Real Time Object Detection Using SSD For Bank Security

Ruhin Mary Saji, Sobhana N V
Department of Computer Science and Engineering, Rajiv Gandhi Institute of Science and Technology, Kottayam, Kerala.
E-mail: ruhimarysji@gmail.com

Abstract. Object detection in real time applications are more challenging and thus faces more difficulties in designing object detection applications. Object detection is a method of identifying predefined objects from images, videos or from other input sources. It consist of different steps for identifying the object from its surroundings. Earlier detection system uses different techniques for objection detection with computationally demanding networks that requires high GPU and needs more power. Considering the current circumstances in the society, it is better to design real time object detection system for security applications. The newly evolved technique such as Tensorflow(TF) along with OpenCV can be used used for object detection. The CCTV footages can be used as input and the monitoring of it can be done inorder to find any security violations. The proposed real time object detection system for bank security mainly aims in providing security to bank vaults. This system can also be used for other security applications where human entry is restricted. The proposed model can be used for different security related applications effectively.

Keywords: SSD, MQTT, OpenCV, TF, TSP, DNN, YOLO, IoU

1. Introduction
Object detection is considered as a method for choosing the instances of real world objects such as face, buildings, bicycles etc in images or videos. It allows detection, recognition and localization of multiple objects within an image and provides a much understanding of images as a whole. The different object detection applications include face recognition, self-driving cars, security related aspects such as face unlock in mobile phones, industry etc.

Many of the object detection application gains its advantage in real time situations. One of the application is the face recognition used in our mobile phone cameras. While taking pictures, it usually shows a bounding box around the faces. Also the security in mobile phones is provided using face lock - unlock mechanism. In that application, the owner has to save the image of his/her face. After saving the image, if the saved face is shown in mobile phone front camera, then it will check the features of current input image with the already saved image. If it found matching, then the phone will get unlocked. If no match found, then it is not possible to unlock it. This method is effective for aged people if they are not remembering the unlock patterns or passwords saved to unlock the phone. Thus providing a good way for ensuring security.

As the technology is improving day by day, the need for proper security in public sector is increasing behind our scope. The invention of new technologies are becoming a challenge to the society. Now the society is witnessing a rising wave of crimes these days. It is necessary to build some methods for restricting security violations. Deploying the advanced technology in detecting a possible crime event and preventing it is gaining more and more attention now a days. The robbery crimes especially in banks
are increasing day by day and it has to be completely controlled for keeping our valuable things safer. It is at this juncture, the need for efficient real time object detection for bank security gains its advantage.

Consider an example, if there arises a robbery situation after normal working hours of bank inside the locker room. Even though the bank authority appointed security man outside the bank, he will not be able to know what is happening inside the bank or locker room. In such situation where human intervention for security is not possible, he will not be able to report the robbery at right time. If able to report the robbery at right time, then the valuable things can be made safer. In such situations, if CCTV footage’s are available, then it is easier to monitor these situations easily. Thus a new idea have to be proposed for efficient security maintenance in banks especially after normal working hours.

The main objectives behind the proposal of Real Time object detection using SSD for Bank Security is to provide the following functions.

• To design a system for providing security for industries, banks etc.
• To facilitate proper reports on security tasks.

The scope of the system is in providing the overall visualization and security for different applications.

The society is witnessing a rising wave of violent crimes these days. If not properly checked and managed, it could become a threat to the socio-economic development of the country. Developing advanced technology in detecting a possible a crime event and preventing it is gaining more and more attention now a days. The valuable items in the bank lockers should be in safe condition. Even the RBI says that, they are not responsible for the valuable items in lockers. It is at this juncture, that the proposed model becomes relevant. It is particularly useful where manual intervention is difficult.

This paper is organized as follows. Section I includes a general introduction of the project with its objective and motivation. Section II describes the existing system with the detailed study of the topic. Section III gives an overview of the system. It explains the core idea of the proposed model and the methodology used with the help of block diagram. Section IV details the experimental set up and result analysis. Section V describes the advantages and applications. Finally Section VI concludes the work explaining along with limitations and possible extensions of the work.

