essential bits of an input picture are compressed (and encoded) by the network.

The second part of U-Net employs up sampling instead of pooling layers to ensure that the output dimensions match the input visuals. the fully Convolutional Regression Network was offered as a solution (FCRN). The structure is similar to that of the U-Net.. In the up-sampling phase of the downsampling section, information from higher resolution levels is not directly transferred to equivalent layers, which is an important distinction. FCRN-A and FCRN-B are two networks proposed in the study, each with distinct down sampling intensities. All convolutional layers are pooled in FCRN-A, but not in FCRN-B.

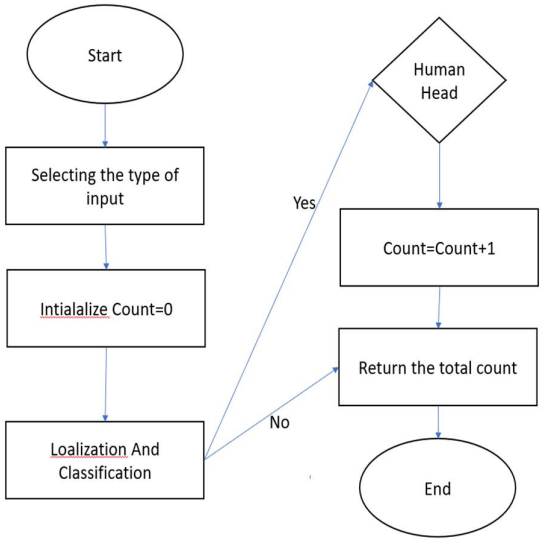
B. Limitation Of Existing System

These methods function poorly in low-density circumstances because the count is underestimated. Second, these systems are incapable of detecting the presence of a human head. Of the head counting methods mentioned above, most algorithms use Ada Boost and LBP-based solid state drives to determine the head for previous extraction. SSDs have lower computing performance compared to faster R-CNN. Due to their high sensitivity to light, these methods are particularly vulnerable to limitations or situations. If the image is of poor quality, the SSD will not give reliable results.

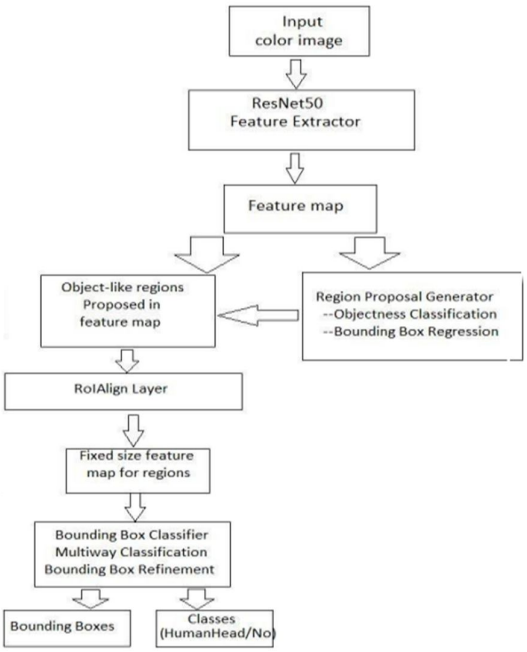
C. Proposed System

The introduced system uses the fastest R-CNN object reconstitution system to identify human heads in the plural and provide the total number. The ideal method is to utilise a three-button graphical user interface (GUI). Users can provide both static and dynamic input while utilising these.

the fastest R-CNN object reconstitution system to identify human heads in the plural and provide the total number. The best way is to use a graphical user interface (GUI) with three input buttons.



The initial stage in our procedure is to remove marks. An extracted primary map is published after the image is passed to a deep convolutional neural network. These characteristics will be discussed in detail in the coming volumes The main feature of our model is based on the restoration of the production network. The core Resnet-50 network in our model is divided into two parts: first, it contains layers from conv1 to conv4x, second, it contains layers from conv5x and above, and it removes the final regression properties and classification suggestions.



III. APPLIED TECHNOLOGIES

A. Python

Python is an interactive programming language. Python does not have a compiler, so the compiler executes the code directly at runtime. This is similar to how JS and PHP work. Python is an object-oriented programming language that makes extensive use of technology and programming.

B. NumPy

NumPy is a Python module that uses arrays to perform a wide range of functions. To ensure faster execution, it can be implemented using the C and C ++ libraries. It works faster with the NumPy module than similar performance without it, it will be not added with python must be installed using pip numpy command.

C. Open CV

Open-CV module may be a Python library that is viable for managing the images and recordings. it's unambiguously supposed for the image analysis and takes the images in sort of NumPy arrays. the images square measure taken as 3d array wherever every array addresses the BGR esteems singly. various tasks like stacking photos, winnow picture element esteems and doing any kind of tasks like applying channels, resizing and any remaining things is handily finished utilizing Open CV.

