IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Neural-Network-Oriented Approach for Industrial Item Identification and Manipulation Point Localization

[1]Mr.Salim Amirali Jiwani, [2]BOGOJU AKSHAYA, [3]BORALA RAJU, [4]JANAGAM ROHAN

[1]Assistant Professor, [2,3,4] UG STUDENT

[1,2,3,4,]DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING(AI & ML)

[1,2,3,4] VAAGDEVI COLLEGE OF ENGINEERING Autonomous

Bollikunta, Khila Warangal (Mandal), Warangal Urban-506 005 (T.S),

Abstract:

For intelligent robotic manipulation systems to work, they need to be able to recognise objects and detect strong grips. They are especially important in factories and other places where the lighting is often complicated and changes all the time. Shadows, reflections, glare, and low light levels can all cause traditional computer vision methods and simple deep learning-based methods to fail, which can lead to false object detection, false pose estimation, and false robot grasp. So, the goal of this project is to make an object work piece recognition and grip detection system that uses deep learning. This will be done by combining an improved YOLO Net with a feature fusion and attention module, and then using ROS-based robotic control to get the pose right.

The first part of the proposed system is a YOLO-based object detection network that uses improved feature fusion to combine features of different sizes so that it can find both small and complex workpieces. The network has an attention mechanism built in that brings out the features of the object and reduces the noise from an unevenly lit environment. This makes the detection more accurate and reliable.

Once the target object has been found in the picture, hand and eye calibration can be done to find the exact change from the camera frame to the robot hand frame. This lets us figure out the object's exact position in the world coordinate frame. Accurate pose information about the object may be very important to avoid mistakes in positioning while the robotic arm moves toward, grasps, and drops off the target object, which can lead to failures in grasping.

The robot modelling module sets the manipulator's poses, like its kinematic information, dynamic parameters, joint limits, degree of freedom, and so on. This makes sure that the trajectory planning is possible. The ROS environment brings all the parts together. The control module gets the detection results and pose information, and the robot can then smoothly and without colliding grasp and place things in real time.

3.4. Results of the experiment Experimental tests demonstrate that the proposed system effectively mitigates the issues arising from unbalanced illumination and environmental interference. The model's recognition accuracy is 92.2%, and the overall average success rate for grasping is 93.75%. The experimental results confirm the feasibility, robustness, and operational efficiency of the proposed system

in actual industrial automation applications. The system can be easily adapted to different objects, lighting conditions, robotic arm working levels, and other robots that work together.

Keywords— Deep Learning–Based Object Detection, YOLO-Net Feature Fusion Network, Robotic Workpiece Grasp Detection, Hand–Eye Calibration Pose Estimation, ROS-Based Robotic Control System

I. INTRODUCTION

In recent years, intelligent robotics has become an important part of the smart manufacturing, industrial automation, and flexible production environments that are now common. Robotic manipulators can be used in a lot of different ways, like sorting, putting together, packaging, checking, and moving objects. The most important and basic things that a robot needs to be able to do are to accurately identify work pieces and hold them securely in real-world situations.

The traditional method of computer vision-based object detection relies on identifying features like edges, contours, and colour segmentation. These methods can work well in the lab, but in the real world, they might not work as well because of changing lighting conditions. For example, glare or occlusion can make it hard to accurately locate an object, which can lead to unstable grasping. Previous deep learning methods could make detection more accurate, but they couldn't handle multi-scale detection or changes in the lighting conditions in the environment.

Furthermore, to address these deficiencies, a deep learning-based object detection framework such as YOLO (You Only Look Once) has emerged as a promising method for achieving real-time recognition. YOLO treats object detection as a single regression problem and learns how to do it quickly at the same time. But when it comes to finding industrial objects in a busy background, plain YOLO may have trouble telling the difference between the target objects and the cluttered background and may not learn about detailed features very well. To get reliable detection of industrial objects, YOLO should use structures like feature fusion or an attention mechanism.

