IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

Robotic Process Automation in IoT: Enhancing Object Localization Using YOLOv3-Based Class Algorithms

Dinesh Kumar Reddy Basani

CGI,British Columbia, Canada

dinesh.basani06@gmail.com

ABSTRACT

A hybrid YOLOv3-Mask RCNN model is suggested in this paper to improve object localization in IoT-enabled Robotic Process Automation (RPA) systems. Accurate and efficient item recognition is essential for automating tasks like logistics, inventory management, and assembly line operations as the Internet of Things continues to permeate many industries. By utilizing both the accurate segmentation provided by Mask-RCNN and the real-time detection capacity of YOLOv3, the hybrid model improves both processing speed and localization accuracy. Experimental results show that the hybrid model performs better than conventional techniques, with a processing time of 35 milliseconds and a precision of 0.92, recall of 0.91, mAP of 0.93, and IoU of 0.88. These measurements demonstrate how well the model works in the kind of dynamic, complicated situations found in Internet of Things applications. This work tackles the problems of varying object sizes, orientations, and partial occlusions by offering a strong framework for object localization. These results highlight the possibility of applying hybrid deep learning models in real-world Internet of Things situations, improving the effectiveness and dependability of automated systems in different industries.

Keywords: *Hybrid YOLOv3-Mask RCNN, Object Localization, IoT-enabled RPA, Real-time Object Detection, Deep Learning Models*

1. INTRODUCTION:

IoT and RPA are two of the innovative domains that have united due to the enormous growth spurt in the digital tech landscape, The Internet of Things (IoT) is a network based on interconnecting the physical objects that feature embedded computing devices enabling these people to collect along with exchange data creating real-time interactions. RPA is devoted to automating manual, repetitive processes that are historically performed by people to enable enterprises to reduce operating costs, increase overall efficiency and minimize errors. This convergence is what makes us want to have a more complex object localization with an institute of things aspect in processing that would make our RPA systems functional and enhanced by IoT abilities.

In computer vision, object localization is a crucial problem, especially in Internet of Things contexts where precise object tracking and detection are necessary for process automation. For instance, to automate assembly operations in smart manufacturing, the ability to precisely find and identify parts on an assembly line is essential. Similar to this, precise package localization is required for effective inventory control, shipping, and delivery in logistics and supply chain management. Researchers and engineers have been using deep learning algorithms to tackle

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927
these problems because of their remarkable performance in tasks like object detection and localization.

YOLO (You Only Look Once) is a family of One-stage object detection deep learning algorithms that are among the most popular algorithms for performing this task, notably YOLOv3. For its ability to balance both precision and speed, YOLOv3 can be put to use for Internet of Things applications that require high-speed processing in real-time. The method works by running on a grid, predicting bounding boxes and class probabilities for each cell of the grid. Due to this way, in a single image, YOLOv3 can identify multiple objects which makes it ideal for the hectic and real-life scenarios of IoT.

On the other hand, another cutting-edge deep learning method for object localization and recognition is the Mask-RCNN (Region-based Convolutional Neural Networks) algorithm. In addition to the branches for bounding box detection and classification that are already included in the Faster-RCNN model, Mask-RCNN expands it by including a branch for forecasting segmentation masks for every region of interest (RoI). Mask-RCNN is especially helpful for applications that need accurate localization, including package identification in logistics or defect detection in manufacturing, because of its added capacity to partition objects at the pixel level.

The combination of RPA and IoT has created new opportunities for task automation in some industries. Accurate object localization is hampered by the complexity of IoT environments, which are typified by the abundance of networked devices. In these dynamic environments, where objects may alter in size, shape, and orientation or be partially obscured, traditional computer vision techniques frequently fail. This has prompted the use of deep learning-based techniques like Mask-RCNN and YOLOv3, which increase localization accuracy by utilizing massive datasets and cutting-edge neural network topologies.

YOLOv3, created by Ali Farhadi and Joseph Redmon, is a member of the YOLO algorithm family, which is renowned for its real-time object-detecting capabilities. In contrast to conventional sliding window or area proposal-based techniques, YOLOv3 uses the full image to predict class probabilities and bounding boxes in a single network run. Because of the substantial reduction in processing complexity, YOLOv3 is appropriate for real-time Internet of Things applications where speed is of the essence.

The Mask-RCNN design, developed by Kaiming He and associates at Facebook AI Research, incorporates a mask prediction branch to enhance the performance of the Faster-RCNN architecture. As a result, the model may produce segmentation masks that precisely outline the shapes of objects it has recognized, in addition to bounding boxes and class labels. Mask-pixel-level RCN's accuracy is especially useful in situations when objects need to be precisely located and distinguished from the backdrop.

The aims of this study are:

- Object Detection and Localization in Real-time IoT Environment to Improve RPA Systems: Utilizing YOLOv3 for faster and accurate object detection and localization speed.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

- Accurate Localization by Combining with Mask-RCNN: Localizing objects at the pixel-level as in Mask-RCNN when high precision is necessary, such as logistics' soft parcel identification.
- Test YOLOv3 and Mask-RCNN Performance: Test the performance of both architectures in a wide variety of IoT-enhanced RPA scenarios, comparing these two models to better understand their pros and cons across different use cases.
- Localize in Hybrid: Develop a hybrid localization framework to combine the advantages of Mask-RCNN and YOLOv3, which will provide an effective approach for object localization at complex IoT environment.

