Deep Learning Based Object tracking and Detection for Autonomous Drones using YOLOv3

G. Naga Leela
Assistant Professor
Dept of CSE
Srinivasa Ramanujan Institute of Technology
Anantapur
Nagaleelag.cse@srit.ac.in

U. Varun
UG Scholar
Dept of CSE
Srinivasa Ramanujan Institute of Technology
Anantapur
varunu5c0@gmail.com

M. Uma
UG Scholar
Dept of CSE
Srinivasa Ramanujan Institute of Technology
Anantapur
204g1a05b5@srit.ac.in

P. Srinath
UG Scholar
Dept of CSE
Srinivasa Ramanujan Institute of Technology
Anantapur
214g5a0511@srit.ac.in

A. Srinivasa Sree Sharan
UG Scholar
Dept of CSE
Srinivasa Ramanujan Institute of Technology
Anantapur
204g1a05a3@srit.ac.in

Abstract-Object detection plays a vital role in enabling drones to perceive and interact intelligently with their surroundings. The process involves data collection, selecting appropriate deep learning architectures for accurate detection. In our proposed system, we explore the integration of advanced deep learning techniques into autonomous drones for object detection. The trained models combined with onboard software integration efficiently predict object presence in real-time imagery. The project aims to significantly improve the accuracy and reliability of object recognition, enabling drones to distinguish between different objects and navigate their environment with increased precision. The project's outcomes hold promising possibilities for enhancing drone applications in fields such as surveillance, search and rescue.

Keywords: Object Detection, Autonomous Drones, Computer Vision, Real-time Processing.

I . Introduction

The project aims to implement a cutting-edge solution for real- time object detection using deep learning algorithms, specifically leveraging the YOLOv3 algorithm, on drone- captured imagery. Object detection plays a pivotal role in various applications, ranging from surveillance and security to environmental monitoring and search-and-rescue operations. By harnessing the power of deep learning and the YOLOv3 model, we endeavor to equip drones with the capability to autonomously detect and identify objects in their field of view, enabling swift and informed decision-making during drone operations. With the increasing integration of drones into various industries, the ability to detect and respond to objects in real- time becomes crucial for enhancing the efficiency effectiveness of drone-based applications.

The YOLOv3 algorithm, known for its speed and accuracy, is well-suited for deployment on drones, enabling rapid and precise object detection without compromising computational performance. Throughout the implementation process, we will navigate the complexities of collecting and annotating drone-captured datasets, training the YOLOv3 model, optimizing it for deployment on drone platforms, and integrating the object detection results seamlessly with the drone's control system. The project places a strong emphasis on adaptability to environmental factors, scalability, and real-time processing, ensuring that the deployed system is robust, versatile, and capable of addressing a wide range of scenarios.

As drones continue to revolutionize industries with their versatility and mobility, the integration of advanced deep learning techniques for object detection adds a layer of intelligence that can significantly enhance the capabilities of these unmanned aerial vehicles. This project seeks to contribute to the evolution of drone technology, bringing fortha solution that holds immense potential for applications in security, agriculture, disaster response, and beyond. Through the successful implementation of deep learning-based object detection on drones, we aim to propel the boundaries of what is achievable in autonomous and intelligent drone systems.

During the transmission process, the drone is tasked with locating and identifying the target object. Hence, an object detection module, grounded on a camera-equipped drone, has been developed. Presently, deep learning is revolutionizing the swiftly advancing domain of computer vision. Consequently, this study endeavors to devise a deep learning-driven algorithm for human body recognition, diverging from conventional machine learning paradigms.