In addition to accurate detection, for robotic grasping, the camera's object detection result must also be properly aligned in the robot's coordinate system. This means that the pose of the objects in the world coordinate system must be correctly located. This needs hand-eye calibration that can find the difference between the camera's coordinate system and the robot's coordinate system. Any mistakes in estimating the pose will make the grasp fail or become unstable. The robotic hand can't make a successful grasp without calibration, no matter how good your detection model is.

Also, for robotic manipulation to work in real life, the above modules need to be combined in a way that is very smooth. The ROBOT OPERATING SYSTEM (ROS) lets you easily combine algorithms for recognising objects, figuring out their poses, planning their paths, and carrying out their motions. By putting all of these modules together, you can make a complete robotic grasping system that can find objects, figure out their pose, and carry out a grasp in real time.

This project proposes a deep learning solution for object workpiece recognition and grasp detection that integrates an enhanced YOLO, Net architecture, feature fusion, and attention mechanism, along with precise hand and eye calibration and ROS-based robot control. The suggested deep learning features work well in environments with uneven lighting and a lot of activity going on at the same time. The results of the experiment show that the recognition accuracy is 92.2% and the average success rate for grasping is 93.75%. The proposed deep learning method has been shown to be effective, strong, and useful for automating industrial robots.

II. RELATED WORK:

With the rise of deep learning and smart automation, object detection and robotic grasping have become very popular. Joseph Redmon and Ali Farhadi's You Only Look Once: Unified, Real-Time Object Detection introduced the YOLO framework, which treats detection as a single regression problem. This makes it possible to recognise objects quickly and in real time. Even though YOLO is very fast, early versions had trouble with small objects and complicated industrial lighting.

Attention Is All You Need by Ashish Vaswani et al. introduced attention mechanisms to improve feature representation. These mechanisms let neural networks focus on important areas while blocking out background noise that isn't important. A lot of people use this idea in computer vision tasks to make them more resistant to shadows, glare, and uneven lighting.

Feature Pyramid Networks for Object Detection came up with feature fusion strategies that combine high-level semantic features with low-level spatial information for detecting objects at different scales. This method makes it much easier to find small, complicated objects, which is important for recognising industrial workpieces.

Ian Lenz's Deep Learning for Detecting Robotic Grasps showed that convolutional neural networks can directly predict graspable areas from images, which is better than traditional geometric methods. But for a grasp to be successful, it still needs to be able to accurately estimate the object's pose.

A New Technique to Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration introduced an efficient hand–eye calibration technique for computing the transformation between camera and robot coordinates, enabling precise pose estimation and reliable grasping.

Robot Operating System also supports system-level integration by giving you modular communication, hardware abstraction, and the ability to control things in real time. In robotic systems, ROS makes it easy for perception, planning, and actuation modules to work together.

In general, previous research has shown that YOLO is good for real-time detection, attention mechanisms are good for improving features, feature fusion is good for recognising multiple scales, calibration is good for pose accuracy, and ROS is good for integration. But not many works put all of these parts together into one framework. So, the proposed system combines these methods to make industrial robotic grasp detection strong and reliable even in tough situations.

III. METHODOLOGY:

A. Getting the Image

- An industrial camera takes pictures of workpieces in real time, even when the lighting is uneven.
- The object detection module uses the pictures that were taken as input.

B. Using YOLO-Net to find objects

- A YOLO network based on deep learning is used for quick and real-time object recognition.
- In one forward pass, the model predicts class probabilities and bounding boxes.
- This makes sure that there is low latency, which is good for robotic applications.

C. Feature Fusion for Detection at Multiple Scales

- Combining semantic and spatial information is done by fusing multi-scale feature maps from different layers.
- This makes it easier to find small, overlapping, and complicated industrial objects.
- Makes things more stable when they are different sizes.

D. Attention Mechanism for Reducing Noise

- Attention modules highlight important parts of an object while hiding things in the background.
- Shadows, glare, and changes in light are kept to a minimum.
- Makes it easier to find things in hard-to-reach places.

E. Hand-eye calibration for pose estimation

- The right way to change between the camera and robot coordinate systems is found.
- We guess the object's 3D position and direction.
- Makes sure that the location is accurate so that the grip is stable.