2. Related Works
The real time tracking of each object remains challenging due to large appearance variations. The goal of object tracking is to estimate the unknown states of target in the subsequent frame. The main aim of the proposed model is to achieve the above goal in a secure way for bank application. Various proposals were put forward and some of them have been already implemented. But they cannot be considered as effective ones. Together with all the challenges in object detection, many ideas have been published and that continue to appear.

Many techniques have been proposed for object detection in literatures. Most of the currently available detection systems uses different classifiers for object recognition. This kind of system uses a classifier for an object and performs different evaluation on it. The detection systems such as deformable parts models (DPM) uses sliding window method where the chosen classifier is run at specific portions in an image [1]. The techniques like R-CNN uses region proposal methods for generating bounding boxes in an image and uses a classifier for classification [2]. Region Proposal Network (RPN) [3] provides convolutional features along with cost-free region proposals.

YOLO [4] is another approach used for detecting objects from an image which frames detection as regression. While comparing with other region proposal classification networks (fast RCNN) which perform detection on various region proposals and thus end up performing prediction multiple times for various regions in an image. The architecture splits the input image in mxm grid and for each grid
generation 2 bounding boxes and class probabilities for those bounding boxes. The bounding box is more likely to be larger than the grid itself.

CSSD [5] is the short word for context-aware single-shot multibox object detector which is built on top of SSD, with additional layers modelling multi-scale contexts. Synthetic aperture radar (SAR) [6] images used for ship detection for providing marine surveillance and transportation security. This method uses SSD along with transfer learning for providing ship detection in complex areas. OverFeat [7], another version of the sliding window technique, predicts a bounding box directly from each location based on the confidences of the underlying object categories.

The DeepMultiBox [10] is used to train a detector for generating a small number of bounding boxes as object candidates. The bounding boxes are created using DNN. The first step is to define object detection as a regression problem to the coordinates of bounding boxes. For each training example, solves an assignment problem between the current predictions and the ground truth boxes and update the matched box coordinates, their confidences and the underlying features through back propagation.

Most existing detection pipelines treat object proposals independently and predict bounding box locations and classification scores over them separately. New EM-like group recursive learning approach [11] is proposed to iteratively refine object proposals by incorporating the context of surrounding proposals and optimal spatial configuration of object detections. This model is not effective for object localisation and error identification.

Feature maps from different levels within a network are known to have different (empirical) receptive field sizes [12]. Object detectors classify scene categories using edges, textures, object, and scene without having more outputs or networks. To address the issue of implementing automatic obstacle detection and tracking, a Kernalized Correlation Filter (KCF) [16] is used for fast and robust detection. The tracker has to be initialized all the time which is an overhead and it reduces robustness.

Traffic scene perception (TSP) [17] aims to extract accurate real-time on-road environment information, which involves three phases: detection of objects of interest, recognition of detected objects, and tracking of objects in motion. The identification and locating depends on the results from detection module, the detection of objects plays an important role in TSP. The method mainly focuses on three important classes of objects such as traffic signs, cars, and cyclists. The main disadvantage of this model is the low detection speed and large intra class variation.

3. System Design
The object detection plays an significant role in maintaining secure events. The security related issues are increasing day by day. It can be due to different reasons. One of the main reason for these kinds of problems is due to the lack of proper security maintenance by official authority. Proper attention should be provided in concern with security violations. Visual tracking helps in continuously tracking an object in different scenes. It is necessary to identify the things that are harmful to our application. So in-order to identify such things, it is necessary to track them from available sources. Now a days, there exists different technologies to lively monitor different places. By properly analysing those sources, it is easier tract things that results in security violations using available techniques.

3.1 Problem Definition:
The problem under consideration is to exploit the rich features of Deep learning and Tensorflow to develop efficient object detection model for bank security.

