**Journal of Artificial Intelligence, Machine Learning and Neural Network**
Vol: 03, No. 01, Jan-June 2023
ISSN: 2799-1172, DOI: 10.55529/jaimlnn.32.1.13 

**Research Paper**

# Convolutional neural networks for object detection and recognition

**Ms. Archana Karne[1], Mr. Radha Krishna Karne[2*], Mr. V. Karthik Kumar[3], Dr. A. Arunkumar[4]**
[1]UG Student, CMR Technical Campus, Hyderabad, India.
[2*]Assistant Professor in ECE, CMR Institute of Technology, Hyderabad, India.
[3]Assistant Professor in ECE, BITS Narsampet, Hyderabad, India.
[4]Professor in CSE, MLRITM, Hyderabad, India.

**Article Info**

***Article History:***
Received: 27 October 2022
Revised: 15 January 2023
Accepted: 22 January 2023
Published: 08 March 2023

***Keywords:***
Convolutional Neural Network
Deep Learning
Object Detection
Deep Neural Networks
Object Recognition
Object Classification

**ABSTRACT**

One of the essential technologies in the fields of target extraction, pattern recognition, and motion measurement is moving object detection. Finding moving objects or a number of moving objects across a series of frames is called object tracking. Basically, object tracking is a difficult task. Unexpected changes in the surroundings, an item's mobility, noise, etc., might make it difficult to follow an object. Different tracking methods have been developed to solve these issues. This paper discusses a number of object tracking and detection approaches. The major methods for identifying objects in images will be discussed in this paper. Recent years have seen impressive advancements in fields like pattern recognition and machine learning, both of which use convolutional neural networks (CNNs). It is mostly caused by graphics processing units'(GPUs) enhanced parallel processing capacity. This article describes many kinds of object classification, object racking, and object detection techniques. Our results showed that the suggested algorithm can detect moving objects reliably and efficiently in a variety of situations.

*Corresponding Author:*
Mr. Radha Krishna Karne
Assistant Professor in ECE, CMR Institute of Technology, Hyderabad, India.
Email: krk.wgl@gmail.com

## 1. INTRODUCTION

The difficult yet important task of detecting and following moving objects in a video surveillance system. In today's society, a quick video surveillance system is necessary. The purpose of the video surveillance system is to be focused on the identification, categorization, and tracking of moving objects. Almost every industry, including the military, workplaces, banks, and schools, uses moving object detection and tracking for security purposes. The use of automatic video surveillance is crucial in the security industry [1]. The use of radar and image processing technologies for the purpose of accurately identifying and following moving objects in this study, moving object recognition and tracking are accomplished using image processing technologies. For every object detection and tracking system, detection, categorization, and tracking are the three crucial processes. The items of interest are initially removed from the series of frames, separated from the backdrop, and then tracked while retaining their identity in subsequent frames. The object recognition and tracking method is applicable to more than only video surveillance systems [2], [3], it may also be used in multimedia databases, virtual reality, video compression, human-machine interfaces, and other areas.

The finest fundamental quality of a person is frequently their capacity for seeing. Our capacity to see with our eyes is thought of as a gift and is crucial to how we go about our daily lives. The fact that many visually impaired persons are constrained by their eyesight and unable to be fully independent is a significant difficulty. People who are visually impaired have difficulty with these activities, thus object recognition will be a crucial characteristic they may rely on frequently. They frequently encounter difficulties moving around and recognizing objects, especially while going on streets. The majority of visually-capable people reach middle age at 50 [4]. A small number of applications to help the blind have entered the market. However, there is still a lack of modernization in the "visually impaired" people's lack of real-time continuous article acknowledgement and object identification with speech output. With the usage of IoT [5], [6], [7], [8], [9], [10], some of these apps are focused on detecting obstacles close to the user and warning them via alarms or blaring noises. Numerous gadgets are needed by consumers to hold for various reasons. For example, navigation aid requires smart sticks with obstacle detectors, cell phones, navigators, etc. These gadgets are pricey and may occasionally cause the user difficulty.

Computer vision has several exciting challenges, such classifying images and identifying objects, both of which fall under the umbrella of object recognition. In recent years, there has been significant scientific progress for these kinds of problems, mostly because CNN, DL approaches, and the rise in parallelism processing power provided by GUPs have all advanced. The goal of the image classification issue is to select a label from a predetermined list of categories to apply to an input image. The labelling of images of skin cancer [11] and the use of high-resolution images to detect natural disasters like floods, volcanoes, and severe droughts while noting the impacts and damage caused by these events are just two examples of the many practical applications and uses of this classification problem, which is central to computer vision. The characteristics that are fed into image classification algorithms have a critical role in how well they function [12]. This indicates that the development of machine learning-based image categorization approaches strongly depended on the engineering of identifying the crucial aspects of the photos that made up the database. As a result, getting these resources has grown to be a difficult undertaking, increasing complexity and expense. When seen as a component of the supervised learning strategy, the support vector machines (SVM) algorithm is frequently employed for tasks including classification, regression, and outlier identification [13]. The system's learning process for numerous objects can be mathematically examined more easily than standard neural network design, which makes it possible to make sophisticated changes to the algorithm with predictable results [14]. A nonlinear separation barrier in the input space is produced by an SVM's basic mapping of the training data to higher-dimensional feature space and separation hyperplane with maximum margin [15]. Deep learning architectures with several specialized layers for automating the filtering and feature extraction process are used by today's most reliable object categorization and detection systems. A prediction

