An Object Detection Technique For Blind People in Real-Time Using Deep Neural Network

Ashwani Kumar
Department of CSE, Vardhaman college of Engineering,
Hyderbadad, India
ashwani.kumarcse@gmail.com

S S Sai Satyanarayana Reddy
Department of CSE, Vardhaman college of Engineering,
Hyderbadad, India
saisn90@gmail.com

Vivek Kulkarni
Department of CSE, Vardhaman college of Engineering,
Hyderbadad, India
vivekdk1971@gmail.com

***Abstract*—** This paper proposed an object detection technique for blind people in real time to detect objects on any device running this model. We use convolutional neural network along with single shot multi-box detector algorithm to develop the proposed model. This model is composed of multiple layers to classify the given objects into any of the defined classes. Due to recent advancement in deep learning with image processing, enable us to develop this model. Our model takes color images as input and train the mode until the error rate is less. We have tested our model by supplying some sample images. To increase the computational performance of the model we have used single shot multi-box detector algorithm along with the help of architecture of faster region convolutional neural network. We also have calculated the accuracy for detecting the objects and found that the accuracy is acceptable compare to other existing scheme. For that, we have considered different parameter like mean average precision (mAP) and frames per second (FPS). The single shot multi-box detector algorithm uses standard VOC and COCO datasets. This model integrates the audio device, which will be helpful for the blind people.

***Keywords*—** *Accuracy; Frame per second; Mean average precision*

I. INTRODUCTION

Our brain recognized the object in the color image instantaneously, but the machines take a lot of time for training and testing to identify the objects. Machines cannot do this task easily. For object detection, artificial neurons are used in deep neural networks they are similar to humans composed of neurons and process forwarding. People are trying hard to solve this problem, but they are able to achieve 65% of accuracy only. It is so hard for the machines to categorize and recognize objects like humans. Each object in a picture or scene is recognized with the help of computer/software called as object detection. Face detection, driver less cars, vehicle detection and few other technologies uses object detection. The three major methods widely adopted in this field are: You only look once (YOLO), single shot detector (SSD) and faster region CNN [1].

II. RELATED WORK

This section represents the early works based on object detection models. Reagan L. Galvez et al. [2] have shown that classification and detection of objects is now accurately possible with the recent advancements in the field of deep neural networks in image processing. Wei Xiang et al. [3] focus on the single shot detection (SSD) which is considered as newest algorithms for detecting the object. Thus, it is broadly observed that this SSD algorithm has comparatively smaller amount of precision in identifying little objects as compared to bigger objects. This is since it does not pay heed to the context from out of the proposal boxes. The paper presents shorthand for single shot multi-box detector i.e. CSSD. Two variants of CSSD have been discussed in this paper. The demonstration results show how multi-scale context modeling significantly enhances the precision in detection.

Adnan Othman et al. [4] recommended OpenCV libraries and deep learning technique for recognition for real-time video and an object detection. The authors have utilized Raspberry Pi 3 for the implementation of this system which helps in monitoring and capturing the frames and thus detecting and recognizing the objects. The application of few enhancements in the proposed method such as multi scale features, default boxes and depth wise separable convolution allow the our approach to achieve a higher accuracy in detecting and recognizing objects. Hui Eun Kim et al. [5] focus on data augmentation specific to particular domain. In this paper, there is remarkable improvement shown by using the proposed method which is particular to on road detection of objects and it upgrades the average accuracy by 30%. In this paper, the authors have used CNN to detect the objects in the live environment [6]. Outputs very clearly show that former model is ideal for applications in real-time, the reason being its speed.

Chengcheng Ning et al. [7] focused on Single Shot Multi-Box Detector (SSD) as one of the fastest algorithms in the field of object detection. It makes use of a single convolutional neural network for detecting the objects in an image. Although SSD algorithm of object detection is fast but still a big gap has been observed when the comparison is done.

978-1-7281-0899-5/19/$31.00 ©2019 IEEE

The authors propose a technique to enhance algorithm for increasing its accuracy of classification without any affect on its speed. Nashwan Maria Jones and Viola [8] suggested that integral image is the new image representation, which allows the detector to detect the objects faster. AdaBoost algorithm from a larger dataset which picks least features and gives extremely efficient classifiers. Cascade method used for complex classifiers combining which enables quickly discard image background regions.

