IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 9, Issue 12, December 2020

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.512

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 12, December 2020||

The Multi Object Detection method based on Deep Learning using YOLO (You Only Look Once)

Vaishnavi Mandlik

Research Student, Department of Information Technology, B.K.Birla College of Arts, Science and Commerce (Autonomous), Kalyan, Maharashtra, India

ABSTRACT: The object detection approach is based on deep learning in this paper. Classicalobject detection methods which is based on regression models are introduced, and their ability to detect objects are analyzed. The multi-view object detection method is proposed to improve the working performance of these methods, the structure of the model, working principle of these approach are interpreted. The ability for object retrieval and the object detection accuracy of multi-view method and classical method are evaluated and compared based on a test of object datasets. The capability of object retrieval shows in experimental result, Multi-view YOLOv3 (You Only Look Once, Updated version of YOLO) a Real Time Object Detection, Multi-view YOLOv5 (updated version of YOLO) achieves Average score, which is higher than the classical method respectively. When difficult objects are not corporated, the mAP (mean average precision) the score of these multi-view methods are much higher than classical methods. The multi-view object detection methods are much faster while achieving the accuracy mAP (mean average precision).

KEYWORDS:multi-view; object detection; deep learning.

I. INTRODUCTION

Object detection is a computer vision technique that represents an important focus of research in the field of computer vision [1]. Domains of object detection which is well researched includes face detection, pedestrian detection, image retrieval, video surveillance, robotics, and driverless cars [2-5]. Traditional object detection methods are based on establishing mathematical models; such asHistogram of Oriented Gradients (HOG) [6], the frame-difference method [7], the background subtraction method [8], the optical flow method [9], the sliding window model method and the deformable part model method [10]. The first four methods operates on feature extraction and mathematical modelling which utilizes the features of the data to build a mathematical model and after solving the model it obtains the object detection results, where the other two methods operate on feature extraction and classification modelling, which together combines hand-crafted features (Viola-Jones object detection framework based of Haarfeatures [11], Scale-Invariant feature transform (SIFT) [12], Histogram of oriented gradient (HOG) features [13]) with classifier such as Support Vector Machine (SVM) [14] or AdaBoost[15] to obtain the results by classifying the features. Deep neural network can automatically learn different features, because deep learning techniques have transformed the field of object detection by improving the object detection robustness and accuracy. Based on deep learning, the models which are used in object detection methods are subcategorized into detection models based on region proposals and based on regression.Deep-learning based models for object detection based on regression, it divides the feature maps and generate the relationships among the bounding boxes, the default bounding boxes and ground truths for training, such as YOLO (You Only Look Once) [16]. The regression method achieves better real-time performance, but the accuracy of regression method is poorer than the region method. If both the accuracy and real-time performance are taken into consideration, then the regression method will give great performance. This paper mainly proposes with the aim to improve the robustness and performance of Regression based model.

II. RELATED WORK

Earliest methods to achieve detection results with high accuracy was the cascaded detector[1]. This method has been used to implement car detection. Efficient Haar features was proposed with alternatively recourse to images[2]. The aggregate channel features made possible to compute HOG features at about 100 frames per second[6]. The limitation of detector cascades is, it is difficult to implement multiclass detectors under this proposed method. An attempt to leverage the success of deep neural networks for object classification was proposed by R-CNN detector[3]. This combines an object proposal mechanism and a CNN classifier [7]. While the R-CNN excel previous detectors by a

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 12, December 2020||

large margin, It's speed is limited by the need for object proposal generated and repeated CNN evaluation [4] has shown that it can improve with recourse to spatial pyramid pooling (SPP) [12], which allows the computation of CNN features once per image, increasing the detection speed by an order of magnitude. Building on SPP, the Fast R-CNN [11] introduced the ideas of back propogation through the ROI pooling layer and multi-task learning of another interesting work is YOLO [15], which outputs object detection within a 7 x 77 x 7 grid. This network runs at 40 frames per second, but with some compromise of detection accuracy.

