Journal of Machine Engineering, 2025, Vol. 25
ISSN 1895-7595 (Print) ISSN 2391-8071 (Online)

Received*: 30 September 2025* / Accepted: *02 November 2025* / Published online: *1 December 2025*

transparent soft skin,
vision–based tactile sensing,
multi-modal sensing,
object detection

Son Tien BUI[1*], Duy Ngoc LE[2],
Tu Anh NGUYEN[2], Son Hong NGUYEN[3],
Son Anh TRAN[2], Luc Phi TRAN[1],
Thong Huy PHAM[4]

DIGITEYE: A TRANSPARENT SOFT TACTILE SENSOR FOR ROBUST MULTI-MODAL PERCEPTION

Tactile sensing remains fundamental for enabling dexterous robotic manipulation and safe human–robot interaction. Existing visuotactile sensors often compromise either deformation depth or optical transparency, limiting their ability to capture both contact forces and external scene information. This paper presents DigitEye, a transparent soft tactile sensor with a hollow box-shaped silicone rubber skin that deforms at the centimeter scale while preserving high optical clarity. A one-shot molding process with inner-frame grooves ensures robust adhesion and modular replacement of the soft skin, while dark-blue markers embedded through CNC-machined molds enable reliable tracking under varied conditions. To validate the design, we constructed two benchmark datasets: a force-sensing dataset linking images to indentation depth and ground-truth force, and an object detection dataset of fruits under varying distances and lighting. Experimental evaluations demonstrate reliable force estimation across multiple contact geometries, together with YOLO-based recognition, achieving a precision of 0.95, a recall of 0.87, and an mAP@0.5 of 0.689. These results highlight DigitEye as a practical platform for transparent visuotactile sensing, supporting both fine-grained contact perception and safer robotic operation in unstructured environments.

1. INTRODUCTION

Tactile sensing has been widely recognized as a critical capability enabling robots to perform dexterous manipulation, adaptive grasping, and safe human–robot interaction in unstructured environments. In recent years, diverse tactile sensing technologies have been developed, including capacitive and piezoresistive arrays, optical fibers, and vision-based approaches. Array-based sensors enable distributed force measurement, yet their use is

[1] Department of Science and Technology, Hanoi University of Industry, Viet Nam
[2] School of Mechanical and Automotive Engineering, Hanoi University of Industry, Viet Nam
[3] HaUI Institute of Technology, Hanoi University of Industry, Viet Nam
[4] School of Information and Communication Technology, Hanoi University of Industry, Viet Nam
[*] E-mail: sonbt@haui.edu.vn
https://doi.org/10.36897/jme/213851

restricted by issues such as wiring complexity, fragility, and limited spatial resolution. By contrast, vision-based tactile sensors are increasingly favored, as they combine high resolution with comparatively simple fabrication.

Several vision-based tactile sensors have demonstrated remarkable performance. GelSight [1] and its successors, such as GelSlim 3.0 [2], employ opaque elastomers with internal cameras to reconstruct contact geometry and forces. The open-source DIGIT sensor [3] further improved compactness and accessibility, enabling widespread adoption. However, these designs rely on opaque skins, preventing simultaneous access to external vision cues. Efforts to integrate transparency were later introduced. FingerVision [4] embedded markers into a transparent silicone layer, supporting multimodal perception but remaining vulnerable to illumination noise and color interference. More recently, Vi2TaP [5] adopted polarization-based switching between tactile and proximity sensing, but this sequential approach requires hardware complexity and prevents real-time fusion. Other attempts explored bio-inspired micropatterned skins [6, 7] or controllable-transparency links [8], focusing on grip enhancement rather than vision integration.

Despite this progress, none of the current designs combine centimeter-scale deformation, bulk optical transparency, and straightforward integration with modern vision models. In addition, benchmark datasets for transparent visuotactile sensors are still limited, which constrains reproducibility and makes comparative studies difficult.

This paper presents DigitEye, a new transparent visuotactile sensor created to address these limitations (Fig. 1). The core innovation lies in its hollow box-shaped silicone rubber skin, which enables large-scale deformation for richer shape encoding while preserving high optical clarity for simultaneous observation of external objects. A modular fabrication approach with one-shot molding secures reliable skin–frame adhesion and allows plug-and-play replacement. To evaluate its capabilities, we constructed two datasets: a force-sensing dataset linking images to precise indentation and force measurements, and an object detection dataset of fruits captured under varying conditions. Combined with a machine learning pipeline, DigitEye achieves robust force estimation and YOLO-based detection with high accuracy.

The primary contributions of this work are:

1. Design of a transparent box-shaped soft skin that combines centimeter-scale deformation with optical clarity for multimodal perception.
2. Integration of a simple architecture–transparent skin and camera–together with a machine learning pipeline for force estimation and YOLO-based object detection, without requiring complex switching mechanisms.
3. Creation of benchmark datasets for force sensing and object detection tailored to transparent tactile skins, supporting both the evaluation of DigitEye and future research on visuotactile sensing.

The remainder of this paper is structured as follows. Section II looks at earlier studies on tactile sensing and visuotactile integration. Section III explains how DigitEye was designed and made. Section IV shows the vision-based multimodal sensing framework, and Section IV describes the data collection experiments, the object detection and force sensing multi-modal. Section V presents the results and discussion. Section VI concludes the paper, and Section VII offers suggestions for future work.