is made, a correction is received, and the prediction mechanism is adjusted based on the correction. At a high level, this is very similar to how a human learns. Machine learning algorithms like linear regression, support vector machines, and decision trees all have their own peculiarities in the learning process. The advent of deep learning has introduced a fresh perspective on the issue, one that aimed to address past limitations by learning abstraction in data via a stratified description paradigm built on a nonlinear transformation [16]. Due to the widespread usage of CNN, DL-based algorithms may learn the feature extraction stage (ConvNet or CNN). Convolution is a specific kind of linear operation that may be viewed in this sense as the straightforward application of a filter to a predetermined input [17]. By adjusting the convolution's parameters, the same filter is applied to an input repeatedly to produce a feature map that shows the positions and intensities of any discovered features. As a result, the network may learn the ideal parameters to extract pertinent data from the database by adjusting itself to decrease error. There have been several deep neural network (DNN)-based object detectors developed in recent years [18], [19]. In order to demonstrate how state-of-the-art DNN models of SSD and Faster RCNN function in scientific research, the YOLO network was trained to solve the mice tracking issue. The algorithms were taught to recognize a variety of animals in images for the traditional detection challenge.

## 1.1 Object Detecting, Classification and Tracking Methods

### 1.1.1 Object Detection Methods

Finding and recognizing objects in an image or video sequence is the challenge of object detection. Even when partially obscured from vision, objects may still be identified. Computer vision systems are still having trouble with this task. For object detection, several techniques have been developed throughout the years.

### 1.1.2 Background Subtraction Method

The foreground of an image is retrieved for further processing using the technique known as background subtraction, which is also known as foreground detection. When morphological object localization is needed as post processing following the step of picture preprocessing, which includes image noise removal, this approach may be used. A common technique for identifying moving objects in films taken with a stationary camera is background removal. Backdrop images or background models are used to identify moving objects by comparing the differences between the current frame and a reference frame. The approach of background subtraction is weak at blocking interference and sensitive to changes in the environment.

### 1.1.3 Optical Flow Method

The distribution of the objects' apparent velocities inside a picture is known as optical flow. The velocities of the objects in the movie may be calculated by calculating optical flow between video frames. In general, moving objects will appear to move more visibly the closer they are to the camera and the faster they are travelling than farther away. For motion-based object recognition and tracking systems, optical flow estimation is used in computer vision to characterize and quantify motion of objects in video streams.

### 1.1.4 Frame Differencing Method

By using a technique called frame differencing, the computer can determine if two video frames differ from one another.

## 1.2 Single Shot Multibox Detector (SSD)

One of the best detectors in terms of speed and accuracy is the SSD [20], which uses convolutional filter applications and feature map extraction as its two primary processes for object detection. The VGG-16 network [21], on which the SSD design is based, has great performance in high-quality image classification tasks and

enjoys widespread use in challenges involving transfer learning. The convolutional kernel application to an input picture in the SSD architecture is shown in Figure 1. A collection of auxiliary convolutional layers replace the initial VGG fully connected layers in the model, enabling the extraction of features at various scales and progressively reducing the size of the input to each succeeding layer. The use of precomputed, fixed-size bounding boxes known as priors to the initial distribution of ground truth boxes is taken into account during the bounding box creation. These priors are chosen to maintain an intersection over union (IoU) ratio of 0:5 or above.

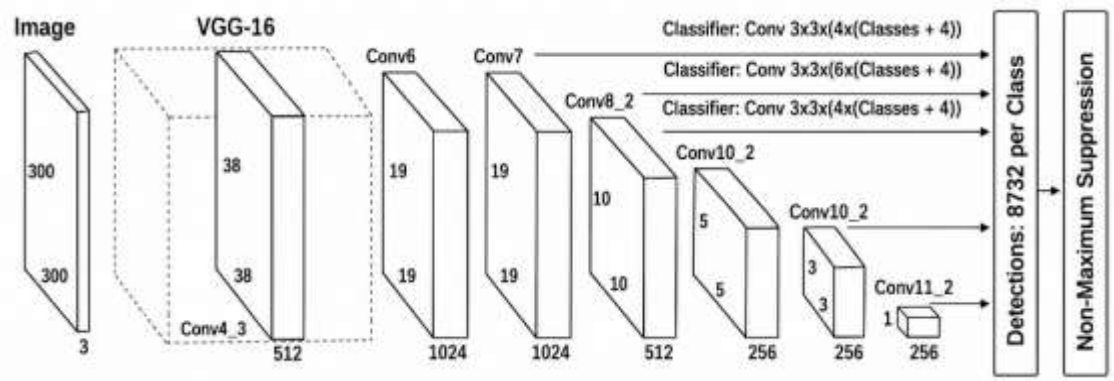



Figure 1. To the End of the Basic Network, the SSD Network Contains Various Feature Levels

### 1.3 Faster Region Convolutional Neural Network (Fast RCNN)

Another cutting-edge CNN-based deep learning object identification method is the Fast RCNN [22]. In this design, a convolutional network is used to create a convolutional feature map from the input picture. A different network is utilized to learn and forecast these regions rather than utilizing the selective search algorithm to detect the region recommendations made in earlier rounds [23], [24]. A region of interest (ROI) pooling layer is then used to reshape the projected region proposals, categorize the image within the proposed region, and forecast the offset values for the bounding boxes. The region proposal network (RPN) training technique uses a binary label for each anchor, with one denoting an object's existence and zero denoting its absence. According to this strategy, any IoU over 0:7 identifies an object's presence, while any value below 0:3 denotes an object's absence. The unified network for object detection used in the Faster RCNN architecture is shown in Figure 2. The region proposal network module instructs the Fast RCNN module where to seek using the currently fashionable language of neural networks with "attention" processes [25].