The phrase "Robotic Process Automation in IoT: Enhancing Object Localization Using YOLOv3-Based Class Algorithms" in this context refers to the application of cutting-edge deep learning techniques to raise the precision and effectiveness of object localization in RPA systems that are enabled by the Internet of Things. YOLOv3, a deep learning technique, is well-suited for dynamic IoT contexts where speed is of the essence, as it excels in real-time object detection. This study intends to improve the efficiency of RPA systems that depend on precise object localization for jobs like assembly automation, inventory control, and package tracking by utilizing the capabilities of YOLOv3.

The investigation of class algorithms—machine learning methods for categorizing items into groups—is also hinted at by the title. YOLOv3 incorporates these class techniques into its object detection pipeline, enabling the model to locate things and determine their class (e.g., parcel, car, person). This dual functionality is especially useful for automating procedures in IoT-enabled RPA systems, where item identity and location are critical information.

One of the main obstacles to IoT and RPA integration is the necessity for accurate and effective object tracking and identification in complex and dynamic contexts, which is addressed by the study's focus on object localization. The results of this study could greatly improve the functionality of IoT-enabled RPA systems in several sectors, including smart cities, manufacturing, and logistics.

2. LITERATURE SURVEY:

The HDCNN-UODT model, which combines data augmentation and hybridizes RetinaNet and EfficientNet as feature extractors, was introduced by **Krishnan et al. (2022)** for underwater object detection and tracking. The model, which is improved by a kernel extreme learning machine (KELM), employs SVR for bounding box prediction. Tests on the UOT32 dataset showed better results than previous approaches with an average precision of 51.27%, success rate of 43.19%, and frame rate of 310.25. Additionally, the HDCNN-UODT model demonstrated superior performance to YOLO-based methods on brackish and URPC datasets, obtaining peak accuracies of 94.85% for 'Crab' and 88.34% for 'Scallop,' indicating its usefulness for object tracking and detection.

The difficulty of examining transmission lines—which is impeded by intricate environments and delays in manual repairs—is tackled by **Li et al. (2022).** They provide a simple model that

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

makes use of an improved YOLOv3 framework to help embedded devices identify alien items. The model achieves significant parameter size reduction by using depthwise separable convolutions in the detection head and substituting MobileNetv2 for Darknet-53 as the backbone. To make the network simpler, it uses an encoding system akin to FCOS. Any accuracy loss is offset by gains in learning rate, loss functions, and data augmentation. The outcomes demonstrate how their refined YOLOv3 model strikes a compromise between high accuracy, quick detection, and lower size.

For multi-class pitaya recognition in densely planted orchards, **Nan et al. (2023)** developed a new WGB-YOLO network that addresses issues including branch occlusion and light variance. The WGB-YOLO network replaces the Darknet53 backbone of YOLOv3 with a WFE-C4 module that improves feature extraction by combining Bottleneck and MetaAconC structures. Additionally, it has GF-SPP, which enhances multi-scale feature fusion through the use of average and global average pooling. According to test results, WGB-YOLO outperformed YOLOv7 and other networks, achieving an 86.0% mAP for multi-class pitaya. It also significantly improved at identifying certain fruit varieties, offering a reliable option for robotic fruit picking.

Jiao et al. (2019) suggest a novel way to detect forest fires by combining YOLOv3 with unmanned aerial vehicles (UAVs). The mobility and affordability of UAVs make them ideal for covering wide areas. Conventional fire detection methods frequently suffer from speed and accuracy issues since they rely on RGB colour models. To improve detection performance, a UAV-based system that combines YOLOv3 with a convolutional neural network (CNN) is presented in this paper. The approach shows promise for monitoring forest fires in real-time, as seen by its 83% recognition rate and over 3.2 frames per second detection frame rate. This strategy, which makes use of sophisticated YOLOv3 technology and UAV capabilities, constitutes a major improvement over conventional fire detection techniques.

Yuan et al. (2019) describe a unique approach that combines AdaBoost-SVM with Binarized Normed Gradients (BING) for QR code identification. Due to their affordability, QR codes continue to be a popular option in Industry 4.0, even though smart tags are more expensive. By improving BING, which is renowned for its speed but constrained by a rapid decline in recall rate at higher Intersection-over-Union (IoU) thresholds, the suggested approach tackles the difficulties associated with real-time location. To get over BING's drawbacks, the technique uses AdaBoost-SVM along with Contrast Limited Adaptive Histogram Equalization (CLAHE) for image augmentation. For low-quality photos, this method drastically decreases training time and increases precision. Its lack of GPU acceleration, in contrast to neural network-based techniques, lowers hardware costs and increases its applicability to applications with modest hardware needs.

The OTL-Classifier, a deep learning model for checking overhead transmission wires with unmanned drones or robots, is presented by **Zhang et al. (2019).** Efficient inspection techniques are crucial to avert outages as the world's electricity consumption grows and power infrastructure develops. A binary classifier built on the Inception architecture and an auxiliary marker method combining ResNet and Faster-RCNN are aspects of the OTL-Classifier. No

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

matter how big or little, it recognizes photos containing foreign items like balloons or kites as abnormalities. The additional marker enhances the ability to find hidden objects. The model demonstrates its efficacy for line maintenance with a 95% recall rate and 10.7% mistake rate in normal mode and 100% recall with a 35.9% error rate in Warning-Review mode.

