2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)
Karnataka, India. Dec 23-25, 2022

2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON) | 978-1-6654-5499-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/SMARTGENCON56628.2022.10083782

Machine Learning based Object Detection and Classification for Future Multi-Media Application

K.N.V.Suresh Varma
Research Scholer, ECE Department,
Sathyabama Institute of science and Technology,
Tamilanadu
Email: knvsureshvarma@yahoo.co.in

Dr. Lalithakumari.S
Associate professor, ECE Department,
Sathyabama Institute of science and Technology,
Tamilanadu
Email: lalithavengat@gmail.com

Abstract: **Object detection is a very important aspect of multimedia applications. The recent innovations that are being deployed in the current era with the latest trends in technological advancements have made researchers and scientists develop systems that are capable of identifying objects using various machine learning algorithms. The objects that are being identified are further classified for identifying the particular class of the object. This process seems to be simple but is quite complex and also has various complexities and complications that need to be analyzed before deploying any algorithm into the machine for automated detection and classification. This paper explores the recent research that is carried out, in particular in the area of object detection and classification using machine learning algorithms. The performance measures like accuracy, sensitivity, recall, and F-measures have been compared with the available model.**

Keywords – Object Detection, Object Classification, Machine Learning.

I. INTRODUCTION

Machine learning is the enthralling area of research for the past decade which has made the evolution of various machine learning algorithms that are defined for a particular application. These applications are been modified to improve the accuracy of working for that particular application. This topic is extended with the development of various applications that are relevant to the detection of objects and classify them based on the features that are extracted. In the current scenario machine learning algorithms are been developed for machine-to-machine communication (M2M) which is also called as self-learning of machines. Figure 1 shows the basic block diagram of any machine learning algorithm. The prime and most primitive part is to extract the features from the dataset. The dataset may be images or any relevant information regarding the quantity that is to be trained. It is very prime to consider the size of the dataset. The size of the dataset that contains the objects which need to be trained will be of increasing exponentially as the timeline varies[1]. The computational complexity of the algorithm that includes classification as the main motive need to consider the dataset size which increases the complexity of the algorithm with the increase in the size of the dataset[2]. Hence, it is the need of the hour to apply various optimization techniques on the dataset so that computational overhead on the algorithm reduces. The next phase is the pre-processing stage wherein the objects are tuned or enhanced using various image processing techniques that suits for a particular application so that standard objects are created. These will be of standard size and quality. The crucial stage is the feature extraction phase that clearly and must extracts optimal features from the objects. Again, optimization comes into picture in this juncture. The features are being optimized so that they accurately represent the object.

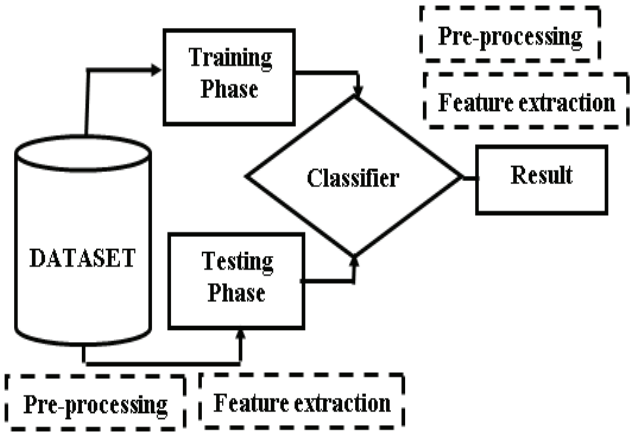

Fig. 1. Block Diagram of basic machine learning based classification.