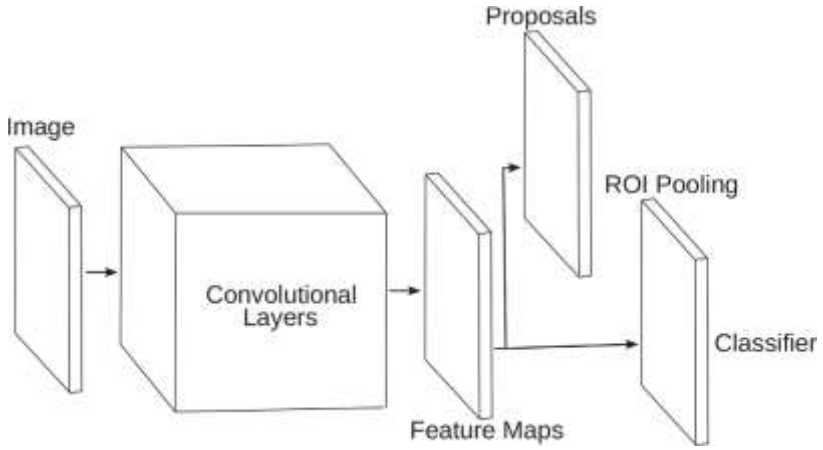



Figure 2. For Object Detection, Faster RCNN Functions As a Single, Integrated Network [22]

### 1.4 Object Classification Method

The object is classed by assigning each object to a class based on characteristics once the object of interest has been identified from a sequence of frames. Features are defined in a broad variety of ways. The feature specifies the target image's form, size, color, and motion for this purpose. The following features are employed in object tracking.

- **Edges:** Image intensities frequently undergo significant fluctuations at object borders. This change's identification is done using edge detections. The fact that edges are less responsive to changes in light than colour characteristics is a significant characteristic of edges [26].
- **Motion:** A major indication for classifying moving objects has been the periodic behaviour of non-rigid articulated item motion [27]. Object motion may also be tracked using optical flow.
- **Color:** All video frame formats are built on a concept of several colour spaces. Different frame data can be recorded in several colour formats, including grayscale, RGB, YCbCr, and HSB. The letters read (R), green (G), blue (B), or RGB are used to indicate colour pictures.
- **Texture:** Texture is utilized to help identify the subject or object of interest. It evaluates characteristics like smoothness and regularity of a surface by measuring intensity variation of that surface.

### 1.5 Object Tracking Method

Target tracking looks for an object's location in each frame of video in order to build the root for that object above time. There are three types of object tracking: silhouette-based tracking, point tracking, and kernel tracking.

**Point Tracking Method:** Veenman created point tracking, a dependable, strong, and precise tracking technique. Their feature points serve as a representation for moving objects. Point tracking techniques are further separated into two groups: deterministic and statistical. The foundation of object tracking is a point that is depicted in an object that is detected in a subsequent frame, and the association of points is based on the prior state of the item. To identify an item in every frame, an external mechanism is needed.

**Kernel Tracking:** Kernel describes an object's representation of its elliptical or rectangular form and appearance. From frame to frame, the object's motion is calculated. In succeeding frames, the object's motion is represented by parametric motion or dence flow field computation. Simple template matching, mean shift technique support vector machines, and layering-based tracking are further categories for kernel tracking. Tracking may be done for a single item in a video using basic template matching, which verifies a reference picture with a frame that has been taken out of the film. Tracking the motion of the item is done using translation and scaling, and the object of interest is defined using a rectangular frame. After that, the tracked item and backdrop are separated. SVMs need a training set of values. Positive values are used to contain these, whereas negative values are used to contain targets that are not being monitored. In layering-based tracking, many objects may track.

**Silhouette Tracking:** Tracking using silhouettes is utilized when a whole object region is needed. Numerous items, such as the human body, head, and hand, have complicated geometries that may be correctly represented using a silhouette-based technique. By utilizing an object model created from previous frames, the silhouette tracker seeks out the object region in each frame. Shape matching and Contour tracking are two subcategories of silhouette tracking.

### 1.6 You Only Look Once (YOLO) Algorithm

YOLOis a cutting-edge object identification method designed for real-time applications; in contrast to some of its rivals, it is not a conventional classifier used for object detection [28]. In order for YOLO to function,

the input picture is split up into a grid of SxS cells, where each cell is in charge of five bounding box predictions that characterize the rectangle around the item. Additionally, it generates a confidence score, which expresses how confidently an object was contained. Therefore, just the form of the box affects the score; the type of object in the box has no bearing on it. As with a typical classifier, a class is predicted for each anticipated bounding box, yielding a probability distribution over all potential classes [29], [30], [31]. One final score that indicates the likelihood that each box contains a certain type of item is created by combining the bounding box confidence score with the class prediction score. Due to these design decisions, the majority of the boxes will have low confidence ratings; thus, only the boxes with final scores that are higher than a certain threshold are maintained. How the YOLO network processes a picture is shown in Figure 3. The input is first processed through a CNN, which creates the bounding boxes with its viewpoints' confidence ratings and creates the class probability map. The final forecasts are created by combining the outcomes of the earlier processes.