3.2 Proposed System:
The object detection plays an important role in many real time applications. The object detection in real time applications is more challenging. An efficient and improved method should be used for real time based object detection. Single shot multibox detector is now considered as an object detection algorithm
for different applications. The newly evolved techniques such as TensorFlow, OpenCV can be used for effective object detection. Open Source Programming (OpenCV) is a library which consist of different functions for computer vision applications. It support the deep learning framework such as TensorFlow to perform better. The TensorFlow is an open source library designed by the Google team to run large-set of numerous computations. It allows to train the models faster and reduces the chances of errors. The block diagram for the proposed model is as shown in Figure 1.

![Block Diagram](image.png)

Figure 1: Block Diagram

The input will be a video stream or images captured in real time using laptop web-cams or can be CCTV footage. The OpenCV will use the initialized camera object to create a new window named as object detection of the size 800x600. Then it will wait for 25 milliseconds for the webcam to show the images else will close the window. The input will be converted into frames using OpenCV for further processing. The frames obtained are resized into 300x300x3 image size which is the required input size for SSD. The pre-processing of image is done with the help of tensorflow in-order to avoid blurriness of image. The SSD module will perform the feature extraction and object detection. SSD uses a pretrained model for object detection since it is difficult to train a new model from a scratch. Upon extracting the features, then a classification model is used. The pretrained models are used to extract visual features. Image classification can be done using different deep learning methods. They extract the features from the input image and uses these features to identify the class of the image. In the final post processing step, taking all the extracted features and then move on to the model to see what all features are matching and which type of class. The results obtained from SSD module includes category scores and box offsets. This result is again filtered using mean value scores above 0.75 to produce the accurate result for object detection. Then it is given to the MQTT for publishing the object identified and to send the result to the subscribed ones and finally the application is developed.

### 3.3 Methodology & Design

The methodology and design of the proposed system is as follows.

#### 3.3.1. Preprocessing

The preprocessing of input is done inorder to handle the input data properly. The input image or input frame is accepted and is converted to required input size. The TensorFlow object detection API is configured inorder to accept the input. The image is preprocessed and is converted to RGB color space. The location of both training and evaluation files have to specify properly. Also need to specify the label map, which is used for mapping between a class id and their class name. The label map must be unique between training and evaluation data sets.
3.3.2. SSD
SSD is considered as the most efficient object detection algorithm due to its accuracy and ease of its implementation. It is used as the base network for the system. The SSD architecture is shown in Figure 2.

SSD mainly consist of two sections, they are feature extraction and object detection. The feature map extraction is done with the help of VGG-16. Conv4 3 is used to detect objects. Object detection is done using convolutional filters. The extracted features using VGG-16 are used for object recognition. The key features of SSD [8] are mainly three.

- Detection using Multi-scale feature maps.
- Detection using Convolutional predictors.
- Collection of default boxes.

SSD uses multiple layers to detect the objects in the input image immediately. It uses lower resolution feature layers (i.e. 4 X 4 feature layer) for detecting large scale objects while higher resolution layers (i.e. 8 X 8 feature layer) for smaller objects. SSD includes six more auxiliary convolutional layers other than VGG16. Among them, five of them are used for object identification purpose. As theses layers are combined, SSD makes a prediction of 8732 per class using these six layers.

3.3.2.1. VGG-16
VGG-16 [15] is also known as OxfordNet is a convolutional neural network architecture. Simonyan and Zisserman put forward the CGG network architecture in 2014. The network gains its advantage by its simplicity, only 33 convolutional layers stacked on top of each other in increasing order of their depth. The term 16 refers the number of weight layers used in the network. The depth of the layer also plays an important role. The features are extracted using convolutional layers. The extracted features such as edge, texture and shape are used for further processing in object detection. The fully connected layers of VGG-16 are not used as such but are converted into convolutional layers.