Identification of cars in a densely traffic road is of current interest as the authorities want to identify the car based on the faulty driving or some other crime that is been carried out. Such an identification is more difficult than that of the sign identification as the similar class will have a wide heterogeneity. such a heterogeneity is observed due to the diverse perspectives. Another major problem alongside of identification of cars is the detection of cyclists which is new area of research that has attracted various domain application based on the perception of the traffic scene. Currently, few methods for detection of cyclists are available and are so designed specifically. This technique involves the detection, recognition and monitoring of different objects of concern in three phases. [3]. Object detection in the computer vision community is a challenging yet have significant application. In many practical applications, such as face detection and pedestrian detection, it has achieved effective performance. The identification features of the objects viz., cars vary dramatically as perspectives change for object classes with a wide intra-class variance. Qichang Hu et. Al [1] presented an object subcategorization approach that aims to group the object class into visually homogeneous subcategories in order to deal with these variations that cannot be tackled by the traditional VJ system[4]. To create multiple subcategories, the proposed subcategorization method applies an unsupervised grouping technique to one unique feature space of the training samples. By splitting it into several sub-problems, this approach simplifies the original learning

problem and enhances the efficiency of model generalization[5]. The main aim of the any object detection system it to recognize the segmented object such that it can be identified as the visually distinctive object in an image. Various segmentation techniques exist such as semantic segmentation that will not result in better segmentation in real time. Several applications such as image and video compression, fragmentation, image editing, tracking of objects, rendering, discovery of new objects, recovery etc., but not limited [6].

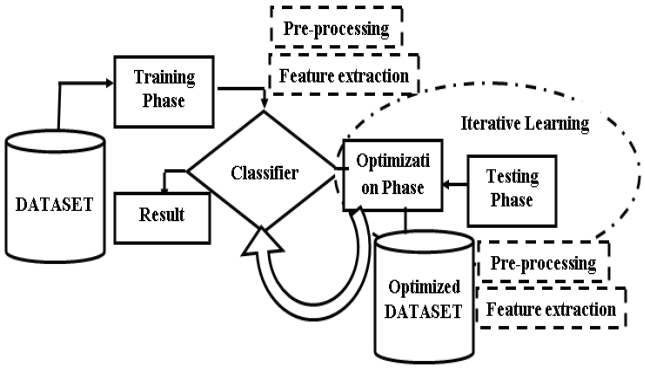

Fig. 2. Block diagram of Iterative Learning

Figure 2 shows the block diagram of iterative learning. In such a learning mechanism, the system tries to learn the features of the object in an iterative manner such that they are optimized every time to further learn from the optimized features. This is done using feedback-based learning. The optimized features have the advantage of the reducing the space complexity and also improves the time in recognition[7]. Hence, such a learning-based approach is widely used and has become recent trend in the machine learning.

Object detection has wide area of applications in forensics. If the same object detection is extended further for the analysis of the video-based object detection, such a system can be used for security purpose. Research on the intelligent video surveillance system that automatically analyses video without constant surveillance by humans has in recent years actively been performed to develop methods to detect and alert unique incidents, such as interference, drought, drop-out, crime and fire detection[8]. Smart monitoring system reduces human errors by reducing human reliance. This also increases the response time by creating alarms once events take place. While systems based on conventional computer vision technologies only suffered accuracy or performance limitations, recent advancements in machine learning have opened up practical ways to develop different applications. But no technique is itself ideal, how to correctly use it for particular applications[9].

The identification of objects in static images has made substantial progress as deep neural networks emerge (CNNs). However, video object detection poses other challenges due to degraded image quality such as motion blur and video defocus, resulting in unreliable video classification for the same object. Many research efforts have therefore been directed to the detection of video objects by exploiting temporary contexts particularly after the ImageNet video object detection (VID) issue has been introduced[10]. Many methods used earlier take advantage of temporal contexts by connecting the same entity to tubules and aggregating tubelet classification ratings. You use static picture detectors first to detect

Then connect objects within each framework by checking the boxes of objects between neighbouring frameworks or predicting object movements among neighbouring frames, in accordance with a spatial overlap among object boxes in different frames[11]. These methods yield very promising results. But the same object changes its locations and appearances in the neighboring frame due to the movement of the object, which may insufficiently cover the spatial overlap between the boxes of the same object and the expected moves of the object[12]. This affects the quality of the object linked, particularly for fast moving objects. In comparison, it is apparent in the same frame that two boxes belong to the same object when they have sufficiently spatial overlapping[13]. In recent years, with impressive developments in deep learning, object detection efforts are exploiting the proposals of objects to facilitate classifications with strong, hierarchical graphic illustration[14]. In the event of an image, the objective of an object detection/location system is to identify and find objects of interest. It is one of the most commonly studied computer vision problems with a number of applications. Most object detectors are highly supervisioned by studying object category appearance models, i.e. by using images with bounding boxes and their labels in the category[15].

