International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Real Time Object Detection Classification Using Cnn Algorithm

Prashant Kumar Tripathi, Ms. Namrata Sahayam
Department Of Electronics and Telecommunication Engineering, Jabalpur Engineering College M.P.

ABSTRACT
Real-time object detection and classification have become essential in various applications such as surveillance, autonomous vehicles, and smart systems. This paper presents a robust approach for real-time object detection and classification using Convolutional Neural Networks (CNN). The proposed method leverages deep learning techniques to automatically extract spatial features from input images and accurately identify objects within dynamic environments. A pre-trained CNN model is fine-tuned and integrated with a detection framework to achieve high accuracy and low latency. The system is evaluated on standard datasets, demonstrating improved performance in terms of precision, recall, and processing speed compared to traditional methods. The results indicate that CNN-based models provide an efficient and scalable solution for real-time object detection tasks in complex scenarios.
Keywords — YOLO, CNN, Real-time object detection, Deep Learning, RNN.

I. INTRODUCTION

Evolving technological landscape, computer vision plays a pivotal role in various applications, ranging from autonomous vehicles to surveillance systems, robotics, and beyond. Real-time object detection, a subfield of computer vision, has gained significant prominence due to its ability to identify and locate objects within live video streams or images in real-time. This capability has wide-ranging applications, from enhancing safety and security to enabling autonomous decision-making in machines. This project, titled "Real-Time Object Detection Using Convolutional Neural Networks (CNN) in MATLAB," is dedicated to developing a robust and efficient system that can perform real-time object detection with accurateness and speed. At its core, this project harnesses the power of deep learning, specifically CNNs, to revolutionize the way objects are detected and localized within visual data. The primary objective of this project is to create a real-time object detection system that operates seamlessly on live video feeds or images, allowing for the immediate identification and localization of objects within the frame. We aim to achieve state-of-the-art accuracy and reliability in object detection, ensuring that the system can make correct decisions with minimal false positives and false negatives. Real-time performance is of paramount importance.

II. DEEP LEARNING

Deep learning, a branch of machine learning, use neural networks to discover intricate patterns and connections in data. Deep learning aims to make it possible for machines to learn from massive amounts of data and base predictions or judgements on those predictions. Artificial neural networks that mimic the functioning of human brains are frequently used to create deep learning models. These networks are made up of layers of interconnected nodes, often known as "neurons," which are used to process and change data. Large amounts of labelled data are fed into the neural network as part of the deep learning learning process, and the weights and biases of the neurons are changed to reduce the error between the expected output and the actual output. Backpropagation is the procedure that enables the network to steadily enhance its performance over time. State-of-the-art performance in a variety of tasks, including picture and audio recognition, natural language processing, and autonomous vehicles, has been attained using deep learning. Additionally, it has facilitated innovations in fields like robotics, computer vision, and drug development. The ability of deep learning to automatically learn complex features and representations from data, its scalability to big datasets and complicated models, and its generalizability to new and unexplored data are some of its main benefits. Deep neural networks can be difficult to train and optimise, and deep learning also needs a lot of data and computer power. Despite this, deep learning is a science that is always evolving and growing thanks to the constant development of new architectures, methods, and applications.. Convolutional Neural Networks (CNNs): A class of neural network that is frequently employed in the processing of images and videos. CNNs are frequently used for tasks like image classification, object identification, and image segmentation because they extract features from pictures using convolutional layers. Recurrent neural networks (RNNs) are a subclass of neural networks that are employed in tasks involving sequence modelling, including speech recognition, language translation, and natural language processing. RNNs can represent sequences of different lengths because they employ recurrent connections to convey information from one time step to the next.

