JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

OBJECT DETECTION AND RECOGNITION USING FEATURE PYRAMID NETWORKS(FPN) IN ARTIFICIAL INTELLIGENCE

1st Mrs. Gowtham S,B.E.,M.E.,(Ph.D.),

Assisstant Professor,
Department of Information Technology
K. S. Rangasamy College of Engineering
Tiruchengode, India

2nd Nithish N
Department of Information Technology
K. S. Rangasamy College of Engineering
Tiruchengode, India

3rd Saran M S
Department of Information Technology
K. S. Rangasamy College of Engineering
Tiruchengode, India

4th Vipin S
Department of Information Technology
K. S. Rangasamy College of Engineering
Tiruchengode, India

***Abstract*— Efficient and accurate object detection is crucial for advancing computer vision systems, with broad applications in fields such as self-driving cars. Over the past fifty years, object detection methods have continuously evolved, yielding numerous promising approaches. These systems identify objects in digital images or videos, spanning diverse classes such as humans and cars. To detect objects, a system relies on several components: a model database, a feature detector, a hypothesizer, and a hypothesizer verifier, which collaborate to achieve accurate detection. Techniques like localization, categorization, and feature extraction are employed to extract appearance characteristics from images and videos, thereby enhancing the accuracy and efficiency of object detection systems.**

Keywords— Artificial Intelligence (AI); Computer Vision (CV); Convolution Neural Network (CNN); You Look Only Once (YOLOv3); Urban Vehicle Dataset; Common objects in Context (COCO); Object detection; object tracking

I. INTRODUCTION

AI-powered object detection is a cutting-edge computer vision approach that's transforming a number of sectors by allowing robots to locate and identify items automatically in photos or videos. By utilizing deep learning methods, specifically convolutional neural networks (CNNs), these models acquire the ability to properly detect objects of interest by learning hierarchical representations of data. Object detection is essential for a wide range of applications, from surveillance systems keeping an eye on public areas to autonomous cars traversing challenging settings. Benefits like high accuracy, scalability, and real-time performance are offered; however, there is still ongoing research being done to address issues like bias, interpretability, and data scarcity. Nevertheless, as object detection continues to improve, it is pushing the limits of what robots can see and comprehend, opening the door for ground-breaking discoveries.

By enabling computers to automatically recognize, locate, and classify things inside images or videos with amazing precision and efficiency, object detection with AI transforms computer vision. Using deep learning methods, such as convolutional neural networks (CNNs), object identification models are able to generalize well across a wide range of circumstances and environments by learning complex patterns and features from large datasets. Object detection finds widespread and revolutionary uses throughout industries, from autonomous vehicles navigating intricate roadways to surveillance systems monitoring public spaces and retail businesses optimizing inventory management. Even though there are still issues with data scarcity, model interpretability, and ethical considerations, continual advances in AI are pushing the frontiers of object identification and pointing to a time when machines will be able to sense and comprehend the visual world with previously unheard-of precision.

A.Objective

The primary objectives of AI-based object detection include developing highly accurate models capable of swiftly and precisely recognizing and categorizing items in photos or videos in real-time while ensuring scalability across diverse datasets and deployment scenarios. To instill confidence and accountability, these models must exhibit robust generalization to novel situations while maintaining interpretability. Furthermore, efforts are concentrated on enhancing resilience against adversarial attacks, safeguarding privacy, mitigating biases, and facilitating continual advancement through ongoing research and innovation, with the overarching aim of pushing the boundaries of performance and applicability across various sectors and domains.

Regarding connectivity, a system referred to as Personal Assistant with Voice Recognition Intelligence is introduced, which accepts user input in the form of voice or text and processes it, providing output in various formats such as actionable instructions or dictated search results. Moreover, this proposed system has the potential to revolutionize interactions between end users and mobile devices. The system is being designed to enable end users to access all services provided by their mobile devices through voice commands.