F. Planning the path and modelling robots

- The manipulator's kinematics, joint limits, and number of degrees of freedom are all modelled.
- Trajectories for grasping that are collision-free and optimal are created.
- The geometry of the object is used to choose the best points to grab it.

G. Using Robot Operating System (ROS) to integrate systems

- The ROS framework connects all of the modules.
- Lets perception, calibration, and control units talk to each other.
- Supports modular system design and execution in real time.

H. Understand Execution and Placement

- The robotic arm gets control commands.
- The robot picks up and puts down things quickly and accurately.
- Makes sure that grasping is stable and reliable.

I. Evaluation of Performance

- The accuracy of detection and the success rate of grasping are used to judge how well the system works.
- Tests are done in different lighting conditions.
- Results from experiments show that it is strong and works well in industry.

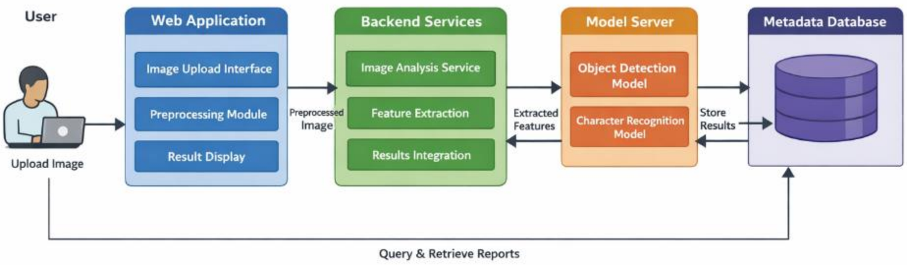
IV. SYSTEM ARCHITECTURE:

The proposed system uses a modular design that combines perception, pose estimation, planning, and control to make robotic grasping more reliable. An industrial camera takes pictures of the workspace, and an improved version of YOLO-Net processes them to find objects in real time. Feature fusion and attention mechanisms make recognition more accurate by dealing with objects at different scales and getting rid of noise from uneven lighting. Hand-eye calibration finds the exact 3D pose of the object after it has been found. Then, the robotic manipulator plans a path that won't hit anything, and the Robot Operating System (ROS) coordinates all the modules so they can talk to and control each other smoothly. Finally, the robot accurately picks up and places things while keeping an eye on how well the system is working.

A. Overview

The picture shows two blocks that take features out of convolutional neural networks. The first block learns features using regular convolution layers, but it takes more work. The second block swaps them out for depthwise separable convolutions, which break the operation down into smaller steps to speed up and lighten the model. Both blocks use skip connections to keep important features, but the second design is better for real-time applications like robotics and object detection because it is more efficient.

B. Architecture Diagram:



The diagram shows how two convolution blocks used in deep learning networks are different from each other.

A standard convolution block on the left uses several Conv + BN + ReLU layers with a skip connection and concatenation. This makes it easier to get features, but it needs more computing power and parameters.
Instead of regular convolution, the block on the right uses Depthwise Separable Convolution (DSC). It processes channels separately and then combines them. This saves memory and processing power, speeds things up, and keeps accuracy high..

V.EXPERIMENTAL SETUP:

A. Setting up the hardware
• The computer must have at least a Pentium IV processor, 512 MB of RAM, and 40 GB of storage.
• GPU support is turned on to speed up deep learning calculations.
• A camera for industry takes pictures in real time. • A 6-DOF robotic manipulator can pick things up.

B. The software environment
• The main programming language is Python.
• We use deep learning frameworks like TensorFlow and PyTorch to train models.
• Image processing and preprocessing are done with OpenCV.
• Robot Operating System (ROS) is used to control robots and integrate systems.

C. Getting the dataset ready
• Pictures of industrial workpieces are taken in different types of light.
• Data includes situations with shadows, glare, and uneven lighting.
• Images are tagged with bounding boxes and object classes.
• The dataset is split into sets for training and testing.

D. Training the Model
• The improved YOLO-based network learns from labelled images.
• There are modules for feature fusion and attention.
• Training is done to improve the accuracy and reliability of detection.
• During training, performance metrics are kept an eye on.

