Journal of Theoretical and Applied Information Technology
15th April 2023. Vol.101. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

BURGLARY DETECTION IN THE RESIDENTIAL AREAS USING YOLO AND SSD ONE-STAGE DETECTION ALGORITHM: A COMPARATIVE PERSPECTIVE

PAVITHRA.S[1], Dr.B. MURUGANANTHAM[2]

[1] Research Scholar, Department of Computing Technologies, School of Computing, SRMIST, Kattankulathur,Chennai, Tamilnadu, India, 603203

[2] Associate Professor, Department of Computing Technologies, School of Computing, SRMIST, Kattankulathur,Chennai, Tamilnadu, India, 603203

E-mail : [1]ps8335@srmist.edu.in, [2]muruganb@srmist.edu.in

ABSTRACT

Deep learning facts are becoming increasingly important and popular. It is a significant factor in object detection. Moreover, object detection, one of the most critical and challenging parts of deep learning, is mainly used in people's everyday lives in areas like video surveillance and self-driving cars. This research is to identify and detect burglars in residential areas. To get a complete and deep understanding of the primary development status of the object (burglary) detection system, we started by looking at the existing algorithms for detection models and setting up our dataset. This paper explains how single-shot multibox detectors and Yolo use deep-learning algorithms to find burglars in residential areas. This research evaluates the deep learning approach for cutting-edge burglary detection systems. This research paper compares two algorithms of one-stage object detection, i.e., you only look once, and single-shot multibox detectors to determine which algorithm is the most effective and fastest. In this comparison, we used the datasets for each algorithm to look at how well they worked, where they were weak, and where they were strong. Based on the results, it says that in the testing environment, Yolo does better than SSD.

Keywords: *Yolo, SSD, Burglary Detection, Deep Learning, Residential*

1. INTRODUCTION

Residential burglary is when someone breaks into an occupied home without permission and plans to steal. Even though the number of burglaries has decreased worldwide in the last 20 years, it is still a common crime. Most people will become victims of a burglary at some point. Burglary is a property crime, but for many families of burglary victims, the fact that someone broke into their home is more important than the property they lose. Many victims contacted the police. The police, victims, and offenders contribute to our understanding of burglary.

Recently, burglary detection has become more popular because it can be used in many ways and because of new scientific discoveries. Academics have done much research on the detection task, and it has real-world uses like crime detection, face detection, vehicle detection, and burglary detection. Object detection methods have evolved quickly because of many different factors and efforts. Two of these are the improvement of Yolo (you only look once) and the power of Nvidia GeForce computation. Deep learning methods are now used by almost everyone in areas like domain-specific and general object detection that are specific to deep learning and computer vision. Deep learning networks are at the heart of the most advanced object detectors and are also used to find things. Every day, more burglaries and more burglars are causing many people to worry about security. So, the primary goal of this investigation is to determine who is breaking into homes in residential areas. Most homes and public places now use new technologies, primarily video surveillance, to stop burglaries.

However, they need to be watched by someone in residential areas. A human's ability to monitor multiple screens at once is not possible. In this research paper, burglary detection is defined as the detection of burglars and activities happening in residential areas.

Moreover, these events come from the surveillance videos the residential panoramic camera recorded. Furthermore, those videos are converted into several frames using SSD and Yolo algorithms. Burglary detection tries to find one or more real objects (burglars) in an image and mark where they are. The purpose of this research article is to compare Yolo and SSD. The first algorithm to be compared in this work is the single-shot multibox detector. This algorithm adds many layers from different parts to the final network and makes it possible to detect. You Only Look Once, which was developed by R. Joseph, is an end-to-end network. In this article, the results of the two algorithms we have talked about so far, which have different structures, were compared using their own data sets to compare them and the same performance measures for each implementation. By comparing the current effectiveness of these methods on their respective datasets, you can learn about the unique features of each algorithm, understand how much they differ from one another, and figure out if the detection method is the best for every situation. Following the background descriptions of CNNs, this work will explain a new way of resolving the above work using Yolo. This paper makes use of the standard Yolo version. When their accuracy, speed, and average accuracy are compared, the pros and cons of Yolo and SSD will become clear. Finally, the Yolo and SSD performances will be summarized. The system makes use of panoramic video surveillance camera footage. This video will be used to find out what happened if there was a break-in or violence. So, this system is more interesting because it can tell in advance when something is not right and then uses a method to let the right person know. It can also be implemented in both outdoor and indoor residential areas. The following article is organized: Section II summarizes relevant detection research work. Section III provides a comparison of both two-stage and one-stage detection. Section IV describes the drawbacks of two-stage detection. In Section V, we talk about how to find Yolo and SSD in a single step. Sections VI and VII discuss how the results were analyzed, what we learned from them, and what we plan to do next.