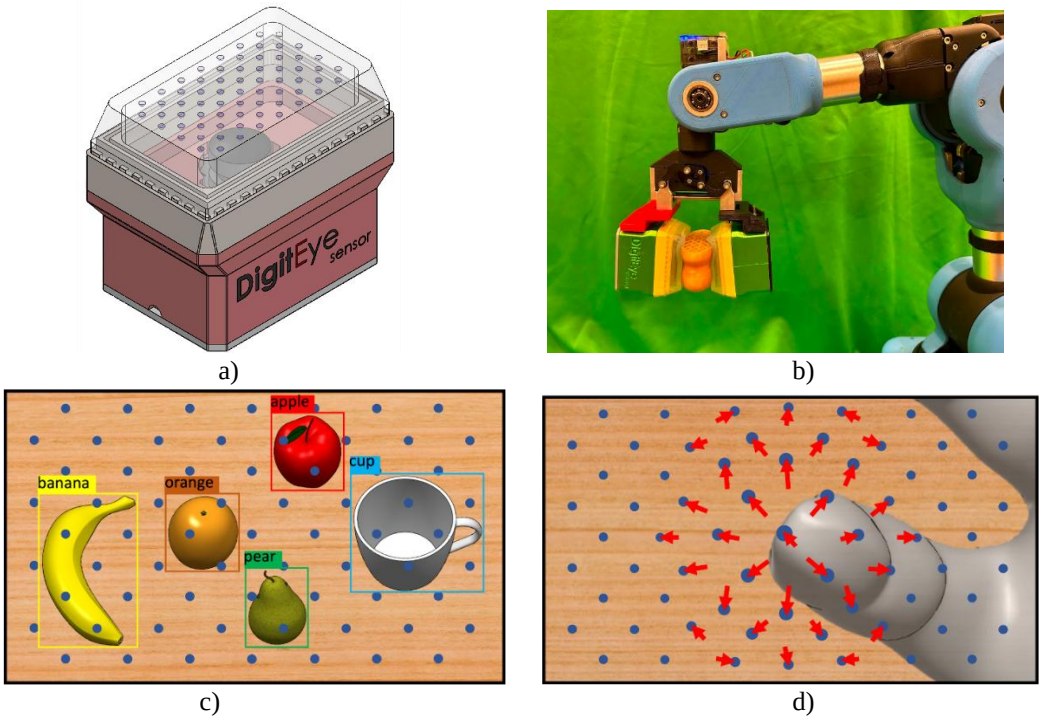


Fig. 1. Overview of the proposed DigitEye tactile sensor: a) The DigitEye prototype with a hollow box-shaped transparent soft skin; b) Integration on a robotic gripper performing a tabletop grasping task; c) Vision-based object detection enabled through the transparent skin; d) Force sensing capability using marker deformation and learning-based estimation

2. RELATED WORKS

Tactile sensing in robotics has been studied using different methods such as resistive and capacitive arrays, optical fibers, and vision-based systems. Among these, vision-based tactile sensors are drawing more interest because they can provide high-resolution contact data while using fairly simple hardware.

Opaque visuotactile sensors were among the earliest successful designs. GelSight [1] pioneered the use of an opaque elastomer with embedded markers and an internal camera to reconstruct contact geometry and force distribution. Later iterations, such as GelSlim 3.0 [2] improved compactness and introduced slip detection, while the open-source DIGIT [3] enabled wide adoption in robotic manipulation research. However, these designs are constrained to internal deformation sensing since their opaque skins block external vision.

To overcome this limitation, transparent skins were used. FingerVision [4] placed markers inside a clear silicone layer, allowing both deformations tracking and outside observation at the same time. However, it was still affected by light noise and colour interference. Vi2TaP [5] employed cross-polarization to switch between tactile and proximity sensing; while effective at separating modalities, this sequential approach prevented true real-time fusion. Other efforts focused on bio-inspired surfaces, such as hexagonal micropatterns [6] and torrent-frog-inspired adhesives [7], which enhanced friction under lubrication but did not provide transparency. A controllable-transparency robotic link [8] further demonstrated integration of optical modulation into soft robotics, but without a dedicated tactile skin.

More compact visuotactile solutions have recently emerged. Shimonomura and Nakashima [9] combined tactile and proximity sensing using a compound-eye camera, offering early demonstrations of multimodality. Zhang et al. [10] proposed an improved thin-format tactile skin equipped with a focus-adjustable imaging system, enabling robust performance under varying contact depths. Li and Peng [11] introduced a monocular visual–tactile fingertip for robust manipulation, emphasizing efficiency and integration. Chen et al. [12] developed a thin-format tactile sensor with a microlens array, achieving high spatial resolution in a miniaturized structure. Similarly, Duong [13] introduced BiTac, a soft vision-based tactile sensor capable of bidirectional force perception, further confirming the trend toward high-resolution, learning-compatible visuotactile designs. Collectively, these studies highlight an evolution toward thin, robust, and multimodal tactile sensors.

At the same time, object detection methods have grown quickly, becoming the main perception backbone for visuotactile fusion. Recent surveys of detectors [14] show the move from two-stage to single-stage pipelines, pointing out the balance between speed and accuracy. Transformer-based models have further enhanced multi-scale feature aggregation, especially for recognizing small objects [15]. Extensions such as ARS-DETR [16] introduced aspect-ratio-sensitive labels and rotated deformable attention to handle oriented targets, while reviews of DETR variants [17] summarized advances in query design and convergence speed. Other work proposed disentangling positional and content information in transformers to achieve higher accuracy with multi-task training [18]. YOLO-derived architectures remain widely used: ESF-YOLO [19] enhanced performance with cross-scale feature fusion and attention modules, while application-driven studies integrated YOLOv5 detection with impedance control for safe human–robot collaboration [20].

Tactile sensing research has moved toward more complex, learning-based methods. DenseTact 2.0 [21] applied an optical fingertip and deep learning to rebuild contact geometry and measure six-axis force/torque, showing good generalization. GTac [22] used a biomimetic two-layer structure with piezoresistive and Hall sensors to detect both normal and shear forces. Vision-based Tac3D [23] relied on binocular imaging to estimate force distribution and friction. Other material innovations include hemispherical protrusion arrays for 3D vector force sensing with angular resolution below 15° [24], multilayered skins for repeatable grasping and >95% classification accuracy [25], and scalable 3D magnetic tactile sensors for grasping and slip detection [26]. Energy-autonomous tactile devices have been reported, including self-powered multidimensional sensors [27]. Additional approaches have leveraged piezoresistive thin films [28] and electrical impedance tomography (EIT) [29] to achieve large-area tactile skins with distributed force mapping.

3. DESIGN AND FABRICATION

3.1. DESIGN

The DigitEye sensor is conceived as a small visuotactile module that combines a transparent soft skin, a rigid frame, and a built-in fisheye camera in one unit (Fig. 2). Its

purpose is to measure deformations at the centimeter scale while keeping the skin highly transparent, so it can capture tactile and visual data at the same time.