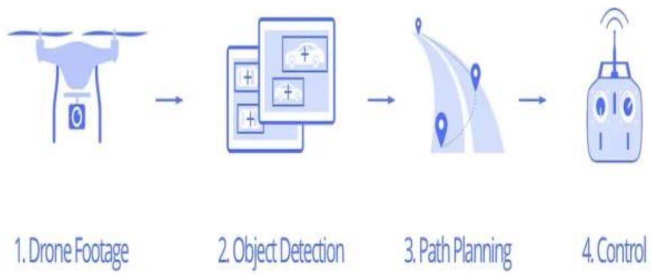

Fig 1: Basic object detection process

Despite the multitude of detection methods proposed in literature, few have explored person detection via deep networks. In [12] and [13], a sequence of exceptionally high-resolution images is captured utilizing UAV sensors for automatic vehicle counting. [12] and [13] undertake the feature extraction process based on scale-invariant feature transformation and gradient histograms, respectively. [8] employs deep convolutional neural networks on raw micro-Doppler spectrograms for human detection and activity classification. Meanwhile, [6] concentrates on human detection, integrating Kernelized Correlation Filter (KCF) and faster R-CNN for real-time detection on embedded and desktop GPUs, catering to drone vision. This concept is termed "deep drone.

II. Literature Survey

Deep learning arises from the convergence of artificial neural networks (ANN), artificial intelligence, graphical modeling, optimization, pattern recognition, and signal processing. ANNs are a specialized class of machine learning algorithms tailored for pattern recognition, drawing inspiration from the brain's structure and function. The fundamental components of ANNs are neurons, which compute a weighted sum of inputs and apply an activation function. Neural networks typically feature multiple layers of interconnected neurons. In the initial layer, known as the input layer, each neuron corresponds to an input feature, such as a pixel. Subsequent layers receive inputs from preceding layers, forming a hierarchical representation. Training a neural network involves determining optimal weights for each neuron connection, enabling it to learn from the provided data[21].

The weights are adjusted through an iterative process known as backpropagation. Deep learning architectures encompass hidden layers of artificial neural networks and a collection of sentence formulas. Additionally, deep generative models incorporate latent variables structured in layers, exemplified by Deep Belief Networks and Deep Boltzmann Machine nodes.

Deep learning stands as a swiftly evolving realm within machine learning, primarily harnessed to tackle computer vision challenges. It encompasses a class of algorithms characterized by a cascade of numerous layers of nonlinear processing. This approach falls within the broader scope of machine learning, which revolves around learning data representations to facilitate end-to-end optimization. Notably, deep learning excels in learning multiple layers of representations, mirroring a hierarchy of conceptual abstractions.

An exemplary application of deep learning lies in object localization and detection based on video streams, a pivotal task in computer vision. Contemporary advancements in object detection predominantly leverage deep learning methodologies like region-based convolutional neural networks (R-CNN).

Previous research highlights the efficacy of traditional machine learning techniques in object detection, including rapid algorithms employing SIFT (Scale-Invariant Feature Transform) for keypoint detection and FLANN (Approximate Nearest Neighbor Fast Library)-based matchers. Such approaches span various object detection methodologies in computer vision, including appearance-based, template-based, part-based, region-based, and contour-based techniques.

The prevalent fusion of object detection and machine learning in computer vision adopts the sliding window approach, extensively utilized in face and person detection tasks. Despite its utility, the sliding window method is time-consuming. To address this, researchers have proposed numerous feature-based approaches for human recognition, encompassing Haar wavelet features, Haar-like features with motion information, implicit shape models, histograms of directional gradients (HOG), covariance descriptors, and others. Our aim is to surmount the limitations posed by gradient histograms (HG)[22]

III . Problem Statement

The project addresses critical challenges associated with the deployment of drones in dynamic environments, where swift object recognition is imperative for obstacle avoidance and timely decision-making. Drones, operating within confined computational capabilities, demand the development of highly efficient object detection models to ensure real-time responsiveness. The adverse effects of changing light and weather conditions on object appearance further complicate accurate detection, necessitating robust adaptation mechanisms. A notable challenge is the need for the system to effectively handle rare objects, emphasizing the importance of comprehensive object detection capabilities beyond common scenarios.

The accurate identification of objects is pivotal for the safe navigation and interaction of drones with the environment, highlighting the urgency of developing a sophisticated and adaptable object detection system to meet these multifaceted challenges.