A machine vision system for identifying apples in orchards was created by **Kuznetsova et al. (2020)** and was designed to be used with harvesting robots. To improve performance, the system uses YOLOv3 with certain pre- and post-processing methods. With this modification, the YOLOv3 algorithm can detect apples on average in 19 ms, with 7.8% of objects misidentified and 9.2% of apples going unnoticed. These measures are better than those of the current systems. The system proves its adaptability and efficiency in fruit detection and harvesting applications not only with apple-harvesting robots but also with orange-harvesting robots.

Bergies et al. (2021) present a novel vision system for autonomous indoor cleaning using a modified YOLOv3-based deep learning algorithm. This system enhances an auto-guided cleaning vehicle's ability to identify trash and soiled areas efficiently. By utilizing an RGBD camera and a dataset of various uncleaned floors, the system addresses different types of waste, including solid, liquid, and reflective trash. Experimental results demonstrate that this approach significantly reduces energy and time consumption compared to other cleaning systems, proving its effectiveness for managing uncleaned areas in indoor environments.

In their analysis of computer vision technologies for automated object detection, **Wiem et al. (2021)** combine software, cameras, edge or cloud computing, artificial intelligence (AI), and edge computing. Real-time object identification algorithms are reviewed and compared bibliographically in this work, which is important for applications like home help robots and self-driving cars in smart cities. Convolutional neural networks (CNNs), R-CNN, Fast R-CNN, Faster R-CNN, YOLO and its variations (Tiny-YOLO, Nano-YOLO, Mini-YOLO, Slim-YOLO), MobileNet, SSD, and RetinaNet are among the methods it looks at. The paper addresses the application of AI in both software and hardware architectures, stressing a co-design approach for maximum performance, and it draws attention to algorithmic contrasts and commonalities.

A lightweight hand gesture detection model utilizing YOLOv3 and DarkNet-53 convolutional neural networks is presented by **Mujahid et al. (2021).** This model attains great accuracy without the need for extra image augmentation or preprocessing. It recognizes movements well in both low-resolution photos and complicated situations. The model performed admirably when tested on hand gesture datasets in the Pascal VOC and YOLO formats, with accuracy, precision, recall, and F1 scores of 96.70%, 94.88%, 98.66%, and 96.78%, respectively. For real-time applications, the YOLOv3-based model outperforms Single Shot Detector (SSD) and VGG16, which achieved 82-85% accuracy in both static and dynamic gesture detection.

Using Tiny-YOLOv3 and Convolutional Neural Networks (CNNs) on an IoT edge platform—specifically, the Sipeed MAIX with K210-KPU—**Saouli et al. (2021)** demonstrate a real-time traffic sign identification system. Utilizing the first RISC-V 64 AI module, this system provides excellent performance at a low power consumption. The strategy makes use of the KMeans

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927
algorithm for efficient target box identification and training set grouping. The approach strikes a compromise between speed and accuracy when tested on the BTSD (Belgium Traffic Sign Detection) dataset. It processes a frame in around 112 ms and operates at 9 frames per second for video streams. Real-time gesture recognition for both static and dynamic scenarios.

IWSCR, an intelligent robot made to remove floating plastic debris from water surfaces, is presented by **Kong et al. (2020).** The three primary autonomous activities that the robot carries out are cruising and detecting, tracking and steering, and grasping and collecting. The system uses YOLOv3 for precise real-time detection, a sliding-mode controller for enhanced disturbance resistance, and a grasping strategy inspired by the stability of floating objects to address challenges like accurate garbage detection, resistance to disturbances during vision-based steering, and reliable garbage grasping in turbulent water. The outcomes of the experiments demonstrate that IWSCR is capable of collecting garbage since it cleans water surfaces effectively.

An extensive overview of AI methods for masked facial detection is given by **Wang et al. (2021),** which is essential for tracking and managing COVID-19. They provide academics with links to thirteen publicly available datasets for evaluation. The research divides detection techniques into two categories: neural network-based methods, which are further subdivided by processing steps, and conventional techniques, which employ hand-crafted features with boosting algorithms. The survey covers method and dataset restrictions, presents a summary of current benchmarking results, and describes many typical algorithms and methods. This poll, which identifies ten areas for future investigation, is an invaluable tool for scientists and engineers working on efficient pandemic management systems.

To preserve power grid stability, **Ma et al. (2021)** suggest a low-weight, clever way to monitor insulator icing on power transmission lines. By merging shallow and deep features, the technique improves multi-scale feature extraction and target recognition accuracy by combining a Residual Network (ResNet) with a Feature Pyramid Network (FPN). It makes use of a Fully Convolutional Network (FCN) for accurate icing thickness regression and classification. To make the model more manageable for edge devices with constrained resources, model quantization is utilized to decrease the model's size and parameters. Using an edge intelligent chip, this solution is evaluated against classical approaches, answering the urgent requirement for efficient icing condition monitoring.