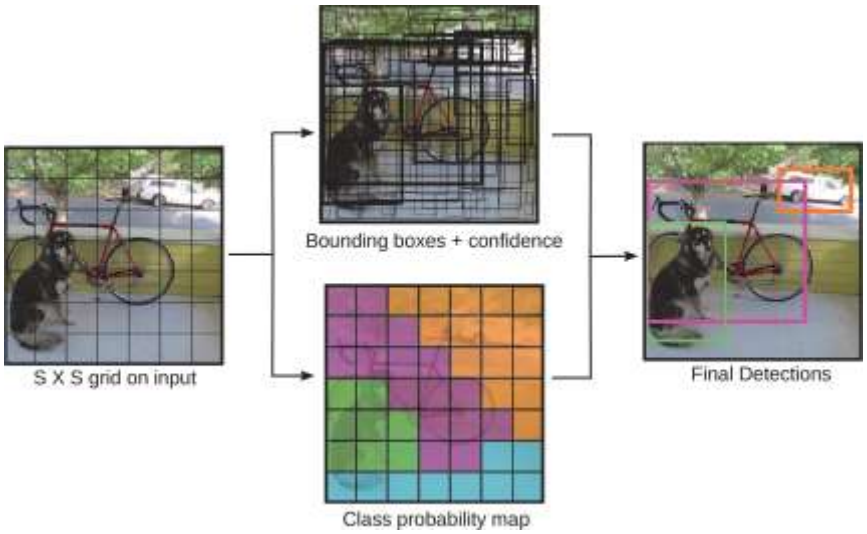



Figure 3. Model Detection for YOLO as a Regression Problem [32]

YOLO is an engaging article openness evaluation. Even if it isn't currently the most cautious article straightforwardness figure, it is a fantastic option when an unsurprising, undeniable need is needed without losing a significant amount of accuracy. YOLO employs a lone CNN to organize things utilizing swaying boxes for both gathering and limiting them. YOLOv2 provides high accuracy and dependable getting ready, but it has more confinement errors and a lower survey response than other area-based pointer checks. A resurrected version of YOLO, known as YOLOv2, beats the lower study response and produces accuracy with vivacious openness. The improvements in YOLOv2 are quickly examined under: The completely related layers that were submitted for the cutoff box expectation were discarded. To change the spatial yield of the framework from 7x7 to 13x13, one pooling layer has to be removed. Since classifiers anticipated that yield names would be fully distinct, the yield object classes were mostly unimportant. YOLO has the capacity to convert the aforementioned scores into probabilities as much as possible. YOLOv3 has a multi-name strategy. A score that is extraordinary can be seen in non-prohibitive yield inscriptions. Instead of utilizing the soft max work, YOLOv3 enrolls the likelihood of the data having a location with a certain cutting by applying free essential classifiers. YOLOv3 determines the framework disappointment using a scene that is shaped using cross-entropy for each name rather than the mean settled slip. Maintaining a few essential charming strategies from the soft max work reduces the complexity of the check. A common, quicker, and more grounded version of YOLO is YOLO-9000. The YOLO algorithm was initially proposed by Joseph Redmon and his colleagues. In 2015, he released a paper on YOLO with the working title "You Only Look Once" Real-Time item recognition, and it became an instant hit. CNN is followed by YOLO. When making predictions, the algorithm only "looks once" at

the image since there is only one propagation that occurs throughout the neural network. Compared to other methods of object identification, the YOLO model is the fastest and most effective. The main benefit of YOLO is its quickness. There are 45 frames per second in this. The model is constructed in a concise manner that allows its network to become used to abstract descriptions of things [33].

### 1.7 Evaluation Parameters

In this paper, we evaluated the effectiveness of the moving object identification method using the precision (P), recall (R), and F1 measures. The performance of the object identification model was assessed using the mean average precision (mAP). Based on the true category and the detection category, the detection results were split into four cases: true positive (TP), false positive (FP), false negative (FN), and true negative (TN).

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2PR}{P + R}$$

## 2. RESULTS AND DISCUSSION

The use of the SSD and RCNN object detection techniques is demonstrated using a portion of the PASCAL VOC [34] dataset. Six classes, out of the 20 offered, were chosen as a sample. The sample size used for each class is shown in Figure 4. The dataset's photos were randomly divided into 1911 for training, which corresponds to 50%, 1126 for validation, which corresponds to 25%, and 1126 for test, which also corresponds to 25%. The dataset utilized for the YOLO network published in [35] was also examined to further highlight the uses of these algorithms in academic study. The dataset, which is presented in [36], is made up of photographs from three studies that involve behavioural trials on mice. The sample size chosen from each of the datasets utilized in this work is shown in Figure 6. A total of 3707 frames from a top view of the mouse social interaction experiment chamber were utilized for the ethological evaluation [36]. A sample of 3073 frames was chosen for the automated home-cage [37] from a side perspective of behavioural studies. A selection of 6842 frames from the CRIM13 [38], including 3492 from the side view and 3350 from the top view, were chosen.