IV . Architecture and Methodology

The deep learning process begins with a well-defined problem statement, where the nature of the task and the target variable are explicitly outlined. Subsequently, the collection and preparation of a relevant dataset take place, emphasizing diversity and proper labeling. Data preprocessing follows, involving cleaning, scaling, and encoding to render the dataset suitable for model training. The selection of an appropriate deep learning architecture, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), is a crucial step. Constructing the model involves defining its architecture, layers, and connections, aligning its complexity with the problem's intricacy. Model compilation, where optimization parameters are specified, precedes the actual training phase. Deployment brings the trained model into real-world applications, be it in an application, a web service, or an edge device. Continuous monitoring and occasional retraining ensure the model remains adaptive, addressing evolving patterns. Comprehensive documentation throughout the process facilitates reproducibility and collaboration, contributing to the success of the deep learning endeavor.

YOLOV3

YOLO (You Only Look Once) is a popular object detection algorithm that falls under the category of real-time object detection algorithms. YOLO is known for its speed and accuracy, making it suitable for various applications, including surveillance, autonomous vehicles, and robotics. YOLO version 3, or YOLOv3, is the third iteration of the YOLO series, introducing several improvements over its predecessors.YOLOV3 uses Darknet 53 architecture.

Utilizing YOLOv3 for object detection with drones involves implementing the YOLOv3 algorithm to analyze images or video frames captured by drones in real-time. The process begins with the collection of a diverse dataset of images or video frames captured by drones, encompassing various scenarios and object types relevant to the intended application. Following dataset collection, the data must be annotated, specifying bounding box coordinates and class labels for each object of interest. Subsequent data preprocessing steps involve resizing images, normalizing pixel values, and applying augmentation techniques to enhance the model's generalization capabilities.

The YOLOv3 model, chosen for its real-time performance and accuracy, is then trained on the annotated dataset. Fine-tuning may be applied to adapt the model to drone-specific scenarios, and training metrics such as loss and accuracy are monitored. Validation on a separate dataset ensures the model generalizes well to unseen data, with adjustments made to hyperparameters or model architecture as necessary

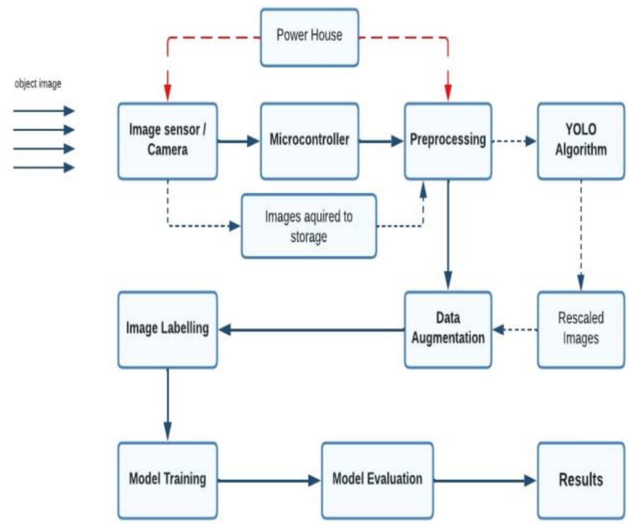

Fig 2:.Block diagram of object detection

Once trained, the YOLOv3 model is deployed to the drone's onboard computing system. Real-time inference is then implemented, with the model identifying and localizing objects of interest in the drone's field of view. Integration with the drone's control system facilitates applications such as navigation or surveillance. Ongoing performance monitoring is crucial, addressing issues related to false positives, false negatives, or changes in environmental conditions.

Consideration for environmental factors such as lighting conditions, weather, and terrain variations is paramount, necessitating adaptation of the model or the use of additional sensors for enhanced robustness. Security and privacy concerns associated with drone-based object detection must also be addressed to ensure compliance with regulations and protect sensitive information.