Romdhani et al. reduced set vectors are computed from the original set vectors in this model making the rejections as early as possible. Fleuret and Geman Starting from coaching examples, we have a tendency to recursively notice larger and bigger arrangements that are "decomposable" which means the chance of an appointment showing decays slowly on an object with its size. Schneiderman and Kande detects face expression and ignores anything like buildings, trees and alternative elements of body. Sung and Poggio uses the means of a view-based face and non-face model it distributes the human face patterns [9]. Rowely et al. proposed small frames of the pictures are identified by the connected neural networks and checks whether it has face or not. Fischler and Elschlager has given a scale checking and altering formula is projected to mechanically adjust the attitude scales throughout the chase method. Later, classifiers were introduced such as neural networks, SVM, AdaBoost, Bayes, etc.

Osuna et al. Introduced training an SVM is admiring determination a linearly unnatural quadratic programming (QP) downside in numerous variables adequate to the quantity of knowledge points. This optimization downside is thought to be difficult once information points exceed few thousands. This initial flourishing family of object detectors, all of them supported applied math classifiers, set the bottom for many of the subsequent analysis in the terns of coaching and analysis procedures and classification techniques. In the first section, the introduction of object detection and deep learning process are shown. Second section demonstrates related work. In third section, we have reviewed the existing techniques of object detection. Section four, put a light on the proposed model. Fifth section represents the experimental results. Finally, section six concludes the proposed research work.

III. Reviewing Existing Systems

The deep neural network are basically consist two different model first is convoluted and the other is non-linear relationships. In both model the object considered as a layered configuration of primitives. In history there are numerous architectures and algorithms for implementing the concept of deep learning these network includes belief network, stacked network, gated recurrent unit etc. The first CNN was constructed by LeCun et al. [10]. The different application domain of convoluted neural network are image-processing, handwriting character recognition etc. Object detection is performed by estimating the coordinates and class of a particular objects in the picture. The presence of these object in an picture may be in random positions. Hence the first and foremost task is calculate the positions of objects. Conventional object detection technique uses selective search algorithm which creates regions by detail features of pictures [11]. In this section we have discussed only faster RCNN and YOLO v3 architecture.

A. Faster RCNN:

Region Proposal Network for generating regions and detecting objects uses two methods of Faster-RCNN [12]. The first method proposes regions and uses the proposed regions respectively. Fig. 1. demonstrates the architecture of Faster-RCNN [13].

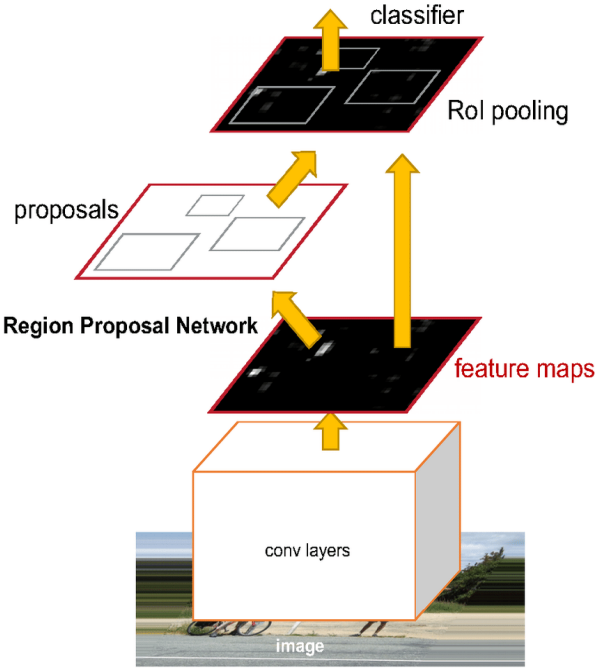

Fig. 1. Architecture of Faster-RCNN [13]

In faster R-CNN the author [13] has used 16 architecture in convolution layers to achieve detection and classification accuracy on datasets as shown in figure 1. Kumar, A. et al. proposed different buyer seller watermarking protocol to provide secure and private transaction between the communicating parties [14-20].