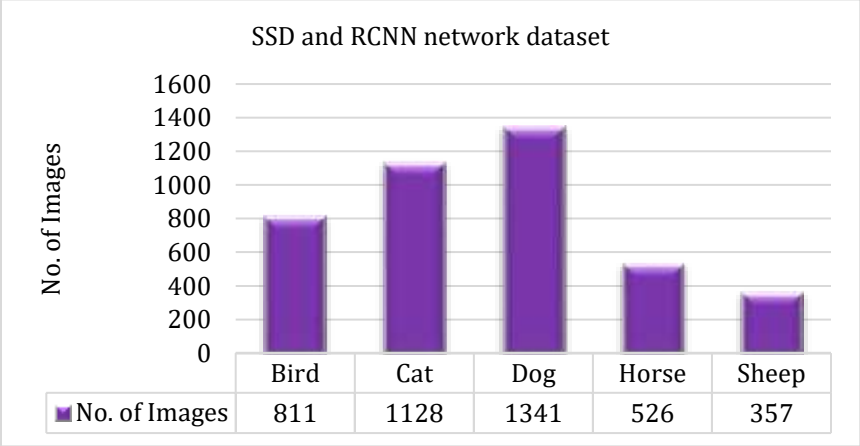



Figure 4. Description of the SSD and RCNN Network Datasets

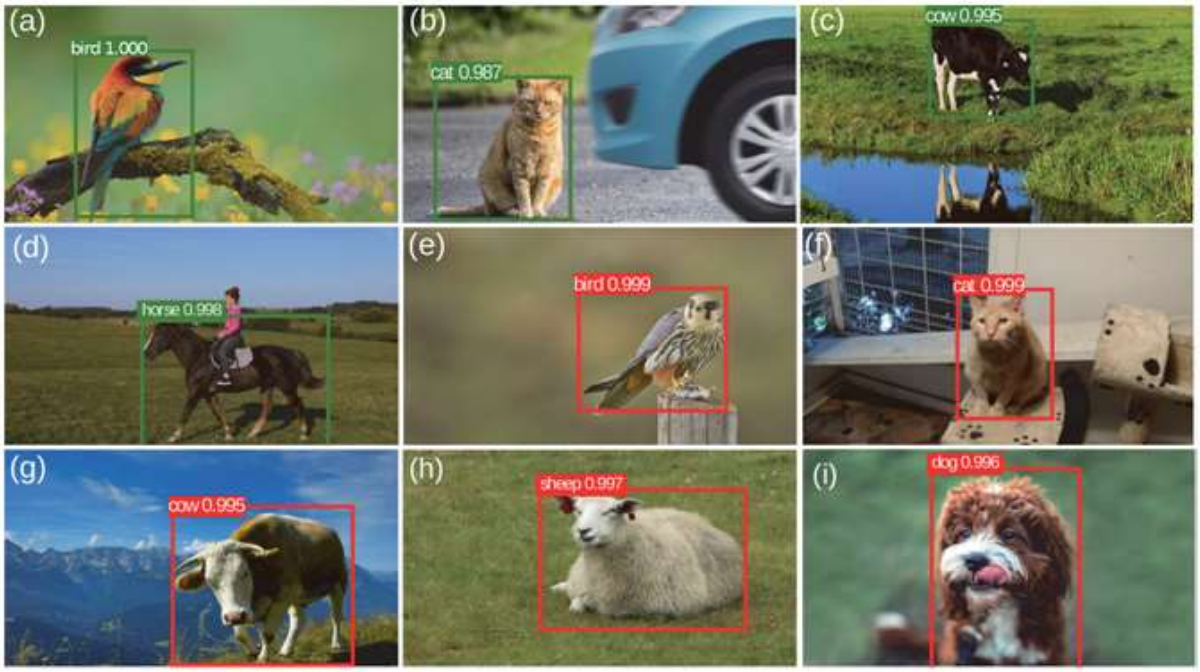



Figure 5. Examples of the SSD (a-d) and Faster RCNN (e-i) Networks' Output

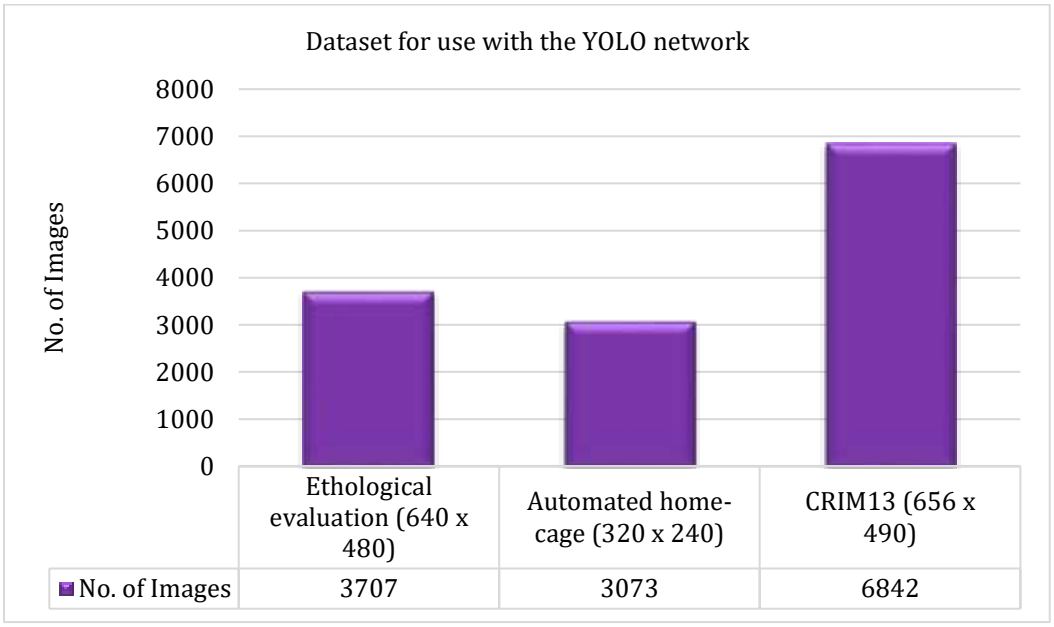



Figure 6. An Explanation of the Dataset Used With the YOLO Network

More results on the effectiveness of object detection is shown in Figure 7. In the beginning, it displays the mean average precision, which is the mean value of the average precisions for each class. Average precision is the average value of 11 points on the precision-recall curve for each potential threshold, or all the probabilities of detection for the same class. We examined the models' performance in terms of accuracy, speed, and model size, shown in Figure 8.