The sensor consists of six main components: (1) a hollow box-shaped transparent silicone rubber skin embedded with an array of dark-blue markers, which deform proportionally to external contact forces; (2) a rigid frame with inner grooves that interlock with the silicone during molding, ensuring strong adhesion and enabling modular replacement of the skin; (3) a housing that provides structural support and alignment; (4) a wide-angle fisheye USB camera (ELP, 150° field of view, 1080p@30 fps, 720p@60 fps, 480p@100 fps) positioned beneath the soft skin to capture both marker displacements and external objects; (5) a screw system that secures the optical unit to the housing; and (6) a protective cover that seals and protects the assembly.

This modular design has three main benefits. First, the hollow box-shaped soft skin can bend on the scale of centimeters, which helps the sensor pick up detailed geometric features of the objects it touches. Second, the silicone rubber's high optical transparency gives a clear path for light, so the camera can detect hand-sized objects from as far as 2 meters away—an important factor for safe robotic operation. Third, the frame–skin interlocking design produced via one-shot molding facilitates plug-and-play replacement of the skin in case of scratches or loss of clarity, thereby improving maintainability and extending the sensor's lifetime.

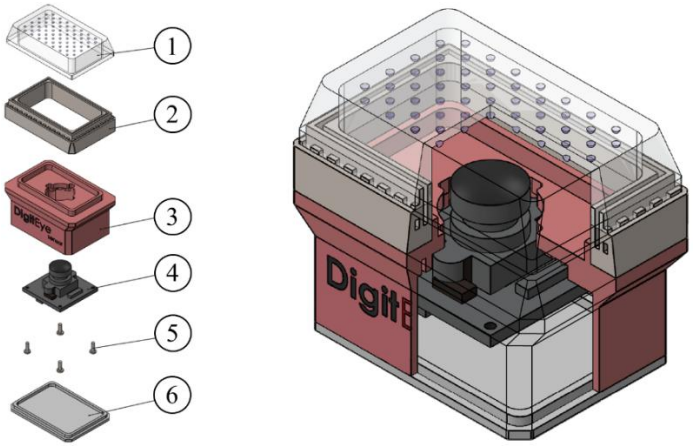


Fig. 2. Exploded view and assembled DigitEye sensor: Left - Exploded view showing the six main components: (1) transparent soft skin with embedded markers; (2) rigid frame with grooves for anchoring; (3) housing; (4) fisheye camera module; (5) screw system; and (6) protective cover; Right - Sectional perspective view of the complete assembly, illustrating the external appearance of DigitEye and the internal arrangement of the optical module beneath the transparent soft skin

As a whole, DigitEye features a simple yet effective design that unites transparency, deformability, and modularity in one compact unit, providing the foundation for multimodal visuotactile sensing.

3.2. FABRICATION

The fabrication process of DigitEye is shown in Fig. 3. The rigid parts, frame, housing, and cover, along with the casting molds, were first created in CAD software (Fusion 360) and

then produced using fused deposition 3D printing with PLA. For the mold core, a mica plate was CNC-milled to obtain a flat, smooth surface and patterned with circular recesses (1 mm ball-end cutter) to keep the geometry consistent.

Markers were manually embedded by filling these recesses with transparent silicone mixed with dark blue dye. After curing, the mold was assembled with the frame, and vacuum-degassed transparent silicone rubber (Zoukei-Mura Co. Ltd., Japan, low-durometer < 40 Shore A) was cast into the cavity. Grooves in the frame allowed the silicone to flow in and interlock, achieving both strong adhesion and modularity of the skin. After demolding, the transparent box-shaped skin with embedded markers was obtained.

Finally, the optical module, consisting of a wide-angle fisheye USB camera (ELP, 150° FOV, 1080p@30 fps, 720p@60 fps, 480p@100 fps), was mounted into the housing, secured with screws, and enclosed with the protective cover. The completed prototype is shown in Fig. 3.

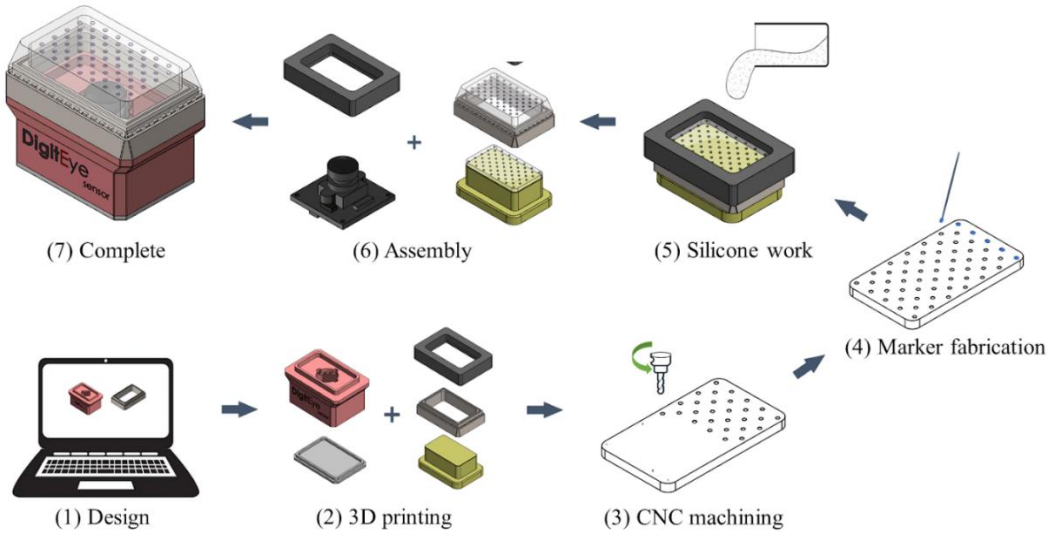
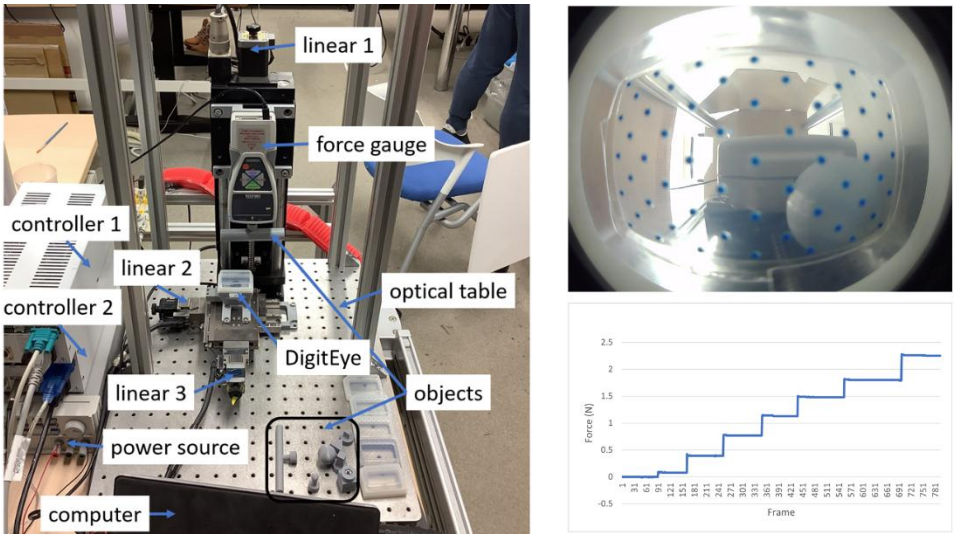