Machine learning algorithms have expanded their span of applications into several domains. In this juncture, remote sensing applications are one such domain where machine learning applications have paved way to improvements in the existing classification of distant objects from satellite imagery. [16] The use of hyper-spectral images (HSIs) in remote sensing applications like target detection, land cover categorization, and anomaly detection is becoming more prevalent. The wavelength spectra of a typical HSI are known to range from 400 nm to 1040 nm and are typically divided into 256 bands with a spacing of 2.5 nm per band. However, each band can be represented by a two-dimensional visual image, so that an HSI is formed by a series of images, which correspond to the wavelength bands in question, one after another[17]. An open research area in remote sensing applications, however, is the use of hyper-spectral features for object-oriented HSI categorization. An initial set of studies used semi-supervised methods, such as neural networks and machine learning-based transductive support vector mechanisms (TSVMs), to classify HSIs[18]. Hughes phenomenon is alleviated in non-parametric and kernel based classification to many classes with TSVM. For neural network training with stochastic gradient descent, provided a flexible embedding regularisation method with additional balancing constraints. Several scholars examined the hidden Markov random field integrated with support vector machine in order to increase HSI classification accuracy (SVM)[19]. Among its advantages is the emphasis it places on object spatial dependence, particularly in the classification map. HSI is typically treated by conventional approaches as an array of spectral data, which results in a significant loss of classification precision. An key prerequisite for developing explainable artificial intelligence (XAI) systems for human users is that multi-sensor networks for object classification grow increasingly sophisticated. Due to the lack of transparency and interpretability of most intelligent algorithms to classify objects, it is critically necessary to explain the decision support of multi-sensor systems, such as military situational awareness situations[20]. Semantic

explanations, rather than facts, are meant to describe the underlying thinking process in words that humans can understand. As a result, researchers in multi-sensor networks from academia and business face a significant difficulty in explaining the predictions of intelligent systems as a fundamental component.

II. Object Detection

One of the fast-emerging fields in the intelligent transportation system is vision-based traffic scene perception. Over the past decade, this area of study has been actively studied. Three steps are involved in TSP: identification, recognition, and monitoring of different objects of interest. Since identification and tracking often rely on detection performance, the ability to effectively detect objects of interest plays a crucial role. Three major objects need to be identified viz., traffic signs, vehicles, and bicycles[21]. The purpose of the identification of traffic signals is to alert the driver to changed traffic conditions. In different traffic settings, the job is to correctly localize and identify road signs. Prior techniques use knowledge about color and shape. However, under harsh weather conditions and lighting conditions, these methods are not adaptive. Additionally, because of the weather and damage caused by collisions, the appearance of traffic signs will physically change over time. Instead of using color and shape characteristics, texture or gradient characteristics such as local binary patterns (LBP) and histogram of directed gradients are used in most recent approaches (HOG)[22]. These characteristics are partly invariant to the modification of image distortion and lighting, but they are also unable to cope with significant deformations. Due to its wide intra-class variability induced by various points of view and occlusion patterns, car detection is a more difficult problem compared to traffic sign detection. While sliding window-based methods have shown impressive outcomes in face and human detection, because of a wide variety of viewpoints, they often fail to detect vehicles. The deformable parts model (DPM), which has gained a great deal of interest in the detection of generic objects, has recently been successfully adapted for vehicle detection. In addition to the DPM, approaches based on visual subcategorization were applied to boost the generalization efficiency of the detection model[23].