As drone operations unfold, iterative improvement becomes essential. New data collected during operations can be used for iterative model retraining, enhancing accuracy and adapting the YOLOv3 model to changing conditions. In summary, deploying YOLOv3 for object detection with drones requires a comprehensive approach, encompassing data preparation, model training, deployment, integration, continuous monitoring, and iterative refinement to ensure optimal performance in dynamic real-world environments.

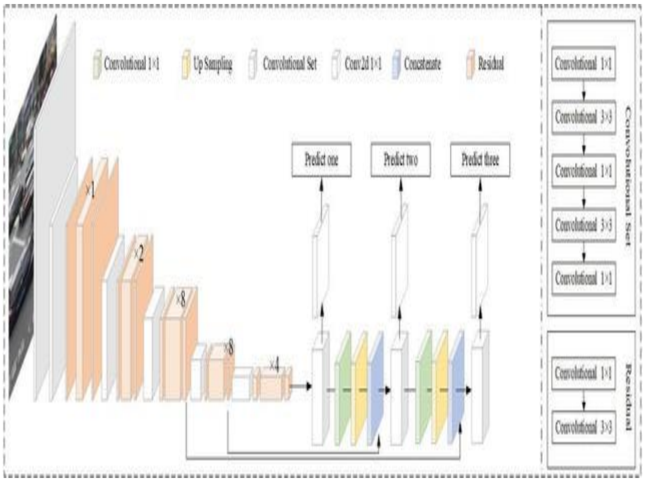

Fig 3: How YOLOV3 works

V .Implementation

Implementing a deep learning-based object detection project using drones involves several key steps. Here is a comprehensive overview of the entire process:

1. Problem Definition and Dataset Collection:

Clearly define the object detection problem for your drone application. Collect a dataset of images or video frames captured by drones, ensuring it covers a variety of scenarios and objects relevant to your use case.

2. Data Annotation:

Annotate the dataset by labeling objects of interest in the images. Specify bounding box coordinates and class labels for each annotated object. Tools like LabelImg or RectLabelcan assist in this process.

3. Data Preprocessing:

Preprocess the annotated data to make it suitable for training. Resize images to a consistent size, normalize pixel values, and apply data augmentation techniques such as rotation or flipping to increase dataset diversity.

4. Model Selection:

Choose YOLOv3 as the object detection model for your mplementation. YOLOv3 is known for its real-time processing capabilities and high accuracy in detecting objects.

5. Model Configuration:

Configure the YOLOv3 model for your specific dataset. Adjust parameters such as the number of classes, anchor boxes, and input dimensions according to the characteristics of your annotated data.

6. Training:

Train the YOLOv3 model on the annotated dataset. Use a machine with GPU support for faster training. Monitor training metrics, and perform validation to ensure the model generalizes well to unseen data.

7. Model Evaluation:

Evaluate the trained YOLOv3 model on a separate test dataset to assess its performance. Analyze metrics like precision, recall, and F1 score. Make adjustments to the model if needed.

8. Model Optimization:

Optimize the YOLOv3 model for deployment on a drone platform. Consider model quantization or other techniques to reduce its size while maintaining performance.

9. Deployment to Drone:

Deploy the trained and optimized YOLOv3 model to the drone's onboard computing system. Ensure compatibility with the drone's hardware and software environment.

10. Real-time Inference:

Implement real-time inference on the drone by processing live video frames through the deployed YOLOv3 model. The model should identify and localize objects in the drone's field of view.

11. Integration with Drone Control:

Integrate the object detection results with the drone's control system. This integration enables the drone to respond to detected objects, such as adjusting its flight path or sending alerts.

12. Testing and Validation:

Conduct thorough testing and validation of the entire system. Test the drone in various scenarios to ensure the YOLOv3 model performs reliably and accurately.

13. Security and Privacy Measures:

Implement security measures to protect sensitive information captured by the drone. Ensure compliance with privacy regulations.