Fig. 3. The fabrication process of DigitEye sensor: (1) CAD design of rigid parts and molds. (2) 3D printing of frame, housing, and mold structures. (3) CNC machining of the mica mold core with patterned recesses. (4) Marker fabrication by embedding dyed silicone into recesses. (5) Casting of vacuum-degassed transparent silicone to form the hollow soft skin. (6) Assembly of the optical module and rigid components. (7) Completed DigitEye prototype

4. VISION-BASED MULTI-MODAL SENSING

4.1. DATA COLLECTING EXPERIMENT

Force Sensing Dataset. The experimental setup for collecting force-sensing data is shown in Fig. 4, where DigitEye was mounted on a two-axis XY stage for lateral positioning of contact points, while a vertical Z-axis carriage carried a digital force gauge (IMADA ZTA-5N, IMADA Co., Ltd., Japan). The gauge indented the soft skin downward in increments of 0.1 mm, up to a maximum depth of 20 mm. At each step, the embedded fisheye camera of DigitEye captured images of marker displacements, synchronized with the force signals measured by the IMADA sensor.



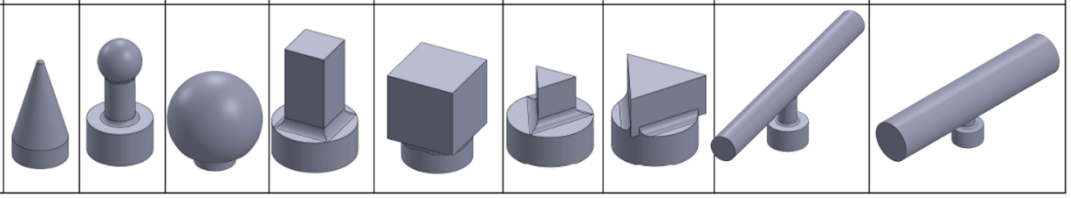
a) b)

Fig. 4. Data collection experiment for force model training: a) Main components of DigitEye data collection experiment with different types of objects: sphere head, triangle head, needle head, cylinder shape; b) Image data recorded by DigitEye (top) and force data recorded by IMADA force gauge (bottom)

To ensure comprehensive coverage, indentations were made on a 39-point grid spread over the sensing surface. Three indenter shapes were used, spherical, cylindrical, and triangular prism tips. This setup generated varied deformation patterns for training. Two ways of collecting data were applied. In the static mode, the indenter stopped for 0.5 s at each point so that steady images and force values could be recorded. In the dynamic mode, indentation and release happened continuously, letting the dataset include both stable contact states and short-term deformation changes.

During each trial, indenters with different geometries (Table 1) were used. The resulting dataset, therefore, consists of synchronized triplets: (i) image of the soft skin deformation, (ii) ground-truth indentation depth, and (iii) measured contact force. With repetitions under varied lighting conditions, the dataset contains several thousand labeled samples, providing a robust basis for training the proposed force sensing model.

Table 1. Table of test heads for data collection experiment

No	1	2	3	4	5	6	7	8	9
Shape	Needle	Sphere	Sphere	Rectangular	Rectangular	Triangle	Triangle	Cylinder	Cylinder
Size	-	d10	d50	10x10	20x20	10	20	d10	d20
Hardness	Rigid	Rigid	Rigid	Rigid	Rigid	Rigid	Rigid	Rigid	Rigid
Image									

Object detection dataset. To evaluate the transparency of DigitEye for external perception, three representative fruits—orange, banana, and apple–were placed at distances

of 20–50 cm. Images were collected under varied illumination and background conditions. Each image was annotated with bounding boxes in YOLO format. The dataset covers several thousand annotated samples, supporting the training and evaluation of the YOLO-based object detection model.

4.2. OBJECT DETECTION AND FORCE SENSING MULTI-MODAL

To realize robust multimodal perception, DigitEye integrates two complementary learning-based models: (i) an object detection model for external vision through the transparent tactile skin, and (ii) a force sensing model for estimating contact forces from marker deformation patterns. Although trained on separate datasets, both models operate jointly in a unified pipeline (Fig. 6), enabling simultaneous visual recognition and tactile force estimation in real time.

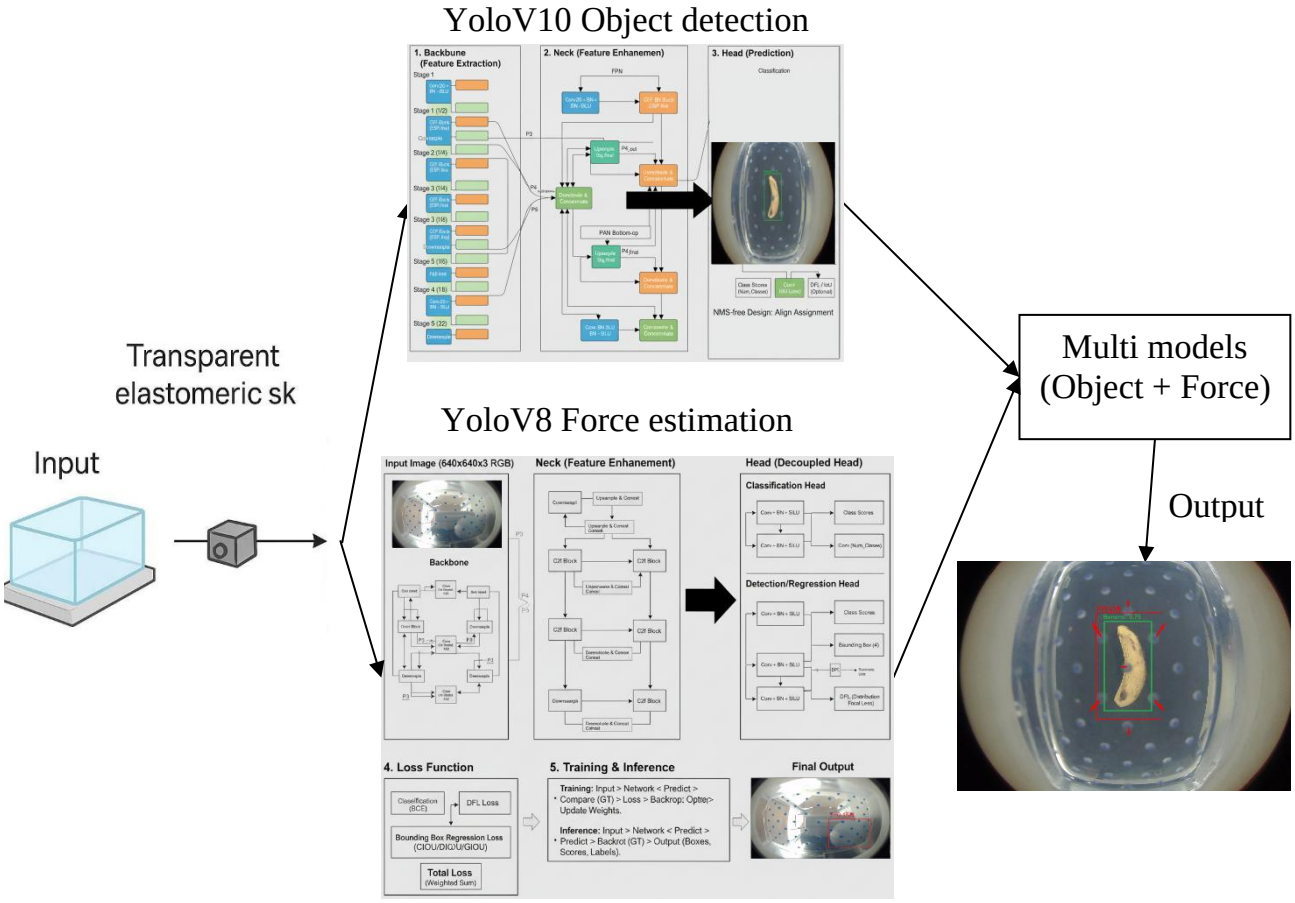


Fig. 6. DigitEye sensing multi-modal: Pipeline of the YOLO models applied for fruit recognition and force detection through the transparent skin

For vision-based recognition, the YOLOv10 framework was used because it offers a good balance between inference speed and detection accuracy [30]. This makes it suitable for

real-time robotic manipulation. The dataset included images of fruits (banana, orange, and apple) placed 20–50 cm from the sensor. To test the robustness of through-skin vision, the data were collected with different backgrounds and lighting conditions. To further enhance generalization, augmentation techniques were applied, including random brightness adjustment, color jittering, and Gaussian noise injection to mimic distortions caused by light scattering and reflections through the transparent elastomer. During inference, YOLOv10 produced bounding boxes, class labels, and confidence scores for detected objects. This data-driven approach aligns with recent advances in industrial visual inspection using YOLO-based anomaly detection pipelines [31], highlighting its adaptability to real-world environments. Experimental results confirmed that the model maintained stable accuracy despite transparency-induced distortions, achieving high mean average precision (mAP) with low latency. These outcomes verify that the transparent skin provides a viable optical pathway for reliable external object recognition, thereby extending the perceptual capability of tactile sensors to the surrounding environment.

We built a supervised learning pipeline with YOLOv8 to estimate contact forces from tactile deformation [32]. The input comes from a fisheye camera inside the sensor, which records how dark-blue markers move within the transparent silicone skin. Because the raw images also show background clutter and lens distortion, we applied a multi-stage preprocessing pipeline to clean them. This included fisheye distortion correction, Gaussian blurring, adaptive thresholding, and morphological filtering to suppress background noise and enhance marker contrast. Connected component analysis was subsequently applied to identify valid marker candidates, resulting in binarized deformation maps suitable for model training (Fig. 6).

Originally designed for object detection, the YOLOv8 network was adapted to perform two functions: identifying and locating markers, and predicting contact force magnitudes and directions from marker displacement patterns. A regression head was attached to the architecture, and training was carried out using a combined loss that integrated detection loss (bounding box localization and classification) with regression loss (mean squared error relative to the ground-truth force labels). Ground-truth annotations were obtained from a calibrated IMADA ZTA-5N force sensor during controlled indentation experiments. Training on several thousand image–force pairs produced a compact model capable of estimating contact forces directly from unseen deformation images with high robustness against noise, lighting variation, and transparency artefacts.

5. RESULT AND DISCUSSION

5.1. DESIGN AND FABRICATION

The DigitEye sensor was built successfully following the proposed design. Its hollow, box-shaped silicone rubber skin bent by several centimetres when touched, as planned. Tests using spherical, cylindrical, and triangular prism indenters (Fig. 5) showed clearly different deformation patterns for each shape, suggesting that DigitEye could be used for shape

recognition based on contact. Moreover, because the silicone rubber skin is soft, the sensor could handle fragile items such as ripe fruits with thin skins, making it suitable for delicate manipulation.

The transparency of the Zoukei-Mura silicone rubber was preserved throughout the fabrication process. The embedded fisheye camera was able to observe hand-sized objects at distances up to 2 m through the transparent skin, a capability that is highly relevant for enhancing safety when DigitEye is mounted on robotic grippers by enabling early detection of nearby objects and humans.

The one-shot molding technique, combined with the inner grooves of the frame, ensured robust adhesion between the skin and the rigid structure. At the same time, this design modularized the skin as a replaceable unit, making it possible to easily exchange the skin in case of scratches or reduced optical clarity. This plug-and-play maintainability distinguishes DigitEye from earlier visuotactile skins that are permanently bonded to rigid substrates.

The fabrication results confirmed that surface quality is important. The CNC-milled mica mold core created flat, smooth surfaces that reduced light scattering and helped keep the optical clarity high. Tests with embedded markers in different colors (black, white, red, and dark blue) showed that dark blue gave the most consistent detection across different lighting and background settings. Consequently, dark blue markers were adopted in the final prototype. The resulting transparent box-shaped skin, combined with uniform marker visibility and strong frame adhesion, provides a robust foundation for subsequent force sensing and object detection experiments.

5.2. VISION-BASED SENSING MULTI-MODAL

The DigitEye sensor was tested for two functions: recognizing external objects and estimating internal forces, both supported by vision-based machine learning models. In this setup, YOLOv10 handles object detection while YOLOv8 predicts forces, creating a single framework that takes advantage of the soft skin’s transparency and the use of embedded deformation markers.

YOLOv10-Based Object Detection. The object detection experiments were designed to validate the feasibility of vision through the transparent silicone skin. In [33], three representative fruits—banana, orange, and apple—were used as test objects due to their varied shapes, textures, and reflective properties, which introduce different levels of optical distortion when viewed through the elastomer. Despite the presence of embedded markers and occasional reflections from the skin surface, YOLOv10 successfully detected and classified objects with high consistency.

The model stayed robust under different lighting conditions, both natural and artificial, and also handled cluttered backgrounds well. There was a slight drop in confidence scores when marker shadows overlapped with object edges, but overall detection accuracy was steady. The fact that the model kept a high mean average precision (mAP) with low inference latency shows it can work in real-time robotic tasks. This result also confirms that the transparent tactile skin does not seriously affect visual object recognition, supporting the multimodal design idea of DigitEye.

YOLOv8-Based Force Estimation. The force-sensing experiments were conducted using deformation images paired with ground-truth labels acquired from the IMADA ZTA-5N force gauge. The YOLOv8-based pipeline was trained to detect marker positions and regress contact force magnitudes and directions from marker displacement patterns.

In Fig. 7, training over 200 epochs demonstrated stable convergence across all loss components, including bounding box regression, classification, and distribution focal losses. Both training and validation losses approached low values, indicating good generalization to unseen samples. As shown in the evaluation metrics, the model achieved a precision of approximately 0.95 and a recall of 0.87. The mean Average Precision at IoU 0.5 (mAP@0.5) reached 0.689, while the more stringent mAP@0.5–0.95 exceeded 0.60. These results highlight that the model not only detects relevant features reliably but also maintains predictive accuracy across varying force levels and contact geometries.

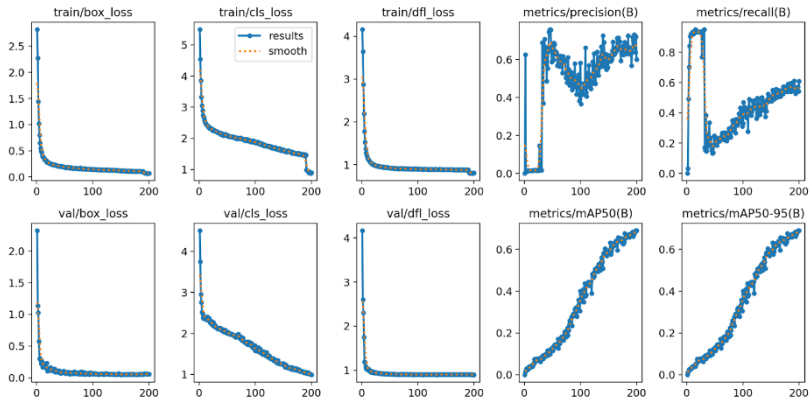


Fig. 7. Results for training and testing steps for Yolov8 force prediction

From Fig. 8, visualization of the F1-confidence and precision–recall curves revealed a balanced operating point around a confidence threshold of 0.37, ensuring both high sensitivity and precision. Confusion matrix analysis further confirmed that most predictions aligned with the true force categories, although limited confusion remained for underrepresented classes. Quantitative evaluation followed a multi-criteria perspective similar to that used in decision-making frameworks such as TOPSIS [34], allowing balanced assessment of precision, recall, and F1 trade-offs. This suggests that dataset balancing could further improve performance, particularly for rare contact scenarios.

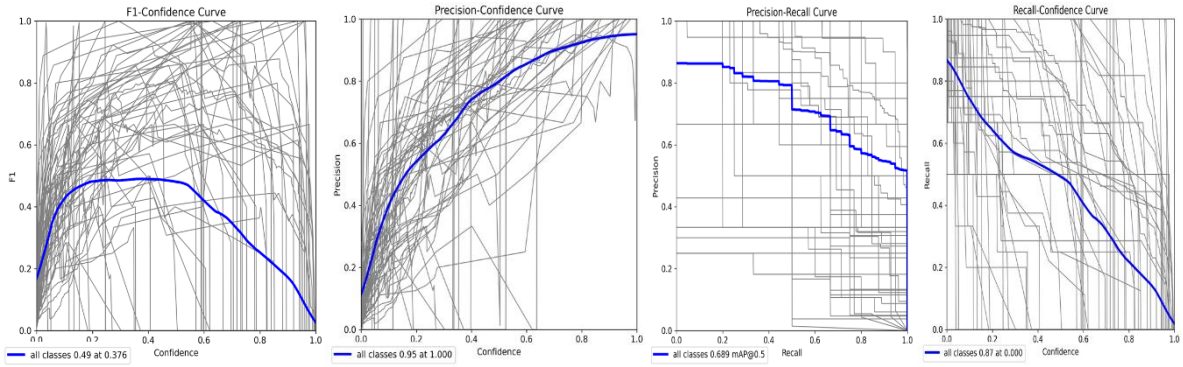


Fig. 8. Results for training and testing with Yolov8 force prediction in the relationship between precision, recall, F1 score, and confidence index

Integration of Object Detection and Force Estimation. The combined results demonstrate that DigitEye can concurrently recognize external objects and estimate applied forces in real time. This integration is particularly valuable for robotic manipulation tasks: the object detection module informs the robot about the nature and approximate position of the object, while the force sensing module provides continuous feedback about contact stability and safety. Such a dual-modality framework allows for adaptive control strategies, such as adjusting grip strength based on the detected object type or releasing fragile items before damage occurs.

The experimental results confirm that DigitEye's transparent visuotactile architecture enables reliable object recognition and force estimation in a unified framework. The YOLOv10-based object detection model sustained high accuracy despite transparency-related artefacts such as light scattering, marker interference, and background clutter, which typically degrade vision performance in transparent media. Similarly, the YOLOv8-based force sensing model demonstrated strong predictive capability across diverse contact scenarios, showing that marker displacement patterns provide rich cues for learning-based force estimation. Nonetheless, several challenges remain. The detection accuracy for objects with colors or textures similar to the embedded markers showed slight degradation, and the force model exhibited reduced precision in classes with fewer training samples, reflecting dataset imbalance. Moreover, the evaluation was limited to relatively simple test objects and controlled contact conditions; real-world manipulation tasks often involve irregular geometries, dynamic motions, and occlusions that could further stress the system. Despite these limitations, the results underscore the potential of DigitEye as a compact multimodal sensor capable of supporting dexterous robotic interaction. Future improvements may include expanding the dataset with more diverse object categories, refining preprocessing algorithms to enhance marker visibility under challenging lighting, and integrating temporal modelling to capture dynamic force evolution during manipulation.

The experimental results show that DigitEye's vision-based multimodal framework is both practical and effective. Using the YOLOv10 object detection model, the study confirmed that the transparent skin allows reliable visual recognition. At the same time, the YOLOv8 force sensing model proved highly accurate in predicting tactile force. Taken together, these findings support the proposed design as a strong option for safe, flexible, and adaptive robotic manipulation in unstructured settings.

6. CONCLUSION

DigitEye utilizes a hollow, box-shaped silicone rubber skin that enables centimeter-scale deformation, allowing the sensor to capture rich contact information while supporting the gentle manipulation of fragile objects. The transparent skin keeps its optical clarity, allowing vision-based detection of hand-sized objects up to 2 m away, which helps improve robotic safety. The frame, designed with inner grooves, provides strong adhesion during one-shot molding and makes the skin a replaceable unit, which improves both maintenance and usability. The fabrication process consistently produced a transparent silicone rubber skin with strong bonding and clear marker visibility, forming a reliable base for multimodal

visuotactile sensing. In addition, the integration of YOLO-based models demonstrated effective performance in both object recognition and force prediction. The detection pipeline achieved high precision and recall, while the force estimation model accurately captured contact dynamics, confirming the capability of DigitEye to support reliable grasping and holding tasks such as the ROSE mechanism [35]. Overall, the proposed sensor establishes a simple yet powerful platform for advancing multimodal tactile–visual perception in robotic manipulation.

7. FUTURE WORKS

Future work will aim to improve both the design and the making of DigitEye. Possible directions are to adjust the geometry of the soft skin to make it last longer, test other transparent silicone mixes to get better optical clarity, and create simpler molding methods for easier large-scale production. On the sensing side, keeping the markers visible is still difficult when objects have similar colors or when lighting is very strong, so image restoration methods will be explored. Furthermore, accurate force estimation under complex contact conditions, such as objects with multiple angles or distributed contact points, will be studied to improve the reliability of interaction modelling. These efforts aim to expand the applicability of DigitEye in more diverse and demanding robotic manipulation scenarios.

DECLERATION

AI use: The authors declare that artificial intelligence (AI) tools were employed solely to assist in smoothing the language and minimizing spelling or grammatical errors during manuscript preparation. The use of AI was limited to linguistic refinement; all ideas, designs, analyses, results, and conclusions presented in this paper are entirely the authors' own work.

Open-source: All datasets, trained models, and design files related to DigitEye will be released as open-source on a public GitHub repository once this paper is accepted for publication, to support transparency and facilitate future research on transparent soft-skin tactile sensors.

ACKNOWLEDGEMENTS

This research was funded by Hanoi University of Industry under Project No. 06-2024-RD/HD-ĐHCN. The authors would also like to thank Soft Haptics Laboratory, especially Prof. Ho Anh Van, Japan Advanced Institute of Science and Technology (JAIST), for their valuable support in this research, and Mr. Nguyen Van Thoan for his technical assistance during the development of DigitEye.

REFERENCES

[1] YUAN W., DONG S., ADELSON E.H., 2017, *Gelsight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force*, Sensors, 17/12, 2762.

[2] TAYLOR I.H., DONG S., RODRIGUEZ A., 2022, *Gelslim 3.0: High-Resolution Measurement of Shape, Force and Slip in a Compact Tactile-Sensing Finger*, Proc. IEEE Int. Conf. Robot. Autom., (ICRA), 10781–10787.

[3] LAMBETA M., CHOU P.W., TIAN S., YANG B., MALOON B., MOST V.R., CALANDRA R., 2020, *Digit: A Novel Design for a Low-Cost Compact High-Resolution Tactile Sensor with Application to In-Hand Manipulation*, IEEE Robot. Autom. Lett., 5/3, 3838–3845.

[4] YAMAGUCHI A., ATKESON C.G., 2016, *Combining Finger Vision and Optical Tactile Sensing: Reducing and Handling Errors While Cutting Vegetables*, Proc. IEEE-RAS Int. Conf. Humanoid Robots, Cancun, Mexico, 1045–1051.
[5] NGUYEN N.H., LE N.M.D., LUU Q.K., NGUYEN T.T., HO V.A., 2025, *Vi2TaP: a Cross-Polarization Based Mechanism for Perception Transition in Tactile–Proximity Sensing with Applications to Soft Grippers*, IEEE Robot. Autom. Lett., 6288–6295.
[6] VARENBERG M., GORB S.N., 2009, *Hexagonal Surface Micropattern for Dry and Wet Friction*, Adv. Mater., 21/4, 483–486.
[7] ITURRI J., XUE L., KAPPL M., GARCIA-FERNANDEZ L., BARNES W.J.P., BUTT H.J., DEL CAMPO A., 2015, *Torrent Frog-Inspired Adhesives: Attachment to Flooded Surfaces*, Adv. Funct. Mater., 25/10, 1499–1505.
[8] LUU Q.K., NGUYEN D.Q., NGUYEN N.H., HO V.A., 2023, *Soft Robotic Link with Controllable Transparency for Vision-Based Tactile and Proximity Sensing*, Proc. IEEE Int. Conf. Soft Robot. (RoboSoft).
[9] SHIMONOMURA K., NAKASHIMA H., 2013, *A Combined Tactile and Proximity Sensing Employing a Compound-Eye Camera*, Proc. IEEE Sensors, Baltimore, MD, USA
[10] ZHANG S., SUN Y., LIU N., SUN F., YANG Y., FANG B., 2024, *An Improved Vision-Based Tactile Skin with Imaging Adjustment System to Reduce Defocusing Caused by Contact Depth Changes*, Sens. Actuators A: Phys., 374, 115495.
[11] LI R., PENG B., 2022, *Implementing Monocular Visual–Tactile Sensors for Robust Manipulation*, Cyborg and Bionic Syst., Article ID 9797562, 1–10.
[12] CHEN X., ZHANG G., WANG M.Y., YU H., 2022, *A Thin-Format Vision-Based Tactile Sensor with a Microlens Array (MLA)*, IEEE Sens. J., 22/22, 22069–22076.
[13] DUONG L.V., 2023, *Bitac: A Soft Vision-Based Tactile Sensor with Bidirectional Force Perception for Robots*, IEEE Sens. J., 23/9, 9158–9167, https://doi.org/10.1109/JSEN.2023.3257645.
[14] JIAO L., ZHANG F., LIU F., YANG S., LI L., FENG Z., QU R., 2019, *A Survey of Deep Learning-Based Object Detection*, IEEE Access, 7, 128837–128868.
[15] REKAVANDI A.M., RASHIDI S., BOUSSAID F., HOEFS S., AKBAS E., BENNAMOUN M., 2025, *Transformers in Small Object Detection: A Benchmark and Survey of State-of-the-Art*, ACM Comput. Surv., 58/3, 1–33.
[16] ZENG Y., CHEN Y., YANG X., LI Q., YAN J., 2024, *ARS-DETR: Aspect Ratio-Sensitive Detection Transformer for Aerial Oriented Object Detection*, IEEE Trans. Geosci. Remote Sens., 62, 1–15.
[17] SHEHZADI T., HASHMI K.A., STRICKER D., AFZAL M.Z., 2023, *Object Detection with Transformers: A Review*, arXiv preprint, arXiv:2306.04670.
[18] WANG Y., HA J.E., 2024, *Improved Object Detection with Content and Position Separation in Transformer*, Remote Sens., 16/2, 353.
[19] TAO W., WANG X., YAN T., LIU Z., WAN S., 2024, *ESF-YOLO: An Accurate and Universal Object Detector Based on Neural Networks*, Front. Neurosci., 18, 1371418.
[20] LUU Q.K., NGUYEN D.Q., NGUYEN N.H., DAM N.P., HO V.A., 2025, *Vision-Based Proximity and Tactile Sensing for Robot Arms: Design, Perception, and Control*, IEEE Trans. Robot., early access, 1–12.
[21] DO W.K., JUREWICZ B., KENNEDY III M., 2022, *Densetact 2.0: Optical Tactile Sensor for Shape and Force Reconstruction*, arXiv preprint, arXiv:2209.10122.
[22] LU Z., GAO X., YU H., 2022, *Gtac: A Biomimetic Tactile Sensor with Skin-Like Heterogeneous Force Feedback for Robots*, IEEE Sens. J., 22/14, 14491–14500.
[23] ZHANG L., WANG Y., JIANG Y., 2022, *Tac3D: A Novel Vision-Based Tactile Sensor for Measuring Forces Distribution and Estimating Friction Coefficient Distribution*, arXiv preprint, arXiv:2202.06211.
[24] HAN C., CAO Z., HU Y., ZHANG Z., LI C., WANG Z.L., WU Z., 2024, *Flexible Tactile Sensors for 3D Force Detection*, Nano Lett., 24/17, 5277–5283.
[25] SHI Y., LÜ X., WANG W., ZHOU X., ZHU W., 2024, *A High-Repeatability Three-Dimensional Force Tactile Sensing System for Robotic Dexterous Grasping and Object Recognition*, Micromachines, 15/12, 1513.
[26] LE SIGNOR T., DUPRE N., DIDDEN J., LOMAKIN E., CLOSE G., 2023, *Mass-Manufacturable 3D Magnetic Force Sensor for Robotic Grasping and Slip Detection*, Sensors, 23/6, 3031.
[27] ZHANG W., XI Y., WANG E., QU X., YANG Y., FAN Y., LI Z., 2022, *Self-Powered Force Sensors for Multidimensional Tactile Sensing*, ACS Appl. Mater. Interfaces, 14/17, 20122–20131.
[28] LIU Y., HAN H., MO Y., WANG X., LI H., ZHANG J., 2022, *A Flexible Tactile Sensor Based on Piezoresistive Thin Film for 3D Force Detection*, Rev. Sci. Instrum., 93/8.
[29] CHEN H., YANG X., WANG P., GENG J., MA G., WANG X., 2022, *A Large-Area Flexible Tactile Sensor for Multi-Touch and Force Detection Using Electrical Impedance Tomography*, IEEE Sens. J., 22/7, 7119–7129.
[30] WANG A., et al., 2024, *Yolov10: Real-Time End-to-End Object Detection*, Adv. Neural Inf. Process. Syst., 37, 107984–108011.

[31] NGUYEN D.H., TRONG H.D., NGUYEN H.L.P., NGUYEN Q.K., TRAN D.T., BUI T.S., NGUYEN V.T., 2024, *A Solution for Anomaly Detection of Red Beans in a Product Processing Line*, Proc. Asia Pacific Signal and Information Processing Association Annual Summit and Conf. (APSIPA ASC), Macao, 1–5, https://doi.org/10.1109/APSIPAASC63619.2025.10849036.

[32] SAFALDIN M., ZAGHDEN N., MEJDOUB M., 2024, *An Improved Yolov8 to Detect Moving Objects*, IEEE Access, 12, 59782–59806.

[33] COCO data set, *https://cocodataset.org/#home*.

[34] TRUNG D.D., 2021, *Application of TOPSIS and PIV Methods for Multi-Criteria Decision Making in Hard Turning Process*, J. Mach. Eng., 21/4, 57–71, https://doi.org/10.36897/jme/142599.

[35] BUI S., KAWANO S., HO V.A., 2023, R*ose: Rotation-Based Squeezing Robotic Gripper Toward Universal Handling of Objects*, Proc. Robotics: Science and Systems XIX (RSS2023), Daegu, South Korea, 1–9, https://doi.org/10.15607/rss.2023.xix.090.