Salient detection of objects is a major trend in saliency modeling which aims to locate the most interested objects in a scene. Recently, it is increasingly attracted by its preprocessing for many applications, such as image segmentation and retargeting. Conventional methods or techniques that measure contrast saliency that measure locally or globally are used for edges and noise detection [24]. The former local processes are sensitive to high degree of contrast for the edges and the noise and try to attenuate smooth areas of the objects, thus making them feasbile for the identification or detection of smaller objects. Hence, a model for segmentaion of important part of the objeect such as pre-object map and form, ie., an outstanding object has a well-determined closed boundary. The previous shape is derived by the combination of salience with the limit information obtained from an edge detector. The salient object is approximated by contouring a new patch-based technique[25]. This method removes the contour in a simple manner without the need of the boundary detection. The picture is the spilt into two parts with one inside and the other outside the contour which are marked based on the object and the background regions. As the contours are observed to be rough the contrasts between the internal and the external patches are measured to remove them by assigning weights.

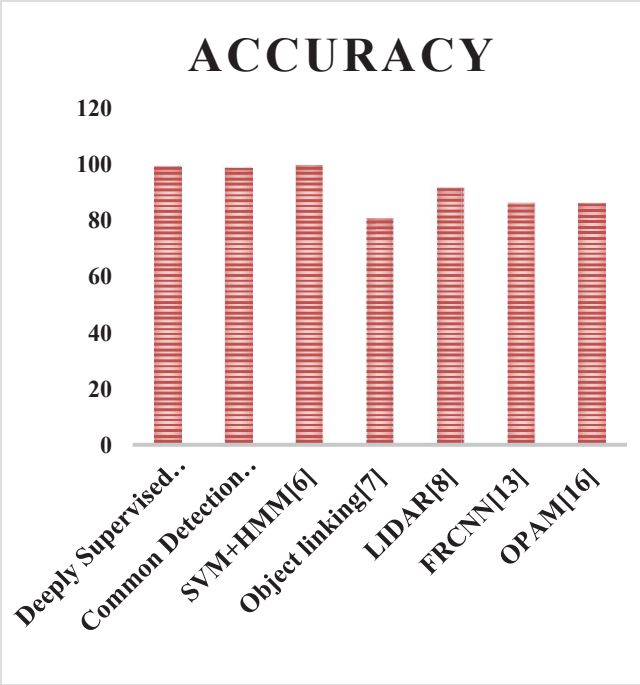

Fig. 3.

TABLE I. SHOWS THE COMPARATIVE STUDY OF RECENT WORKS CARRIED OUT BY SEVERAL RESEARCHERS IN THIS CONTEXT BY EXPLOITING THE PROBLEMS.

Author and Citation	Method	Features	Drawbacks
Qichang Hu et. al[1]	dense feature extractor and common detection framework	Since all dense features need to be evaluated only once during the testing phase, switching to one common framework speeds up detection process.	System run time increases.
Qibin Hou et. al[2]	deeply supervised network for salient object detection	holistically-nested edge detector	Computationa l Compex
Shuze Du and Shifeng Chen[3]	Random Forest	Edge Contouring	Limited Accuracy
Xiaozhi Chen et. Al[8]	CNN	Minimization of energy function by encoding the objects	Time complexity analysis need to be improved
Yunhang Shen et. Al[10]	Object Specific Pixel Gradient	Mapping of pixels	Computationa l complex for heavy data

A. mathematical equation

$$(Zx', Zy', 1) = [Zx, Zy, 1]\begin{bmatrix} \text{cosec}\,\theta & \sec\theta & 0 \\ -\sec\theta & \cos ec\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$HBP\_b = \sum_{n=1}^{N} T(g_n - g_c) * 2^n \quad (2)$$

$$HBP\_w = \sum_{n=1}^{N} T(w_n - w_c) * 2^n \quad (3)$$

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}} \cdot for\ i = 1, \ldots, K \quad (4)$$

$$Loss(y, z) = -\sum_{i=1}^{K} Z_i \cdot \log(y_i) \quad (5)$$

The above all equation 1 to 5 explains about object detection mathematical modeling. By using this technology can be performed in effective manner.