B.Applications of Detection

Numerous industries use object detection extensively. These industries include autonomous vehicles, which use it to identify pedestrians, cyclists, and other vehicles and ensure safe navigation; surveillance systems, which use it to detect suspicious activity and intruders in public spaces; retail, which uses it to manage inventory, optimize stock levels, and improve the shopping experience; healthcare imaging, which uses it to more accurately diagnose conditions like tumours and fractures; industrial automation, which uses it to perform tasks like quality control and robotic assembly processes; agriculture, which uses it to monitor crop health and optimize resource usage; natural disaster management, which uses it to prioritize rescue efforts; and environmental monitoring

Through the use of traffic management systems, object detection technology optimizes traffic flow and ensures road safety. It also facilitates natural human-computer interaction through gesture recognition. It improves sports analytics by tracking player movements, helps with document analysis by automatically extracting information from scanned documents, and supports wildlife conservation initiatives by keeping an eye on endangered species and reporting poaching activity. Additionally, it drives environmental monitoring to evaluate ecosystem health and sustainability, provides fire detection for prompt response and evacuation, and powers retail analytics to comprehend customer behaviour. These many uses demonstrate the adaptability and importance of object detection technology in solving challenging problems and spurring creativity in a range of fields.

C.Detects Object

Object detection using AI operates by training deep learning models on large datasets of annotated images, enabling them to learn to recognize and classify objects within images or videos. During training, the models adjust their parameters to minimize prediction errors based on ground truth annotations. Once trained, these models are deployed for inference, where they analyze new data and predict the presence and location of objects. Post-processing techniques are then applied to refine the detection results, ensuring accuracy and reliability. Through iterative evaluation and refinement, object detection models can achieve high accuracy and robustness, enabling their deployment in various real-world applications such as autonomous vehicles, surveillance systems, healthcare, and retail analytics, among others.

Once the object detection model meets the desired performance criteria, it can be deployed in real-world applications. This may involve integrating the model into software systems, embedded devices, or cloud-based services, depending on the specific deployment requirements.

II. LITERATURE REVIEW

Finding objects is aided by the use of the Tile convolution neural network and the recursive mode of the same network in driver assistance systems (DAS). To assist in learning and modifying weights depending on a wide range of training data, the approach utilizes unsupervised training. The purpose of including obstacle validation techniques is to lower the number of valid detections.

[1] The analysis of motion of things that are not visible to the naked eye involves the application of concepts such as optical flow and magnitude histogram. Classification and localization are used to detect normal and aberrant events, which helps the campus environment distinguish between the two objects.

[2] Pretrained networks are used to extract features, and SVM is used to distinguish between the classified outputs. Approach aids in ITS route guidance.

[3] Numerous methods, such as feature extraction based on colour and gradients fail to give spatial positioning in the image. The challenges are overcome by employing Analysis of principal components by PCANet.

[4] Pipeline that uses coordinates and velocities for picture undistortion, registration, classification, and detection. Method employs FAST and FREAK descriptors as detectors, and Squeeze Net categorization comes next.

[5] The process of creating candidate targets, taking features out of them, and placing ground truth boxes around objects all help with tracking. VGGNet is used to classify the items.

[6] Originally intended for image classification, CNN was modified to recognize objects. To bound objects identified, the method handles object detection as a regression for the object class. Gradual advancements have been observed, starting with RCNN, moving on to Fast RCNN and Faster RCNN, and ultimately arriving at YOLO. Frames per second are processed more quickly since a picture is scanned once rather than repeatedly evaluated, as in CNN. YOLO is trained based on losses, unlike the traditional classification approach.

[7] The paper discusses the area of road traffic video analytics. Vehicle counting is one of the primary

application areas, in addition to vehicle detection and tracking. The Single Shot Detector (SSD), a cutting-edge algorithm, is utilized. Features like Binary big objects are handled via algorithms. In applications like object categorization, it produces superior outcomes. Concepts like the virtual coil technique and background subtraction are used in object tracking. SSD performs better than YOLO versions in terms of precision. When choosing an object detection algorithm, there are always trade-offs between speed and precision. A performance metric with a speed of 58 frames per second and an accuracy of over 85% is considered optimal.