III. METHODOLOGY

YOLO is a single neural network that performs object region detection and object classification. Without dividing the detection process into different stages, YOLO achieves end-to-end object detection. SSD is a Single Shot Multibox Detector, which combines the regression approach of YOLO with the anchor mechanism of Faster R-CNN. The anchor mechanism is useful for extracting features at different ratios and different scales. On the other hand, regression approach can reduce the complexity of the neural network to improve the performance in real-time. The feature extraction method of SSD is more effective than the feature extraction method of YOLO regarding object detection. The feature representations is related to different scales are different, a multi-scale feature extraction method is applied in SSD which enhances the robustness of object detection at different scales. YOLO, it divides each image into a fixed grid and there will be limited number of detected objects. For eg. Considering a grid a scale of SxS, where S=7. Here each grid yields 2 predicted bounding boxes, it means there will be 98 predicted bounding boxes are observed in a single detection, so it means that not more than 98 objects can be detected at one time. When two or more class objects are located in the same grid cell, they cannot be identified simultaneously. For eg. If the image scale input is 446x446, then not more than two objects can be identified simultaneously in a single 64x64 grid scale, and their must be the same classes. YOLO has relatively less capability to detect small objects, because YOLO, if once the image is passed through twenty four convolutional layers and four pooling layers, less information can be observed in the resulting feature map.

SSD is a multi-scale approach, the feature maps are used to detect different objects at different scale. Because at the same level, the feature map is produced from convolution results. The convolution open fieldsof the different levels, it must be different in size. The open field of high-level convolution layer is larger than the lower layers, and the extracted information which is related to a high-level feature layer is more abstract. If there will be more abstract feature extraction information, then there will be less detailed information.

The formula for calculating the convolution receptive field is as follows,

$$SRF(i) = (SRF(I - 1) - 1)Ns + Sf \quad (1)$$

Where SRF(i) is the size of the convolution receptive field of the i-th layer, Ns is the step length, and Sf is the size of the filter.

SSD takes on the anchor mechanism, for which the results which are detected are directly related to the size of the bounding box which is produced from the anchors. The ratio of the bounding box can be calculated according to the formula for the input image. The relationship between the bounding box and the image input can be used to examine the region mapped to the image input.

The mapping of the default bounding box coordinates to the original image coordinates on the feature map is as follows:

$$Xmin = Cx - Wb/2 \,/\, Wfeature \times Wimg = (i+0.5/\,|Fk|- Wk/2)\, Wimg \quad (2)$$

$$Ymin = Cy - Hb/2 \,/\, Hfeature \times Himg = (j+0.5/\,|\,Fk\,|\,-Hk/2)\, Himg \quad (3)$$

$$Xmax = Cx + Wb/2 \,/\, Wfeature \times Wimg = (i+0.5|Fk| + Wk/2)\, Wimg \quad (4)$$

$$Ymax = Cy + Hb/2 \,/\, Hfeature \times Himg = (j+0.5\,|Fk| + Hk/2)\, Himg \quad (5)$$

Where *(Cx,Cy)* , it denotes the centre coordinates of the bounding box, *Hb* is the height of the bounding box, *Wb* is the width of the bounding box,*Hfeature* is the height of the feature map, *Wfeature* is the width of the feature map, |*fk*| is the size in the *k-th* feature map, *Himg* is the height of the original image, *Wimg* is the width of the original image, and

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 12, December 2020||

(Xmin, Ymin, Xmax, Ymax) it denotes the mapping coordinates of the bounding box, centred at*(i+0.5|Fk|), (j+0.5 |Fk|)* and scaled to a height of *Hk* and the width of *Wk* in the *k-th* feature map.

IV. EXPERIMENTAL RESULTS

Inobject detection, the detection results are expected to have high classification and location accuracy. The classification accuracy can be measured using confidence of the predicted bounding boxes, and the location accuracy based on coordinate information of the predicted bounding boxes. The proposed paper's main aim is to improve the object detection performance. The object detection performance will be compared between the Multi-view YOLO and classical YOLO methods and between the Multi-view SSD and classical SSD methods to verify the effectiveness of this approach.