III. APPROACHES OF IMPLEMENTATION OF YOLO AND CNN

Object detection is a technology that detects the semantic Objects of a category in virtual snap shots and films. Certainly, one of its Actual-time packages is self-riding automobiles. In this, our challenge is to locate a couple of items from a photo. The maximum common Item to come across on this utility is the car, motorcycle, and Pedestrian. For locating the gadgets within the photograph, Object localization and should find multiple item in real-time structures. There are various techniques for item Detection, they can be break up into classes, first is the Algorithms primarily based on classifications. CNN and RNN come below this category. On this, pick out the involved Regions from the picture and ought to classify them the use of Convolutional neural community. This technique may be very sluggish Due to the fact it should run a prediction for

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

every decided on Vicinity. The second one class is the algorithms primarily based on Regressions. Yolo approach comes below this category. In This, it might not choose the fascinated regions from the photograph. Instead, it expect the training and bounding containers of the complete picture at a single run of the algorithm and detect a couple of gadgets using an unmarried neural community. Yolo Set of rules is rapid as compared to other classification Algorithms. In actual time our algorithm technique 45 frames consistent with 2d. Yolo algorithm makes localization mistakes however Predicts less fake positives in the background [3]. Humans can without problems come across and pick out gadgets of their environment without attention in their instances, irrespective of what position they're in and whether or not they are the wrong way up, one-of-a-kind in shade or texture, in part occluded, and so forth. To extract information about the objects and shapes in a picture, the same item detection and recognition on a computer requires a lot of processing. Identifying an object in an image or video is referred to end as an object detection in computer vision. The main steps in object detection or future extraction, which is imported for surveillance, cancer reduction, car identification, and underwater object detection, among other applications. Different approaches had been used to accurately and correctly identify the object for specialized packages. These suggested solutions still have an issue with inaccuracy and inefficiency, though. Device learning and deep neural network approaches are more successful in addressing these issues of object detection [4].

IV. YOLO IN IMAGE AND VIDEO PROCESSING

YOLO, which stands for "You Only Look Once," is a real-time object detection algorithm widely used in image processing and computer vision. Developed by Joseph Redmon and Santosh Divvala, YOLO approaches object detection in a unique way compared to traditional methods.

Here are key aspects of YOLO:

Single Forward Pass:

YOLO performs object detection in a single forward pass through the neural network. This is in contrast to traditional methods that often involve multi-stage processing.

Grid-Based Approach:

The image is divided into a grid, and each grid cell is responsible for predicting bounding boxes and class probabilities for objects contained within it. This grid-based approach allows YOLO to efficiently process the entire image.

Bounding Box Prediction:

Each grid cell predicts multiple bounding boxes along with their confidence scores and class probabilities. YOLO uses anchor boxes to improve the accuracy of bounding box predictions.

Class Prediction:

YOLO predicts class probabilities for each object class within the grid cell. The algorithm is capable of detecting multiple object classes in a single image.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Non-Maximum Suppression (NMS):

After predictions are made, YOLO uses NMS to eliminate redundant or low-confidence bounding boxes, keeping only the most relevant ones.

V. RESEARCH MOTIVATION

People do lead a normal life with their own style of doing things. But, they definitely face troubles due to inaccessible infrastructure and social challenges. The biggest challenge for a person, especially the one with the complete loss of vision, is to navigate around places. Obviously, roam easily around their house without any help because they know the position of everything in the house. People have a tough time finding objects around them. So we decided to make a REAL TIME OBJECT DETECTION System. We are interested in this project after we went through few papers in this area. As a result we are highly motivated to develop a system that recognizes objects in the real time environment.

RESEARCH AIM

The aim of this synopsis is to evaluate the classification performance of the suitable deep learning models for real-time object recognition and tracking. The capability of machines to identify the suspicious object and further identify their activities in a specific environment is an important part of permitting a machine to interact with humans in effective and easy manner. The current approach for analyzing and detecting the suspicious object usually needs exceptional markers connected to the suspicious object that prevents the extensive technology application. In this synopsis, to study as well as analyze the previous approach towards object tracking using video sequences through different phases. Three key steps in video analysis are discussed as follows:

1. Identification of targeted object in moving sequence.
2. Object recognition based on one frame to another frame.
3. Tracking of the object from camera.
4. Output gives in Text form.

RESEARCH OBJECTIVES

The following objectives have been identified to fulfil the aim of this synopsis work:

• To identify suitable and highly efficient deep learning models for real-time object recognition and tracking of person variable object.

• Evaluate the classification performance of the selected deep learning models.

• Compare the classification accuracy performance of the selected models among each other and present the results.

• Implementation of simulation model object name gives output.