B. Yolo v3:

It elaborates to for you only look once. This is a detector of objects which makes use of features learned by a deep convolutional neural network for detecting object in real time. It consists of 75 convolutional layers, with up-sampling layers and skip connections for the complete image one neural network is applied. Regions of the image are made. Later bounding boxes are displayed along with probabilities. The most noticeable feature of v3 is that the detections at three different scales can be done with the help of it. Fig. 2. represent the architecture of Yolo v3.

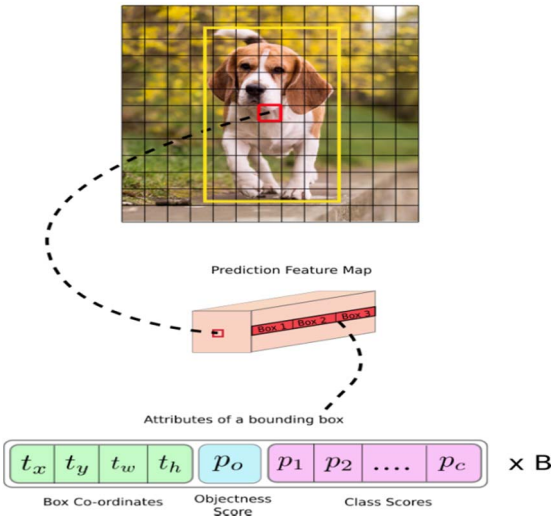

Fig. 2. Architecture of Yolo v3[1]

IV. Proposed Approach

Our proposed approach detect the small objects from the color images in real-time by using convolutional neural network for the blind people with the help of an audio device integrated into the model. We have used Faster R-CNN with deep neural network and single shot detector (SSD) algorithm with additional layers. The SSD algorithm uses feature maps extraction and convolutional filters for detection of small objects. During the training, our model should get a high class confidence score. Our model is having multi-Box on multiple layers, which leads to significant results in detection. The SSD algorithm uses the default boxes and truth boxes for identification of objects. The problems with the previous technique were how to recover the fall in precision. This fall can be reduced by using multi-scale feature map and default boxes of the SSD algorithm. Selection of default box, matching of boxes and loss function are the key point of the proposed scheme. Fig. 3 represents the entire proposed scheme.

A. SSD algorithm

In order to understand the role of SSD algorithm, the following points are very important:

Single Shot: this means that the tasks of the thing localization and classification are exhausted one passing play of the network.

Multi-Box: Ground Truth Box and Predicted Box are the boxes in MultiBox. This is introduced by Szegedy.

Detector: The network is associate degree object detector that conjointly classifies those detected objects.

Default the size of the boxes: The selection of boxes is based on the minimum value of convolution layer and maximum values of change in intensity [14]. The first algorithm represent the procedure of producing specified feature maps F(m).

Truth Boxes: After finding the size of boxes, the next phase is matching of the boxes with the corresponding truth boxes. For a specific given picture to identify the truth Boxes is explained in the second algorithm.

Loss Function: The loss function is unbelievably simple and it is a methodology of evaluating how well your role models your dataset. If your predictions are entirely of your loss function can operate next range. If the output range is less, it means that the model is good. The main objective is to minimizing loss function. The loss function is also depends upon the sum of weighted localization and classification loss functions [26].

A. Steps in SSD Algorithm

When a color image is fed into the input layer, SSD does the following:

Step 1: Image is passed through large number of convolutional layers extracting feature maps at different points.