The number of images containing objects were selected based on object scale.The images contains ground-truth objects. So, the comparison is between Multi-view YOLO and classical YOLO methods and between the Multi-view SSD and classical SSD methods. An updated version of YOLOv5 has a detection speed of 140 frames per second and mean Average Precision(mAP) of 84.7%. Experiments were conducted on YOLOv3 [4], YOLOv5 [8] , and SSD to obtain a very high set of comparison results. The objects classes include chair, monitor, person, and car. The YOLOv3 detection threshold was set to 0.7, and the threshold detection for YOLOv5 was set to 0.9. The detection results are shown in **Figure 1 and Figure 2**. The characteristics after compared with the classical methods.

1. The multi-view object detection method, it detects more objects.
2. They achieve more amount of confidence when recognizing the same object.
3. They identify objects correctly, that are identified incorrectly by the classical methods.

Multi-view YOLOv3, YOLOv5, and SSD, it detects all the objects in the images successfully. The results show that multi-view methods achieve better performance in object detection.

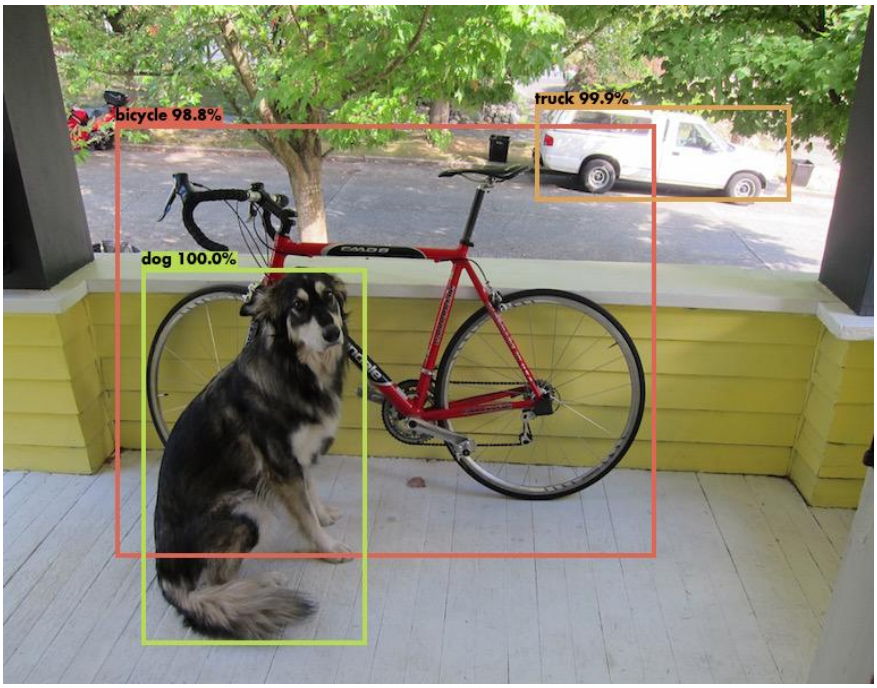

Figure 1. Comparison of object detection results between the classical (You Only Look Once) YOLOv3 method and Multi-view YOLOv3 method shown in figure.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 12, December 2020||

Figure 2. Comparison of object detection results between the classical (You Only Look Once) YOLOv5 method and Multi-view YOLOv5 method is shown in the figure.

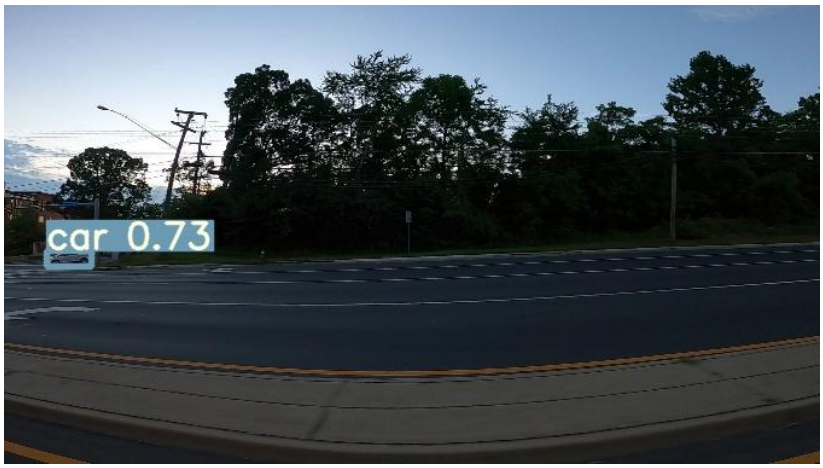

Figure 3.