VI. METHODOLOGY

In this section are describe general design principles, system requirements, feature extraction, system architecture and image quality improvement and image denoising.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

GENERAL DESIGN PRINCIPLES

Here we are proposed YOLOV3 and CNN models, One of the important fields of Artificial Intelligence is Computer Vision the science of computers and software systems that can recognize and understand images and scenes. Computer Vision is also composed of various aspects such as image recognition, object detection, image generation, image super-resolution and more. Object detection is probably the most profound aspect of computer vision due the number of practical use cases.

Object detection refers to the capability of software systems to locate objects in an image/scene and identify each object. It has been widely used for face detection, vehicle detection, pedestrian counting, web images, security systems and driver less cars. There are many ways object detection can be used as well in many fields of practice. Like every other computer technology, a wide range of creative and amazing uses of object detection will definitely come from the efforts of computer programmers and software developers. Getting to use modern object detection methods in applications and systems, as well as building new applications based on these methods is not a straight forward task.

Early implementations of object detection involved the use of classical algorithms, the popular computer vision library. However, these classical algorithms could not achieve enough performance to work under different conditions.

Advantages of proposed system

- High Accuracy
- Very Effective Models

SYSTEM REQUIREMENTS

Software Requirements: The functional requirements or the overall description documents include the product perspective and features, operating system and operating environment, graphics requirements, design constraints and user documentation.

The appropriation of requirements and implementation constraints gives the general overview of the project in regards to what the areas of strength and deficit are and how to tackle them.

- Python del 3.7 version (or)
- Anaconda 3.7 (or)
- Jupiter (or)
- Google colab

FEATURE EXTRACTION

In image processing, feature extraction refers to the process of selecting and representing essential information or features from an image. The goal is to capture relevant patterns, structures, or characteristics that are crucial for a particular analysis or task while reducing the dimensionality of the data. Feature extraction is a fundamental step in image analysis, computer vision, and pattern recognition.

SYSTEM ARCHITECTURE

In Convolutional Neural Networks (CNNs), feature extraction refers to the process of automatically identifying important patterns or features from input data. CNNs are particularly effective in tasks such as image recognition and computer vision, where the hierarchical structure of features is crucial for understanding the content of an image.

Here's a breakdown of the feature extraction process in CNNs:

Convolutional Layers: These layers apply convolution operations to the input data. Convolution involves sliding a filter (also known as a kernel) over the input image to detect patterns such as edges, textures, or colors. Multiple filters are used to capture different features.

Activation Function: After the convolution operation, an activation function (commonly ReLU - Rectified Linear Unit) is applied element-wise to introduce non-linearity. This helps the network learn more complex relationships in the data.

Pooling (Subsampling) Layers: Pooling layers downsample the spatial dimensions of the input data. Common pooling operations include max pooling or average pooling. This reduces the computational load and makes the network more robust by retaining the most salient features.

Flattening: After several convolutional and pooling layers, the data is flattened into a vector. This vector serves as the input for the fully connected layers.

Fully Connected Layers: These layers are typically found towards the end of the network and connect every neuron to every other neuron in the preceding and succeeding layers. Fully connected layers help in learning global patterns and relationships in the data.

LeNet-5: One of the earliest CNN architectures, developed by Yann LeCun for handwritten digit recognition. It consists of convolutional and subsampling layers, followed by fully connected layers.

VII. IMAGE QUALITY IMPROVEMENT AND IMAGE DENOISING

Image segmentation is a computer vision task that involves dividing an image into distinct and meaningful segments or regions. The goal is to partition the image into regions that share similar visual properties, such as color, intensity, texture, or other relevant features. Each segment or region is expected to represent a coherent and semantically meaningful part of the image. Image segmentation is a crucial step in various computer vision applications in object recognition.