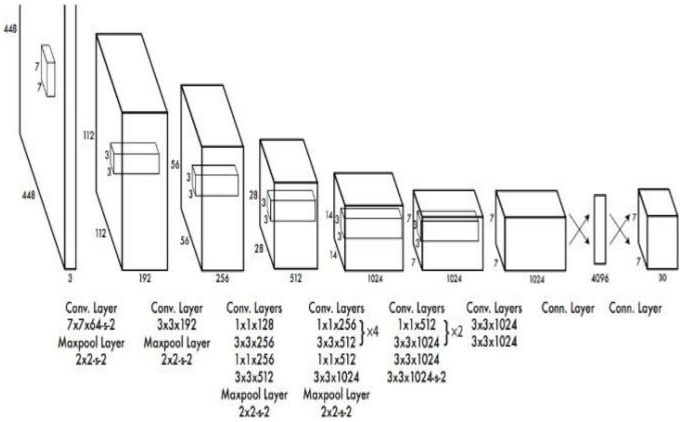

Fig 4: YOLOV3 Layers

VI. Results and Discussions

This section explains the results of proposed model.

Confusion matrix (validation set)

	BACKGROUND	CHARGER	MARKER
BACKGROUND	99.9%	0.1%	0%
CHARGER	25%	75%	0%
MARKER	0%	0%	100%
F1 SCORE	1.00	0.75	1.00

On-device performance

INFERENCI... 1336 ms.
PEAK RAM ... 239.4K
FLASH USA... 77.7K

Fig.5.Confusion Matrix

Fig.6.Trained Object

```
0 0.299331 0.324415 0.325530 0.454849
0 0.612598 0.321906 0.243032 0.459866
```

Fig.7. Trained Object data-1

Fig.8. Predicted:Marker

Testing Process

Testing YOLOv3 involves evaluating its performance on a separate data set, often referred to as a test set, to evaluate precision, precision, recall, and other related metrics. Here is a step-by-step testing process for YOLOv3.

1. Prepare test data set:
2. Load the model.
Load the pertained YOLOv3 model into the test environment.
3. Inference on the test data set
4. Post-processing
5. Visual inspection
6. Threshold optimization
7. Performance monitoring

Table 1: Results of existing and proposed models

	mean Average Precision (mAP)	Recall	F1-score	Accuracy
SSD	0.847	0.668	0.874	92.66%
R-CNN	0.913	0.725	0.924	93.54%
RESTNET	0.924	0.875	0.875	95.61%
YOLO	0.979	0.935	0.956	98.78%