2. RELATED WORKS

The earlier research suggests several ways to discover what people are doing in a video. The current plan uses a video surveillance system to catch burglars, break-ins, and other burglary activity in residential areas. Human behavior identification has many real-world applications, like intelligent video surveillance and customer behavioral analysis. Surveillance video has diverse applications, especially in indoor and outdoor scenarios. Video surveillance is an essential aspect of safety and security. Security cameras are now vital to provide protection and safety in daily life. Skill India, a training program run by the Indian government, gives e-surveillance top priority, and surveillance is still used [1]. The benefits of surveillance videos include keeping an eye on things, needing less staff, doing an internal audit without spending much money, and following new security trends.

The Advanced Motion Detection algorithm has been used to detect unauthorized access to a security zone [2]. In the first phase, the object has been detected entirely through a background subtraction algorithm, and in the second phase, the object has been extracted from the sequence of frames. In the second phase, the system detects unusual activity. The algorithm used in the system is intended to improve real-time video processing at a low computational cost. Moreover, the amount of data storage has restricted the system currently offered, and it can also be used in surveillance areas that employ a sophisticated video recording method.

A semantic method has been proposed in [3]. So, the background subtraction algorithm was used to find an area of interest in the captured video sequence. After subtracting the background, the Haar cascade algorithm sorts the objects into two groups: those that live there and those that do not. The objects were tracked using the real-time blob tracking algorithm. This article detects fire alarm systems as well.

All the incidents were found in [4] based mostly on how the objects moved. The linguistic approach was used to identify suspicious activities. Correlation and object detection methodologies were implemented to detect and track objects [3]. The activities are grouped based on how they move and when they happen, and this framework has a lower computation cost.

The input image enhancement in each area of unusual activity at a university has been predicted using the Lucas Kanade algorithm. They then generated a histogram of the image's sequential feature vector of magnitude. Software algorithms manipulate video data to classify activities as normal or abnormal [5].

A system was developed to differentiate abnormal activities from daily events by analyzing motion information from consecutive frames of the surveillance videos. The Hidden Markov model was used to train the frequency distribution of optical flow orientations in the video frames. This compares the frames of the video favorably to historically prevailing average frames and defines their level of similarity. Different data sets, like the pedestrian attribute dataset [6], have been used to test and implement the method. Unusual events in surveillance video can be detected using people tracking. An image-processing background subtraction algorithm is used to detect people in videos. The features were extracted using a convolutional neural network and fed into a discriminative deep belief network. Labeled video content of several suspected activities is also fed into the deep belief network method, as are their characteristics. Using a DBN, the convolutional neural network features were compared to those of suspicious activities that had been clearly labeled and put into categories or to those of the video that showed more than one incident [7]. Identifying an object inside a picture close to its confinement and order is known as "object detection." Identifying the scenario of at least each object in the image and marking a bounding box around their extent is referred to as "object limitation." Image sequences could be a tactic for handling huge numbers and predicting the classification of a particular object in an image. These two procedures are combined in object detection by restricting and organizing at least each object in the image [8].

The problem with RCNN is that training takes too long, so it must classify 2000 regional proposals. Real-time implementation is, therefore, not possible [9]. As a result, there is no need to learn the fundamentals. This may contribute to the development of poor regional proposals. However, single-stage proposals such as Yolo and SSD are an exception. Consider object identification as a simple regression issue in which an image is used to learn about the probabilities of the class. Features with bounding boxes achieve lower frame rates of accuracy but are considerably quicker than objects with two-stage detectors.

The deep convolutional neural network in the paper uses the weight-sharing principle. This notifies us of those critical CNN points. In a recent study [10], Chen et al. used a bounding box and a more accurate regression loss function for facial recognition. They proposed a Yolo-based facial expression detector to solve detection issues with differing face scales. The authors' algorithm outperformed previous Yolo versions and variants [11]. We evaluated the Yolo by comparing it to specific other models in our data analysis.