Image denoising is the process of reducing noise from a digital image to improve its visual quality and enhance the interpretability of its content. Noise in an image is typically unwanted variations in pixel values that arise due to various factors, such as sensor limitations, transmission errors, or environmental conditions during image acquisition. The goal of image denoising is to retain the important structures and details in an image while suppressing or removing the undesirable noise.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

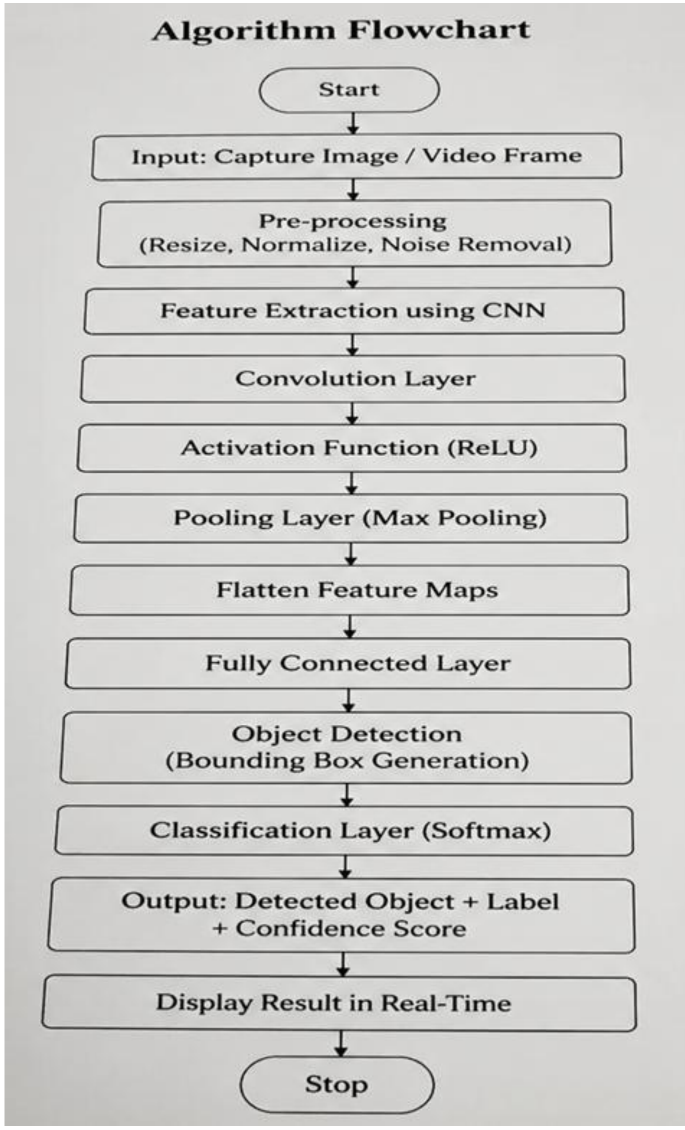


Fig 1. System Flow Diagram

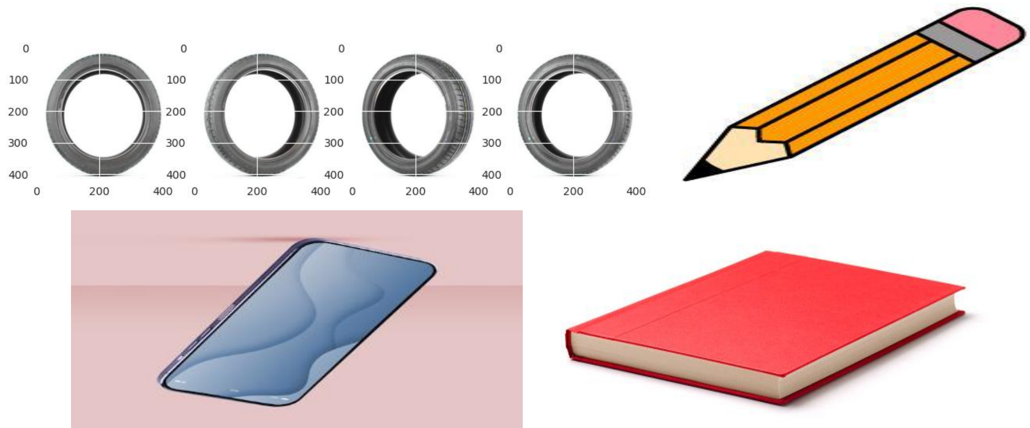
Convolutional Neural Networks (CNN) are used by the YOLO method to recognize items instantly. This technique, as its name suggests, only needs one forward propagation through the neural network to identify an item. Three strategies are used by the YOLO algorithm: Bounding box regression with residual block union (IOU) Remaining Block- The image is first separated into several grids. Each grid has S x S dimensions. The grids created from the input image are displayed in the image below. CNNs are used to simultaneously forecast the probability and bounding boxes for various classes. There are various iterations of the YOLO algorithm. Little

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

YOLO and YOLOv3 are the most popular. Residual Block-Bounding Box Regression Union (IOU), Residual Block, and the YOLO algorithm all function in three different ways.