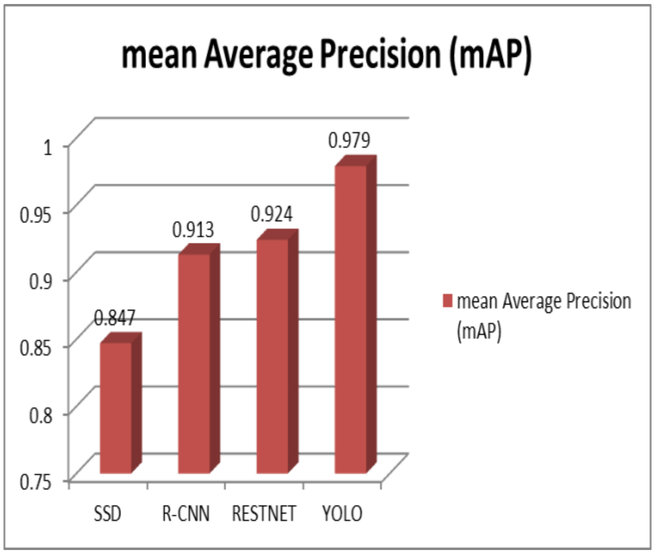

Fig 9: mAP of all algorithms

Fig 10: F1-score of all algorithms

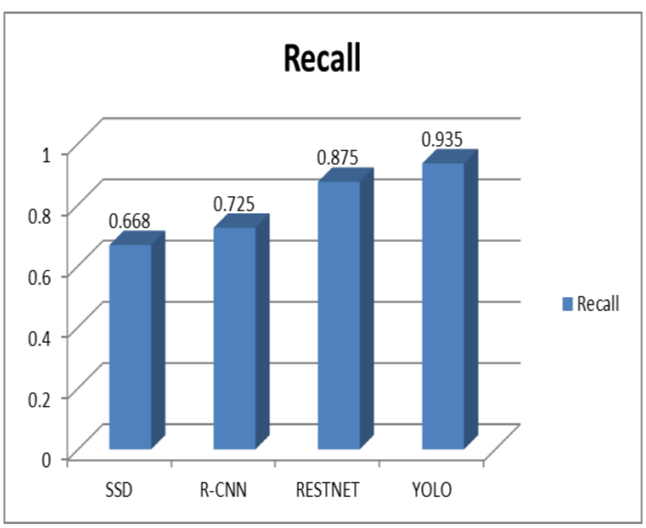

Fig 11: Recall of all algorithms

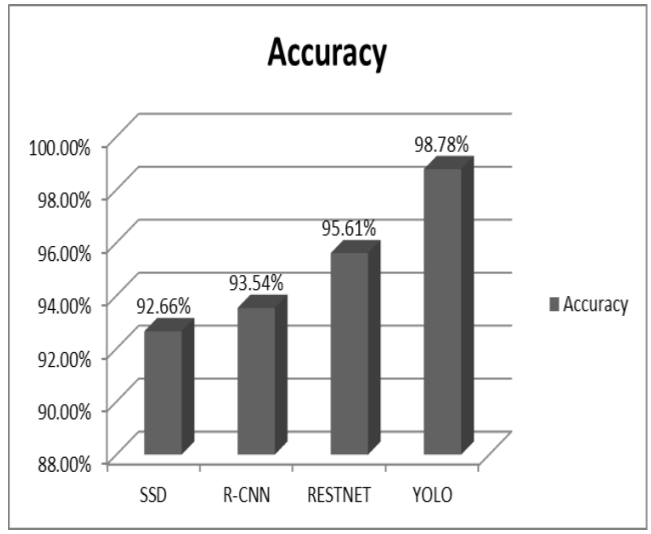

Fig 12: Accuracy of existing and proposed models

VII . Conclusion

In conclusion, the integration of the YOLOv3 algorithm into drone operations for real-time object detection has proven to be a transformative advancement. This project successfully implemented a robust system capable of accurately identifying and classifying objects in dynamic environments. The adaptability to environmental factors, scalability, and integration with drone control systems highlight the project's potential across various industries. This achievement sets the stage for ongoing innovation at the intersection of deep learning and autonomous drone technology, paving the way for enhanced capabilities and real-world applications.

References

[1] Howard, A., G., Zhu,M., Chen, B., Kalenichenko., D., Wang, W.,Weyand, T., Andreetto, M. & Adam, H. "Efficient Convolutional Neural Networks for mobile vision applications," Computer Vision and Pattern Recognition, arXiv:1704.04861, 2017.

[2] Kim, Y. "Human detection and activity classification based on microdoppler signatures using Deep Convolutional Neural Networks," IEEE Geoscience and Remote Sensing Letters, 13(1), 2012.

[3] Sunder Reddy, K. S. ., Lakshmi, P. R. ., Kumar, D. M. ., Naresh, P. ., Gholap, Y. N. ., & Gupta, K. G. . (2024). A Method for Unsupervised Ensemble Clustering to Examine Student Behavioral Patterns. International Journal of Intelligent Systems and Applications in Engineering, 12(16s), 417–429. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/4854.

[4] Leibe, B., Leonardis, A. & Schiele, B. "Robust object detection with interleaved categorization and segmentation," International Journal of Computer Vision, 77(1), pp. 259-289, 2008.