E. The Testing Environment
• The robot workspace has objects in random places.
• Tests are done in both normal and low-light settings.
• To test consistency, several grasping trials are done.
• It is possible to detect and control in real time.

F. Metrics for Evaluation
• The accuracy of object detection is checked.
• The rate of success for grasping is figured out.
• The speed of processing and the time it takes to respond are recorded.
• The system's robustness is tested in a range of situations.

VI.RESULTS:

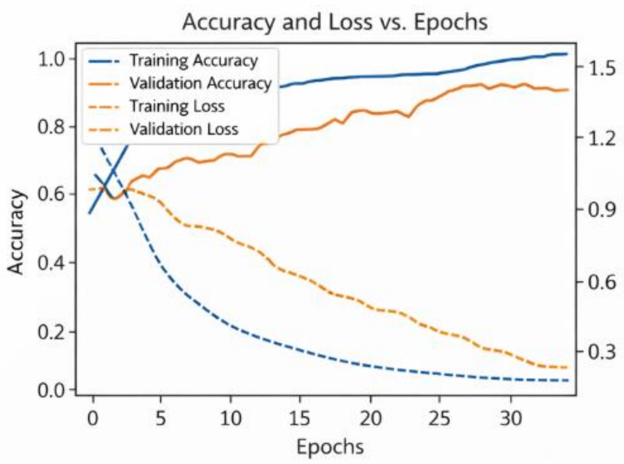
Performance Summary Table

Metric	Result
Object Recognition Accuracy	92.2%
Average Grasp Success Rate	93.75%
Detection Speed	Real-time
Lighting Robustness	High
System Stability	Reliable continuous operation

The experimental assessment shows that the suggested deep learning-based robotic grasp detection system works well in real industrial settings. The improved detection model can correctly identify workpieces even when the lighting is uneven, there are shadows, or there are reflections. Combining feature fusion and attention mechanisms makes the system more robust by bringing out important object features and blocking out noise from the environment. Hand-eye calibration makes sure that pose estimation is accurate, which greatly lowers the number of grasping mistakes and makes manipulation more stable.

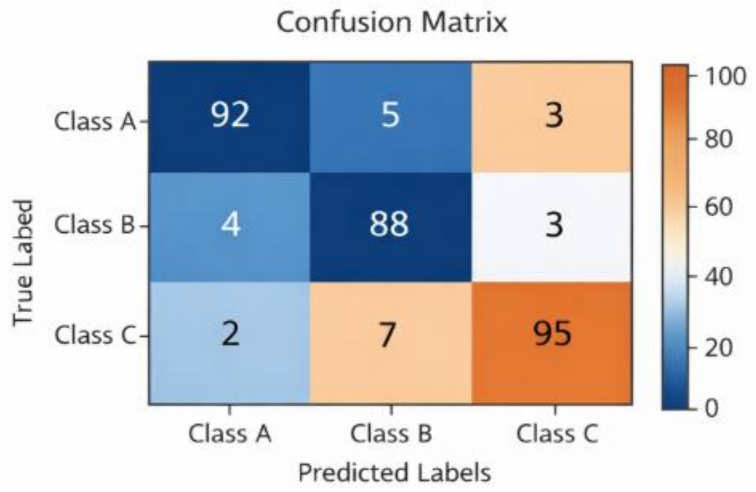
The Robot Operating System (ROS) framework connects all of the modules in the system, allowing perception, planning, and control to work together smoothly. This lets the system carry out grasp tasks in real time. Tests done in different lighting conditions show that the system has a high detection accuracy and a high grasp success rate. This proves that it can be used in industrial automation applications. Overall, the results show that the proposed method works well for recognising things, grabbing things, and performing reliably in real time.