VIII. RESULT AND SIMULATION

Data set image



S.N.	Object	Accuracy
1	Tier	92%
2	Pencil	98%
3	Mobile	100%
4	Book	99%

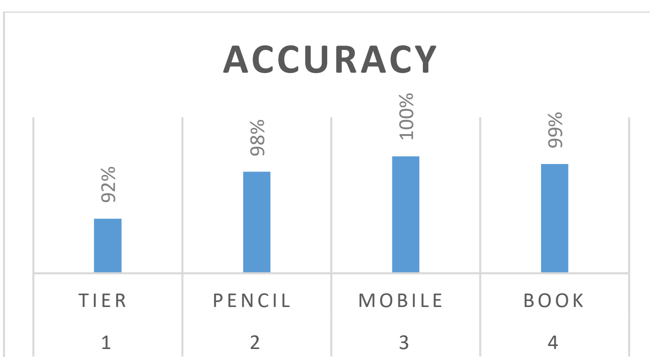


Fig.2. Accuracy.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

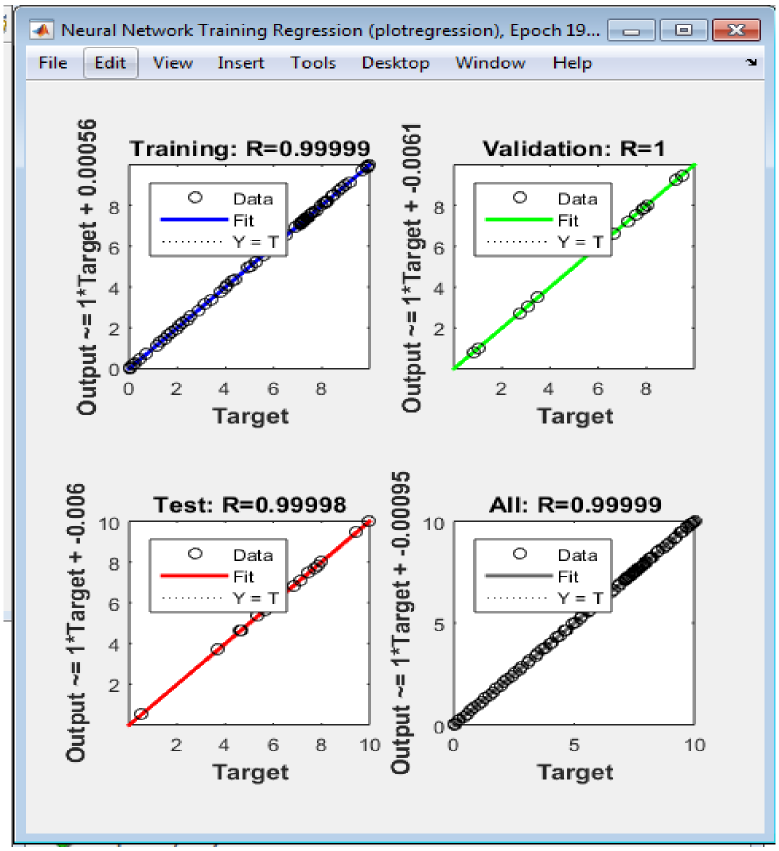


Fig.3. Prediction.

IX. CONCLUSION

Finally the conclusion the progress made in computer vision systems, particularly in object detection and recognition, combining traditional methods with deep learning models. The reliability of these systems in accurately identifying and categorizing objects, even in complex situations, showcases their potential for practical use. The ability for real-time processing further boosts their effectiveness in time-critical tasks like surveillance and autonomous vehicles. While these achievements are significant, recognizing the limitations of these systems opens up opportunities for future research and enhancement. Overcoming obstacles such as challenging lighting conditions, heavy obstruction, and similar object appearances offers avenues for refining these systems. Exploring advanced techniques for feature extraction, integrating contextual details, and utilizing multimodal data fusion methods show promise in improving performance. Looking ahead, continuous improvements in detection network models, with an emphasis on reducing memory usage and increasing speed, will be essential. Broadening the scope of recognizable object classes will expand the applicability of these systems across different fields. Ultimately, these advancements contribute to the progression of computer vision technology, unlocking new possibilities for a variety of applications, including video surveillance. As part of the future enhancements, the model will be custom trained with the other objects to increase its detection capability. With the help of transfer learning, the used network will be trained with other objects to increase the scope of objects the MobileNet can detect.