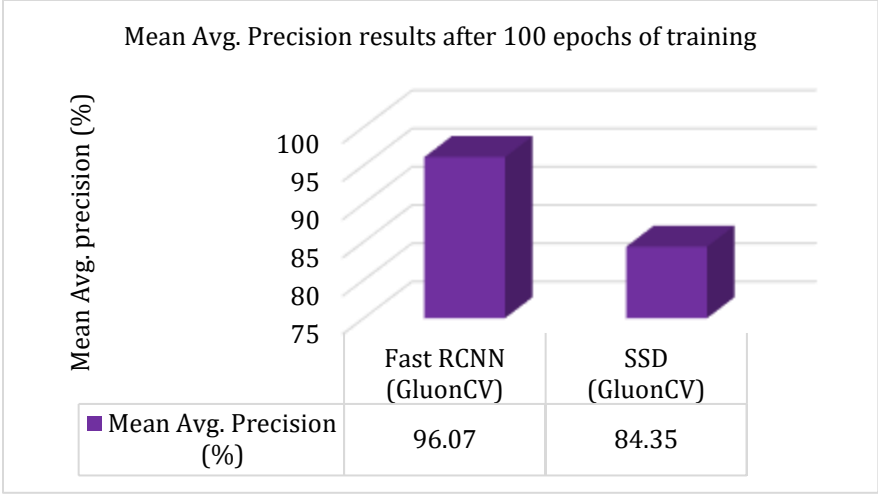
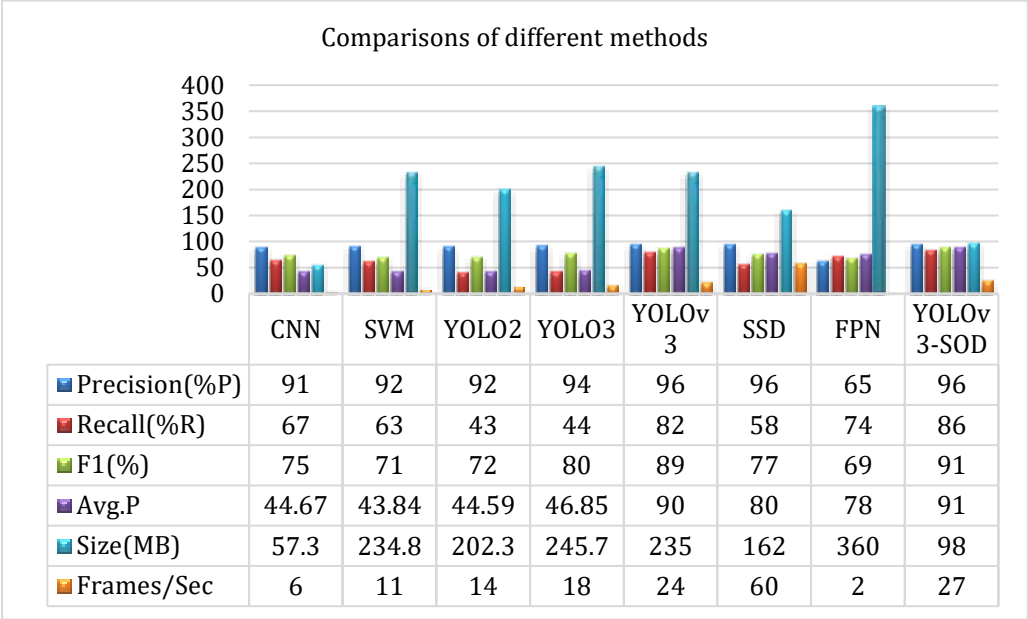



Figure 7. Results of the Mean Average Precision after 100 Training Iterations




| | CNN | SVM | YOLO2 | YOLO3 | YOLOv 3 | SSD | FPN | YOLOv 3-SOD |
|---|---|---|---|---|---|---|---|---|
| Precision(%P) | 91 | 92 | 92 | 94 | 96 | 96 | 65 | 96 |
| Recall(%R) | 67 | 63 | 43 | 44 | 82 | 58 | 74 | 86 |
| F1(%) | 75 | 71 | 72 | 80 | 89 | 77 | 69 | 91 |
| Avg.P | 44.67 | 43.84 | 44.59 | 46.85 | 90 | 80 | 78 | 91 |
| Size(MB) | 57.3 | 234.8 | 202.3 | 245.7 | 235 | 162 | 360 | 98 |
| Frames/Sec | 6 | 11 | 14 | 18 | 24 | 60 | 2 | 27 |

Figure 8. Comparisons of Different Methods

## 3. CONCLUSION

In this paper, we presented a novel approach to the detection and identification of moving objects, and also an overview of the related literature as well as an object tracking literature study are presented in this paper. The three kinds of tracking techniques are object detection, object classification, and object tracking. The future of object detection offers tremendous prospects in a variety of businesses. Based on the resources available, methods for object detection and video processing are presented. The experiment results of CNN, SVM, SSD, FPN, YOLO2 and YOLO3, were compared by means of precision, recall, F1-score, average precision and generating frames per sec. Results of the experiments show that the suggested approach accurately and successfully detects moving objects. As future work,, we will concentrate on developing a moving object detection algorithm that is more reliable and integrating it into embedded surveillance application systems.

## Acknowledgments

The authors have no specific acknowledgments to make for this research.

## Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## Author Contributions Statement

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ms. Archana Karne | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | | | |
| Mr. Radha Krishna Karne | | | | | | | | | | | | | | |
| Mr. V. Karthik Kumar | | | | | | | | | | | | | | |
| Dr. A. Arunkumar | | | | | | | | | | | | | | |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

## Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Informed Consent

All participants were informed about the purpose of the study and their voluntary consent was obtained prior to data collection.

## Ethical Approval

The study was conducted in compliance with the ethical principles outlined in the Declaration of Helsinki and approved by the relevant institutional authorities.

## Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

# REFERENCES

[1] A. Kumar and R. K. Arun, 'IIoT-IDS Network using Inception CNN Model', Journal of Trends in Computer Science and Smart Technology, vol. 4, pp. 126-138, 2022. doi.org/10.36548/jtcsst.2022.3.002

[2] R. Karne and T. K. Sreeja, 'ROUTING PROTOCOLS IN VEHICULAR ADHOC NETWORKS (VANETs)', International Journal of Early Childhood.