[8] The paper describes how the upgrade to YOLO was made. Updates have been made gradually during the course of the YOLOv1, YOLOv2, and YOLOv3 series. State-of-the-art technology is YOLOv3. improvements like slimmer-bounding boxes that don't touch neighboring pixels. The COCO dataset, using YOLOv3, demonstrates that mAP is just as good as SSD. Results from YOLOv3 are three times faster. YOLOv3 claims to be able to detect tiny things.

[9] As the number of vehicles in metropolitan areas increases, single-object tracking will not be able to meet demand. It is possible to track many objects by using a kernelized correlation filter (KCF). Many KCFs are operated concurrently. KCF works well in pictures with occlusions. When background removal is paired with KCF, trustworthy data regarding urban traffic is produced.

[10] More processing power, time, and data are needed for Deep Networks to perform as well as Neural Nets do. Any algorithm's success depends on its parameter adjustment. Algorithms vary depending on the application. Modern Neural Nets can be fine-tuned to reduce training time and increase accuracy. The dataset, algorithm, and network used all affect the results.

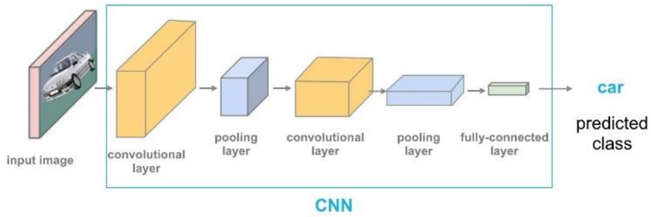
III. Existing System

AI-based object detection has made great strides in a variety of approaches and systems. From the early, ground-breaking work of R-CNN to efficiency-focused approaches such as EfficientDet, the field has advanced to provide real-time processing capabilities with high accuracy. Single-shot detection systems have been developed with the use of techniques like SSD and YOLO, which allow for quick inference without sacrificing accuracy. Furthermore, advances such as Mask R-CNN have expanded object detection to include instance segmentation problems, while datasets such as COCO and Pascal VOC have offered consistent evaluation benchmarks. These developments have spurred applications in a variety of fields, including surveillance, autonomous driving, and healthcare, illustrating the broad reach of object detection in AI-driven systems.

A. R-Convolutional Neural Networks (R-CNN)

Object detection has advanced significantly since R-CNN's original conception and its development into Faster R-CNN. By fusing region proposals with CNNs, Girshick et al. (2014) invented the R-CNN framework and brought about significant gains in accuracy. Later developments, including Girshick's (2015) Fast R-CNN, improved speed and accuracy even more by optimizing computation throughout the entire image. But faster R-CNN was first shown by Ren et al. (2015), who completely changed the field by incorporating the region proposal network (RPN) into the detection pipeline. This breakthrough made it possible to detect objects in real time, raising the bar for object detection systems' effectiveness and performance.

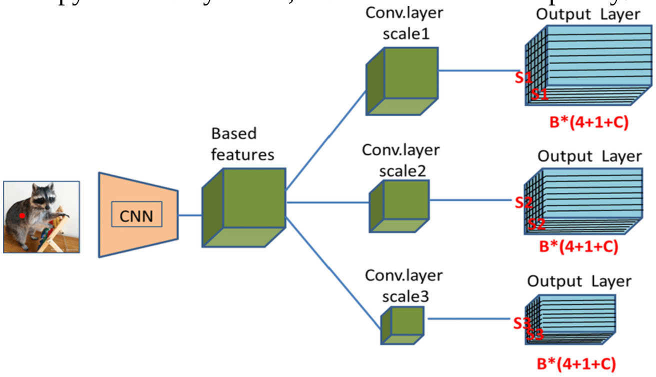
Another significant limitation of the current system is its lack of noise cancellation capabilities. In environments with background noise or disturbances, the voice assistant may struggle to accurately interpret and process user commands.