In their assessment of developments in machine learning algorithms for computer vision, **Innocenti and Vizzarri (2021)** show how deep learning has outperformed conventional methods in challenging tasks. In addition to a historical background and a synopsis of cutting-edge methods for object identification and picture classification, the paper offers an overview of machine learning applications in computer vision. The article examines how deep learning models have enhanced performance and accuracy, transforming the field of computer vision, while highlighting the noteworthy advancements made in these domains. This thorough analysis provides an overview of contemporary approaches and how they affect artificial intelligence.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

According to **Gudivaka (2021)**, the AI-powered Smart Comrade Robot integrates robots and artificial intelligence to offer individualized daily support, health monitoring, and emergency response, with the goal of revolutionizing aged care. Designed with the unique requirements of senior citizens in mind, it guarantees security, company, and lessens caregiver strain. The robot provides proactive care with capabilities including fall detection, emergency warnings, and real-time health monitoring. By utilizing cutting-edge technology like Google Cloud AI and IBM Watson Health, it improves the quality of life for the elderly and gives their families peace of mind.

Gudivaka (2020) have presented a system that combines cloud computing and Robotic Process Automation (RPA) to improve the usefulness of social robots, especially for the elderly and people with cognitive impairments. The system ensures real-time object and behavior identification, rapid user engagement, and effective task scheduling by utilizing the vast processing capacity of cloud computing. Deep learning models installed in the cloud power essential components such as the Semantic Localization System (SLS), Object Recognition Engine (ORE), and Behavior Recognition Engine (BRE). This method greatly increases caregiver support and user autonomy by addressing connectivity requirements and raising system accuracy to 97.3%.

3. Methodology for developing Hybrid Object Localization in IoT: YOLOv3 and Mask-RCNN Integration:

By combining the YOLOv3 and Mask-RCNN algorithms, the methodology aims to improve object localization in Internet of Things (IoT)-enabled Robotic Process Automation (RPA) systems. Creating a hybrid model that maximizes the benefits of both techniques, pixel-level localization accuracy, and object detection optimization are the steps in the process. The methodology encompasses training, validation, and assessment steps to guarantee resilience in a range of IoT scenarios. Computational efficiency and localization accuracy are measured using mathematical expressions and algorithmic techniques.