Figure 4.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 12, December 2020||

Figure 5.

Figure 6.

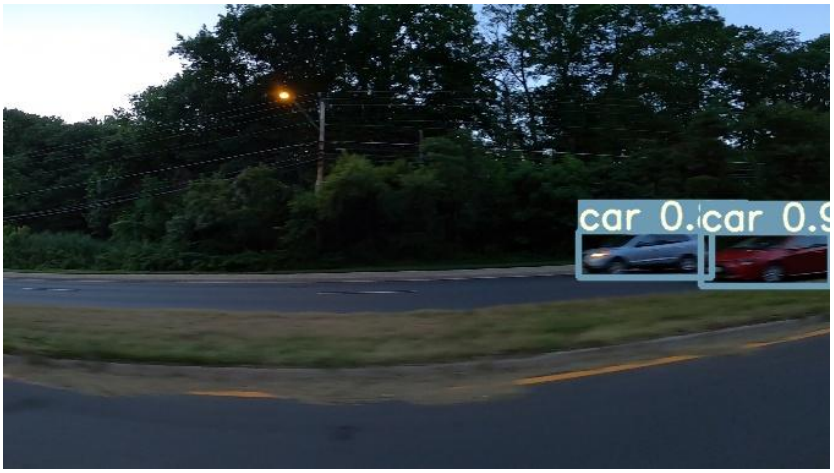

Figure 7.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 12, December 2020||

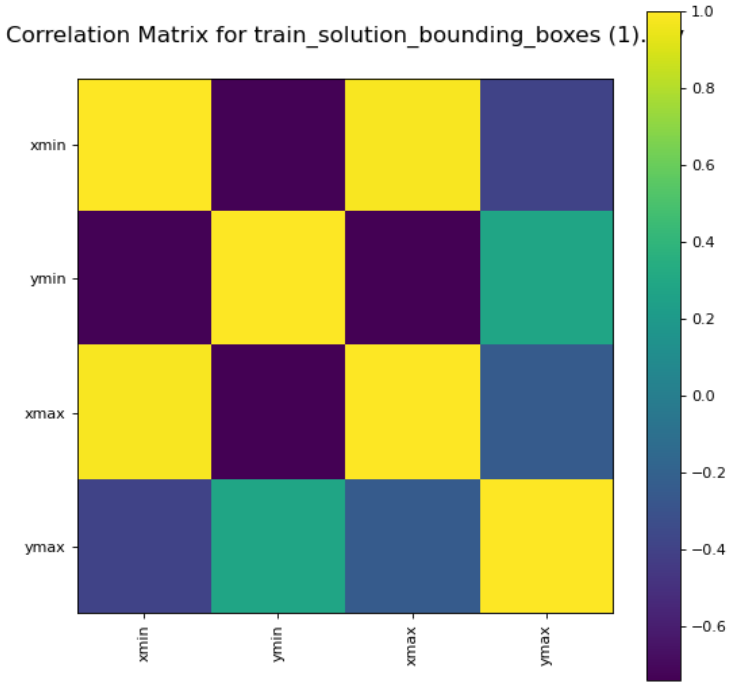

Figure 8. Co-relation matrix for train solution bounding boxes.