III. Object Classification

Aiming to recognise hundreds of subcategories under the same basic-level category, such as hundreds of subclasses of birds, vehicles, pets, flowers and aircrafts is extremely difficult with Good picture categorization. Fundamental picture categorization, on the other hand, merely needs to distinguish between basic categories like bird and car. According to Figure 1, basic-level and fine-grained images are classified differently. This is a critical activity that has a wide range of applications, including automatic driving, biological conservation, and cancer diagnosis. Image patches with high objectness created by a bottom-up technique are commonly used to locate the discriminative objects and pieces. Image patches can be generated using an unsupervised method called selective search, which has been widely used in recent ways. It is necessary to remove the noisy image patches and maintain those that include the object or discriminative sections because the bottom-up approach has high recall but low precision. This can be achieved by top down attention model. There are two levels of attention when it comes to fine-grained picture classification: one at the object level, and the other at the component level (part-level). Part annotation (i.e., the location of individual parts) may seem like an obvious solution to the problem of identifying individual parts, but labelling is extremely time-consuming. The first restriction is this, and the image of shrimp and prawn has been shown in figure 4.

Fig. 4. a) Shrimp b) prawn

When it comes to explainable object classification, it is commonly agreed that not only does a well-functioning system need to be in place, but it is also vital for human users to understand the reasoning behind their decisions. As a general rule, machine learning researchers treat object classification as a fundamental topic when developing classifiers that have been trained on a large number of previously classified cases.

It has been common practise to use hyper-spectral images (HSI) in remote sensing applications such as detecting targets, classifying terrain, and spotting anomalous features. The wavelength spectra of a typical HSI are known to range from 400 nm to 1040 nm and are typically divided into 256 bands with a spacing of 2.5 nm per band. For each wavelength, a two-dimensional visual image can be depicted in order to build up an image sequence that corresponds to that wavelength's corresponding HSI. Hyper-spectrum classification of objects using hyper-spectrum features is still an active research question for remote sensing applications.

TABLE II. Comparison of Results

parameters	Accuracy	Recall	F 1 score	Sensitivity
Method [1]	87.42	89.21	89.14	89.54
Method [2]	88.46	91.17	91.27	90.75
Method [3]	88.37	92.54	93.17	92.53
Method [4]	91.36	97.17	95.86	93.54
Proposed ML	99.76	99.32	98.76	99.76

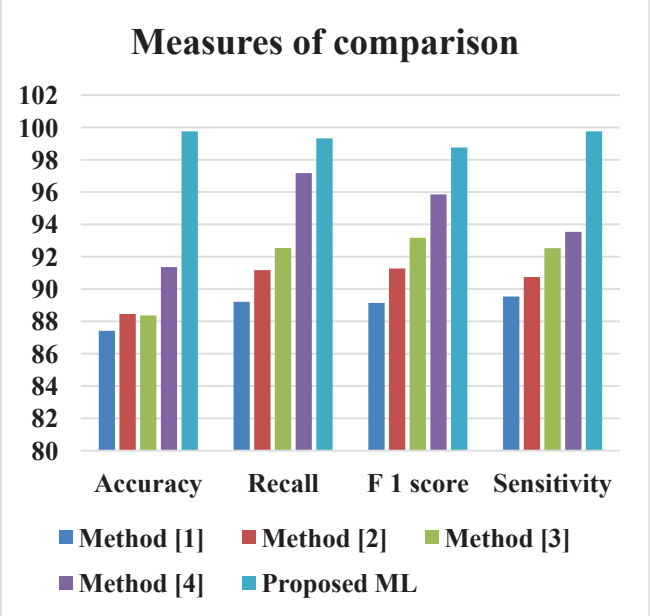

Fig. 5. Measures of comparison

The above table 2 and figure 5 describes that various methods comparision, in this proposed methodology attains more improvement comapred to earlier techniques.

IV. Conclusion

A significant role in video surveillance systems is played by the categorization of moving objects. It is, however, sensitive to the features that are utilised and the amount of classes that are created. Researchers recommended the use of many characteristics in conjunction with a feature selection approach in order to tackle this challenge. Others added additional features, and the majority of them made use of the binary categorization system.