B. You Only Look Once (YOLOv3)

You Only Look Once (YOLO): Redmon et al. (2016) introduced YOLO, a real-time object detection system. In a single neural network pass, it can predict bounding boxes and class probabilities from whole images.

YOLO has undergone several iterations, with each version enhancing its design and training techniques to improve accuracy and speed (YOLOv2, YOLOv3). In YOLO versions 1 and 2, scores are converted into probabilities by using softmax algorithms, which assume mutually exclusive items. On the other hand, YOLOv3 makes use of multiple label classification, employing independent logistic classifiers to ascertain the probability that an input belongs to a given label. By eliminating the softmax function, the loss computation is based on binary cross-entropy for every label, which reduces complexity.



IV. Proposed System

A. *Analysis and design*

The structured approach to analysis and design in object detection using AI involves several critical steps. Initially, understanding the problem scope and requirements sets the stage, followed by meticulous data collection and preprocessing to ensure a high-quality training dataset. Model selection and architecture design involve choosing appropriate detection frameworks and customizing them to suit the specific task requirements. Subsequently, defining training objectives, optimizing strategies, and evaluating performance ensure the model's effectiveness and efficiency. Deployment and integration steps focus on deploying the model in real-world scenarios and seamlessly integrating it with existing systems. Finally, iterative refinement and maintenance ensure continuous improvement and adaptation to evolving requirements and environments, ultimately enabling the creation of robust and effective object detection systems tailored to specific applications and scenarios, leveraging the capabilities of AI for accurate and efficient detection tasks.

B. Benefits of Object Detection

AI-based object detection has several benefits in a variety of fields and uses. First off, compared to human approaches, it saves a significant amount of time and resources by automating the identification and localization of objects inside photos or videos, hence streamlining procedures. Tasks like retail inventory management, manufacturing quality control, and monitoring are made more efficient by this automation. Additionally, AI algorithms are incredibly accurate, particularly when trained on large and varied datasets, which supports decision-making in crucial applications like medical imaging for illness detection. Furthermore, AI-based object detection is scalable, which means it can easily manage high data volumes and rising computational needs. This makes it appropriate for real-time processing in applications such as infrastructure monitoring in smart cities and autonomous navigation. The adaptability of models for object identification driven by AI.

V. SYSTEM DESIGN

A.1.Architecture Selection

Considering the limitations and specifications of the application, select an appropriate object detection architecture. Take into account variables like computational resources, accuracy, and speed. One-stage detectors like YOLO or SSD, two-stage detectors like Faster R-CNN, or more sophisticated designs like Mask R-CNN or EfficientDet are among the available options.

A.2.Data Pipeline

Create a pipeline for data to be ingested, pre-processed, and enhanced for training. Resizing, normalization, and augmentation methods like flipping, rotating, and color jittering can be used as preprocessing processes to increase the diversity of the dataset. To effectively feed data into the training process, use data loaders.

A.3.Model Development

Utilizing a deep learning framework like Tensor Flow or PyTorch, implement the selected object detection architecture. Modify the architecture as necessary to meet the application's unique needs; for example, add attention methods, change the backbone networks, or add context information.

A.4.Training Strategy

Establish training objectives and evaluation criteria, including mAP and loss functions like cross-entropy loss. Divide the dataset into test, validation, and training sets. Then, provide the means to track the model's performance as it is being trained.

A.5.Optimization Techniques

To train the model effectively and prevent convergence problems, use learning rate schedules and optimization techniques (such as Adam and stochastic gradient descent). To accelerate training and boost output, apply strategies like transfer learning and pretrained model fine-tuning.

V. IMPLEMENTATION

A.Modules Description

A.1.Convolutional Neural Networks:

One subclass of deep neural networks called convolutional neural networks (CNNs) is typically used to analyze organized data that is grid-like, like photographs. They are the foundation of many cutting-edge object identification algorithms and have completely transformed the field of computer vision.

A.2.Region Proposal Network:

Region Proposal Networks, or RPNs, are an essential aspect of many modern object identification models, especially those that use a two-stage methodology. For two-stage frameworks like Faster R-CNN, the Region Proposal Network is essential in producing high-quality candidate object areas quickly and accurately, which makes object recognition possible.