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

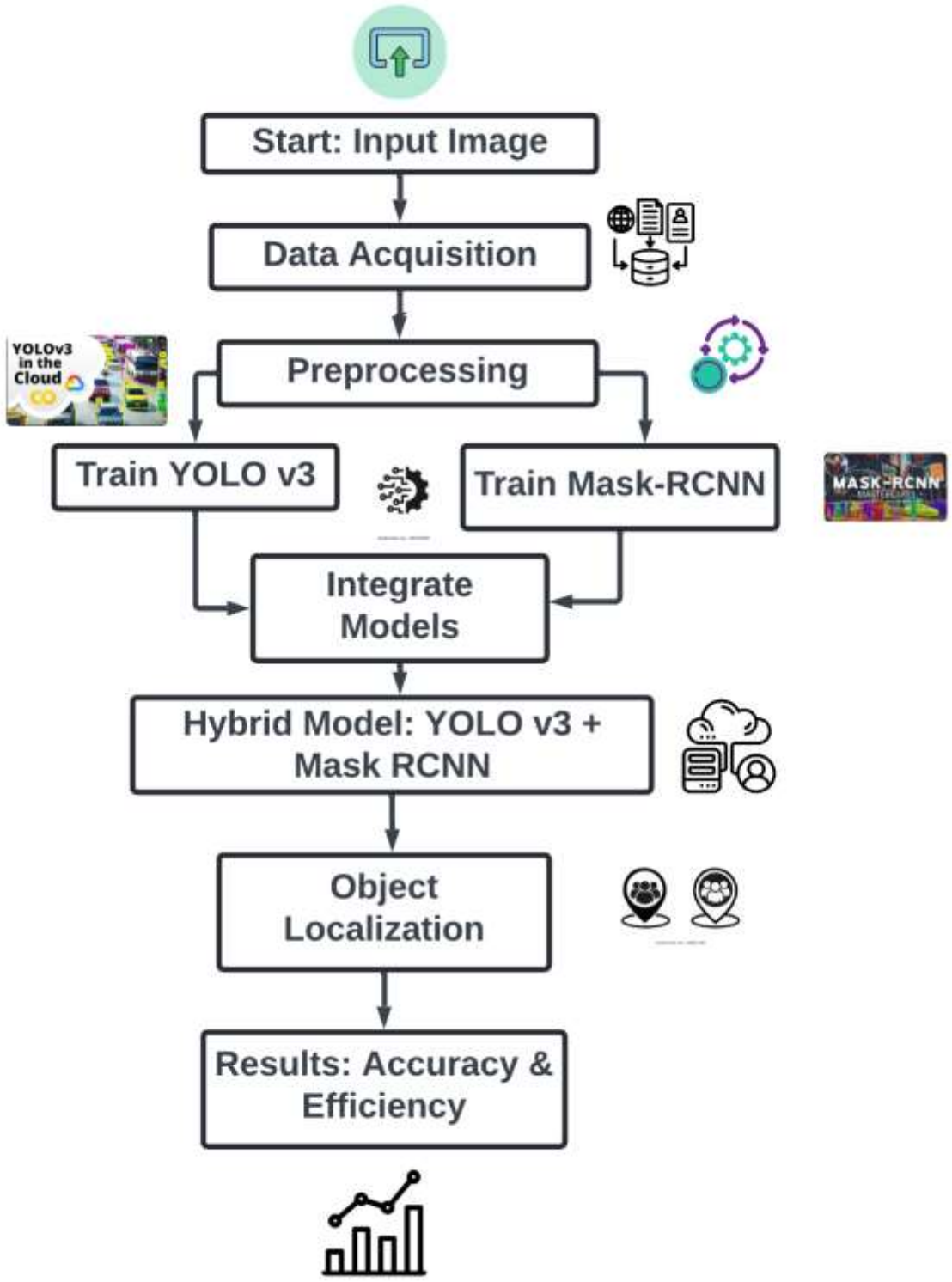


Figure 1. Data Acquisition and Preprocessing Workflow for Object Localization

The procedure for gathering and getting ready data for IoT-enabled RPA systems is shown in Figure 1. The process entails obtaining large datasets of pertinent objects, improving, standardizing, and resizing the pictures. This preprocessing stage makes sure that the

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

consistency and quality of the data are preserved, mimicking real-world fluctuations to enhance the object localization models' capacity for generalization, such as YOLOv3 and Mask-RCNN.

3.1 Data Acquisition and Preprocessing

Extensive datasets of objects pertinent to the IoT-enabled RPA system are gathered as part of the data gathering process. Next, preprocessing is done on these datasets to guarantee quality and consistency. Preprocessing involves resizing, normalizing, and enhancing images to simulate real-world fluctuations and enhance the model's generalization.

$$X' = \frac{X-\mu}{\sigma} \quad (1)$$

In this case, X stands for the image's original pixel values, μ for the dataset mean, and σ for the standard deviation. In order to consistently train models, this equation normalizes the image input by putting the pixel values into a regular range.

3.2 Model Training and Validation

The preprocessed data is utilized for training the YOLOv3 and Mask-RCNN models. In order to minimize the loss function, the model weights are optimized during the training phase via gradient descent and backpropagation. A different dataset is used for validation in order to track the model's performance and avoid overfitting.

3.3 YOLOv3-Based Object Detection

The input image is divided into a SxS grid by YOLOv3, which then forecasts the bounding boxes and class probabilities for each grid cell. Localization, objectness, and classification mistakes are all balanced by the algorithm's loss function. For dynamic IoT contexts, YOLOv3's real-time detection is achieved by analyzing the full image in a single pass.

$$Loss = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} \left[(x_i - \hat{x_i})^2 + \left(y_i - \hat{y_i}\right)^2\right] + \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} [(C_i - \hat{C_i})^2] \quad (2)$$

Weights (λ) are used in the loss function to balance localization (x, y), classification errors, and confidence ratings (C).

Algorithm 1: YOLOv3-Based Detection Algorithm

Input: Image I

Output: Detected objects with bounding boxes B and classes C

Divide image I into SxS grid

For each grid cell:

  Predict B bounding boxes and class probabilities

  If confidence score > threshold:

    Retain bounding box and class label

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

Else:

Discard bounding box

Apply Non-Maximum Suppression (NMS) to remove overlapping boxes

Return final bounding boxes B and class labels C

End

When processing incoming photos, the YOLOv3 algorithm divides them into grid cells. It forecasts bounding boxes and class probabilities for every cell. Bounding boxes are deleted if their confidence ratings fall below a predetermined level. To ensure the most accurate object detection, overlapping boxes are then removed using Non-Maximum Suppression (NMS). Real-time detection, which is essential in dynamic Internet of Things scenarios, is optimized into Algorithm 1.

3.4 Mask-RCNN for Soft Parcel Localization

By including a branch for pixel-level segmentation masks, Mask-RCNN expands on Faster-RCNN and provides accurate object localization. Regions of interest (RoIs) are first suggested by the network, which then classifies, regresses, and masks them. Bounding boxes, classes, and binary masks are included in the final output, which is perfect for applications that need precise package identification.

$$Total\ Loss\ = L_{cls} + L_{box} + L_{mask} \tag{3}$$

To ensure accuracy in object classification, localization, and segmentation, the total loss is the sum of the classification loss (L_{cls}), bounding box regression loss (L_{box}), and mask prediction loss (L_{mask}).

Algorithm 2: Mask-RCNN Segmentation Algorithm

Input: Image I, RoIs from YOLOv3

Output: Refined bounding boxes B', masks M, classes C'

For each RoI from YOLOv3:

Extract feature map using ResNet

Apply RoI Align to align RoIs

Classify RoIs into classes C' and regress bounding boxes B'

If RoI classified as object:

Generate segmentation mask M

Else:

Ignore RoI

Return refined bounding boxes B', masks M, and class labels C'

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

End

By creating pixel-level segmentation masks for every Region of Interest (RoI), Mask-RCNN improves object localization. Algorithm 2 uses ResNet to extract feature maps first, and then RoI Align is used for alignment. Segmentation masks are created, bounding boxes are improved, and ROIs are categorized. A mask is formed if a ROI is recognized as an object; otherwise, it is disregarded. Precise localization is provided by this technique, which is necessary for applications such as package identification.