3.3.2.2. Training Phase of SSD
Training an object detector from scratch can take many days, for speeding up the training phase, it is advisable to reuse the feature extraction parameters from pre-existing image classification or object detection check point. The training phase of SSD requires the ground truth information need to be assigned properly. Once the assignment is completed, calculate the loss function and apply the same. Training phase also includes the selection of default boxes, their scale values as well as negative mining and data augmentation. The criterion for matching the priors with the ground truth box is the jaccard overlap (as in MultiBox [13]) with a threshold of 0.5. More overlap results in better match for the system. The SSD training objective is obtained from the MultiBox objective [13, 14] but an extension is added.
to handle multiple object categories. The objective loss function is calculated as the weighted sum of localization loss (loc) and confidence loss (conf). The localization loss is considered as the inconsistency between the ground truth box and the boundary box that is predicted. The localization loss is calculated as the smooth L1 loss [9]. SSD penalizes from positive matches and negative matches are ignored. The confidence loss is calculated as the loss in making a class prediction. The confidence loss is calculated as the softmax loss over multiple class confidences. For every positive match prediction, it is advisable correct the loss based on the confidence score of the corresponding class. For negative match prediction, the loss is penalized based on the confidence score of class 0.

**Aspect ratio and scales for default boundary boxes:** In order to maintain the complexity of the system low, the default boxes are chosen manually and keeps the default boxes to a minimum of 4 or 6 with one prediction per default box. Feature maps from different levels within a network may have different (empirical) receptive field sizes [12]. In the SSD framework, the default boxes do not necessary need to correspond to the actual receptive fields of each layer. The default boxes are designed in such a way that specific feature maps learn to be responsive to particular scales of the objects. Considering from the left layer, Conv4_3 detects the objects at the smallest scale (Smin) of 0.2 and linearly increases to rightmost layer upto largest scale (Smax) of 0.9. For the layers starting with six predictions, SSD uses five target aspect ratios such as (1, 2, 3, 1/2, 1/3). The width is calculated as the product of scale and square root of aspect ratio. The height is calculated as the quotient obtained by performing the divide operation on scale value and square root of aspect ratio. SSD adds an extra default box with scale calculated as the square root of the product of scale of current level and scale at next level. The aspect ratio for the extra default box is set to 1.

In the **matching phase**, even though positive samples are boosted there can occurs a chance of unmatched priors. The huge number of priors labelled as background can result in making the dataset unbalanced. The concept of hard negative mining is introduced to make the issue balanced. The technique is that, sort the negatives according to the calculated confidence loss and pick the negatives having high confidence loss for making the ratio between positive and negative samples atmost 3:1.

The **data augmentation** has a major role in training phase. It is done inorder to improve accuracy. By augmenting the data by means of flipping, cropping and color distortion. Inorder to deal with the variations in objects sizes and shapes captured in real time data, each training image is randomly sampled. The sampled patch will be having an aspect ratio between 0.5 and 2. After sampling, the image is then resized to a fixed size and flip one-half of the training data. Thus the performance can be improved. Thus the main stages in training phase is completed.

3.3.2.3. **Object Prediction**

The object detection is labelled as a classification problem. The classification deals with predicting the labels of the object present in the input identified while detection is different. The detection deals with finding the location of the objects identified. To have more accurate results for object detection, different feature maps are going through a convolutional layer of 3 x 3. If for example, consider Conv4_3 layer which is of size 38 x 38 x 512. It consist of 4 bounding boxes and when 3 x 3 convolutional filter is applied, each of the bounding box will have an output of (classes + 4). Thus for Conv4_3, it will be 38 x 38 x 4 x (classes + 4). If the number of object classes is 20 with one background class, then the output will be 38 x 38 x 4x (21 + 4) = 144,400. Thus total number of bounding boxes for this layer will be 38 x 38 x 4 = 5776 boxes. Similarly the number of bounding boxes for other layers are 2166, 600, 150, 36 and 4 respectively. Thus there will be a total of 8732 boxes per class which is higher than YOLO algorithm (98 boxes only).

The detection has more relevance when there exist multiple objects to detect in the same image. Thus detection deals with detecting all objects present in the image and predicting their class labels along with their respective position. The respective positions are identified inorder to create a bounding box around the identified object. So the final output after detection will be the class labels and bounding box
coordinates such as \(cx\) (x coordinate of center), \(cy\) (y coordinate of center), \(h\) (height of object), \(w\) (width of object).