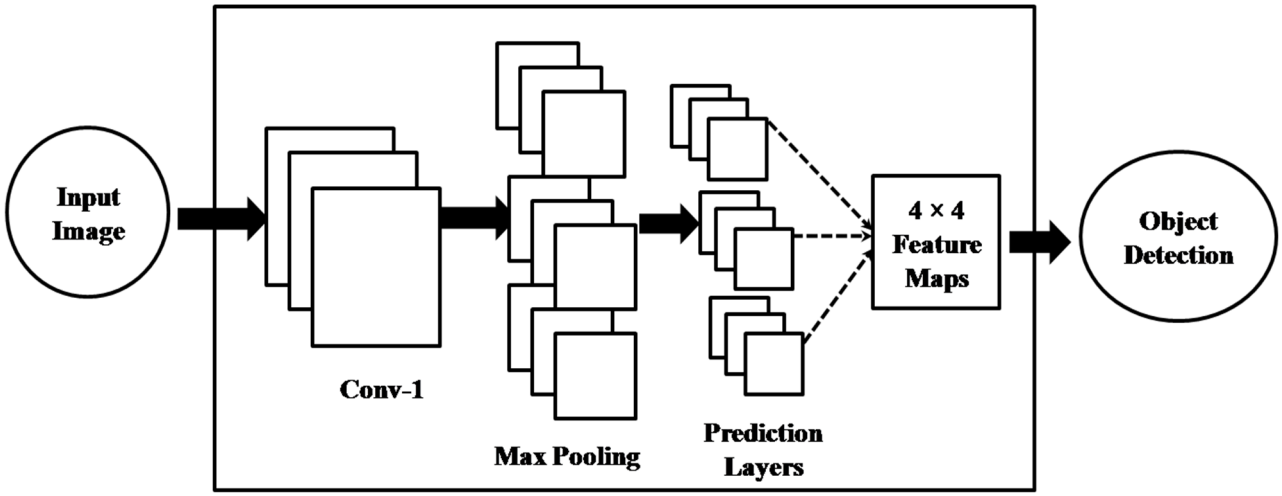

Fig. 3. The proposed approach

Step 2: Every location in each of those feature maps use a 4x4 filter to judge a tiny low default boxes.

Step 3: Predict the bounding box offset for each box.

Step 4: Predict the class probabilities for each box.

Step 5: Based on IOU the truth boxes are matched with the predicted boxes.

Step 6: Instead of exploitation all the negative examples, type the result exploitation the best-assured loss for every default box.

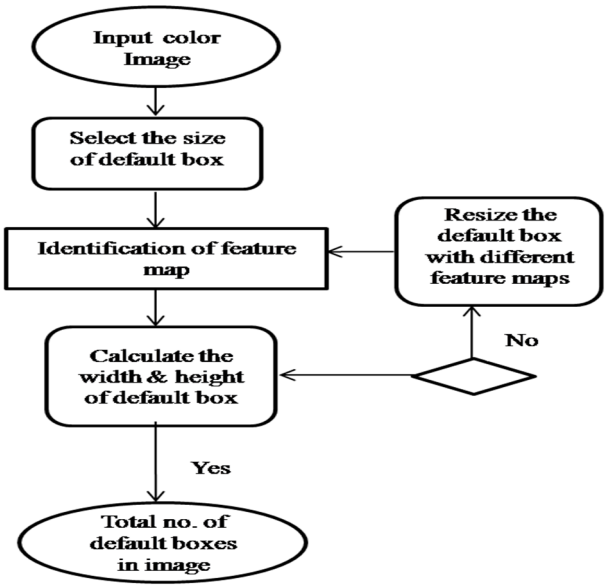

Fig. 4. Process of identification of total number of default boxes

In figure 4 the procedure of identification of total no. of default boxes in a color image are shown. Our proposed SSD algorithm basically consist two phases in the first phase identification of default boxes is done in the second phase the identification of class labels with different color boxes is carried out. Figure 5 represent the process of detecting the objects with respective class labels.

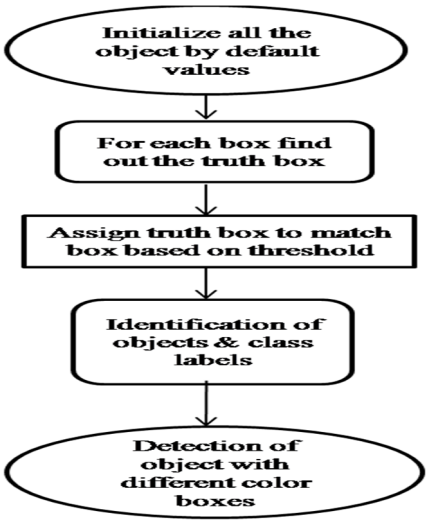

Fig. 5. Process of detection of object with different color boxes

The prime objective of this paper is to use single shot multi-box detector (SSD) algorithm for higher detection of the objects from the color images and label them with their respective class in real-time. Our proposed approach uses a new architecture as a combination of Faster R-CNN with convolutional features with multi-scale contexts in additional layers for detecting the tiny objects which increases the object detection rates. This makes use of feature maps extraction with small convolutional filters for detection of objects with acceptable accuracy.