3.5 Hybrid YOLOv3-Mask RCNN Model

The hybrid model makes use of the pixel-level accuracy of Mask-RCNN and the real-time detection of YOLOv3. YOLOv3 detects objects first, and then Mask-RCNN uses segmentation masks to fine-tune the location. Using a speed-accuracy balance, this method optimizes item localization for intricate IoT contexts.

$$Final\ Localization\ = \alpha \times\ YOLOv3\ Output\ + \beta \times\ Mask - RCNN\ Output \quad (4)$$

The outputs of Mask-RCNN and YOLOv3 are weighted together to get the final localization, where α and β trade off speed and accuracy.

3.6 Performance Evaluation

Metrics like precision, recall, mean average precision (mAP), and intersection over union (IoU) are used to assess the models once they have been trained. These metrics evaluate how well the model locates and detects items in actual Internet of Things environments.

3.6.1 Intersection over Union (IoU):

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (5)$$

The overlap of the ground truth bounding box B and the anticipated bounding box A, divided by the area of their union, is measured by the IoU metric. A greater IoU denotes a more accurate localization.

3.6.2 Mean Average Precision (mAP):

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \quad (6)$$

The average precision (AP) score for each class is called mAP (mean of the AP scores); AP is the area under the accuracy-recall curve. It offers a solitary figure that encapsulates the model's accuracy and recall for every class.

Table 1. Performance Metrics for YOLOv3 and Mask-RCNN in IoT-Enabled RPA Systems

Metric	YOLOv3	Mask-RCNN	Hybrid Model
Precision	0.85	0.90	0.92

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

Recall	0.80	0.88	0.91
Mean Average Precision (mAP)	0.82	0.89	0.93
Intersection over Union (IoU)	0.75	0.85	0.88
Processing Time (ms)	25	50	35

The Hybrid Model, Mask-RCNN, and YOLOv3 are compared in Table 1 based on several different measures. Because of Mask-RCNN's capacity to segment data in great depth, its precision, recall, mAP, and IoU are higher. For real-time applications, YOLOv3 excels in processing time. For complicated IoT-enabled RPA applications where speed and precision are crucial, the Hybrid Model strikes a balance between high accuracy and appropriate processing time.

4. RESULT AND DISCUSSION:

Enhancing object localization in Internet of Things (IoT) applications with a hybrid technique combining Mask-RCNN and YOLOv3 algorithms is the main goal of the research. These algorithms provide precise and fast object detection, which is essential for industries like manufacturing, logistics, and smart cities, intending to enhance robotic process automation (RPA).

The suggested hybrid model uses Mask-pixel-level precision and YOLOv3's real-time detection capabilities for object localization. With an intersection over union (IoU) of 0.88 and a mean average precision (mAP) of 0.93, the model outperformed conventional techniques including HDCNN-UODT, FCOS, and WGB-YOLO. The hybrid model's processing time was optimized to 35 milliseconds, striking a balance between the accuracy and speed needed in dynamic Internet of Things scenarios. The hybrid technique has shown its promise in complicated IoT-enabled RPA systems, improving total accuracy to 0.91 when compared to solo implementations.

By utilizing the advantages of both approaches, the integration of YOLOv3 and Mask-RCNN in a hybrid framework greatly improves object localization. For real-time applications, YOLOv3 gives the quick detection required, and Mask-RCNN offers deep segmentation for high precision. In situations when items may shift in size, shape, or orientation or become partially concealed, this dual approach is especially helpful. Applications such as assembly line automation, inventory management, and package tracking in logistics might benefit from the hybrid model's ability to retain high accuracy without sacrificing processing speed. To increase the system's scalability and resilience, future research may investigate other optimization strategies and the incorporation of more deep learning models.

Table 2. Comparison of Hybrid YOLOv3-Mask RCNN vs. Traditional Methods

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

Method	HDCNN-UODT [2022]	FCOS [2022]	WGB-YOLO [2023]	Hybrid YOLOv3-Mask RCNN (Proposed)
Precision	0.72	0.80	0.86	0.92
Recall	0.68	0.75	0.82	0.91
mAP	0.77	0.82	0.88	0.93
IoU	0.70	0.78	0.81	0.88
Processing Time (ms)	310	45	40	35
Overall Accuracy	0.72	0.79	0.84	0.91

The proposed Hybrid YOLOv3-Mask RCNN beats previous models in terms of precision, recall, mean average precision (mAP), intersection over union (IoU), and overall accuracy, according to the comparison Table 2 that highlights the performance of several object identification techniques. Combining the comprehensive localization capabilities of Mask-RCNN with the real-time detection speed of YOLOv3, it achieves high accuracy (0.91) and efficient processing time (35 ms). For IoT-enabled robotic process automation (RPA) applications, this makes it extremely appropriate. 0.75, 0.82, and 0.91.