3.3.2.4. Non Maximum Suppression (NMS)
The NMS is done in order to improve the accuracy of detection. It is applied to filter and remove duplicate predictions regarding the same object. First step is to sort the predictions by means of confidence scores. Starting from the top confidence predictions, SSD will evaluate whether any predictions made before have an IoU greater than 0.45 with the current prediction for the same identified class, then it will ignore the current prediction. Thus the multiple boxes for the same object class is avoided. The accuracy of SSD normally increases with the increased number of default bounding boxes when cost of speed is considered. The use of multiscale feature maps improve the detection of objects. The SSD has lower localization error and has better accuracy compared to other object detection algorithms.

3.3.3. Filtration
It is done inorder to handle the output from SSD. Filtration is done using the mean value calculated based on the number of detections. If the score obtained after the iteration is above 0.75, then the result is published to the subscribed ones. This is done inorder to handle the real time data since it is changing every time. Filtration helps in attaining good accuracy for the labelled data.

3.3.4. MQTT
The MQTT protocol is a lightweight protocol that is best suited for IoT applications. It is easier to implement in software applications. The messaging technique is used in for minimized data transmission and needs less network and power usage. Since the project deals with real time data processing, MQTT works perfect for internet of things based applications. This protocol works based on the client and server communication with the help of a broker. The illustration of MQTT protocol is shown in Figure 3. The client request for data transmission are handled by MQTT server. MQTT server is acts as MQTT broker and clients as connected devices. When a client wants to communicate with the broker by sending data, then publish operation is performed. The device that publishes data is known as publisher. When a device wants to receive data from the broker, then subscribe operation is performed. The device that receives data from broker are known as subscriber. MQTT broker is responsible for handling the publish subscribe actions to target destination.

![Figure 3: MQTT Protocol](image)

3.3.5. Bank Application
The rising wave of violent crimes not properly checked, it will became a threat to the socio-economic development of the country. Developing advanced technology in detecting a possible crime event and preventing it gains more and more attention now a days. It is at this juncture that the proposed application
for banking sector becomes relevant. It is particularly useful when the manual intervention and detection of crimes is difficult.

The bank application developed as a part of the current scenario mainly focuses on the security of bank vault as well as inner portions of banks. The system provides surveillance for bank locker room. The bank locker robbery issues are increasing day by day and proper surveillance is essential. Most of the bank lockers are under CCTV surveillance. The cameras used are mainly motion detection cameras and burglar alarms are placed in bank lockers. After the normal working hours of bank, the burglar alarms will be active all the time. When any item stuck the door, the alarm will get triggered informing someone has entered. This alarm does not need to get triggered if it is not an intruder. So this case has to be handled properly. For handling this situation, the proposed system works better. The working of the proposed system is as explained below.

The motion detection cameras are easily available and using the CCTV footage from those cameras can be used for intruder detection. The CCTV footage can be monitored using the proposed system. If the camera detects any motion or any person or objects, then the proposed models detects the presence of object(s) and identifies the object. If the detected object is a person, laptop, cell phone, knife or suitcase or any other objects usually a person carries, then an SMS will be send to the bank officials as well as police stations. Thus they can take corrective measures to handle the situation. Otherwise no action is required. This output can be used to trigger the burglar alarm instead of making false predictions. The bank officials can login to see whether any data entry is occurred regarding any human intervention to the bank locker or its surroundings at restricted time. Thus ensuring a proper security for bank lockers all the time.

4. Results and Discussions

An experimental scheme with the following objectives is designed.

- Test the developed prototype with different object cases.
- Compare the test results of the proposed system against on COCO dataset along with other methods for object detection.