This research presents an efficient approach for real-time object detection and classification using Convolutional Neural Networks (CNN). The proposed system successfully demonstrates the capability of CNN models to extract meaningful features and accurately identify objects in dynamic and complex environments. Experimental results confirm that the method achieves high detection accuracy, improved precision and recall, and faster processing speed compared to conventional techniques. The integration of deep learning with real-time processing makes the system suitable for practical applications such as surveillance, autonomous systems, and smart monitoring. Overall, the study highlights that CNN-based object detection provides a reliable, scalable, and high-performance solution for modern computer vision challenges.

REFERENCES

1. Chen, Y., Zhang, H., Wang, L., & Wu, Q. (2023). A survey on visionbased UAV navigation. IEEE Transactions on Intelligent Transportation Systems, 24(3), 2835-2854. https://doi.org/10.1109/TITS.2022.3224325
2. Zhang, Y., Yuan, Y., Feng, Y., & Lu, X. (2023). UAV swarms in urban environments: A survey of emerging trends and challenges. Drones, 7(2), 98. https://doi.org/10.3390/drones7020098
3. Li, K., Ni, W., Tovar, E., & Jamalipour, A. (2023). On-board deep learning in UAV-based applications: Opportunities and challenges. IEEE Internet of Things Journal, 10(4), 3553-3575. https://doi.org/10.1109/JIOT.2022.3215698
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In Proceedings of the International Conference on Learning Representations (ICLR). https://doi.org/10.48550/arXiv.2010.11929
5. Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in vision: A survey. ACM Computing Surveys, 54(10s), 1-41. https://doi.org/10.1145/3505244
6. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637-646. https://doi.org/10.1109/JIOT.2016.2579198
7. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 779-788). https://doi.org/10.1109/CVPR.2016.91

International Journal of Engineering, Science and Humanities

An international peer reviewed, refereed, open-access journal

Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

8. Zhang, Y., Yuan, Y., Feng, Y., & Lu, X. (2021). Cascade det: A universal cascade detection framework for small objects in dronecaptured scenes. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-13. https://doi.org/10.1109/TGRS.2021.3138849
9. Xu, Z., Wu, W., Qi, L., & Lu, X. (2023). Efficient vision transformers for edge devices: A survey. IEEE Access, 11, 28456- 28478. https://doi.org/10.1109/ACCESS.2023.3258741
10. Rizk, Y., Awad, M., & Tunstel, E. W. (2019). Cooperative heterogeneous multi-robot systems: A survey. ACM Computing Surveys, 52(2), 1-31. https://doi.org/10.1145/3303848
11. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (Vol. 25, pp. 1097-1105). https://doi.org/10.1145/3065386
12. Mishra, R. (2024). Raspberry Pi Performance analysis across its Operating System in LED Control Operation. International Journal of Advanced Research and Multidisciplinary Trends (IJARMT), 1(2), 01-11.
13. Mishra, R. (2025). IOT and DSP (combination of hardcore Virtex-5 FPGA and soft core DSP processor) OFDM System PAPR Reduction Using Artificial Intelligence Algorithm. International Journal of Advanced Research and Multidisciplinary Trends (IJARMT), 2(1), 135-149.
14. Mishra, R., & Sharma, A. (2026). Enhanced Trajectory Tracking of a 6-DOF Robotic Manipulator Using GA–PID and ANN–PID Controllers. International Journal of Research & Technology, 14(2), 53-70.
15. Vaswani, A., Ramachandran, P., Srinivas, A., Parmar, N., Hechtman, B., & Shlens, J. (2021). Scaling local self-attention for parameter efficient visual backbones. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 12894-12904). https://doi.org/10.1109/CVPR46437.2021.01270