V. Result Analysis

In this section we have shown the result analysis of our proposed scheme i.e. an object detection technique for blind people in real-time using deep neural network. For showing the results, we use Python programming language and Raspberry Pi 3. We have also used OpenCV 2.4 library to support machine learning functionality. Standard dataset has taken for producing the result like Pascal VOC and COCO datasets. COCO dataset contains 80 classes, 80,000 images and 40,000 validation images. Pascal VOC could be a well-liked dataset for building and evaluating algorithms for image classification and object detection. We have also integrated waves max audio to support the demand from OpenCV 2.4.

To calculate the performance of the SSD algorithm we have used mean average precision (mAP) and frames per second (FPS) along with intersection over union (IOU). The accuracy is calculated using the below equation no. (1) and it could be improved over the original dataset also. It is used to measuring correctness of the classification process. The proposed scheme performs well to achieve high accuracy and FPS with good speed to detect objects real-time for blind people.

$$A_{ccuracy} = \frac{Object\ \ (O_{correct})}{Total\ \ Object\ \ (T_{obj})} \quad (1)$$

Where $O_{correct}$ = number of correctly detected object and total (T_{obj}) number of images. IOU is calculated by Jaccard Index to find out overlap between two bounding boxes [27]. Equation number (2) shows the formula for IOU. Cloud may result to break integrity and confidentiality, and increase the security risk as the usage increases on demand [21]. Energy efficient design needs systematic optimization in all aspects of design from process technology and logic design [22].

$$IOU = \frac{area\ \ of\ \ overlap}{area\ \ of\ \ union} \quad (2)$$

Muni Sekhar et al. [23-25] proposed content-aware reversible data hiding technique to trade-off between embedding capacity and visual quality.

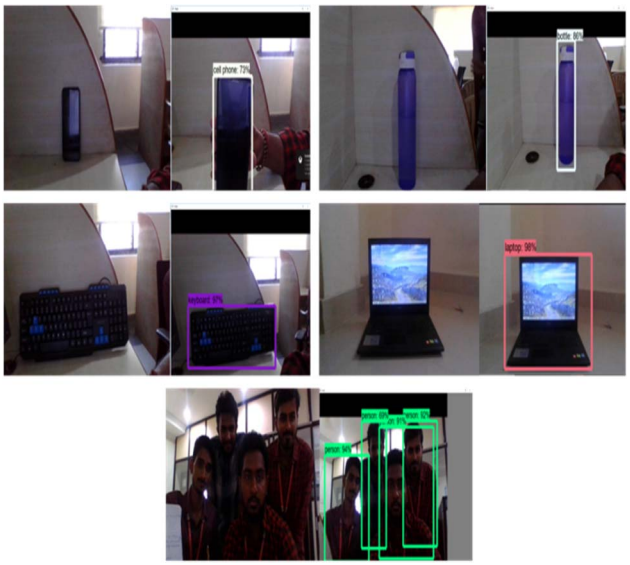

Fig. 6. Object detection results on the test images

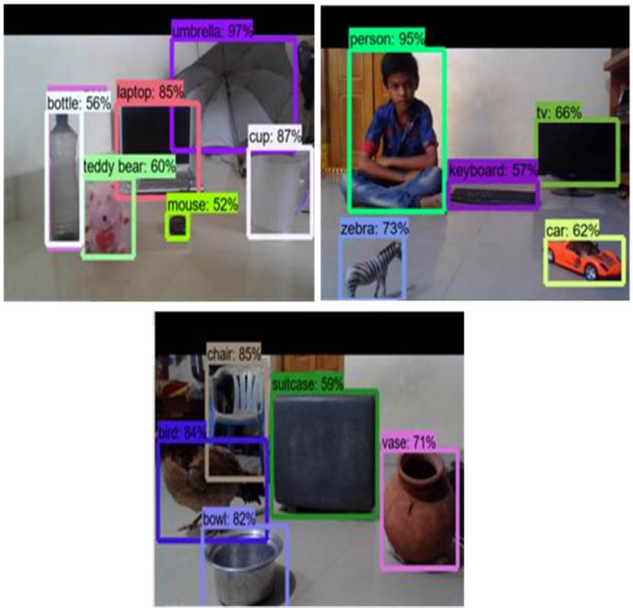

Fig. 7. Detected objects with different color boxes and their respective class labels