4.1. Dataset

COCO is a large dataset used for object detection, their segmentation and for providing captions. It has many features such as object segmentation, Recognition in context, Super pixel stuff segmentation, 330K images (200K labelled), 1.5 million set of object instances, 80 object categories, 91 stuff categories, 5 captions per image etc., The proposed model uses 90 classes of detection and is an optimized form for object detection.

4.2. Evaluation Metrics

The following evaluation metrics are considered to differentiate the proposed system with the existing object detection methods on COCO dataset. All of them are external evaluation measures and their values range from 0 to 1; an index of 1 implies that the method is perfect and an index of 0 means that the model make no sense.

4.2.1. F-Measure

F-Measure also known as Dice Similarity Coefficient is a generic and popular metric representing the accuracy of a test which is computed as the harmonic mean of Precision and Recall values. \( F_\beta \) measure is a special case of F-measure (for non-negative values of \( \beta \)). \( F_\beta \) is used to measures the effectiveness of retrieval of data with respect to a person who gives \( \beta \) times as much importance to recall as precision. This is the evaluation metric used for evaluating the effectiveness of the proposed system.

4.2.2. Jaccard Index
It quantifies the similarity measure between two datasets and is obtained as the number of unique elements which are common to both sets to the total number of distinct elements in both sets. Jaccard index is used for matching strategy of ground truth and default bounding boxes.

4.2.3. Intersection over Union (IoU) Method

IoU is considered as an evaluation metric that is used for measuring the efficiency of an object detector on a specified dataset. The IoU is the ratio of the overlapping area of ground truth and predicted area to the total area. This is the measure used for non-maximum suppression.

4.3. Tests and Results

The prototype is tested against other object detection models on COCO dataset. The test results shown in the table 2 in terms of the above mentioned F- measure captures the reaction of the proposed method with others. It clearly shows that the proposed method obtains better results which itself proves the efficiency of the proposed system as an object detection technique. And for almost all of the test cases in COCO dataset, the system exhibits a better performance over others. The precision and recall values are calculated to find Fβ measure with beta value 2. The recall and precision is calculated by finding TP, TN, FP, FN. The recall and precision values for the given models are calculated.

Different test cases are applied for finding the effectiveness of the system. 25 test cases are done which includes 20 different situation analysis for person and 5 analysis for different objects that can be considered in the application. The different test cases for person are adult, person with short height, eyes covered, body parts, body partially covered, face covered and so on. These test cases are done inorder to check whether the system is able to find an intruder in bank. Intruders can enter the bank in different forms. Also there is a possibility of carrying objects such as laptops, cell phones, knifes, suitcase etc., in their hand. So the test cases regarding these objects are also done. Thus total 25 different test cases are done to calculate the effectiveness of the system. The test result is shown in the table 1.

| Model Name | Model Name | Recall | Precision | Fβ Measure |
|------------|------------|--------|-----------|------------|
| Proposed Model(ssdlite-mobilenet-v2) | 0.95       | 0.86   | 0.93      |
| Faster-rcnn-inception-v2          | 0.44       | 0.66   | 0.33      |
| Faster-rcnn-resnet50              | 0.45       | 0.83   | 0.36      |
| Rfcn-resnet101                    | 0.43       | 0.90   | 0.36      |

The measured Fβ value for the proposed model shows an accuracy of 0.93 which is greater than other models. The graph that shows the recall, precision, and Fβ is plotted on Figure 4 to show the difference for each models. The other models except proposed system is not performing an optimized way of running. They are highly depended on machines with high specification. They consume more time to perform identification. The proposed SSD model works better in optimized format. The system performs...
well for the proposed application which shows the efficiency of system developed compared with other models.

Figure 4: Graph representing Recall, Precision and Fβ Measure Value

5. Advantages and Applications

The Real Time Object Detection System using SSD is a model proposed for providing overall security to the public sector. As technology proliferates, the security issues also increased which is an overhead to the authority. Proper monitoring of banks, industries, and public areas will help somehow to reduce the security related issues. The proposed model aims to do so.