A.3.Data Augmentation:

By performing several modifications on the pre-existing data samples, a technique known as "data augmentation" is frequently used in machine learning and deep learning to artificially increase the diversity of a training dataset. Data augmentation is especially crucial for enhancing model generalization and avoiding overfitting when it comes to AI-based object recognition.

A.4.Backbone Networks:

Many deep learning-based object identification systems are built on backbone networks, commonly referred to as backbone architectures or backbone models. From the input photos, they are in charge of extracting high-level features that are subsequently utilized for other tasks, including segmentation, classification, and object detection.

B.Implementation Details:

B.1.Data Preparation:

Obtain or produce a tagged dataset with pictures of the objects of interest annotated with the appropriate bounding boxes. Preprocess the dataset by increasing diversity and resilience, levelling pixel values, and resizing photos to a consistent size.

B.2.Choose a Model:

Based on the particular needs of the application, such as speed, accuracy, and resource limits, choose an appropriate object detection model architecture. Retina Net, SSD (Single Shot Multibox Detector), YOLO (You Only Look Once), and Faster R-CNN are popular options.

B.3.Model Training:

To make use of transfer learning, initialize the selected model with pretrained weights on a sizable dataset (such as ImageNet). Use bounding boxes and annotated images to fine-tune the model on the target dataset. Establish suitable loss functions (e.g., cross-entropy loss for classification, smooth L1 loss for bounding box regression) for object localization and classification tasks. While keeping an eye on performance on a validation set, train the model using optimization methods such as stochastic gradient descent (SGD) or its derivatives (e.g., Adam).

B.4.Evaluation:

Evaluate the trained model's performance using evaluation metrics like mean Average Precision (mAP), recall, accuracy, and precision on a different test set. Analyze false positives and false negatives and visualize the detection findings to perform a qualitative analysis.

B.4.Monitoring and Maintenance:

Keep an eye on the deployed model's performance in production to identify problems like idea drift or data drift. Get input from stakeholders and users to determine what needs to be improved and where changes should be made. Retrain the model with fresh data on a regular basis to adjust to changing circumstances and requirements.

C. Technology Used

Artificial intelligence and object recognition are used in object detection to quickly and precisely provide the desired outcome for the user. Asking a computer to set a timer may seem straightforward, but the technology that makes it possible is amazing.

C.1.Feature Pyramid Networks:

FPNs, which are frequently combined with object detection models, offer multi-scale feature maps that aid in addressing scale variation. This makes it possible to detect items in an image with greater accuracy at different sizes.

C.2.Anchor Box:

Anchor boxes are pre-built boxes with various aspect ratios and scales that are utilized in a lot of object identification models. They help increase detection accuracy by acting as reference bounding boxes for estimating object positions and sizes.

C.3.Transfer Learning:

Pretrained CNN models are frequently employed as feature extractors for object detection applications. These models were trained on extensive picture datasets, such as ImageNet. Through the use of transfer learning, performance and convergence speed can be increased by bootstrapping training on smaller, task-specific datasets using the learned representations from these models.

VI. CONCLUSION AND FURTHER SCOPE

A.1.Further Integration:

In order to improve the capabilities, effectiveness, and applicability of the system, further integration in object detection using AI entails combining techniques like multi-modal fusion, domain adaptation, uncertainty estimation, attention mechanisms, semantic and instance segmentation, and real-time processing optimizations with interactive interfaces. Finer-grained knowledge of images is achieved by merging semantic and instance segmentation, which enables accurate object delineation. Diverse data sources are integrated through multi-modal fusion to provide richer context, while attention mechanisms allow for dynamic emphasis on pertinent regions. Enhancement of model generalization is achieved by domain adaptation and transfer learning, while prediction confidence is revealed through uncertainty estimation. Optimizing real-time processing speeds up and increases efficiency, while interactive interfaces help users communicate and comprehend one another. By means of this integration, artificial intelligence-based object identification systems may tackle intricate real-world problems in diverse applications and areas.