D. Tkinter

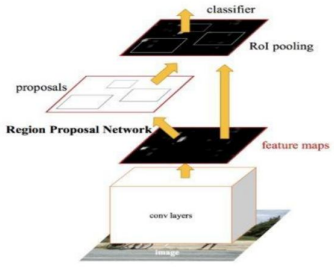
Python offers a lot of features that can be used to build user interfaces (graphical user interface). Tkinter is the most widely used of all interface systems. This is the standard Python community that emphasizes the emotional connection between the Tk and Python interfaces. Python with tkinter is a quick and easy way to create interface programs. Creating a tkinter interface may seem like a simple process.

E. Torchvision

Torchvision is a library of computer vision that is closely related to PyTorch. It includes tools for efficient image and video conversion, one or two commonly used predictions, and some data sets (not included with Torchvision ,Pythorch must be provided separately).The Torchvision cluster includes computer vision databases, design samples, and simple image transfers. His progressive thinking must be directly implemented.

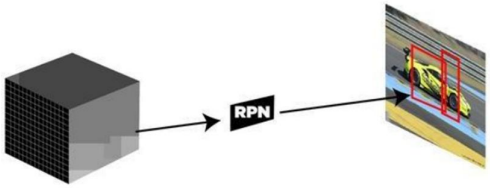
IV. ALGORITHM

The piecemeal technique of the programme or its implementation is shown by the algorithm. In our planned system, we've got used quicker R-CNN Object Detection algorithmic program. This algorithmic program doesn't use any selective search. quicker R- CNN design consists of 2 networks. 1. Region Proposal Network 2. Object Detection Network CNN is the network's backbone, and it's shared by the Region Proposal Network and therefore the Object Detection Network.



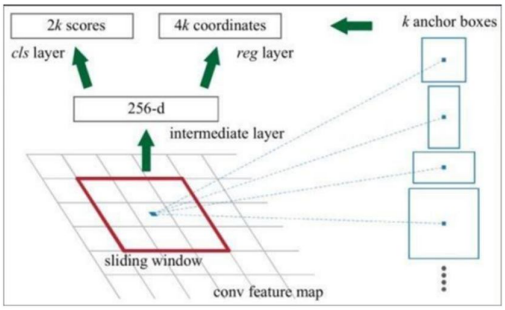
A. Region Proposal Network

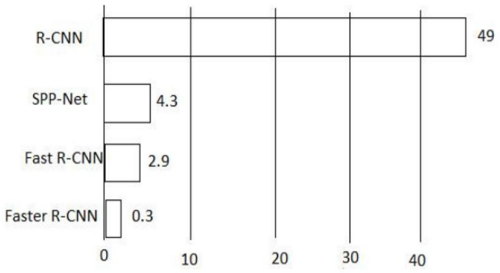
The backbone layer generates a convolution feature map, that RPN uses to come up with anchors victimization window convolution applied to the input feature map.



B. Anchors

We get the most range of k- anchor boxes for every window. for every of the image's varied sliding positions, the default price of k=9(3 scales of (128*128, 256*256, 512*512) 3 facet ratios of (1:1, 1:2, 2:1) be used. for every convolution feature map of W * H, N = W* H* k anchor boxes square measure generated. These region proposals square measure then transmitted through AN intermediate layer with 3*3 convolution and one artifact, still as 256 or 512 output channels (for ZF or VGG-16). The output of this layer is sent to two 1*1 convolution, classification, and regression layers. The regression layer's output parameters are 4*N (W * H * (4*k)) (denoting the coordinates of bounding boxes) and therefore the classification layer has 2*N (W * H *(2*K)) output parameters.





The above diagram shows the prediction time for giving the output when an image is given to it. R-CNN takes high amount of time when compared with Faster R:CNN.

C. ROI Pooling

We send the region proposal's output to the Roi pooling layer, that serves a similar purpose as quick R-CNN converts RPN region suggestions of various sizes into a fixed-size feature map. during this article, we tend to went through RoI pooling in nice detail. This RoI pooling layer's output has a size of 23 (7*7*D) (where D =256 for ZF).

D. Softmax and Bounding Box Regression Layer

The RoI pooled feature map of size (7 * seven * D) is then transmitted to 2 absolutely connected layers, that flatten the feature maps and send the output to two parallel, absolutely connected layers, every with a definite task assigned to it: the primary layer employs a SoftMax layer with N+1 output parameters to predict the objects within the region proposal (N is that the variety of sophistication labels and background). A bounding box regression layer with four* N output parameters might be used as the second layer. The bounding box position of the item in the picture is regressed using this layer.