The advantages of the proposed system are:
• High Accuracy
• Provides more robustness on object detection
• Faster Object Detection
• Live problem reporting
• Enhanced Security Evaluation

Other applications of the system are:
• Security concerned problems
• Traffic related issues
• Surveillance applications
• Terrorism Monitoring
• Drone based applications

6. Conclusion & Future Scope

Object detection plays an important role in security related applications. The security related issues can be handled by proper object detection. The method TensorFlow is one of the most advanced and faster technique which can be used for effective object detection. OpenCV as well as TensorFlow helps in proper object detection for different real time applications. The one of the important factor for obtaining better quality results is the quality of camera used. As technology is advancing day by day, it is possible to use better ways for improving image quality. Thus proper security can be used for providing bank security applications. Since it is necessary to maintain security in all aspects, the proposed method can be efficiently used.

The future work focuses on providing more image enhancement for increasing system accuracy. It will be better if the system is able to identify the situation if intruder tries to cover the camera and further
enhancement work will focus on that issue too. Drone related applications can be integrated into the system for deploying in areas where human entry is restricted.

7. References

[1] Ade P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan., "Object detection with discriminatively trained part-based models". IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9):16271645, 2010.

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik., "Rich feature hierarchies for accurate object detection and semantic segmentation". In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference, page 580587. IEEE, 2014.

[3] Ren, S., He, K., Girshick, R. and Sun, J.;"Faster R-CNN: Towards real-time object detection with region proposal networks". In: NIPS. (2015).

[4] Joseph Redmon , Santosh Divvala , Ross Girshick , Ali Farhadi., "You Look Only Once, Real Time, Unified Object Detection”. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE 2016.

[5] Wei Xiang,Dong-Qing Zhang, Heather Yu and Vassilis Athitsos, "Context-Aware Single Shot Detector", IEEE Winter Conference on Applications of Computer vision, 2018.

[6] Yuanyuan Wang,Chao Wang and Hong Zhang, Combining single Shot Multibox Detector with Transfer Learning for ship Detection using Sentinel-1 images, 2017 SAR in Big Data Era: Models, Methods and Applications (B IGSAR DATA), IEEE, 2017..

[7] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R. and LeCun, Y.;"Overfeat: Integrated recognition, localization and detection using convolutional networks". In: ICLR. (2014).

[8] W Liu, D Anguelov, D Erhan, C Szegedy and S Reed,"SSD: Single Shot Multibox Detector”, European conference on computer vision, 2016.

[9] Girshick, R.;"Fast R-CNN", International Conference on computer vision (ICCV), 2015.

[10] A. Krizhevsky, I. Sutskever, and G. Hinton.,"Imagenet classification with deep convolutional neural networks". In Advances in Neural Information Processing Systems 25, pages 11061114, 2012.

[11] Jianan Li and Xiaodan Liang, Multistage Object Detection With Group Recursive Learning, IEEE Transactions on Multimedia, vol. 20, July 2018.

[12] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. and Torralba, A."Object detectors emerge in deep scene cnns". In: ICLR. (2015).

[13] Erhan, D., Szegedy, C., Toshev, A., Anguelov, and D.;"Scalable object detection using deep neural networks". In: CVPR. (2014).

[14] Szegedy, C., Reed, S., Erhan, D., Anguelov, and D.;"Scalable, high-quality object detection”. arXiv preprint arXiv:1412.1441 v3 (2015).

[15] Simonyan, K.and Zisserman, A;"Very deep convolutional networks for large-scale image recognition”. In: NIPS. (2015).

[16] Sushil Bharati, Yuanwei Wu, Yao Sui, Curtis Padgett, and Guanghui Wang, “Real-time Obstacle Detection and Tracking for Sense-and-Avoid Mechanism in UAVs.” IEEE Transactions on Intelligent Vehicles, 3, 2, Pp. 185-197.

[17] Qichang Hu and Sakrapee Paisitkriangkrai, “Fast Detection of Multiple Objects in Traffic Scenes with a Common Detection Framework”, IEEE Transactions on Intelligent Transportation Systems, (Volume: 17, Issue: 4, April 2016).