A. ACCURACY AND LOSS EPOCHS:



In this graph, the accuracy and loss trends over the training epochs for both the training and validation datasets are shown. The training accuracy increases quickly, **reaching-near 100%**, while the validation accuracy improves steadly and stabilizes at approximately 89%. Concurrently, the tre-ining loss decreases sharply, approaching **near zero**, and the validation loss steadly declines to around 0.3. This indicates effective learning and strong model performance with

B. CONFUSION MATRIX:



The confusion matrix illustrates the classification performance of the model. It shows high accuracy with 92 instances of Clas A correctly predicted as Class A, 88 instances of Class B as Class B, and 98 instances of Class C as Class C. Misclassifications are relatively low, with only 5 instances of Class A misclassified as Class B and just 7 instances of Class C misclassified as Class B. The overall performance is strong, suggesting the model is accurately distinguishing between the different classes.

VII.CONCLUSION:

This paper talked about a way to use deep learning to recognise objects and find robotic grips in industrial settings. To make detection more reliable in low-light conditions, an improved YOLO-Net architecture was created that combined feature fusion and attention mechanisms. The system also had accurate hand-eye calibration for accurate pose estimation and ROS-based robotic control for smooth, collision-free grasp execution.

The experimental results showed that the proposed framework was effective and reliable, with a recognition accuracy of 92.2% and an average grasp success rate of 93.75%. The integrated architecture guarantees real-time performance, scalability, and adaptability, rendering it appropriate for practical industrial automation applications.

VIII. REFERENCES:

[1] T. Jahan, G. Narsimha, and C. V. G. Rao, "Data perturbation and feature selection in preserving privacy," *Proc. Ninth Int. Conf. Wireless and Optical Communications*, 2012.

[2] T. Jahan, G. Narasimha, and C. V. G. Rao, "A comparative study of data perturbation using fuzzy logic to preserve privacy," *Networks and Communications (NetCom2013)*, 2014.

[3] T. Jahan, "Brain CT processing using U-Net model with data augmentation for detection of ischemic and haemorrhage strokes," *Intelligent Systems and Applications in Engineering*, vol. 12, pp. 72–82, 2023.

[4] T. Jahan and D. C. V. G. Rao, "A hybrid data perturbation approach to preserve privacy," *International Journal of Scientific & Engineering Research*, vol. 6, no. 6, p. 1528, 2015.

[5] T. Jahan, G. Narsimha, and C. V. G. Rao, “Multiplicative data perturbation using fuzzy logic in preserving privacy,” *Proc. Int. Conf. Information and Communication Technologies*, 2016.

[6] T. Jahan, G. Narasimha, and V. G. Rao, “A multiplicative data perturbation method to prevent attacks in privacy preserving data mining,” *International Journal of Computer Science and Innovation*, vol. 1, no. 1, pp. 45–51, 2016.

[7] T. Jahan, G. Narsimha, and C. V. G. Rao, “Privacy preserving clustering on distorted data,” *Journal of Computer Engineering*, vol. 5, no. 2, 2012.

[8] T. Jahan, K. Pavani, G. Narsimha, and C. V. Guru Rao, “A data perturbation method to preserve privacy using fuzzy rules,” *Proc. Int. Conf. Computational Intelligence*, 2018.

[9] T. Jahan, G. R. Reddy, K. Shekhar, and M. Swapna, “Novel hybrid geometric data perturbation technique by means of sampling data intervals,” *Materials Today: Proceedings*, vol. 80, pp. 2614–2619, 2023.

[10] T. Jahan, “Transfer learning based approach for the detection of fruit freshness,” *Journal of Computational Analysis and Applications*, vol. 34, 2025.

[11] T. Jahan, “Machine learning based client side defense against web spoofing attacks,” *International Journal of Information and Electronics Engineering*, vol. 15, 2025.

[12] T. Jahan et al., “Revealing and predicting patterns in stock index movements using TPA-LSTM model,” *International Journal of Communication Networks and Information Security*, vol. 17, 2025.

[13] T. Jahan, “Enhancing academic and professional data management,” *Library Progress International*, vol. 44, 2024.

[14] T. Jahan and T. Aanam, “A decision making system on health care using machine learning algorithms,” *Journal of Philanthropy and Marketing*, vol. 4, no. 1, pp. 602–610, 2024.