3. COMPARISON OF ONE-STAGE AND TWO-STAGE OBJECT DETECTION

3.1 Two-Stage Detection

The accuracy of detection is prioritized in this strategy. The evaluation and detection steps are separated by the two-stage method. The recognized objects are trimmed after object detection and then processed by a different subnet for current evaluation. This means that the images must be resampled at least a few times: once to analyze, once to suggest a region, and once to find. That works well but is inconvenient and slow because the classification and detection must be run on several instances. Two-stage algorithms are used in basic models like RCNN, faster RCNN, and fast RCNN.

3.2 One-Stage Detection

In this strategy, speed of inference is prioritized. On the other hand, the proposed approach does not require image resampling and instead uses convolutions to identify the object and forward propagation. This gives a high-speed boost because the image needs to be resized, and the evaluation for finding or locating something is shared. That is much faster and more suitable for mobiles. Yolo, as well as SSD, are the most widely used one-stage detection algorithms. The most widely used benchmark dataset is the COCO dataset. The mAP (mean average precision) is frequently used to evaluate models.

4. CAVEATS OF TWO-STAGE DETECTION

- It allows a substantial period to train the model, generating more group-based regional proposals for every image.
- Every test image requires approximately 48 seconds to complete, but it cannot be implemented in real-time.
- Learning is not required at this stage since the search technique is fixed.
- They lose all sensitive data about the object's location and orientation, and all data is routed to a similar neuron, which is almost definitely incapable of solving these kinds of information.
- The convolutional neural network predicts by examining a picture and determining if specific parts of that image are prevalent. If they are, the image is rearranged accordingly.
- RCNN, Fast RCNN, and Faster RCNN were designed to overcome obstacles. By testing its output with the MS-COCO dataset, the Faster RCNN algorithm outperformed all others in terms of training time and accuracy.
- To overcome this limitation, Faster RCNN was developed. Yolo analysis of errors versus Fast RCNN search shows several location errors in Yolo. However, the SSD's precision was reduced in comparison to the Faster RCNN. Moreover, Yolo object detection methods evolved using darknet images; for example, in terms of accuracy and inference time, Yolo outperformed Faster RCNN and SSD.

5. METHODOLOGY OF ONE-STAGE DETECTION (YOLO & SSD)

5.1 You Only Look Once

Yolo is a convolutional network that detects objects in real-time. The Yolo algorithm performs the image using a neural network, which is then divided into various regions that can detect probability for each class and bounding boxes on each region. While still operating in real-time, such bounding boxes might accurately assess the desired probability values. Yolo was very well for involving only forward propagation across the neuron. This reverts approved objects with frames after NMS (non-maximum suppression), where it ensures that every object is detected precisely once. This yolo algorithm is performed by dividing an image into a grid S×S with frames in every grid. For every frame, the system returns an output class probability. For every frame, the system returns an output class probability. To identify the object more precisely, a frame with a class probability more incredible than the threshold's minimum value is chosen. This frame is then used only within the image. So, at a minimum of 45 frames per second and a maximum of 60 frames per second, Yolo is better than both object detection algorithms in terms of size. The Yolo algorithm has a weakness in that it has trouble finding small objects because of its limitations. Unlike sliding windows as well as area proposal-based approaches, Yolo recognizes artifacts in images very well because it is used to seeing the whole image throughout the testing and training phase so that it gets each input into the overall image and object as well as its presence. The image is divided into grids, and the image classification algorithm is run on each grid cell. Through the grid system, it detects and scores bounding boxes. The confidence score represents the bounding box's accuracy. Because several of the bounding boxes have low-reliability values, by setting a specified threshold, unnecessary bounding boxes or objects can be avoided.

A short version of Yolo:

The input images are divided into an S×S grid by the Yolo model. The grid cell oversees detecting objects, as the center of the objects falls into grid cells.