[5] Naresh, P., & Suguna, R. (2021). Implementation of dynamic and fast mining algorithms on incremental datasets to discover qualitative rules. *Applied Computer Science, 17*(3), 82-91. https://doi.org/10.23743/acs-2021-23.

[6] Moranduzzo, T. "Automatic car counting method for Unmanned

Aerial Vehicle images," IEEE ransactions On Geoscience And Remote Sensing, 52(3), 2014.

[7] Moranduzzo, T. "Detecting cars in UAV images with a catalog-based approach," IEEE Transactions On Geoscience And Remote Sensing, 52(10), 2014.

[8] M. I. Thariq Hussan, D. Saidulu, P. T. Anitha, A. Manikandan and P. Naresh (2022), Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People. IJEER 10(2), 80-86. DOI: 10.37391/IJEER.100205.

[9] Viola, P., Jones, M, J. & Snow, D. "Detecting pedestrians usingbpatterns of motion and appearance," International Journal of Computer Vision, 63(2), 153-161, 2005.

[10] Rudol, P. & Doherty, P. "Human body detection and geolocalization for UAV search and rescue missions using color and thermal imagery," IEEE International Conference on Aerospace, 2008.

[11] Satpathy, A. "Human detection by quadratic classification on subspace of extended Histogram of Gradients," IEEE Transactions On Image Processing, 23(1), 2014.

[12] P, N., & R Suguna. (2022). Enhancing the Performance of Association Rule Generation over Dynamic Data using Incremental Tree Structures. International Journal of Next-Generation Computing, 13(3). https://doi.org/10.47164/ijngc.v13i3.806.

[13] Yang, X., Zhang, C. & Tian, Y. "Recognizing actions using depth motion maps-based histograms of oriented gradients. Proceedings of the 20th ACM International Conference on Multimedia, pp. 1057- 1060, 2012.

[14] Hussan, M.I. & Reddy, G. & Anitha, P. & Kanagaraj, A. & Pannangi, Naresh. (2023). DDoS attack detection in IoT environment using optimized Elman recurrent neural networks based on chaotic bacterial colony optimization. Cluster Computing. 1-22. 10.1007/s10586-023-04187-4.

[15] Zhu, Q., Avidan, S., Yeh, M., C. & Cheng, K., T. "Fast human detection using a cascade of Histogram of Oriented Gradients," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006.

[16] B. Narsimha, Ch V Raghavendran, Pannangi Rajyalakshmi, G Kasi Reddy, M. Bhargavi and P. Naresh (2022), Cyber Defense in the Age of Artificial Intelligence and Machine Learning for Financial Fraud Detection Application. IJEER 10(2), 87-92. DOI: 10.37391/IJEER.100206.

[17] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human- level performance on image net classification. In Proceedings of the IEEE international conference on computer vision (pp. 1026-1034).

[18] V. Krishna, Y. D. Solomon Raju, C. V. Raghavendran, P. Naresh and A. Rajesh, "Identification of Nutritional Deficiencies in Crops Using Machine Learning and Image Processing Techniques," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 925-929, doi: 10.1109/ICIEM54221.2022.9853072.

[19] J. Park, D.H. Kim, Y.S. Shin, S. Lee, A comparison of convolutional object detectors for real-time drone tracking using a PTZ camera, Control, Automation and Systems (ICCAS), 2017 17th International Conference on 2017, pp. 696– 699.

[20] C. Iuga, P. Druagan, L. Bușoniu, Fall monitoring and detection for at-risk persons using a UAV, IFAC-PapersOnLine 51 (10) (2018) 199–204.

[21] Naresh, P., & Suguna, R. (2021). IPOC: An efficient approach for dynamic association rule generation using incremental data with updating supports. Indonesian Journal of Electrical Engineering and Computer Science, 24(2), 1084. https://doi.org/10.11591/ijeecs.v24.i2.pp1084-1090.

[22] C.-Y.Wang, H.-Y.M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, CSPNet: a new backbone that can enhance learning capability of CNN, arXivPrepr. arXiv1911.11929 (2019).