V. OUTPUT SCREENS

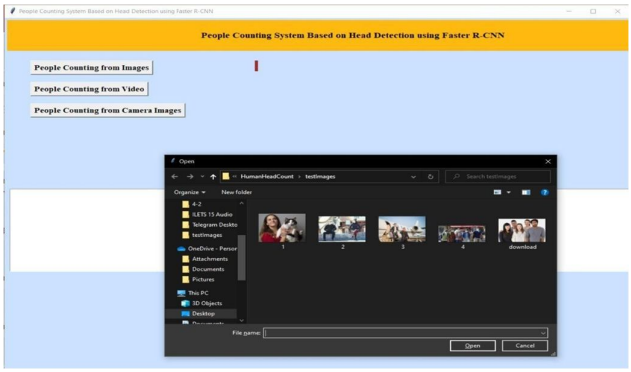


Fig 5.1:Picture Uploading

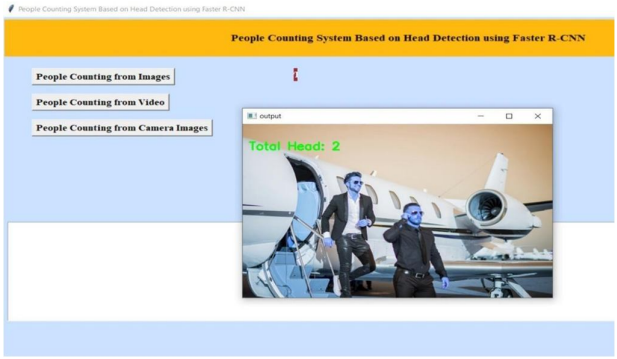


Fig 5.2:Output of selected picture

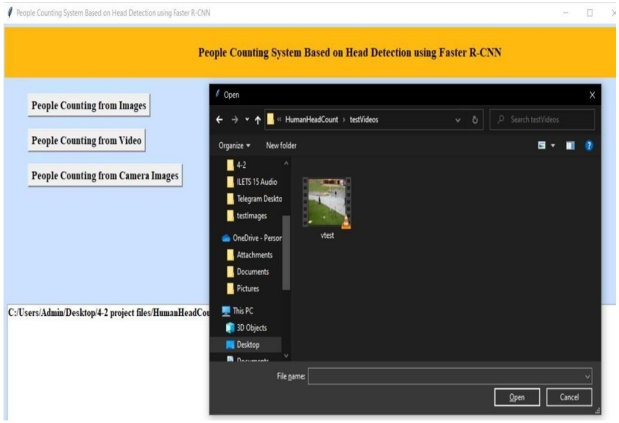


Fig 5.3: Video Uploading

Fig 5.4:Output from video 1

Fig 5.5:output from video 1(2)

VI. TEST CASES

TEST ID	TEST CASE SCENARIO	TEST CASE	PRE CONDITION	TEST STEP	TEST DATA	EXPECTED RESULT	STATUS
TEST_1	Object detection from video	Detect the object	Need a video with minimum clarity	Record the video Find the confidence of object detection	Valid If the object is classified accurate	Yes	Pass
TEST_2	Object detection from video	Detect the object	Need a video with minimum clarity	Record the video Find the confidence of object detection	InValid If the object isn't classified accurate	NO	Pass
TEST_3	Object detection from image	Detect the object	Need a Image with minimum clarity	Upload the image Find the confidence of object detection	Valid If the object is classified accurate	Yes	Pass
TEST_4	Object detection from image	Detect the object	Need a Image with minimum clarity	Upload the image Find the confidence of object detection	InValid If the object isn't classified accurate	NO	Pass

VII. CONCLUSION

We suggested a system for counting individuals in crowded movies based on human head recognition. Its detection accuracy has increased while the time it takes to classify has dropped. We did experimental evaluations on the data set. The results showed significant progress. People counting systems are in high demand in the travel and transportation industry to improve

passenger management systems. Our proposed method can be used in a variety of nations to manage the personnel count in workplaces in compliance with new regulatory standards. Many businesses are turning to advanced people counting technologies to keep track of employee data and provide protection in the event of a COVID-19 emergency.

REFERENCES

[1] Austin Choi-Fitzpatrick and Tautvydas Juskauskas. “Up in the Air: Applying the Jacobs Crowd Formula to Drone Imagery”. In: Procedia Engineering 107 (Dec. 2015), pp. 273–281. DOI: 10.1016/j.proeng. 2015.06.082.