Journal of Theoretical and Applied Information Technology
15th April 2023. Vol.101. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

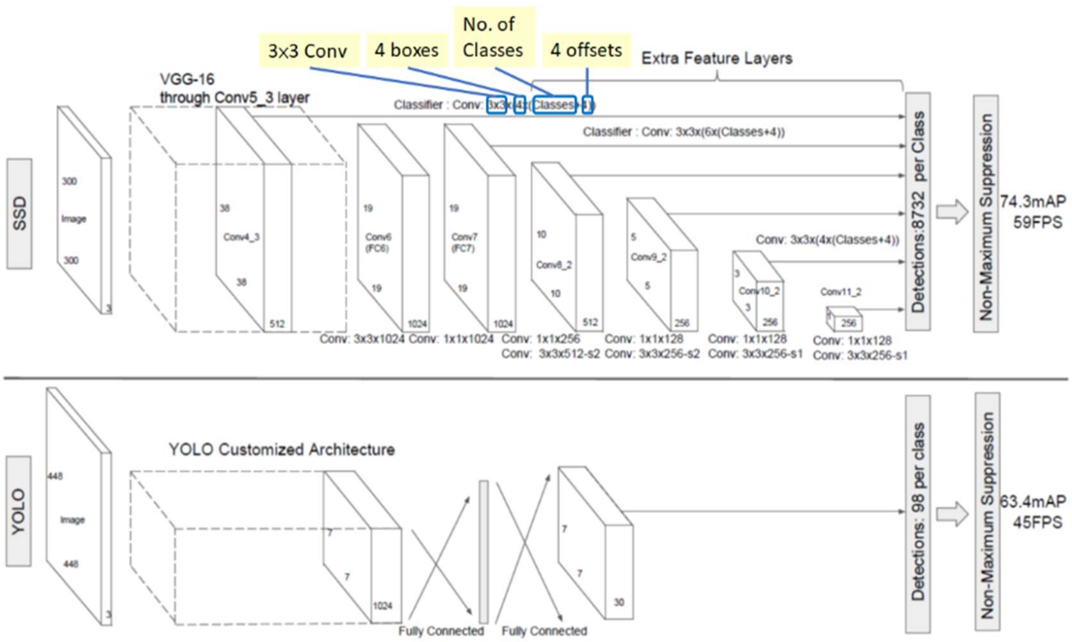

Figure 1. Single Shot Multibox Detector Versus Yolo Architecture [12]

5.2 Single Shot MultiBox Detector

The SSD architecture employs a technique for recognizing multiple object categories in an image by presenting confidence scores linked to the existence of every category of objects. Additionally, it alters how the boxes' contents are shaped. It is ideal for everyday application fields, unlike Faster RCNN, which does not evaluate boundary box inferences. The convolution neural-based single-shot multibox detector detects specific target object classes in two phases: first, extract the feature maps, and then the convolution filter is used to identify the objects. Pattern recognition and object detection in computer vision are unsolved problems. The critical object detection issues, such as transformations, noise resilience, and obstacles, are carried over, but new problems are introduced, such as the detection of various objects, overlapping images, and attempting to identify their positions within an image. Single-shot multibox detector achieves a better balance of accuracy and speed. When an image is entered, the direction of the flow of the traditional network displays a functional diagram.

A short version of SSD:

SSD is split into two sections:

1. Feature maps are extracted.
2. Detecting objects using convolutional filters.

Table 1. compares Yolo and SSD in terms of detection time, accuracy, speed, mAP, fps, training time, and training data and whether they are suitable for real-world scenarios. The table clearly illustrates that yolo outshines the higher precision as well as lower frame rate than SSD algorithms. At 512*512 resolution, Yolo runs through 45fps with 0.36 mAP, which is nearly as precise as SSD but approximately two times faster. A careful compromise is made to obtain this speed. Yolo has an appropriate mean average precision for use in real-world applications, even with a lower mean average precision, and yolo is an efficient algorithm when combined with fps and accuracy [13].

Table 1. Yolo And SSD Algorithms Comparison

Serial Number	Criterion	You only look once	Single shot detector
1.	Model name	Yolo	SSD
2.	Time for training	10 hours	14 hours
3.	More training data	No	Yes

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.	Accuracy of test data	93%	82%
5.	Speed	Low	High
6.	Time for detection	0.80~0.9 sec/frames	0.15~0.24 sec/frames
7.	Frames per second (fps)	45	59
8.	Mean average precision (mAP)	0.36	0.241

5.3 Data Collection

Yolo is involved in this project to find the most efficient and fastest comparative evaluation using their dataset. We employed the algorithms of Yolo and SSD. The whole paper represents one of CNN's most effective processes. Thus, Yolo deviates from the legacy of the convolutional neural network by inventing an entirely new method for resolving detection as simply and effectively as possible. Yolo outperforms the SSD in terms of speed. Yolo achieved 45 fps, while SSD achieved 59 fps. Yolo had the highest accuracy, 93%, and SSD, 82%.