[3] K. Vaigandla, S. Kumar, and R. K. Thatipamula, 'Investigation on Unmanned Aerial Vehicle (UAV): An Overview', IRO Journal on Sustainable Wireless Systems, vol. 4, pp. 130-148, 2022. doi.org/10.36548/jsws.2022.3.001

[4] R. Karne and D. Sreeja, 'COINV-Chances and Obstacles Interpretation to Carry new approaches in the VANET Communications', Communications" Design Engineering, pp. 10346-10361, 2021.

[5] K. K. Vaigandla, 'Communication technologies and challenges on 6G networks for the internet: Internet of things (IoT) based analysis', in 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 2022, pp. 27-31. doi.org/10.1109/ICIPTM54933.2022.9753990

[6] R. Karne and T. K. Dr, 'Review On Vanet Architecture And Applications', Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, pp. 1745–1749, 2021.

[7] R. K. Karne and Muralidharan, 'A novel approach for dynamic stable clustering in VANET using deep learning (LSTM) model', International Journal of Electrical and Electronics Research, vol. 10, no. 4, pp. 1092-1098, Dec. 2022. doi.org/10.37391/ijeer.100454

[8] S. Singh Sengar and S. Mukhopadhyay, 'Motion Detection using Block based Bi-directional Optical Flow Method', Journal of Visual Communication and Image Representation, vol. 49, pp. 89-103, Aug. 2017. doi.org/10.1016/j.jvcir.2017.08.007

[9] S. Singh Sengar and S. Mukhopadhyay, 'Moving Object Detection based on Frame Difference and W4', in Signal, Image and Video Processing, Springer, 2017, pp. 1357-1364. doi.org/10.1007/s11760-017-1093-8

[10] R. Karne, 'Simulation of ACO for Shortest Path Finding Using NS2', pp. 12866-12873, 2021.

[11] Esteva A et al. Dermatologist-levelclassification of skin cancer with deepneural networks. Nature. 2017;542(7639):115 doi.org/10.1038/nature21056

[12] S. Srinivas, R. K. Sarvadevabhatla, R. K. Mopuri, N. Prabhu, and S. Kruthiventi, 'Venkatesh Babu R. An introduction todeep convolutional neural nets forcomputer vision', Deep Learning for Medical Image Analysis, pp. 25-52, 2017. doi.org/10.1016/B978-0-12-810408-8.00003-1

[13] R. De Menezes, L. De Azevedo Lima, O. Santana, A. M. Henriques-Alves, R. M. Santacruz, and H. Maia, 'Classification of micehead orientation using support vectormachine and histogram of orientedgradients features', in 2018International Joint Conference onNeural Networks (IJCNN), IEEE, 2018, pp. 1-6. doi.org/10.1109/IJCNN.2018.8489558

[14] J. Q. Gan and H. Hu, 'Adaptiveschemes applied to online SVM forBCI data classification', in 2009Annual International Conference ofthe IEEE, 2009, pp. 2600-2603. doi.org/10.1109/IEMBS.2009.5335328

[15] M. A. Hearst et al., 'Support vectormachines', IEEE Intelligent Systems andtheir Applications, vol. 13, 1998. doi.org/10.1109/5254.683174

[16] 'Classification of malaria-infected cellsusing deep convolutional neuralnetworks', in Machine Learning:Advanced Techniques and EmergingApplications, 2018.

[17] Goodfellow I, Bengio Y,Courville A. Deep Learning. MIT Press;2016

[18] L. Deng, G. Hinton, and B. Kingsbury, 'New types of deep neural networklearning for speech recognition andrelated applications: An overview', in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2013, pp. 8599-8603. doi.org/10.1109/ICASSP.2013.6639344

[19] N. Kriegeskorte, 'Deep neuralnetworks: A new framework formodeling biological vision and braininformation processing', Annual Reviewof Vision Science, vol. 1, pp. 417-446, 2015. doi.org/10.1146/annurev-vision-082114-035447

[20] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, and C.-Y. Fu, 'SSD: Single shot multibox detector', in European Conference on Computer Vision, Cham: Springer, 2016, pp. 21-37. doi.org/10.1007/978-3-319-46448-0_2

[21] S. Singh Sengar, 'Motion segmentation based on structure-texture decomposition and improved three frame differencing', in 15th International Conference on Artificial Intelligence Applications and Innovations, Crete, Greece: Springer, 2019, pp. 609-622. doi.org/10.1007/978-3-030-19823-7_51

[22] S. Ren, K. He, R. Girshick, and J. Sun, 'Fasterr-cnn: Towards real-time object detection with region proposal networks', in Advances in Neural InformationProcessing Systems, 2015, pp. 91-99.

[23] Girshick R, Donahue J, Darrell T,Malik J. Rich feature hierarchies foraccurate object detection and semanticsegmentation. In: Proceedings of theIEEE Conference on Computer Visionand Pattern Recognition. 2014.p. 580587 doi.org/10.1109/CVPR.2014.81

[24] R. Girshick, 'Fast r-cnn', in Proceedings of the IEEE, 2015, pp. 1440-1448. doi.org/10.1109/ICCV.2015.169

[25] J. K. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, 'Attentionbasedmodels for speech recognition', in Advances in Neural InformationProcessing Systems, 2015, pp. 577-585.

[26] A. Kinjal and D. G. Joshi, 'ASurvey on Moving Object Detection AndTracking in Video Surveillance System', International Journal of Soft Computing andEngineering (IJSCE), no. 2, July 2012.

[27] S. Himani, G. Darshak, and U. K. Thakore, 'A Survey on ObjectDetection and Tracking Methods', International Journal of Innovative Researchin Computer and CommunicationEngineering, vol. 2, no. 2, Feb. 2014.