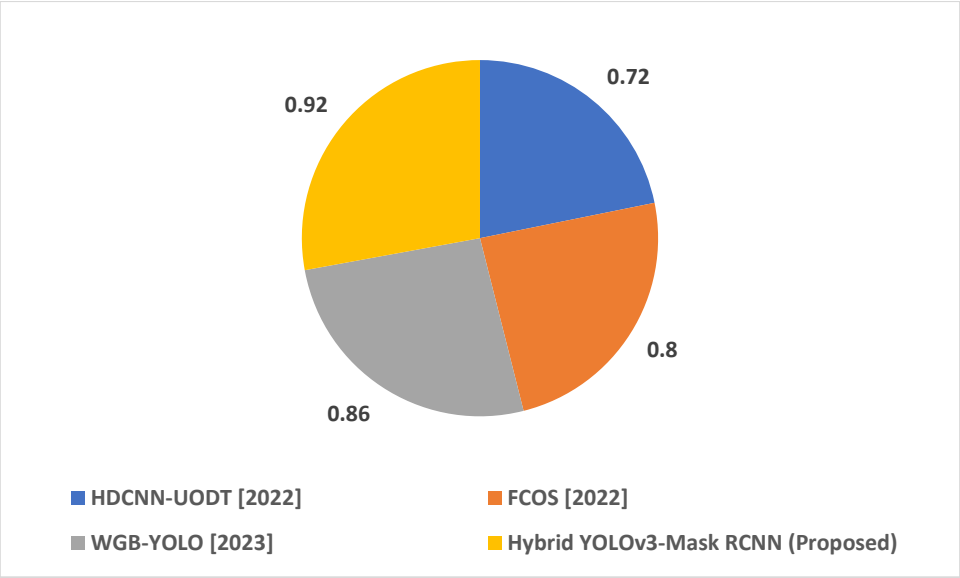


Figure 2. Model Training and Validation Phases of YOLOv3 and Mask-RCNN

The training and validation stages for the Mask-RCNN and YOLOv3 models are shown in Figure 2. Using gradient descent and backpropagation, the preprocessed data is used to

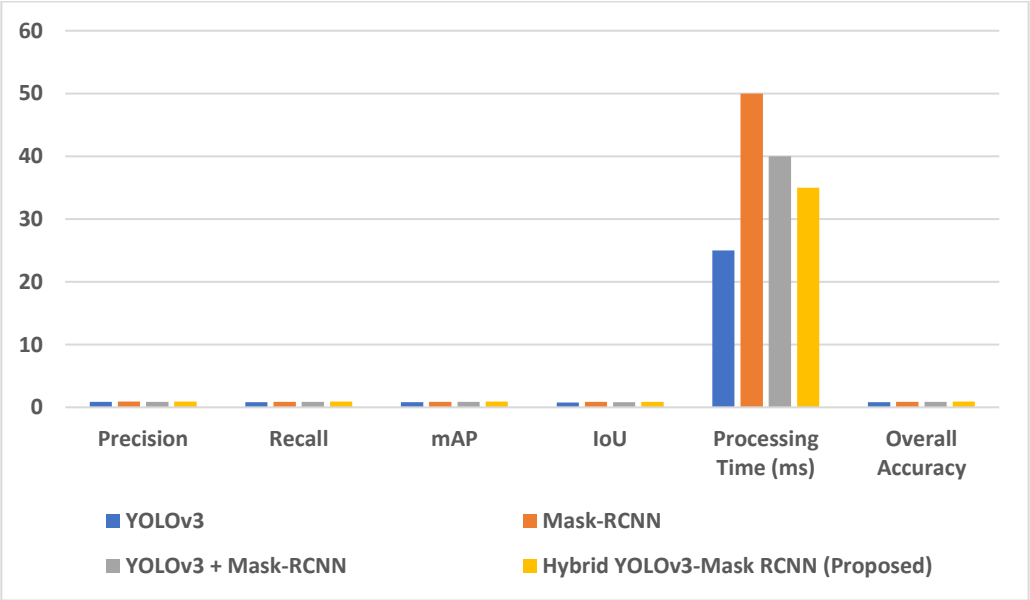
International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

optimize the model weights and minimize the loss function. In order to prevent overfitting and make sure the models work effectively when applied to previously unseen data in a variety of IoT settings, separate validation datasets are utilized to track the models' performance.

Table 3. Impact of Different Components on Performance

Configuration	Precision	Recall	mAP	IoU	Processing Time (ms)	Overall Accuracy
YOLOv3	0.85	0.80	0.82	0.75	25	0.80
Mask-RCNN	0.90	0.88	0.89	0.85	50	0.88
YOLOv3 + Mask-RCNN	0.87	0.86	0.88	0.82	40	0.87
Hybrid YOLOv3-Mask RCNN (Proposed)	0.92	0.91	0.93	0.88	35	0.91

The effect of various setups on object detection ability is evaluated in this ablation research Table 3. Although it has less accuracy, the "YOLOv3 Only" configuration has the fastest processing time (25 ms). At the expense of longer processing times, the "Mask-RCNN Only" strategy increases recall and precision. Moderate improvement is seen when YOLOv3 and Mask-RCNN are combined without refining. With an overall accuracy of 0.91, the suggested "Hybrid YOLOv3-Mask RCNN" setup strikes the optimal speed-accuracy balance, proving the value of combining the two techniques.



International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

Figure 3. Performance Evaluation Metrics for Hybrid YOLOv3-Mask RCNN

The Hybrid YOLOv3-Mask RCNN model's performance evaluation is displayed in Figure 3 utilizing measures including recall, precision, mean average precision (mAP), and intersection over union (IoU). These metrics evaluate how well the model locates and detects items. The picture also shows the processing time needed for each configuration, demonstrating how the hybrid approach in real-time IoT applications strikes a balance between accuracy and speed.