5.4 Data Pre-processing

Data pre-processing takes up roughly 75% of the time. After the data is collected, it needs to be cleaned, pre-processed, and formatted in a way that makes it easy to understand and use. Deep learning techniques and tools should be used to analyze the collected real-world data.

5.5 Data Training and Testing

For the conceptual framework to be correct, it needs to be tested and trained in different situations using both Yolo and SSD parameters. Furthermore, the models' precision is optimal. The information gathered is used to train and test the model. If necessary, provisions must be made to improve just the algorithm being used.

6. RESULT ANALYSIS

Our proposed system is implemented in two phases: the training phase and the testing phase for our two algorithms, Yolo, and SSD. During the training phase, the Yolo model achieved 93% accuracy, and the SSD model achieved 82% accuracy. The models' accuracy can be increased in terms of iterations. The video content is subdivided into images and possibly saved inside a single folder during the testing phase. The method identifies the frames as burglars and burglary activities using our trained model. In the incident of a burglary, the detection of burglar's alert will be sent to the appropriate authorities. Therefore, with our comparative analysis Yolo model outperforms better than the SSD model with achieved accuracy of 93%.

6.1 Burglary Detection using SSD and Yolo Algorithm

When we perform the SSD and Yolo algorithm for burglary detection using the video recorded in the residential surveillance cameras, it is trained and converted into frames in the testing phase using our dataset that detects burglars and burglary activities. A video comparison of SSD and Yolo shows that SSD is significantly faster than Yolo, and yet Yolo is more accurate.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 2. Burglary Detection Using SSD With Accuracy And FPS

Figure 3. Burglary Detection Using Yolo With Accuracy And FPS

6.2 Accuracy Graph of SSD and Yolo

The accuracy of the burglary detection in the residential is shown below using the two algorithms of SSD and Yolo. With this comparative analysis, the Yolo model performs better than the SSD model in catching burglars in residential. Therefore, this system suits better for residential to thwart burglary activities in the residential and gives the perfect accuracy with the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

surveillance video of the residential to catch the burglars.

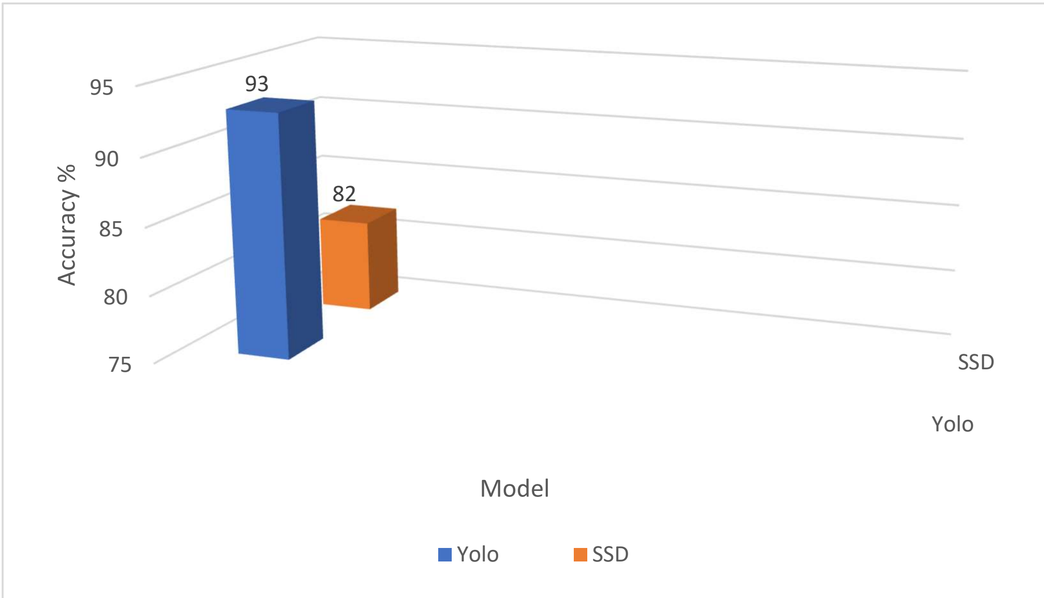

Figure 4. Burglary Detection Accuracy Using Yolo And SSD Model

Image resolutions and feature extractors have a natural effect on speed. The highest and lowest frames per second identified in this paper are listed below. However, because they were measured at different mean average precision mAP, the results below may be highly biased.