A.2.Natural Conversations:

People may want to know more about the capabilities, precision, uses, restrictions, and possible ramifications of AI-based object detection. They may ask questions about how AI systems find and identify items in photos or videos, how accurate they are in comparison to human performance, and how they may be used in real-world scenarios like retail, autonomous cars, and surveillance. In addition, talks may cover the difficulties in gathering enough training data and the possible limits of object detection algorithms when dealing with a variety of objects and environmental circumstances. All in all, these discussions offer a thorough and approachable way to investigate the potential, difficulties, and consequences of object detection by AI.

A.3.Conclusion:

We are able to detect objects more precisely and identify each one individually with its specific location in the image along the x and y axes by employing this thesis and the experimental results. Additionally, it presents experimental findings comparing several approaches to item detection and identification and evaluates the relative efficacy of each approach. Artificial intelligence has fared better than image processing techniques when it comes to computer vision tasks. For daytime photos, the CNN model trained on a dataset of road vehicles for single object detection obtained a validation accuracy of 95.7% for cars, 96.5% for autos, and 96% for heavy vehicles. The large volume of data from each class that it is trained on is what accounts for the excellent validation accuracy. For day, night, and near-infrared photos, performance measures are tabulated. YOLOv3 is used to implement multiple object detection for the KITTI and COCO datasets. For YOLOv3, performance measurements are tabulated for the examined image classes. The mAP value will be higher the higher the class precession value. The image selected for computation determines the mAP value. For tracking and detection, an IoU of 0.5 is optimal. Increasing genuine positive values can improve mAP levels. The image dataset that is employed completely determines the performance metrics' results. With video, additional things can be found by using the region of interest. The performance measures were vehicle speed and color, vehicle type, vehicle movement direction, and the total number of vehicles in the ROI. YOLOv3 and OpenCV are used to perform multiple object tracking for traffic surveillance footage. Various items are identified and monitored across various video frames. Increasing the number of photos and training the models on more potent GPUs will allow you to assess the models on different datasets and, if necessary, adjust the design to make the model more reliable.

VII.REFERENCES:

[1] V. D. Nguyen et all., “Learning Framework for Robust Obstacle Detection, Recognition, and Tracking”, IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 6, pp. 1633-1646, June 2017.

[2] Zahraa Kain et all, “Detecting Abnormal Events in University Areas”, 2018 International conference on Computer and Applications(ICCA), pp. 260-264, 2018.

[3] P. Wang et all., “Detection of unwanted traffic congestion based on existing surveillance system using in freeway via a CNN-architecture trafficnet”, IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 1134-1139.

[4] Q. Mu, Y. Wei, Y. Liu and Z. Li, “The Research of Target Tracking Algorithm Based on an Improved PCANet”, 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, 2018, pp. 195-199.

[5] H. C. Baykara et all., “Real-Time Detection, Tracking and Classification of Multiple Moving Objects in UAV Videos”, 29th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Boston, MA, 2017, pp. 945-950.

[6] W. Wang, M. Shi and W. Li, “Object Tracking with Shallow Convolution Feature”, 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, 2017, pp. 97-100.

[7] K. Muhammad et all., “Convolutional Neural Networks Based Fire Detection in Surveillance Videos”, IEEE Access, vol. 6, pp. 1817418183, 2018.

[8] D. E. Hernandez et all., “Cell Tracking with Deep Learning and the Viterbi Algorithm”, International Conference on Manipulation, Automation and Robotics at Small Scales (MARSS), Nagoya, 2018, pp. 1-6.

[9] X. Qian et all., “An object tracking method using deep learning and adaptive particle filter for night fusion image”, 2017 International Conference on Progress in Informatics and Computing (PIC), Nanjing, 2017, pp. 138-142.

[10] Y. Yoon et all., “Online Multi-Object Tracking Using Selective Deep Appearance Matching”, IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), Jeju, 2018, pp. 206-212.