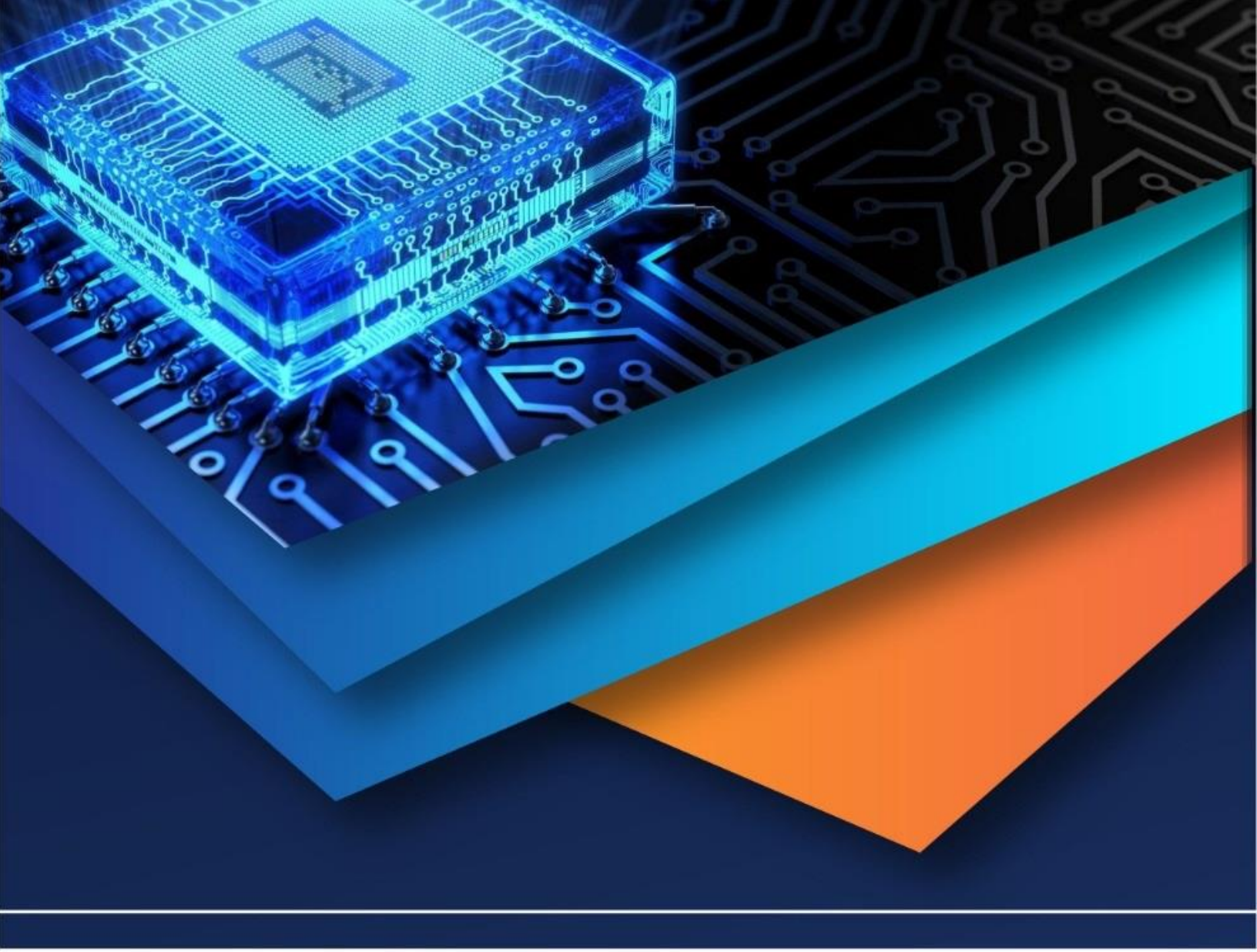
[2] Dan Kong, Douglas Gray, and Hai Tao. “A viewpoint invariant ap- proach for crowd counting”. In: 18th International Conference on Pattern Recognition (ICPR’06). Vol. 3. IEEE. 2006, pp. 1187–1190.

[3] Lu Zhang, Miaojing Shi, and Qiaobo Chen. “Crowd counting via scale- adaptive convolutional neural network”. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE. 2018, pp. 1113– 1121.

[4] Shunqiang Liu, Sulan Zhai, Chenglong Li, and Jin Tang. “An effective approach to crowd counting with CNN-based statistical features”. In: Sept. 2017, pp. 1–5. DOI: 10.1109/ISC2.2017.8090827.

[5] Dumitru Erhan, Christian Szegedy, Alexander Toshev, and Dragomir Anguelov. “Scalable object detection using deep neural networks”. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014, pp. 2147–2154.

[6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. “Ssd: Single shot multibox detector”. In: European conference on computer vision. Springer. 2016,pp. 21–37.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY