FAKULTAS VOKASI
UNIVERSITAS SANATA DHARMA

SURAT TUGAS

Nomor: 7j/Dek/FV-USD/VIII/2025

Pimpinan Fakultas Vokasi Universitas Sanata Dharma Yogyakarta, dengan ini memberikan tugas kepada:

Nama	:	Ir. Sukma Meganova Effendi, M.Tr.T.
Pekerjaan/ Jabatan	:	Dosen Program Studi Teknologi Rekayasa Mekatronika Fakultas Vokasi
Unit Organisasi	:	Universitas Sanata Dharma Yogyakarta
Alamat	:	Kampus III Paingan, Maguwoharjo, Depok, Sleman, Daerah Istmewa Yogyakarta
Telepon	:	0274-883037
Untuk Keperluan	:	Menulis Artikel ilmiah "*Deep Learning-Driven Ogject Detection fore Smart Robotic Manipulation: A YOLOv11 Training Benchmark*"
Waktu	:	1 (satu) Semester pada Semester Gasal 2025

Demikian Surat Tugas ini dibuat untuk dilaksanakan sebaik-baiknya, dan apabila sudah selesai agar segera memberikan laporan.

Yogyakarta, 13 Agustus 2025

Dekan,

Ir. Bernardinus Sri Widodo, M.Eng.

2025 5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA) | 979-8-3315-6140-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICICYTA68677.2025.11362870

Homepage

The 5th International Conference on Intelligent Cybernetics Technology & Applications 2025 (ICICyTA 2025) is scheduled to be held on December 17-19, 2025, in Jogjakarta, Indonesia. This event will be conducted in a hybrid format, accommodating both onsite and virtual participants, and is jointly hosted by Human Centric (HUMIC) Engineering at Telkom University (Tel-U) and supported by Bina Nusantara University (Binus), Pembangunan Nasional "Veteran" Yogyakarta University (UPNVY) and Pembangunan Nasional "Veteran" Jakarta University (UPNVJ), Indonesia. Moreover, this conference is sponsored by Humic Engineering and IEEE Indonesia Section.

Fig. 1. Example documentation photos taken during the conference onsite.

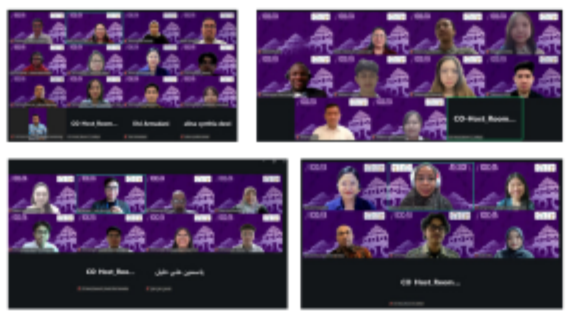

Fig. 2. Example documentation photos taken during the conference online.

This year's conference focuses on the theme *" AI-Driven Transformation: Cybernetic Systems for Sustainability, Security, and Smart Living."* Papers accepted through the rigorous review process will be considered for publication in IEEE Xplore, subject to compliance with its scope and quality standards. We encourage submissions of original and innovative research in fields such as cybernetics, computational intelligence, IoT, biomedical engineering, and related disciplines.

ICICyTA 2025 aims to foster meaningful discussions and collaborations among researchers and practitioners. The selection process for this conference has been highly competitive, resulting in 152 papers has been accepted (Acceptance ratio 56.1%). These paper come from various region such as Asia/pacific, The United Kingdom, The United States of America, The Middle East, Africa and Latin America. We hope these works will inspire further research and practical solutions for sustainable, secure, and smart living. We deeply appreciate the dedication of authors and the meticulous reviews provided by experts, which have ensured the integrity of the program.

In closing, we extend our sincere gratitude to all contributors and participants for their support and active engagement in ICICyTA 2025. Your efforts are integral to the success of this conference.

Media Type	Part Number	ISBN
XPLORE COMPLIANT	CFP25AY3-ART	979-8-3315-6140-6

Organized by Telkom University
Prepared by EDAS Conference Services.

2025 5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA) | 979-8-3315-6140-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICICYTA68677.2025.11362375

Committee

Steering Committee
Prof. Dr. Adiwijaya, Telkom University (Indonesia)
Prof. Dimitrios Georgakopoulos, Swinburne University of Technology (Australia)
Prof. Oshita, Kyushu Institute of Technology (Japan)

General Chairs
Bambang P., Ph.D. (Telkom University, Indonesia)

Co-Chair
Assoc. Prof. Dr. Putu Harry Gunawan (Telkom University, Indonesia)

Technical Program Committee Chair
Dr. Hilal Hudan Nuha (Telkom University, Indonesia)
Dr. Erwin Budi Setiawan (Telkom University, Indonesia)
Dr. Syafiqah binti Saidin (Universiti Teknologi Malaysia, Malaysia)
Assoc. Prof. Dr. Z.K. Abdurahman Baizal (Telkom University, Indonesia)
Prof. Dr. I.K.G. Bendesa, M.A.D.E (Universitas Bali Internasional, Indonesia)
Prof. Haueise (Technische Universitt Ilmenau, Germany)
Prof. Oshita (Kyushu Insititue of Technology, Japan)

Track Chair
Assoc. Prof. Satria Mandala, Ph.D(Telkom University, Indonesia)
Dr. Aditya Firman Ihsan (Telkom University, Indonesia)
Dr. Hilal Hudan Nuha (Telkom University, Indonesia)
Dr. Aizreena Azaman (Universiti Teknologi Malaysia, Malaysia)

Program Committee
Dr. Dian Indiyati (Telkom University, Indonesia)
Dr. Arif Bijaksana (Telkom University, Indonesia)
Dr. Nachwan Mufti Adriansyah (Telkom University, Indonesia)
Dr. Irni Yunita (Telkom University, Indonesia)
Assoc. Prof. Dr. Eileen Su Lee Ming (Universiti Teknologi Malaysia, Malaysia)
Dr. Muhammad Amir As'ari (Universiti Teknologi Malaysia, Malaysia)
Dr. Mariaulpa Sahalan (Universiti Teknologi Malaysia, Malaysia)
Dr. Mohd Azhar Abdul Razak (Universiti Teknologi Malaysia, Malaysia)
Dr. I Ketut Tunas, M.Si. (Universitas Bali Internasional, Indonesia)
I Kadek Arya Sugianta, S.Kom., M.Kom. (Universitas Bali Internasional, Indonesia)

Publication
Dr. Ema Rachmawati (Telkom University, Indonesia)
Dr. Imelda Atastina Pangaribuan (Telkom University, Indonesia)

Brahmantya Aji Pramudita (Telkom University, Indonesia)
Ts. Dr. Nurul Ashikin Abdul Kadir (Universiti Teknologi Malaysia, Malaysia)
Ts. Dr. Johanna Ahmad (Universiti Teknologi Malaysia, Malaysia)
Dr. Nies Hui Wen (Universiti Teknologi Malaysia, Malaysia)
Dr. Pang Yee Yong (Universiti Teknologi Malaysia, Malaysia)

Technical Program Committee Members / Reviewers
Prof. Dr. Tong Boon Tang (Universiti Teknologi Petronas, Malaysia)
Prof. Dr. Mohd Shahrizal bin Sunar (Universiti Teknologi Malaysia, Malaysia)
Prof. Dr. Oshita (Kyushu Institute of Technology, Japan)
Prof. Dr. Ali Selamat (Malaysia-Japan International Institute of Technology, Malaysia)
Prof. Dr. Haueisen (Universitat Ilmenau, Germany)
Assoc. Prof. Satria Mandala, Ph.D (Telkom University, Indonesia)
Assoc. Prof. Dr. Angelina Prima Kurniati (Telkom University, Indonesia)
Assoc. Prof. Dr. Putu Harry Gunawan (Telkom University, Indonesia)
Assoc. Prof. Dr. Rania Al-Ashwal (Universiti Teknologi Malaysia, Malaysia)
Assoc. Prof. Dr. Nasrul Humaimi Mahmood (Universiti Teknologi Malaysia, Malaysia)
Assoc. Prof. Dr. Eileen Su Lee Ming (Universiti Teknologi Malaysia, Malaysia)
Dr. Salim M Zaki (Directorate of Studies, Planning and Follow-up, Iraq)
Rashid Khokhar, PhD (Charles Sturt University, Australia)
Narita Aquarini, PhD (Banque et Gestion de Risque Excelia Group, France)
Assoc. Prof. Dr. Wael M.S. Yafooz (Taibah University, Saudi Arabia)
Dr. Mohd Soperi Mohd Zahid (Universiti Teknologi Petronas, Malaysia)
Dr. Kemas Muslim Lhaksmana (Telkom University, Indonesia)
Dr. Kadek Dwi Hendratma Gunawan (Universitas Sebelas Maret, Indonesia)
Dr. Didit Adytia (Telkom University, Indonesia)
Dr. Rio Guntur Utomo (Telkom University, Indonesia)
Dr. Nurul Ikhsan (Telkom University, Indonesia)
Dr. Eng Mahmud Dwi Sulistyo (Telkom University, Indonesia)
Dr. Guntur Prabawa Kusuma, S.T., M.T. (Telkom University, Indonesia)
Dr. Vera Suryani (Telkom University, Indonesia)
Dr. Agung Toto Wibowo (Telkom University, Indonesia)
Dr. Agustian Taufiq Asyhari (Telkom University, Indonesia)
Dr. Eng. Favian Dewanta (Telkom University, Indonesia)
Dr. Eng. Wikky Fawwaz Al Maki (Telkom University, Indonesia)
Irma Palupi, Ph.D (Telkom University, Indonesia)
Ts. Dr. Nurul Ashikin Abdul Kadir (Universiti Teknologi Malaysia, Malaysia)
Ts. Dr. Johanna Ahmad (Universiti Teknologi Malaysia, Malaysia)
Ts. Dr. Goh Eg Su (Universiti Teknologi Malaysia, Malaysia)
Ts. Dr. Chan Weng Howe (Universiti Teknologi Malaysia, Malaysia)
Dr. Aizreena Azaman (Universiti Teknologi Malaysia, Malaysia)
Dr. Nies Hui Wen (Universiti Teknologi Malaysia, Malaysia)
Dr. Noor Hidayah Zakaria (Universiti Teknologi Malaysia, Malaysia)
Dr. Norhana Jusoh (Universiti Teknologi Malaysia, Malaysia)
Dr. Muhammad Amir As'ari (Universiti Teknologi Malaysia, Malaysia)
Dr. Mariaulpa Sahalan (Universiti Teknologi Malaysia, Malaysia)
Dr. Pang Yee Yong (Universiti Teknologi Malaysia, Malaysia)
Dr. Mohd Azhar Abdul Razak (Universiti Teknologi Malaysia, Malaysia)
Gde Palguna Reganata, S.Si., M.Si. (Universitas Bali Internasional, Indonesia)
Khadijah Febriana Rukhmanti Udhayana, HR. S.T., M.Kom. (Universitas Bali Internasional, Indonesia)
I Gede Wahyu Surya Dharma, S.Kom., M.Kom. (Universitas Bali Internasional, Indonesia)
I Gede Karang Komala Putra S.Kom.,M.Kom. (Universitas Bali Internasional, Indonesia)
Ni Kadek Winda Patriani, S.Kom., M.Kom. (Universitas Bali Internasional, Indonesia)
Muhammad Ridha, S.T., M.Kom. (Astra Polytechnic, Indonesia)

Anni Chen (Amazon, USA)
Sreepal Reddy Bolla (Senior IEEE Member)
Mamatha Adinarayana Swamy (Microsoft Corporation)
Ravindra Reddy Madireddy
Soma Kiran Kumar Nellipudi (Incomm Payments, USA)
Madhu Niranjan Reddy Puduru (Synchronoss Technologies Inc, USA)
Rishab Gupta (Palo Alto Networks, USA)
Karthik Chakravarthy (Microsoft)
Satwik Reddy Jambula (Uber Technologies)
Sai Kaushik Ponnekanti (Meta)
Praveen Chinnusamy (Amazon, USA)
Ajay Averineni (IBM, Indonesia)
Venkaiah Chirumavilla (D Scripts Inc)
Banani Mohapatra (Walmart)
Bhavnish Walia (Amazon, USA)
Sumanth Kadulla (Western Illinois University)
Krishna Dornala (Resource Logistics Inc, USA)
Prabhakar Singh (Meta)
Justin Davis (Castlight Health, Inc, USA)
Ramamohan (Broadcom, USA)
Tapan Vijay (Meta Platform)
Sai Prasad Mukala (Info Keys Inc, USA)
Shakir Poolakkal Mukkath (Walmart)
Ravi Babu Vellanki (Tek Dallas Inc)
Soham Chakraborti (Amazon, USA)
Sriramu Kundoor (Organon LLC, USA)
Sarat Piridi (Hanwah-Qcells, Korea)
Sanjeev Kumar Mourya (Gyansys Inc, USA)
Nidhin Jose (Walmart)
Mandhuri (Microsoft)
Bhaskar Goyal (Google)
Dinesh Boinpally (Goldman Sachs, USA)
Chaitra Vatsavayi (Google)
Amarnatha Reddy Thumalur (JPMorgan Chase & Co)
Lakshmi Srinivasarao Kothamasu (Fidelity Investments, USA)
Kunj Raval (Meta Platforms, Inc)
Vina Sari Yosephine (Universitas Pembangunan Nasional "Veteran" Jakarta, Indonesia)
Ni Putu Devira Ayu Martini (Universitas Pembangunan Nasional "Veteran" Jakarta, Indonesia)
Fajar Rahayu (Universitas Pembangunan Nasional "Veteran" Jakarta, Indonesia)
Silvia Anggraini (Universitas Pembangunan Nasional "Veteran" Jakarta, Indonesia)
Elvi Armadani (Universitas Pembangunan Nasional "Veteran" Jakarta, Indonesia)
Dr. Purba Daru Kusuma (Telkom University)
Dr. Budhi Irawan (Telkom University)
Dr. Andrew Brian Osmond (Telkom University)

Organized by Telkom University
Prepared by EDAS Conference Services.

ICICyTA 2025

2025 5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA) | 979-8-3315-6140-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICICYTA68677.2025.11362391

2025 5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)

Parallel Session 1A (*1A*)

Parallel Session 1B (*1B*)

Parallel Session 1C (*1C*)

Parallel Session 1D (*1D*)

Parallel Session 1E (*1E*)

Parallel Session 2A (*2A*)

Parallel Session 2B (*2B*)

Parallel Session 2C (*2C*)

Parallel Session 2D (*2D*)

Parallel Session 2E (*2E*)

Parallel Session 3A (*3A*)

Parallel Session 3B (*3B*)

Parallel Session 3C (*3C*)

Parallel Session 3D (*3D*)

Parallel Session 3E (*3E*)

Parallel Session 4A (*4A*)

Parallel Session 4B (*4B*)

Parallel Session 4C (*4C*)

Parallel Session 4D (*4D*)

Parallel Session 4E (*4E*)

Parallel Session 5A (*5A*)

Parallel Session 5B (*5B*)

Parallel Session 5C (*5C*)

Parallel Session 5D (*5D*)

Parallel Session 5E (*5E*)

Parallel Session 6B (*6B*)

Parallel Session 6C (*6C*)

Parallel Session 7A (*7A*)

Parallel Session 7B (*7B*)

Parallel Session 7C (*7C*)

Parallel Session 8A (*8A*)

2025 5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)

Deep Learning-Driven Object Detection for Smart Robotic Manipulation: A YOLOv11 Training Benchmark

1st Sukma Meganova Effendi
Mechatronics Engineering
Rajamangala University of Technology Thanyaburi
Pathum Thani, Thailand
sukma_e@mail.rmutt.ac.th

2nd Dechrit Maneetham
Mechatronics Engineering
Rajamangala University of Technology Thanyaburi
Pathum Thani, Thailand
dechrit_m@hotmail.com

3rd Petrus Sutyasadi
Mechatronics Department
Faculty of Vocational Study
Sanata Dharma University
Yogyakarta, Indonesia
peter@usd.ac.id

2025 5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA) | 979-8-3315-6140-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICICYTA68677.2025.11362381

***Abstract*—Robotic manipulation relies heavily on vision systems that have the capability of accurately detecting and locating objects in dynamic environments. However, traditional methods based on vision remain inefficient because of the sensitivity of variations lighting, occlusions, and background noise. Therefore, their use in robots has been limited. Deep learning-based detectors, particularly the YOLO models, have overcome these obstacles through end-to-end training and real-time inference. The possible settings of YOLOv11, in terms of model scale, optimizer, and training duration for detection of color-coded objects in robotic perception, have few studies that systematically assess them. This research analysed the performance of three YOLOv11 variants (Nano, Small, and Medium) in collaboration with robotic manipulation. A dataset of 1,536 images captured from the robot's environment was used for evaluation of two optimizers (Stochastic Gradient Descent and AdamW) along with four training durations (50, 100, 200, and 400 epochs), which created a total of 24 training conditions. The evaluation metrics were precision, recall, mAP@0.5, mAP@0.5–0.95, and training time. The results show that YOLOv11-Medium with SGD for 400 epochs reached the maximum accuracy (mAP@0.5–0.95 = 0.9631), while YOLOv11-Small with SGD for 200 epochs gave the most efficient combination of accuracy and cost in computing resources. The Nano model at 50 epochs was shown to be effective for low-power embedded systems. Qualitative testing of stereo cameras validated proper detection and the models trained with YOLOv11 serve as a reference for selecting the right settings, while evaluation of real-time deployment is reserved to future research.**

Keywords—YOLOv11, training, optimizer, epoch, metrics

I. INTRODUCTION

Robotic arms are essentially the tools of precision, flexibility, and independence in smart manufacturing, industrial automation, and intelligent systems. An intelligent robotic manipulation will entirely depend upon the three module integrations: perception, planning, and control. This will enable the robots to sufficiently sense environment conditions, make decisions, and then execute actions. Amongst the three modules, object detection is the one that performs the function of perception—the one that enables robots to identify and localize objects accurately even in the midst of dynamic environments [1]. In the case of advanced robotic systems, the perception module is the basis for the higher-level processes of path planning and motion control. Object detection is an important function within this perception layer since it gives the robots the spatial awareness they need for making decisions and manipulation. The research mainly focuses on the development and evaluation of YOLOv11 training settings to figure out the best detection model. The future work is expected to involve real-time robotic perception and deployment-level performance analysis.

Vision-based object detection techniques are mostly based on color segmentation, edge extraction, or shape descriptors which are not very strong to handle well the varying light, occlusions, and color changes [2], [3]. Classical algorithms such as Canny or Sobel edge detectors require manual parameter tuning and are highly sensitive to noise and contrast fluctuations, making them unsuitable for unstructured environments typical of smart robotic manipulation [3]. The earlier approaches of single monocular image segmentation and 3D model reconstruction gave manipulations more accuracy but still faced difficulties in precise localization in dynamic environments [4]. One of the detectors in the YOLO (You Only Look Once) family was the original YOLO, which was the first system to surpass the two-step detection and classification processes of image processing to achieve real-time performance with great accuracy, thus revolutionizing robotics perception by one single forward pass of the network [1], [5]. Deep learning-based approaches enable feature extraction directly from images and learn hierarchical representations, eliminating the dependence on hand-crafted features that constrained classical vision methods. On the other hand, the best YOLO settings are determined by the dataset, training settings, and application scope—variables that have not yet been considered in depth for the scenario of robotic manipulation with color-coded objects and different lighting conditions.

Certainly, deep learning is the main method for robotic perception and still gives an excellent chance for the wide ranging application in visual domains as well as handling of the difficult environments. One-stage detectors such as YOLO, Single Shot Multi-Box Detector (SSD), and Efficient-Det have performed exceptionally well in real-time detection tasks and thus have outperformed their two-stage counterparts like Faster R-CNN in terms of performance [5]. Naturally, the YOLO family has resulted in a continuous increase in speed, precision, and architectural efficiency throughout the entire period from YOLOv1 up to YOLOv12 [6], [7], [8], [9]. The C3k2 block and the C2PSA attention module are within the YOLOv11 design and have been previously proven to efficiently detect small and overlapping objects while consuming limited CPU resources [8], [9]. YOLOv11's features make it suitable for embedded and robotic applications that require a balance between speed and precision. However, most YOLO-related studies still focus on generic datasets such as COCO and PASCAL VOC, which do not reflect the structured, color-based environments typical of robotic grasping and manipulation [6]. Unexpectedly, there are only a limited number of studies examining how the choice of optimizer (Stochastic Gradient Descent (SGD) versus AdamW) or varying training times influence YOLO' performance in robotic applications. Existing empirical have shown that adaptive optimizers such as AdamW tend to stabilize convergence in gradients, but it is clear that SGD provides strong generalization when appropriately tuned [11], [12]. Given this context, there remains little comparison of any optimizer applied to any specific YOLOv11 task [10]. This highlights an issue in the benchmarking of YOLOv11 models, such as the nano, small, and medium version, trained in various training setups for object detection in domains relevant to robotics.

To close this gap, the current benchmark systematically tests the YOLOv11-nano/small/medium models trained with SGD or AdamW over multiple training durations—50, 100, 200, and 400 epochs—in a robotic perception environment. A supplementary dataset was created consisting of 1,536 RGB images featuring color-coded obstacles, targets, and grippers to simulate real-world manipulation scenarios. Each training iteration was scientifically analyzed based on mAP and precision-recall measurements, tracking epoch-wise performance, which led to selecting model-optimization pairs that effectively balance accuracy and computational demands, an important aspect for embedded robotics applications. This research therefore establishes a reproducible comparison framework that supports the optimization of deep learning–based perception, while real-time robotic inference remains an avenue for future exploration.

A full YOLOv11 training assessment was performed to close the gap, and it was applied for detecting objects in an environment centered around manipulation tasks. A custom dataset of 1,536 RGB images featuring color-coded obstacles, targets, and grippers was created to simulate the actual conditions of robotic perception. The different YOLOv11 models were trained for 50, 100, 200, and 400 epochs each, using SGD and AdamW optimizers, and were assessed in terms of mAP, precision-recall metrics, and convergence characteristics. The identification of the most productive model-optimizer pairs that weigh accuracy and computational cost—an essential factor for embedded robotic systems—was made possible by the results. Therefore, the present research contributes a comparison framework that is not only reproducible but also applicable in the optimization of deep learning-based perception, whereas real-time robotic inference is still a futurist working direction.

Perception plays a critical role in the efficiency and responsiveness of vision-based robotic applications. The model detection selection affects not only the accuracy but also the timing requirements of subsequent modules like path planning and inverse kinematics. The benchmark insights help developers through the selecting and tuning of the YOLOv11 configuration for intelligent robotic applications. Although real-time deployment is outside the scope of this research, the analysis presented here partially clarifies misconceptions about the role of architectural variations and training approaches in affecting detection performance in future robotic systems.

II. Methodology: YOLOv11-Based Object Detection

This research presents a comprehensive methodology for the assessment of detection performance of YOLOv11 variants in the context of robotic manipulation through color-coded objects. The method was aimed to identify the best combination of model variant, optimizer, and training time that would allow for accurate detection of objects in the SCORBOT-ER 4U robotic arm environment. The procedure comprises dataset preparation, designing the model architecture, and configuring the training, having been executed in an approach systematic as deep learning methodologies that found similar with current robotic vision research [1], [6], [9].

A. Dataset Preparation

A custom RGB image collection comprising 1,536 images was made directly from the SCORBOT-ER 4U robotic workstation through a camera by the authors as shown in Fig. 1. The images were captured in different areas (laboratory, big and small workspace) with different types of light, like sunlight through windows without artificial light, mixing of fluorescent and LED light, and only fluorescent light in the laboratory. The camera views included front, back, side, diagonal, and high angles at different distances, and the blue target, orange gripper, and colored obstacles were placed in different positions throughout the workspace. A lot of images have the robotic arm or other objects partly occluding the targets, hence the dataset incorporates domain shifts in background, viewpoint variation, and realistic occlusions naturally. The visual aspects of the robotic environment were represented in five groups: a blue target, an orange gripper, and three colored obstacles (red, green, and yellow).

Every single image was uploaded to Roboflow, the cloud-based platform utilized for dataset organizing and annotation. The next steps in preprocessing were as follows:

- Auto-orientation, that guarantees all images take the same upright position and have the same orientation;
- Resize to 640×640 to match YOLOv11 input requirements;
- No synthetic augmentation. With no synthetic augmentation, the training behavior remains consistent, thus allowing the benchmark to isolate the effects of model scale, optimizer, and training length without introducing additional stochastic factors.

Roboflow's Smart Polygon and Polygon tools were used to manually annotate each image, and this process produced segmentation masks and YOLO-formatted labels with normalized coordinates and item dimensions. The dataset annotated was then exported and managed on Kaggle's GPU environment, where it was divided into an 80% training set and a 20% validation set, in accordance with common machine learning protocols ensure balanced class representation.

B. YOLOv11 Model Architecture

The detection framework utilized in this research is YOLOv11, an advanced iteration of the YOLO family developed by Ultralytics. YOLOv11 was selected because it is capable of detecting small, overlapping, and color-based objects efficiently, characteristics which are valuable for robotic perceptions and adaptive behavior.

According to [6] and [9], YOLOv11 has a number of architectural improvements that include the C3k2 block and the Cross Stage Partial with Spatial Attention (C2PSA) module that increase the extraction of spatial features and the awareness of the context as shown in Fig. 2. These modules help to get a more precise detection without a significant increase in computation, thereby making YOLOv11 especially appropriate for robotic and embedded applications.

The architecture is based on the traditional backbone–neck–head structure [14]:

- The Backbone takes the visual features from the highest level down to the lowest through convolutional layers and the Spatial Pyramid Pooling-Fast (SPPF) module that combines multi-scale context information.
- The Neck connects network depths from different layers using Cross-Stage-Partial (CSP) connections to improve the consistency of features and the localization of small or clustered objects.
- The Head generates bounding box coordinates, class probabilities, and confidence scores by using an anchor-free detection mechanism that has been optimized for low-latency inference.

The three variants of YOLOv11 that were given the test were:

- YOLOv11-n (Nano): a configuration that offers the most minimal size and very low power consumption, suited for the usage in embedded or low-power robotic systems.
- YOLOv11-s (Small): a compromise between the two extremes that combines fast inference and good detection accuracy.

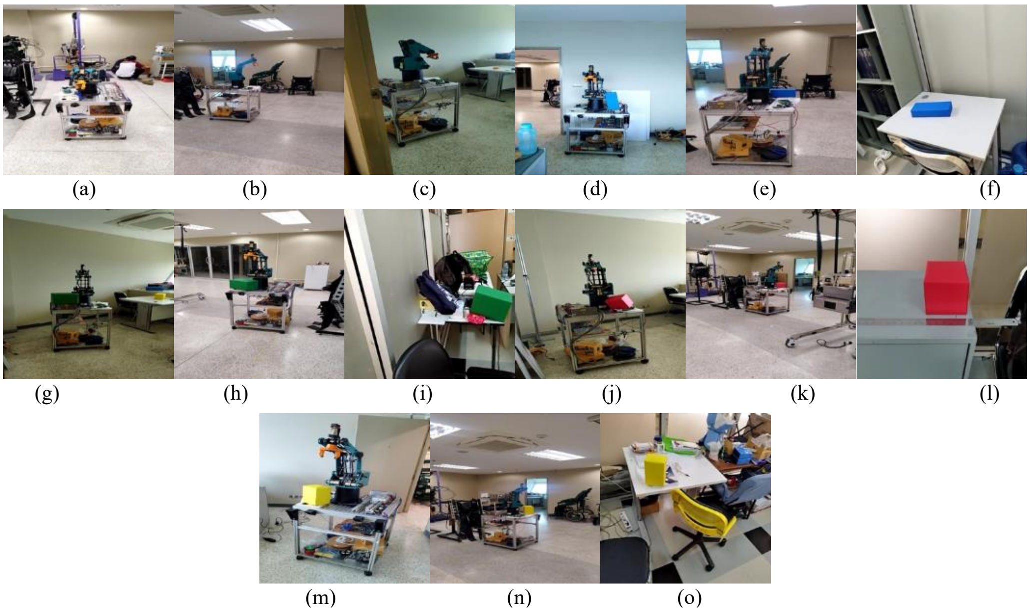

Fig. 1. Example dataset images were taken by author for classified into Gripper (a) – (c), Blue Target (d) – (f), Green Obstacle (g) – (i), Red Obstacle (j) – (l), and Yellow Obstacle (m) – (o), under varying illumination and angles.

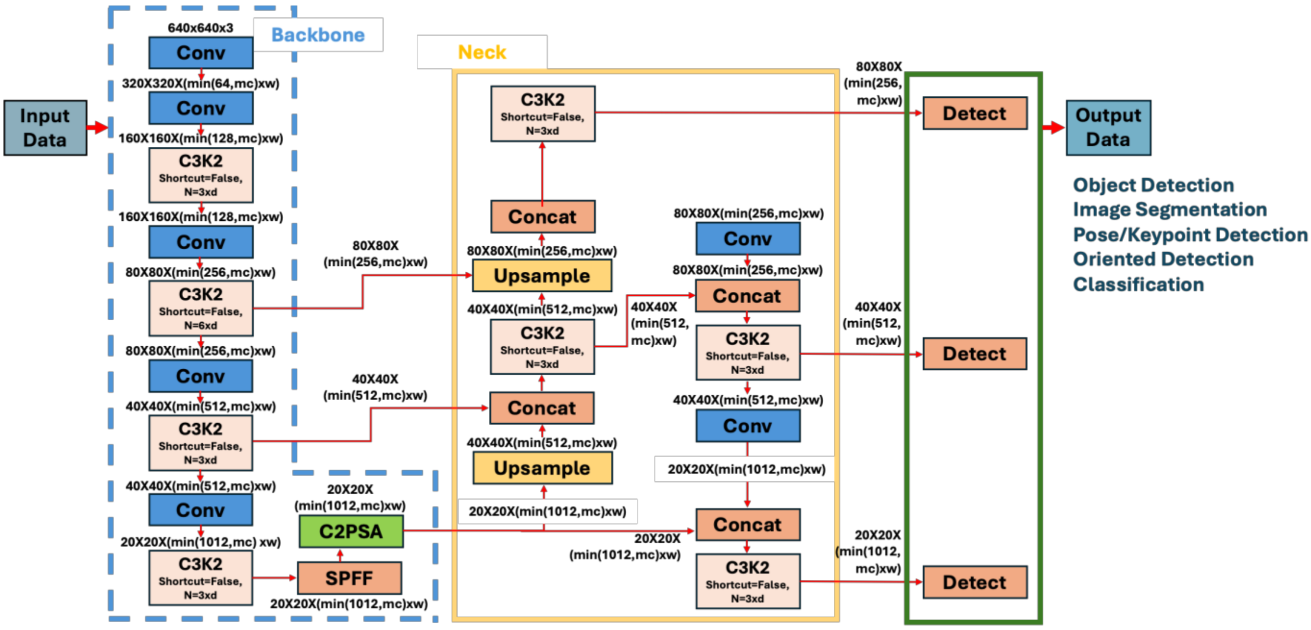

Fig. 2. Architecture of YOLOv11. The enhanced backbone utilises C3k2 blocks in place of C2f, SPPF for multi-scale feature extraction, C2PSA for attention mechanism, and an optimised neck.

- YOLOv11-m (Medium): a model that has the largest size and is the most powerful one demonstrating very high precision in detection on GPU-based platforms.

The existence of these models provided an opportunity to study the scalability of both computational efficiency and performance accuracy, which is representative of the demands of robotic systems that need to trade-off between speed and precision [1].

C. Training and Optimization Configuration

The Kaggle GPU P100 platform was utilized for the training and benchmarking, where the Ultralytics YOLOv11 framework was deployed, thereby confirming that all the experiments were performed under identical computational conditions. The workflow consisted of automatic dataset preparation, model initialization, optimizer selection, and performance evaluation.

The training of all the YOLOv11 variants (n, s, and m) was done using SGD (Stochastic Gradient Descent) and AdamW as the two optimization algorithms and the four different training durations of 50, 100, 200, and 400 epochs at the same time. This way, a total of 24 training configurations were produced and hence a very thorough comparative study was possible. The training hyperparameters setting are shown is Table I. The initial learning rates followed to the Ultralytics defaults, with 0.01 for SGD and 0.001 for AdamW, and both applied with a cosine decay learning rate policy. Momentum for SGD was 0.937, weight decay was 0.0005, and AdamW had a standard decoupled weight decay of 0.01. Other than these defaults, no data augmentations, scheduling modifications, or optimizer-specific calibrations were performed. These parameters ensure that all the model variants were trained under the same, fully reproducible conditions which allowed a fair and controlled comparison of the entire benchmark.

TABLE I. TRAINING PARAMETERS SETTING

Component	Method	Component	Value
Model Variants	YOLOv11-n/ YOLOv11-s/ YOLOv11-m	Training Method	Single Training
Detection Type	Anchor-Free, One-Stage Detector	Epoch	50, 100, 200, 400
Dataset Handling	Manual collection, multi-lighting, multi-angle	Batch size	16
Optimizers	SGD and AdamW	Image Size	640x640
Loss Functions	Box loss, Class loss, Distribution Focal Loss (DFL)	Weight Decay	0.0005 (SGD), 0.01 (AdamW)
Post-Processing	Non-Maximum Suppression (NMS)	Momentum	0.937
Learning Rate	0.01 (SGD), 0.001 (AdamW); cosine decay	Platform	Kaggle P100 GPU

The metrics for model performance assessment were:

- Precision (P): proportion of correct detections among all predicted detections and evaluated in (1).

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \quad (1)$$

- Recall (R): proportion of correctly detected objects among all ground-truth labels and evaluated in (2).

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)} \quad (2)$$

- mAP@0.5 and mAP@0.5–0.95: mean Average Precision across Intersection over Union (IoU) thresholds for global accuracy.

- Training time: total time and time per epoch to assess computational efficiency.

This benchmark concentrates on overall metrics across all classes; per-class average precision and in-depth precision-recall curves are reserved for future work related to deployment-oriented analysis. All results were automatically recorded by the Ultralytics framework and after each experiment, datasets and result directories were archived for reproducibility. Each of the 24 configurations was trained once with a constant random seed. A complete multi-seed variance analysis for all configurations will substantially increase the computational cost and thus it will be a topic of future work; the focus here is on a broad, controlled comparison between model variants, optimizers, and epoch lengths rather than on the stochastic variability that is detailed.

According to [15] and [16], AdamW usually shows quick convergence as well as smooth loss reduction due to learning-rate adjustment, whereas SGD is more likely to generalize when given long training periods. This research has taken the empirical path to assess the performance of both in their detection capability determining the constraint of robotic perception.

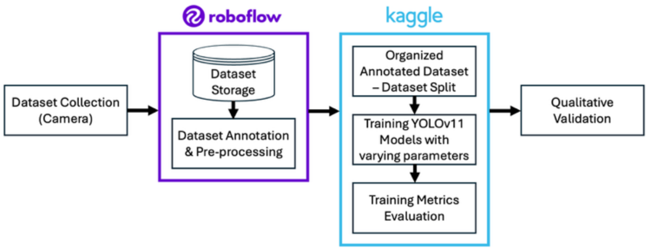

Fig. 3. Methodological pipeline for YOLOv11 training, optimization, and qualitative validation on stereo images.

The methodology shown in Fig. 3 integrates realistic datasets, the use of advanced deep neural network architectures, and systematic benchmarking of optimizers into a single experimental pipeline. The use of controlled dataset preparation, structured model comparison, and multi-metric evaluation not only guarantees reproducibility but also offers precision in the assessment of detection performance. By measuring the effects of YOLOv11 variants, optimizers, and epoch lengths on accuracy and efficiency, this research establishes an essential base for the integration of deep learning-based perception to stereo vision and adaptive path-planning systems for intelligent robotic manipulation.

III. RESULTS AND DISCUSSION

Fig. 4 presents the collective accuracy of the entire experiment comprising 24 configurations (three YOLOv11 n – s – m models, two optimizer SGD and AdamW, and four epochs 50 – 100 – 200 – 400). The number of run index means the value of last row from each results.csv file from training results file index 1 until 24. For the different settings, mAP@0.5–0.95 scores were found in the range of 0.9159 to 0.9631, indicating a gradual increase in performance with the

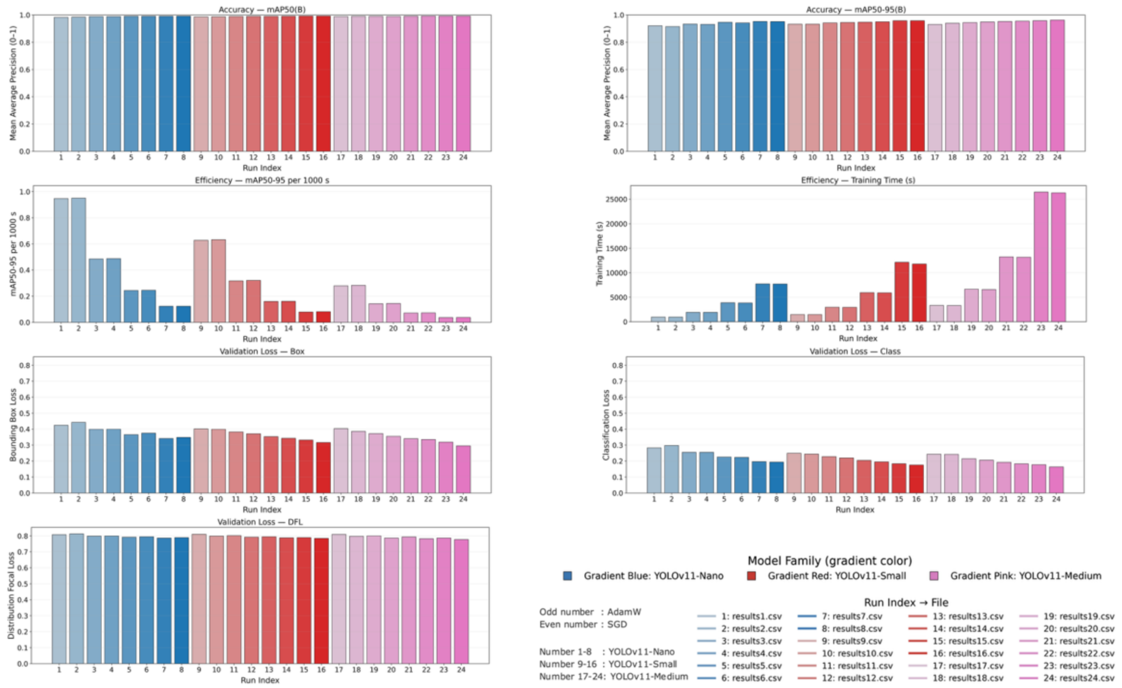

Fig. 4. Comparison final-epoch performance metrics (mAP, loss, precision, recall, and training time) of YOLOv11 models trained with different optimizers and epoch settings have been compared.

increase in the number of epochs. The maximum score by the Medium model with SGD was 0.9631, followed by the Small model with SGD (0.9582) and the Nano model with AdamW (0.9528) at 400 epochs. This illustrates the scale of the model and the selected optimizer highly affect the final number of detection accuracies.

The loss curves (Box loss, Class loss, and DFL loss) as shown in Fig. 5 shows a smooth and steady decrease of both training and validation losses, thus confirming stable optimization without oscillation. For every model, the difference between training and validation losses narrows significantly after 200 epochs, indicating improved generalization. Beyond this point, the curves flatten, providing strong evidence of convergence and highlighting that additional training yields minimal gain. Fig. 6 additionally illustrates the

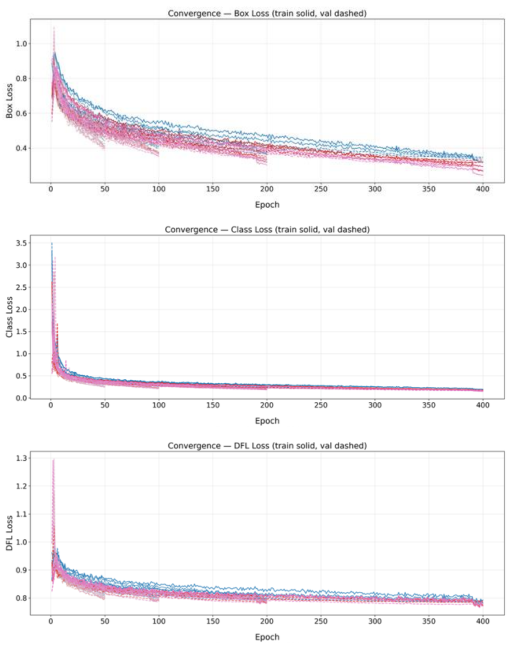

Fig. 5. Learning curve of YOLOv11 showing training and validation loss reduction over 400 epochs (Box loss, Class loss, and DFL loss).

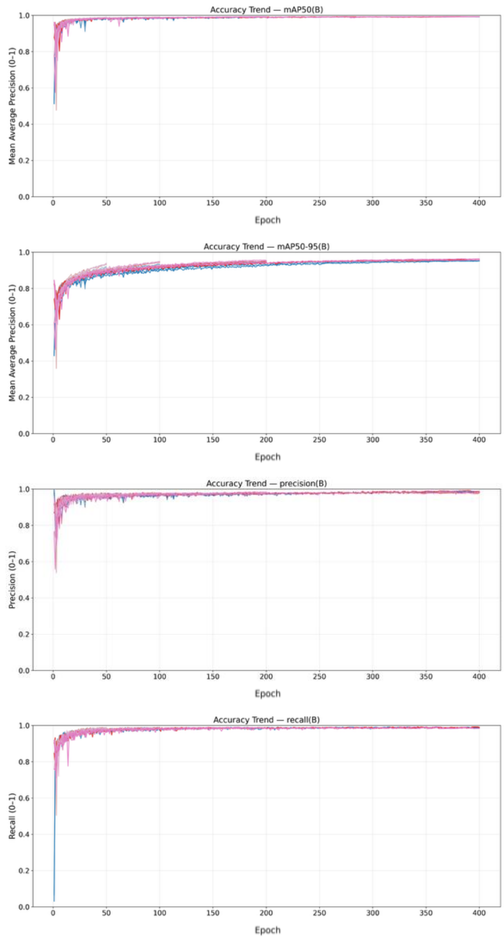

Fig. 6. Accuracy-trend curves for all YOLOv11 setups over 400 training epochs (mAP@0.5, mAP@0.5–0.95, precision, and recall).

(a) (b) (c)

Fig. 7. Validation results from the: (a) YOLOv11-m (SGD 400), (b) YOLOv11-s (SGD 200), and (c) YOLOv11-n (SGD 50).

evolution of the aggregated precision, recall, mAP@0.5, and mAP@0.5–0.95 during the 400-epoch schedule for all configurations. The metrics increase rapidly during the initial epochs, thereafter they continue to be around the saturation level with no indications of divergence or oscillation in the graphs. The reliable detection of the color-coded objects by the trained YOLOv11 models under varying visual conditions is demonstrated by representative validation results from the three key configurations shown in Fig. 7. This uniformity of the behavior indicates that the training setup has a correctly detect over the natural dataset variations in lighting, viewpoints, and partial occlusions.

During shorter training periods of 50 to 100 epochs, AdamW and SGD demonstrated comparable performance, indicating that both optimizers behaved similarly during the early convergence stage marked by a rapid decrease in loss. However, when training extended beyond 200 epochs, SGD consistently demonstrated slightly improved generalization on the Small and Medium models. The result highlights a difference in how the two methods navigate the loss landscape: AdamW quickly converges early on, frequently settling in sharper minima, while SGD's noisy updates and implicit regularization usually lead to wider, flatter minima associated with improved generalization. This behavior has been described in earlier theoretical and empirical studies although it is not very often mentioned in connection with YOLO-based object detectors. The recent findings confirms this trend, indicating that extended training durations enable a comprehensive examination of SGD's ability to generalize, which accounts for its better performance in later epochs compared to AdamW.

The training time analysis in Fig. 8 shows a very clear diminishing-return trend: from 200 to 400 epochs results in less than 1% gain in accuracy whereas total training time nearly doubled. The outcomes of this research indicate that 200 epochs provides an optimal balance between of accuracy and cost when dealing with robotic applications that are limited by GPU power. The Small-SGD-200 setting gives mAP@0.5–0.95 = 0.9500 in 5906 s, which is just 1.3% lower than the absolute best but trained four times faster.

In order to show these trade-offs in a quantitative way, the Table II gives the most representative results: the highest accuracy (Medium-SGD-400), the best balance (Small-SGD-200), and the best efficiency (Nano-SGD-50). The summary helps the readers in quickly assessing the practical trade-offs between the YOLOv11 model sizes and selecting the right configuration according to the different hardware capabilities.

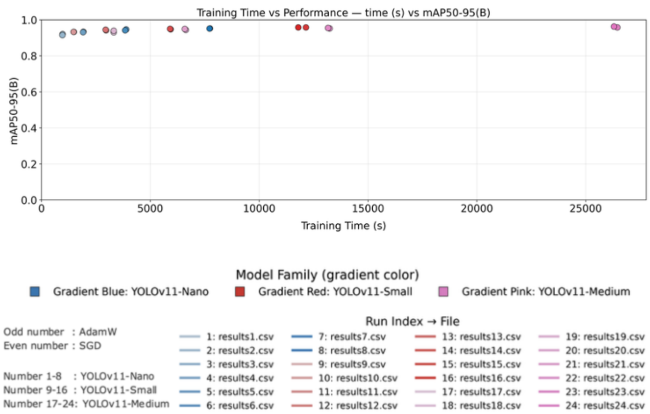

Fig. 8. The trade-off between training time and model performance for YOLOv11, showing the balance between computational cost and detection accuracy (mAP50-95).

TABLE II. SUMMARY OF KEY CONFIGURATION

Parameters	Configuration Type		
	Best Accuracy	***Best Balance***	***Best Efficiency***
Model	YOLOv11-m	YOLOv11-s	YOLOv11-n
Optimizer	SGD	SGD	SGD
Epoch	400	200	50
mAP@0.5–0.95	0.9631	0.9500	0.9159
Precision	0.9886	0.9835	0.9688
Recall	0.9838	0.9864	0.9862
Time (s)	26298	5906	963

Finally, Fig. 9 shows a frame from two stereo camera captured during qualitative validation of the system. The images show that the YOLOv11 models used in the detection process successfully identified and labelled the colored objects (blue target, red/green/yellow obstacles, and orange gripper). Even though real-time performance analysis is not included in the current scope, these demonstrations (qualitative validation) confirm the system's compatibility with the robotic control pipeline and show that the correct detection is still preserved with different rooms, lighting conditions, viewpoints, and partial occlusions.

All the results of the evaluations presented in this paper

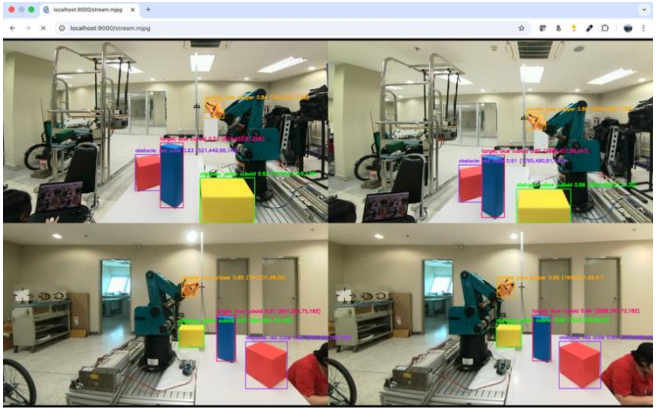

Fig. 9. Visualization of qualitative validation of the stereo camera input, illustrating detection bounding boxes and class labels.

(mAP, precision, recall) were calculated with the validation set. The study did not make use of a different test set. This research does not include per-class AP or class-specific PR curves because the main aim is to compare different YOLOv11 configurations during training rather than providing an in-depth error analysis. In future studies focusing on actual robotic deployment, a more comprehensive distribution of class-wise behavior will be performed.

IV. Conclusion

This research carries out a comprehensive assessment of YOLOv11 model scales (n, s, m), two optimizers (SGD and AdamW), and four epoch durations (50, 100, 200, and 400) for the purpose of finding the best training strategies for the detection of objects by robotic manipulation. Among the different configurations, the YOLOv11-Medium model with the SGD optimizer (400 epochs) achieves the highest accuracy (mAP@0.5-0.95 = 0.9631), whereas the YOLOv11-Small model with the SGD optimizer (200 epochs) is found to be the best option in terms of accuracy (0.9500) against training time, achieves four times the speed with just a 1.3% decrease in accuracy. Loss curves from the training data confirm that the models all smoothly converged, with clear generalization after 200 epochs. The Nano model (50 epochs) allows for rapid prototyping and it is also the case when the GPU is limited. Object detection reliability has been established through qualitative tests performed in various lighting conditions, with occlusions, and from multiple perspectives. The benchmark we proposed serves as a reproducible reference point for subsequent studies that focus on deep-learning-based perception in robotic arms. Future work will focus on extending the analysis set of class-wise performance and its real-time deployment.

Acknowledgement

The authors acknowledge the support of the United Board for Christian Higher Education in Asia and the Sanata Dharma Foundation.

References

[1] A. Vijayakumar and S. Vairavasundaram, “YOLO-based Object Detection Models: A Review and its Applications,” Multimed Tools Appl, vol. 83, no. 35, pp. 83535–83574, Mar. 2024, doi: 10.1007/s11042-024-18872-y.

[2] J. Canny, “A Computational Approach to Edge Detection,” IEEE Trans Pattern Anal Mach Intell, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.

[3] W. Rong, Z. Li, W. Zhang, and L. Sun, “An improved Canny edge detection algorithm,” in 2014 IEEE International Conference on Mechatronics and Automation, Tianjin, China: IEEE, Aug. 2014, pp. 577–582. doi: 10.1109/ICMA.2014.6885761.

[4] J. R. Sanchez-Lopez, A. Marin-Hernandez, E. R. Palacios-Hernandez, H. V. Rios-Figueroa, and L. F. Marin-Urias, “A Real-time 3D Pose Based Visual Servoing Implementation for an Autonomous Mobile Robot Manipulator,” Procedia Technology, vol. 7, pp. 416–423, 2013, doi: 10.1016/j.protcy.2013.04.052.

[5] M. L. Ali and Z. Zhang, “The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection,” Computers, vol. 13, no. 12, p. 336, Dec. 2024, doi: 10.3390/computers13120336.

[6] N. Jegham, C. Y. Koh, M. Abdelatti, and A. Hendawi, “YOLO Evolution: A Comprehensive Benchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions,” Mar. 2025, [Online]. Available: http://arxiv.org/abs/2411.00201

[7] R. Khanam and M. Hussain, “A Review of YOLOv12: Attention-Based Enhancements vs. Previous Versions,” Apr. 2025, [Online]. Available: http://arxiv.org/abs/2504.11995

[8] R. Khanam and M. Hussain, “YOLOv11: An Overview of the Key Architectural Enhancements,” Oct. 2024, [Online]. Available: http://arxiv.org/abs/2410.17725

[9] R. Sapkota et al., “YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series,” Artif Intell Rev, vol. 58, no. 9, p. 274, Jun. 2025, doi: 10.1007/s10462-025-11253-3.

[10] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, “On the importance of initialization and momentum in deep learning,” in Proceedings of the 30th International Conference on Machine Learning, 2013, pp. 1139–1147.

[11] P. Zhou, J. Feng, C. Ma, C. Xiong, S. HOI, and W. Ee, “Towards Theoretically Understanding Why SGD Generalizes Better Than ADAM in Deep Learning,” in NIPS’20: Proceedings of the 34th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada: Curran Associates Inc., Oct. 2020, pp. 21285–21296. doi: 10.48550/arXiv.2010.05627.

[12] A. M. Moraes, L. F. Pugliese, R. F. dos Santos, G. B. Vitor, R. A. da S. Braga, and F. R. da Silva, “Effectiveness of YOLO Architectures in Tree Detection: Impact of Hyperparameter Tuning and SGD, Adam, and AdamW Optimizers,” Standards, vol. 5, no. 1, p. 9, Mar. 2025, doi: 10.3390/standards5010009.

[13] R. Szabó and A. Gontean, “Robotic arm control algorithm based on stereo vision using roborealm vision,” Advances in Electrical and Computer Engineering, vol. 15, no. 2, pp. 65–74, 2015, doi: 10.4316/AECE.2015.02009.

[14] J. Wei, A. As’arry, K. Anas Md Rezali, M. Zuhri Mohamed Yusoff, H. Ma, and K. Zhang, “A Review of YOLO Algorithm and Its Applications in Autonomous Driving Object Detection,” IEEE Access, vol. 13, pp. 93688–93711, 2025, doi: 10.1109/ACCESS.2025.3573376.

[15] A. Abdi and J. H. Park, “A Hybrid AI-Based Adaptive Path Planning for Intelligent Robot Arms,” IEEE Access, vol. 11, pp. 137837–137848, 2023, doi: 10.1109/ACCESS.2023.3338566.

[16] Y. Chen, L. Chen, J. Ding, and Y. Liu, “Research on Real-Time Obstacle Avoidance Motion Planning of Industrial Robotic Arm Based on Artificial Potential Field Method in Joint Space,” Applied Sciences, vol. 13, no. 12, p. 6973, Jun. 2023, doi: 10.3390/app13126973.