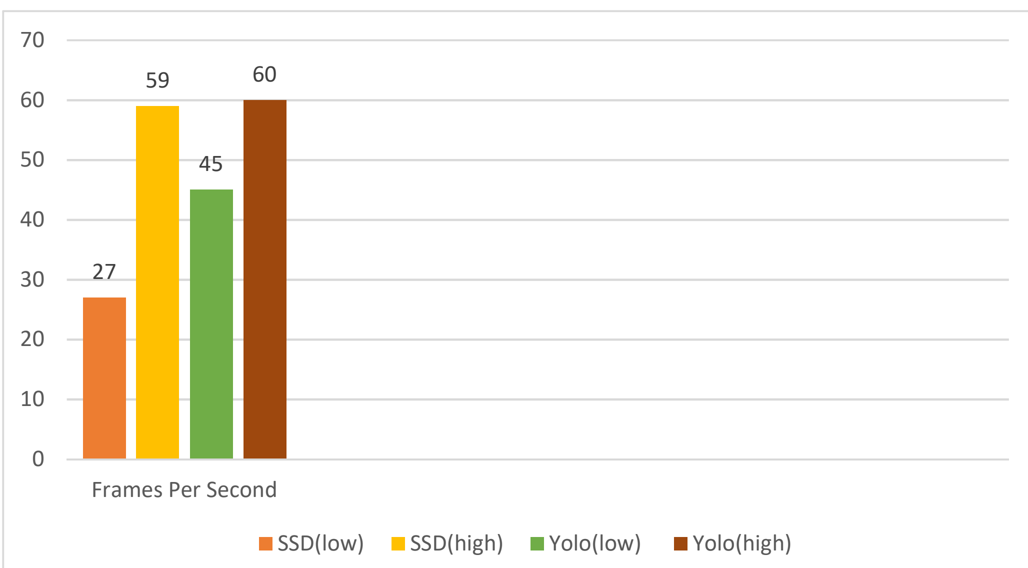

Figure 5. Burglary Detection Frames Per Second Using Yolo And SSD

7. CONCLUSION AND FUTURE WORK

Mostly everyone knows the significance of surveillance footage in today's modern world, and in most instances, the whole video clip is mainly used for the investigative process after a suspected burglary activity happened. This current model offers the benefits of avoiding burglary activities in residential areas. Surveillance video footage of the residential is being monitored and analyzed 24*7. This analysis will alert the appropriate authority in the supervision and indicates that suspicious activity is about to occur or has occurred. As a result, this will be further stopped. Recent advancement in deep learning-based burglary detection systems is compared and analyzed. This paper compares the most modern and sophisticated convolutional neural-based object (burglary) detection algorithms Yolo and SSD. It would have been impossible to identify the millions of images posted on the web every day without detection. SSD was observed to be faster than Yolo. If you require to analyze a real-time video, Yolo is the tool to use. In the meantime, SSD offers a good balance of speed and accuracy. However, Yolo produces more accuracy than SSD. As a result, of the two different object (burglary) detection algorithms analyzed, Yolo has the overall best performance.

Although the current proposal is limited to residential areas, it also has the potential to detect more burglary activities and suspicious behavior in public places, academic institutions, and corporate offices. This framework and techniques can be used when training on suitable burglary activities must be offered. The model can be improved by differentiating the burglars from the burglary activities.

8. CONFLICTS OF INTEREST

According to the authors, there are no conflicting interests.

9. AUTHOR CONTRIBUTIONS

Pavithra.S is investigating under Dr. B. Muruganantham, using their dataset and additional images from Google. The effectiveness of these results is validated using the paper's standard formulas, equations, figures, and tables.

10. ACKNOWLEDGMENT

I want to express my appreciation to my research supervisor Dr.B. Muruganantham for motivating me and encouraging me to learn how to write this comparative perspective paper. Everyone's knowledge and experience, which helped this study in various ways, are also sincerely appreciated.

REFERENCES

[1] Amrutha, Jyotsna, Amudha, "Deep Learning Approach for Suspicious Activity Detection from Surveillance Video," Proceedings of the Second International Conference on Innovative Mechanisms for Industry Applications, ISBN: 978-1-7281-4167-1, pp. 335-339, 2020.