They do, however, have a number of drawbacks. We suggest this work a unique model based on the merging of data that may enhance the high degree of categorization reliability by a factor of two or three. Also, we provide a unique technique to 3-dimensional object identification in the in the topic of self-driving vehicles. If distinguish ourselves from the majority of previous work, we make use of stereo images to develop a collection of two 3-dimensional object suggestions, that are after fed into a CNNs to provide high level quality of 3-dimensional object detections.

Rex Net is a saliency map generator that creates saliency maps from end - end & crisp object edges. The picture is initially divided into two sizes of complimentary areas, which are called super pixel boundaries and edge points, in the proposed framework.

The network after provides a striking result for each area from end - end, and factors across several levels is regarded to be fused with the striking ratings for each region in turn. For the first time, the suggested Rex Net accomplishes both unambiguous detection boundaries and multi-scale contextual resilience at the same time, resulting in an optimal performance.

References

[1] Q. Hu, S. Paisitkriangkrai, C. Shen, A. van den Hengel and F. Porikli, "Fast Detection of Multiple Objects in Traffic Scenes With a Common Detection Framework," in IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 4, pp. 1002-1014, April 2016, doi: 10.1109/TITS.2015.2496795.

[2] Q. Hou, M. Cheng, X. Hu, A. Borji, Z. Tu and P. H. S. Torr, "Deeply Supervised Salient Object Detection with Short Connections," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 4, pp. 815-828, 1 April 2019, doi: 10.1109/TPAMI.2018.2815688.

[3] S. Du and S. Chen, "Salient Object Detection via Random Forest," in IEEE Signal Processing Letters, vol. 21, no. 1, pp. 51-54, Jan. 2014, doi: 10.1109/LSP.2013.2290547.

[4] P. Araújo, J. Fontinele and L. Oliveira, "Multi-Perspective Object Detection for Remote Criminal Analysis Using Drones," in IEEE Geoscience and Remote Sensing Letters, vol. 17, no. 7, pp. 1283-1286, July 2020, doi: 10.1109/LGRS.2019.2940546.

[5] H. Park, S. Park and Y. Joo, "Detection of Abandoned and Stolen Objects Based on Dual Background Model and Mask R-CNN," in IEEE Access, vol. 8, pp. 80010-80019, 2020, doi: 10.1109/ACCESS.2020.2990618.

[6] V.V.S. Tallapragada and E.G. Rajan, "Improved Kernel based Iris Recognition system in the Framework of Support Vector Machine and Hidden Markov Model", IET Image Processing, vol. 6, no. 6, pp. 661-667, 2012.

[7] P. Tang, C. Wang, X. Wang, W. Liu, W. Zeng and J. Wang, "Object Detection in Videos by High Quality Object Linking," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 5, pp. 1272-1278, 1 May 2020, doi: 10.1109/TPAMI.2019.2910529.

[8] X. Chen, K. Kundu, Y. Zhu, H. Ma, S. Fidler and R. Urtasun, "3D Object Proposals Using Stereo Imagery for Accurate Object Class Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 5, pp. 1259-1272, 1 May 2018, doi: 10.1109/TPAMI.2017.2706685.

[9] H. Guan, Y. Yu, J. Li and P. Liu, "Pole-Like Road Object Detection in Mobile LiDAR Data via Supervoxel and Bag-of-Contextual-Visual-Words Representation," in IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 4, pp. 520-524, April 2016, doi: 10.1109/LGRS.2016.2521684.

[10] Y. Shen, R. Ji, C. Wang, X. Li and X. Li, "Weakly Supervised Object Detection via Object-Specific Pixel Gradient," in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 12, pp. 5960-5970, Dec. 2018, doi: 10.1109/TNNLS.2018.2816021.

[11] S. Matteoli, G. Corsini, M. Diani, G. Cecchi and G. Toci, "Automated Underwater Object Recognition by Means of Fluorescence LIDAR," in IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 1, pp. 375-393, Jan. 2015, doi: 10.1109/TGRS.2014.2322676.