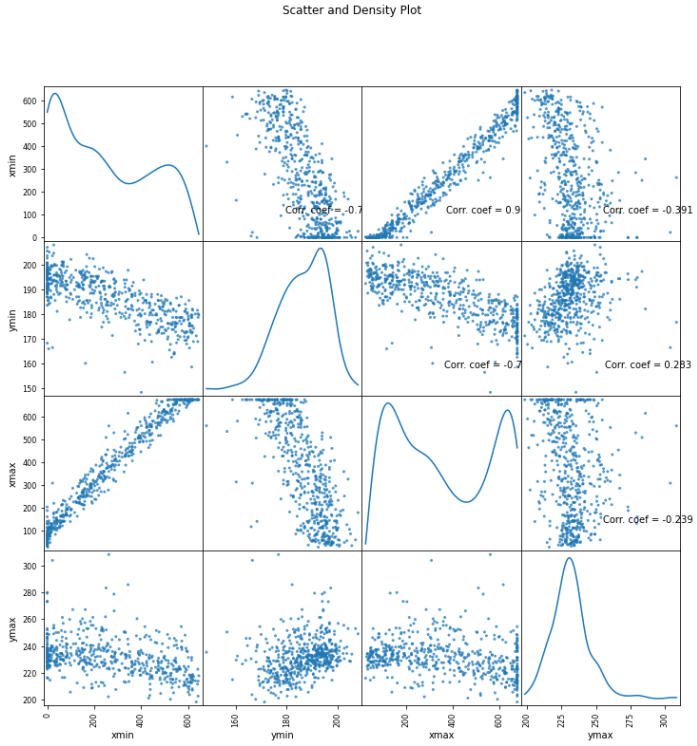

Figure 9. Density Plot

The retrieval abilities of the multi-view method achieves more accuracy than the classical methods. After calculating the average score results for all the images, the Average score of multi-view YOLOv3 and classical YOLOv3 is 0.84

and 0.72, respectively. The Average score of multi-view YOLOv5 and classical YOLOv5 is 0.91 and 0.77, respectively.

In the object detection, the mAP (mean Average Precision) is taken on for evaluating the detection accuracy for all the images. The mean Average Precision (mAP) values of the Multi-view methods and classical methods are calculated with the confidence threshold for YOLOv3 and YOLOv5 is set to 0.7 and 0.9 respectively. The results are shown in **Table 1.**

Table 1. Object Detection results based on regression method.

Method	mAP (%)	AP (%)			
		chair	monitor	person	car
YOLOv3	74.6	57.7	61.6	67.1	69.1
Multi-view YOLOv3	84.2	74.6	73.2	71.5	77.2
YOLOv5	78.7	65.2	68.5	78.3	77.3
Multi-view YOLOv5	96.4	84.3	94.2	84.7	91.1

The above **Table 1.** shows the mAP (mean Average Precision) scores of the multi-view methods which are higher than the classical methods. The mAP of Multi- view YOLOv3 is 84.2% which is higher than the classical YOLOv3 by 74.6%; the mAP of Multi-view YOLOv5 is 96.4% which is higher than the classical YOLOv5 by 78.7% respectively. Hence, the multi-view methods achieves the higher accuracy than the classical methods.

V. CONCLUSION

This paper proposes a multi-view object detection approach. Experiments were performed using a real world object datasets. The experimental object detection results show that the capability of object retrieval is higher in Multi-view YOLOv3 and YOLOv5, and also it achieves higher Average score than the classical methods. The final results shows that the Multi-view object detection and retrieval capability are much more higher than the classical object detection method. The applicability of the proposed multi-view detection method is not restricted to YOLO, it can be applied in combination with some other regression based detection models. It will further improve the robustness and performance of the regression based detection models. The proposed approach multi-view method, it gives an excellent results for improving the object detection capabilities of regression based detection models. In future work, the overall image information will be combined with multi-view method which will enhance the robustness and performance of the model.

VI.ACKNOWLEDGEMENT

I would like to thank Prof. Swapna Augustine Nikale, Department of Information Technology, B.K.Birla College Kalyan for guiding throughout the research work.