In figure 6 and figure 7, we have shown various object detected with their respective class labels by the proposed algorithm. We have chosen different colors of boxes to show different class labels [28-29]. Figure 6 and figure 7 shows that our proposed approach correctly detects and recognizes bottle, laptop, mouse, cup, teddy bear, and umbrella, person, keyboard, TV, Zebra, toy car, bowl, chair, bird, vassal, suitcase. Table I represent the different parameter like mean average precision (mAP), frames per second (FPS), no. of default boxes and resolution of the image of the proposed method by using VOC and COCO test dataset.

Table I. represents the results on Pascal VOC and COCO test.

System Model	mAP	FPS	No. of Boxes	Resolution
F-CNN	73.2	7	6000	1000×600
YOLO	66.4	155	98	448×448
SSD512	76.8	19	24564	512×512
SSD300	74.3	46	8732	300×300
Proposed Approach	78.68	89	5988	1024×1024

VI. Conclusion

In this paper, an object detector model using deep learning neural networks for detecting the object for blind people is proposed. This model can be used to detect the objects in images, videos or even webcam feeds. The accuracy of the model is more than 75%. The training time for this model is about 5-6 hours. This model uses deep neural networks to extract feature information and then perform feature mapping, the SSD MobileNet FPN performs localization and classification of neural networks and hence is fast compared to other models. In this work, we have used a single shot multi-box detector (SSD) algorithm to achieve high accuracy and IOU in real time for detection of the objects for the blind person.

A convolutional neural network is used to extract feature information from the image and then perform feature mapping to classify the class label. This paper points out that the algorithm uses truth box to extract feature maps. The metrics of our approach indicates that the algorithm achieves a higher mAP and acceptable accuracy for detecting the objects from color images for blind person.

References

[1] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: CVPR. (2016).

[2] R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra, J. M. Z. Maningo, "Object Detection Using Convolutional Neural Networks", TENCON 2018-2018 IEEE Region 10 Conference, 2018.

[3] P. Poirson, P. Ammirato, C.-Y. Fu, J. Liu, Wei Koeck, A. C. Berg, "Fast single shot detection and pose estimation", International Conference on 3DVision(3DV), 2016.

[4] N.A. Othman, I. Aydin, "A new deep learning application based on movidius ncs for embedded object detection and recognition", 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2018.

[5] Wei Xiang Dong-Qing Zhang Heather Yu Vassilis Athitsos, Context-AwareSingle-ShotDetector, 2018 IEEE Winter Conference on Applications of Computer Vision, pp. 1784-1793, 2018.

[6] D. Alamsyah and M. Fachrurrozi, Faster R-CNN with Inception V2 for Fingertip Detection in Homogenous Background Image, IOP Conf. Series: Journal of Physics: Conf. Series 1196 (2019) 012017.

[7] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, "SSD: Single shot multibox detector", 2015.

[8] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, vol. 1, pp. 511-518, 2001.

[9] C. Papageorgiou, M. Oren, T. Poggio, "A General Framework for Object Detection", Proc. IEEE Int'l Conf. Computer Vision, 1998.

[10] LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:1–10.

[11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, "SSD: Single shot multibox detector", 2015.

[12] R. Girshick, "Fast R-CNN," in IEEE International Conference on Computer Vision (ICCV), 2015.

[13] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pages 91–99, 2015.

[14] Kumar, Ashwani, S. P. Ghrera, and Vipin Tyagi. "An ID-based Secure and Flexible Buyer-seller Watermarking Protocol for Copyright Protection." Pertanika Journal of Science & Technology 25.1 (2017).