[12] X. Chen, J. Guan, G. Wang, H. Ding and Y. Huang, "Fast and Refined Processing of Radar Maneuvering Target Based on Hierarchical Detection via Sparse Fractional Representation," in IEEE Access, vol. 7, pp. 149878-149889, 2019, doi: 10.1109/ACCESS.2019.2947169.

[13] Y. R. Pandeya, B. Bhattarai and J. Lee, "Visual Object Detector for Cow Sound Event Detection," in IEEE Access, vol. 8, pp. 162625-162633, 2020, doi: 10.1109/ACCESS.2020.3022058.

[14] X. Wang, H. Ma, X. Chen and S. You, "Edge Preserving and Multi-Scale Contextual Neural Network for Salient Object Detection," in IEEE Transactions on Image Processing, vol. 27, no. 1, pp. 121-134, Jan. 2018, doi: 10.1109/TIP.2017.2756825.

[15] Y. Tang et al., "Visual and Semantic Knowledge Transfer for Large Scale Semi-Supervised Object Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 12, pp. 3045-3058, 1 Dec. 2018, doi: 10.1109/TPAMI.2017.2771779.

[16] Y. Peng, X. He and J. Zhao, "Object-Part Attention Model for Fine-Grained Image Classification," in IEEE Transactions on Image Processing, vol. 27, no. 3, pp. 1487-1500, March 2018, doi: 10.1109/TIP.2017.2774041.

[17] L. Zhang, H. Lu, H. Zhang, Y. Zhao, H. Xu and J. Wang, "Hyper-Spectral Characteristics in Support of Object Classification and Verification," in IEEE Access, vol. 7, pp. 119420-119429, 2019, doi: 10.1109/ACCESS.2019.2936130.

[18] Z. Hao, J. Wu, T. Liu and X. Chen, "Leveraging Cognitive Context Knowledge for Argumentation-Based Object Classification in Multi-Sensor Networks," in IEEE Access, vol. 7, pp. 71361-71373, 2019, doi: 10.1109/ACCESS.2019.2919073.

[19] H. Zhou, A. Liu, W. Nie and J. Nie, "Multi-View Saliency Guided Deep Neural Network for 3-D Object Retrieval and Classification," in IEEE Transactions on Multimedia, vol. 22, no. 6, pp. 1496-1506, June 2020, doi: 10.1109/TMM.2019.2943740.

[20] H. -U. Kim, Y. J. Koh and C. -S. Kim, "Online Multiple Object Tracking Based on Open-Set Few-Shot Learning," in IEEE Access, vol. 8, pp. 190312-190326, 2020, doi: 10.1109/ACCESS.2020.3032252.

[21] S. Steyer, C. Lenk, D. Kellner, G. Tanzmeister and D. Wollherr, "Grid-Based Object Tracking With Nonlinear Dynamic State and Shape Estimation," in IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 7, pp. 2874-2893, July 2020, doi: 10.1109/TITS.2019.2921248.

[22] Mythreya, S., Murthy, A. S. D., Saikumar, K., & Rajesh, V. (2022). Prediction and Prevention of Malicious URL Using ML and LR Techniques for Network Security: Machine Learning. In Handbook of Research on Technologies and Systems for E-Collaboration During Global Crises (pp. 302-315). IGI Global.

[23] Saikumar, K., Rajesh, V., Babu, B.S. (2022). Heart disease detection based on feature fusion technique with augmented classification using deep learning technology. Traitement du Signal, Vol. 39, No. 1, pp. 31-42. https://doi.org/10.18280/ts.390104

[24] Kailasam, S., Achanta, S.D.M., Rama Koteswara Rao, P., Vatambeti, R., Kayam, S. (2022). An IoT-based agriculture maintenance using pervasive computing with machine learning technique. International Journal of Intelligent Computing and Cybernetics, 15(2), pp. 184–197.

[25] Saikumar, K., Rajesh, V. A machine intelligence technique for predicting cardiovascular disease (CVD) using Radiology Dataset. Int J Syst Assur Eng Manag (2022). https://doi.org/10.1007/s13198-022-01681-7