REFERENCES

[1] Lu, J., Ma, C., Li, L., Xing, X., Zhang, Y., Wang, Z., & Xu, J. (2018). A Vehicle Detection Method for Aerial Image Based on YOLO. *Journal of Computer and Communications*, *06*(11), 98–107. https://doi.org/10.4236/jcc.2018.611009

[2] Peng, H., & Chen, S. (2019). BDNN: Binary convolution neural networks for fast object detection. *Pattern Recognition Letters*, *125*, 91–97. https://doi.org/10.1016/j.patrec.2019.03.026

[3] Wu, Y., Wang, Y., Zhang, S., &Ogai, H. (2020). Deep 3D object detection networks using LiDAR Data: A Review. *IEEE Sensors Journal*, 1. https://doi.org/10.1109/jsen.2020.3020626

[4] H C, D. (2020). An Overview of You Only Look Once: Unified, Real-Time Object Detection. *International Journal for Research in Applied Science and Engineering Technology*, *8*(6), 607–609. https://doi.org/10.22214/ijraset.2020.6098

[5] Jaehn, B., Lindner, P., &Wanielik, G. (2015). Multi-view point cloud fusion for LiDAR based cooperative environment detection. *Advances in Radio Science*, *13*, 209–215. https://doi.org/10.5194/ars-13-209-2015

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 12, December 2020||

[6] Yang, Y., Chen, F., Wu, F., Zeng, D., Ji, Y.-, & Jing, X.-Y. (2020). Multi-view semantic learning network for point cloud based 3D object detection. *Neurocomputing*, *397*, 477–485. https://doi.org/10.1016/j.neucom.2019.10.116

[7] Vural, S., Mae, Y., Uvet, H., & Arai, T. (2012). Multi-view fast object detection by using extended haar filters in uncontrolled environments. *Pattern Recognition Letters*, *33*(2), 126–133. https://doi.org/10.1016/j.patrec.2011.10.004

[8] LU, W.-H., LI, Y.-L., WANG, S.-J., & DING, X.-Q. (2012). Improvements of 3D Object Detection with Part-based Models. *Acta AutomaticaSinica*, *38*(4), 497–506. https://doi.org/10.3724/sp.j.1004.2012.00497

[9] Kim, J., & Cho, J. (2019). YOLO-based Real-Time Object Detection Scheme Combining RGB Image with LiDAR Point Cloud. *The Journal of Korean Institute of Information Technology*, *17*(8), 93–105. https://doi.org/10.14801/jkiit.2019.17.8.93

[10] Srazhdinova, A., Ahmetova, A. '., &Anvarov, S. (2020). Detection and tracking people in real-time with YOLO object detector. *Challenges of Science*, 69–75. https://doi.org/10.31643/2020.010

[11] Fang, W., Wang, L., & Ren, P. (2020). Tinier-YOLO: A Real-Time Object Detection Method for Constrained Environments. *IEEE Access*, *8*, 1935–1944. https://doi.org/10.1109/access.2019.2961959

[12] Object Detection Method Based on YOLOv3 using Deep Learning Networks. (2019). *International Journal of Innovative Technology and Exploring Engineering*, *9*(1), 1414–1417. https://doi.org/10.35940/ijitee.a4121.119119

[13] Changgeng Yu, Liu, K., & Zou, W. (2020). A Method of Small Object Detection Based on Improved Deep Learning. *Optical Memory and Neural Networks*, *29*(2), 69–76. https://doi.org/10.3103/s1060992x2002006x

[14] John, P. J. (2019). Review Paper on Object Detection using Deep Learning- Understanding different Algorithms and Models to Design Effective Object Detection Network. *International Journal for Research in Applied Science and Engineering Technology*, *7*(3), 1684–1689. https://doi.org/10.22214/ijraset.2019.3313

[15] Qu, H., Zhang, L., Wu, X., He, X., Hu, X., & Wen, X. (2019). Multiscale Object Detection in Infrared Streetscape Images Based on Deep Learning and Instance Level Data Augmentation. *Applied Sciences*, *9*(3), 565. https://doi.org/10.3390/app9030565

[16] Xiao, F., Deng, W., Peng, L., Cao, C., Hu, K., & Gao, X. (2018). Multi-scale deep neural network for salient object detection. *IET Image Processing*, *12*(11), 2036–2041. https://doi.org/10.1049/iet-ipr.2018.5631

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.512

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

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