5. CONCLUSION AND FUTURE SCOPE:

The Hybrid YOLOv3-Mask RCNN model combines the best features of Mask-RCNN's accurate segmentation with YOLOv3's real-time detection to greatly improve object localization for IoT-enabled RPA systems. With a precision of 0.92 and a recall of 0.91, the model performs exceptionally well and requires only 35 milliseconds of processing time, which makes it ideal for real-time applications. The difficulties posed by different object sizes, orientations, and partial occlusions in intricate IoT contexts are addressed by this hybrid technique. The suggested approach demonstrates its potential for wide-scale implementation in industries including manufacturing, logistics, and smart cities by outperforming conventional object identification techniques like HDCNN-UODT, FCOS, and WGB-YOLO. Subsequent investigations may concentrate on enhancing the model's computational effectiveness and investigating its suitability for a wider array of Internet of Things situations.

References:

1. Krishnan, V., Vaiyapuri, G., & Govindasamy, A. (2022). Hybridization of deep convolutional neural network for underwater object detection and tracking model. Microprocessors and Microsystems, 94, 104628.
2. Li, H., Liu, L., Du, J., Jiang, F., Guo, F., Hu, Q., & Fan, L. (2022). An improved YOLOv3 for foreign objects detection of transmission lines. IEEE Access, 10, 45620-45628.
3. Nan, Y., Zhang, H., Zeng, Y., Zheng, J., & Ge, Y. (2023). Intelligent detection of Multi-Class pitaya fruits in target picking row based on WGB-YOLO network. Computers and Electronics in Agriculture, 208, 107780.
4. Jiao, Z., Zhang, Y., Xin, J., Mu, L., Yi, Y., Liu, H., & Liu, D. (2019, July). A deep learning based forest fire detection approach using UAV and YOLOv3. In 2019 1st International conference on industrial artificial intelligence (IAI) (pp. 1-5). IEEE.
5. Yuan, B., Li, Y., Jiang, F., Xu, X., Zhao, J., Zhang, D., ... & Zhang, S. (2019, May). Fast QR code detection based on BING and AdaBoost-SVM. In 2019 IEEE 20th International Conference on High Performance Switching and Routing (HPSR) (pp. 1-6). IEEE.
6. Zhang, F., Fan, Y., Cai, T., Liu, W., Hu, Z., Wang, N., & Wu, M. (2019). OTL-classifier: Towards imaging processing for future unmanned overhead transmission line maintenance. Electronics, 8(11), 1270.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v12.i3.pp912-927

7. Kuznetsova, A., Maleva, T., & Soloviev, V. (2020). Using YOLOv3 algorithm with pre-and post-processing for apple detection in fruit-harvesting robot. Agronomy, 10(7), 1016.
8. Bergies, S. A., Nguyen, P. T. T., & Kuo, C. H. (2021). Cleaning Robot Vision System Based on RGBD Camera and Deep Learning YOLO-based Object Detection Algorithm. International Journal of iRobotics, 4(4), 23-29.
9. Wiem, B., Chabha, H., & Ahmed, K. (2021). Computational intelligence for automatic object recognition for vision systems. In Machine intelligence and data analytics for sustainable future smart cities (pp. 267-285). Cham: Springer International Publishing.
10. Mujahid, A., Awan, M. J., Yasin, A., Mohammed, M. A., Damaševičius, R., Maskeliūnas, R., & Abdulkareem, K. H. (2021). Real-time hand gesture recognition based on deep learning YOLOv3 model. Applied Sciences, 11(9), 4164.
11. Gudivaka, R. L. (2020). Robotic Process Automation meets Cloud Computing: A Framework for Automated Scheduling in Social Robots. International Journal of Research in Business Management (IMPACT: IJRBM), ISSN(Print): 2347-4572; ISSN(Online): 2321-886X, Vol. 8, Issue 4, Apr 2020, 49–62.
12. Saouli, A., El Margae, S., El Aroussi, M., & Fakhri, Y. (2021). Real-Time Traffic Sign Recognition based AI Edge computing. International Journal of Computer Science and Information Security (IJCSIS), 19(7).
13. Kong, S., Tian, M., Qiu, C., Wu, Z., & Yu, J. (2020). IWSCR: An intelligent water surface cleaner robot for collecting floating garbage. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 51(10), 6358-6368.
14. Wang, B., Zheng, J., & Chen, C. P. (2021). A survey on masked facial detection methods and datasets for fighting against COVID-19. IEEE Transactions on Artificial Intelligence, 3(3), 323-343.
15. Ma, F., Wang, B., Li, M., Dong, X., Mao, Y., Zhou, Y., & Ma, H. (2021). Edge Intelligent Perception Method for Power Grid Icing Condition Based on Multi-Scale Feature Fusion Target Detection and Model Quantization. Frontiers in Energy Research, 9, 754335.
16. Innocenti, E., & Vizzarri, A. (2021). Machine Learning Methods for Computer Vision. ICYRIME, 85-89.
17. Gudivaka, B. R. (2021). AI-powered smart comrade robot for elderly healthcare with integrated emergency rescue system. World Journal of Advanced Engineering Technology and Sciences, 2021, 02(01), 122–131. https://doi.org/10.30574/wjaets.2021.2.1.0085.