[2] P. Bhagya Divya, S. Shalini, R. Deepa, Baddeli Sravya Reddy, "Inspection of suspicious human activity in the crowdsourced areas captured in surveillance cameras," International Research Journal of Engineering and Technology, December 2017.

[3] Jitendra Musale, Akshata Gavhane, Liyakat Shaikh, Pournima Hagwane, Snehalata Tadge, "Suspicious Movement Detection and Tracking of Human Behavior and Object with Fire Detection using A Closed-Circuit TV (CCTV) camera," International Journal for Research in Applied Science & Engineering Technology, Volume 5 Issue XII, December 2017.

[4] U.M. Kamthe, C.G. Patil "Suspicious Activity Recognition in Video Surveillance System," Fourth International Conference on Computing Communication Control and Automation, 2018.

[5] Zahraa Kain, Abir Youness, Ismail El Sayad, Samih Abdul-Nabi, Hussein Kassem, "Detecting Abnormal Events in University Areas", International conference on Computer and Application,2018.

[6] Tian Wanga, Meina Qi, Yingjun Deng, Yi Zhou, Huan Wangd, Qi Lyua, Hichem Snoussie, "Abnormal event detection based on analysis of movement information of video sequence," Article-Optik, volume 152, 2018.

[7] Elizabeth Scaria, Aby Abahai T and Elizabeth Isaac, "Suspicious Activity Detection in Surveillance Video using Discriminative Deep Belief Network," International Journal of Control Theory and Applications, Volume 10, Number 29, 2017.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[8] Dr. S. V. Viraktamath, Ambika Neelopant, "Comparison of YOLOv3 and SSD Algorithms", International Journal of Engineering Research & Technology, Vol. 10 Issue 02, February 2021.

[9] Kanishk Wadhwa, Jay Kumar Behera, "Accurate Real-Time Object Detection using SSD," International Research Journal of Engineering and Technology, 2020.

[10] object detection YOLO algorithm https://towardsdatascience.com/yolo-you-only-look-once-3dbdbb608ec4 (date accessed: August 26, 2021).

[11] Object detection SSD algorithm https://developers.arcgis.com/python/guide/how-ssd-works/ (date accessed: August 19, 2021).

[12] Object Detection SSD and Yolo algorithm https://dl.asmaamir.com/basics-of-object-detection/ssd-and-yolo.

[13] Preethi Srinivasalu, Aishwarya Palaniappan, "Brain Tumor Detection by Modified Particle Swarm Optimization Algorithm and Multi-Support Vector Machine Classifier," International Journal of Intelligent Engineering and Systems, Volume 15, Number 6, 2022.

[14] Andi Sunyoto, Hanafi1, "Enhance Intrusion Detection (IDS) System Using Deep SDAE to Increase Effectiveness of Dimensional Reduction in Machine Learning and Deep Learning," International Journal of Intelligent Engineering and Systems, Volume 15, Number 4, 2022.

[15] Chandramani Kumar, Kartik Chawla, Shreya Arora, "The comparison between various object detection algorithms," International Journal of Advance Research, Ideas, and Innovations in Technology, Volume 4, Issue 2, 2018.

[16] Peddinti Mounika, Lakshmi Narayana, Kondapalli Venkata Ramana, "A Comparative Study of Yolo and SSD Object Detection Algorithms," International Journal of Creative Research Thoughts, Volume 9, Issue 11, November 2021.

[17] Anand John, Divyakant Meva, "A Comparative Study of Various Object Detection Algorithms and Performance Analysis," International Journal of Computer Sciences and Engineering, Volume 8, Issue 10, October 2020.

[18] Ethala Sandhya, Annapurani Kumarappan, "Enhancing the Performance of an Intrusion Detection System Using Spider Monkey Optimization in IoT," International Journal of Intelligent Engineering and Systems, Volume 14, Number 6, 2021.

[19] Agus Zainal Arifin, Safri Adam, Avin Maulana Mohammad, "Detection of Overlapping Teeth on Dental Panoramic Radiograph," International Journal of Intelligent Engineering and Systems, Volume 12, Number 6, 2019.

[20] Cahya Rahmad, Kohei Arai, Rosa A. Asmara, "Comparison of Geometric Features and Color Features for Face Recognition," International Journal of Intelligent Engineering and Systems, Volume 14, Number 1, 2021.