[28] Redmon J, Farhadi A. Yolov3: AnIncremental Improvement. arXiv; 2018

[29] S. Singh Sengar and S. Mukhopadhyay, 'Foreground Detection via Background Subtraction and Improved Three-frame Differencing', Arabian Journal for Science and Engineering, vol. 42, no. 8, pp. 3621-3633, June 2017. doi.org/10.1007/s13369-017-2672-2

[30] P. Shyam, S. Singh Sengar, K.-J. Yoon, and K.-S. Kim, 'Robust Video Enhancement by Adversarial Evaluation of Inter-Frame consistency and Integrated within Camera-ISP', in the 32nd British Machine Vision Conference, 2021.

[31] P. Shyam, S. Singh Sengar, K.-J. Yoon, and K.-S. Kim, 'Exploring Data Efficient Techniques for Image Restoration and Enhancement', in International Joint Conference on Artificial Intelligence Workshop - Artificial Intelligence for Autonomous Driving, Montreal, Canada, 2021.

[32] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, 'You only look once: Unified,real-time object detection', in Proceedings of the IEEE Conference onComputer Vision and PatternRecognition, 2016, pp. 779-788. doi.org/10.1109/CVPR.2016.91

[33] S. Thorat and M. Nagmode, 'Detectionand Tracking of Moving Objects', International Journal of Innovative Researchin Advanced Engineering (IJIRAE), vol. 1, no. 1, Apr. 2014.

[34] M. Everingham, 'The Pascalvisual object classes (VOC) challenge', International Journal of ComputerVision, vol. 88, no. 2, pp. 303-338, 2010. doi.org/10.1007/s11263-009-0275-4

[35] H. M. Peixoto, R. S. Teles, J. Luiz, A. M. Henriques-Alves, and S. Cruz, 'Mice Tracking Using the YOLOAlgorithm', Mice Tracking Using the YOLOAlgorithm, vol. 7, pp. e27880-27881, 2019. doi.org/10.7287/peerj.preprints.27880v1

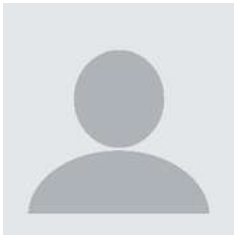
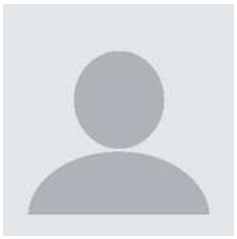
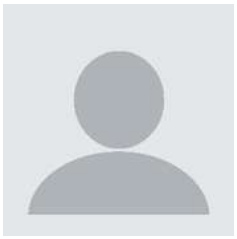
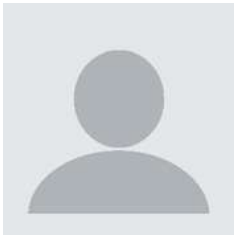
[36] A. M. Henriques-Alves and C. M. Queiroz, 'Ethological evaluation of the effects ofsocial defeat stress in mice: Beyond thesocial interaction ratio', Frontiers inBehavioral Neuroscience, vol. 9, 2016. doi.org/10.3389/fnbeh.2015.00364

[37] H. Jhuang, 'Automated homecagebehavioural phenotyping of mice', Nature Communications, vol. 1, 2010. doi.org/10.1038/ncomms1064

[38] X. P. Burgos-Artizzu, P. Dollár, D. Lin, D. J. Anderson, and P. Perona, 'Social behaviorrecognition in continuous video', in 2012 IEEE Conference on ComputerVision and Pattern Recognition, IEEE, 2012, pp. 1322-1329. doi.org/10.1109/CVPR.2012.6247817

**How to Cite:** Ms. Archana Karne, Mr. Radha Krishna Karne, Mr. V. Karthik Kumar, Dr. A. Arunkumar. (2023). Convolutional neural networks for object detection and recognition. Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 3(1), 55–67. https://doi.org/10.55529/jaimlnn.32.1.13

**BIOGRAPHIES OF AUTHORS**

| | |
|---|---|
|  | **Ms. Archana Karne,** is currently studying undergraduate degree programs at CMR Technical Campus. She investigates artificial intelligence and deep learning and computer vision and object detection methodologies. She has led projects which developed intelligent surveillance systems through her work with convolutional neural networks. Her academic research centers on creating real-time object tracking systems which utilize current machine learning techniques. |
|  | **Mr. Radha Krishna Karne,** works as an Assistant Professor in the Department of Electronics and Communication Engineering at CMR Institute of Technology. His research work focuses on deep learning and vehicular ad hoc networks and Internet of Things and wireless communication and computer vision. He has published several research papers in reputed international journals and conferences. His research investigates intelligent communication systems together with object detection technology and real-time systems that utilize artificial intelligence. Email: krk.wgl@gmail.com |
|  | **Mr. V. Karthik Kumar,** At BITS Narsampet he serves as an Assistant Professor in the Department of Electronics and Communication Engineering. His research interests include image processing and machine learning and embedded systems and signal processing. He has contributed to multiple educational and scientific projects which focused on artificial intelligence and object tracking system development. His current research work focuses on deep learning algorithms which he applies to practical situations in computer vision and surveillance systems. |
|  | **Dr. A. Arunkumar,** works as a Professor in the Department of Computer Science and Engineering at MLR Institute of Technology and Management. He has developed substantial expertise through his work in teaching and researching artificial intelligence data science computer vision and deep learning fields. He has supervised multiple student projects while he has published research papers in well-known academic journals and conferences. His research investigates intelligent systems and object recognition together with advanced machine learning methods that solve real-world challenges. |