[15] Ashwani Kumar et. al. "A Lightweight Buyer-Seller Watermarking Protocol Based On Time-Stamping and Composite Signal Representation", International Journal of Engineering & Technology, 7 (4.6) 39-41, (2018).

[16] Ashwani Kumar, Satya Prakesh Ghrera, & Vipin Tyagi, "Modified Buyer Seller Wa-termarking Protocol based on Discrete Wavelet Transform and Principal Component Analysis", Indian Journal of Science and Technology, 8(35), 1-9, 2015.

[17] Ashwani Kumar, Satya Prakesh Ghrera, & Vipin Tyagi, "A comparison of buyer-seller watermarking protocol (BSWP) based on discrete cosine transform (DCT) and discrete wavelet transform (DWT)", Emerging ICT For Bridging The Future-Proceedings of The 49th Annual Convention Of The Computer Society Of India (CSI) Volume 1: Springer. 401–408, 2015.

[18] Ashwani Kumar, Satya Prakesh Ghrera, & Vipin Tyagi, "A new and efficient buyer-seller digital Watermarking protocol using identity based technique for copyright protection", Third International Conference on Image Information Processing (ICIIP), IEEE 2015.

[19] Ashwani Kumar, Satya Prakesh Ghrera, & Vipin Tyagi, "Implementation of wavelet based modified buyer-seller watermarking protocol", WSEAS Trans. Signal Process. 10, 212-220, 2014.

[20] Kumar, Ashwani. "Design of Secure Image Fusion Technique Using Cloud for Privacy-Preserving and Copyright Protection." International Journal of Cloud Applications and Computing (IJCAC) 9.3 (2019): 22-36.

[21] Anandita Singh Thakur, P. K. Gupta, "Framework to Improve Data Integrity in Multi Cloud Environment", International Journal of Computer Applications, Vol. 87, Issue 10, 2014.

[22] P. K. Gupta, G. Singh," Energy-Sustainable Framework and Performance Analysis of Power Scheme for Operating Systems: A Tool," IJISA, vol. 5, no. 1, pp. 1-15, 2013.

[23] V.Muni Sekhar, K.V.G.Rao, N SambasivaRao and M. Gopichand. "Comparing the Capacity, NCC and Fidelity of Various quantization Intervals on DWT", Published, International Conference on ICICSE2015, August 2015, Hyderabad, India, DOI: 10.1007/978-981-10-0419-3.

[24] V Muni Sekhar, K.V.G.Rao, N. Sambasiva Rao, Ch Sravan Kumar, "A Reversible RIE based Watermarking scheme", IEEE 7th International conference on IACC2017, 5-7 January, 2017, Hyderabad, India, DOI: 10.1109/IACC.2017.179.

[25] V Muni Sekhar, N. Sambasiva Rao, P. Buddha Reddy, "Content-aware Reversible Data Hiding: URHS", Springer FICTA2017 October, 2017, Bhubaneswar, India, DOI: 10.1007/978-981-10-7566-7_5.

[26] Viral Thakar, Walid Ahmed, Mohammad M Soltani, Jia Yuan Yu, "Ensemble-based Adaptive Single-shot Multi-box Detector" in the Proceedings of the ISNCC 2018, 19-21 June 2018, Rome, Italy.

[27] Borui Jiang, Ruixuan Luo, Jiayuan Mao, Tete Xiao, and Yuning Jiang. Acquisition of localization confidence for accurate object detection. In Proceedings of the European Conference on Computer Vision (ECCV), 2018.

[28] Raman Dugyala, Bruhadeshwar Bezawada, Romanch Agrawal, Sai Sathyanarayan, Rajinikanth Tatiparthi, "Application of Information Flow Tracking for Signature Generation and Detection of Malware Families", International Journal of Applied Engineering Research (IJAER). ISSN 0973-4562 Volume 9, Number 24 (2014), pp. 29371–29390.

[29] Dugyala Raman, Bruhadeshwar Bezawada, T. V. Rajinikanth, Sai Sathyanarayan, "Static Program Behavior Tracing for Program Similarity Quantification", Proceedings of the First International Conference on Computational Intelligence and Informatics, 2017.