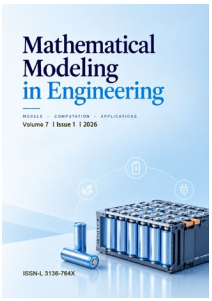
UNIVERSITY OF WEST ATTICA

eISSN: 3136-764X

**Mathematical Modeling in Engineering**

Vol. 7, No. 3, July, 2026




# Perceiver IO Model for Visual Servoing of Reflective Parts Based on Polarization Camera Feedback

Penta Hirano[a], Venta Kawai[a,*], Asuka Miyazaki[a]

[a] *Graduate School of Information Technology, Saitama University, Saitama, 338-8570, Japan*

Received 27 May 2026; Revised 20 July 2026; Accepted 25 August 2026; Published 30 August 2026

**Abstract**

Reflective surfaces challenge industrial visual servoing because specular highlights can obscure true edges, distort pose cues and destabilize closed-loop robot correction. This paper introduces a Perceiver IO model for visual servoing of reflective parts based on polarization camera feedback. The method constructs polarization intensity difference, angle-sensitive response, reflection confidence and valid-edge masks, then combines these visual cues with servo-error vectors and robot-state tags. Perceiver IO maps heterogeneous image and state tokens into latent representations for pose correction, confidence estimation and review-priority output. A reflection-aware confidence branch is added to reduce unstable control updates when highlights, saturation or partial occlusion weaken geometric evidence. A synthetic dataset of 6,055 servo window assemblies is constructed under weak reflection, medium reflection, strong reflection, partial occlusion and mixed glare. The model achieves a Servo Success Rate of 94.8%, a mean final Pose Error of 0.62 mm and an average convergence time of 1.04 s. Compared with the compact convolutional servo baseline, final pose error under strong reflection is reduced by 28.6%, and the unstable update rate decreases by 16.4%. These results indicate that polarization feedback, when fused with robot-state information through latent attention, improves reflective-part servo reliability while retaining traceable pose correction, confidence and review information for safe robot assembly.

*Keywords:* Polarization Camera; Reflective Part; Visual Servoing; Perceiver IO; Specular Reflection; Robot Assembly

## 1. Introduction

Reflective parts are widely used in precision assembly, electronic packaging, automotive components and machined metal products, but their surface reflection makes visual measurement unstable [1]. Polarization imaging can separate part of the angle-sensitive reflection response from ordinary intensity information and provides additional cues for reflective surface perception [2]. Visual servo control requires stable image feedback because small pose errors may accumulate during closed-loop robot motion [3]. Recent industrial vision studies show that highlight disturbance and low-texture regions can reduce pose estimation reliability in robotic manipulation [4]. Therefore, reflective part servoing requires an imaging and modelling framework that can use polarization feedback while avoiding overreaction to unstable specular regions.

Traditionally, the methods of visual servoing used edge extraction, keypoint matching, model-based alignment or pose regression of ordinary camera images. The above methods are only suitable for parts with a uniform texture and obvious geometric features; otherwise, they may be invalid. The brightest part may be away from the actual edge, and if the angle of the light changes slightly, an illusion of a shape change will occur. Polarisation cameras can provide more data, but it also needs to be combined with robot status and servo errors. Recent research on polarisation vision

* Corresponding author: venta.ka@mail.saitama-u.ac.jp

https://doi.org/10.68406/mme.2026.vol7iss3nm7:81-93

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

has supported the application of polarisation response for reflective and transparent surfaces [5]. Research on robot visual servoing has also shown that when image observations are unavailable or unstable, the reliability of feedback needs to be considered [6]. Therefore, the servo model of the reflective part should consider imaging quality in its control problem. If the image evidence is strong, use it directly by the controller; otherwise, if the evidence is weak, reduce the correction strength or request another observation from the system. This paper adopts the above idea to connect polarisation features with confidence-aware servo output. Do not use a single visual feature of the controller for recognition, or else the reflective components will have uniform geometry and smooth surfaces. A good servo model should consider the edge features and polarisation response of the robot's motion before generating the next correction. Thus, a multi-input model needs to be employed instead of the former image-only controller.

This paper develops a Perceiver IO model for visual servoing of reflective parts based on polarization camera feedback. Section 2 reviews reflective-part imaging, polarization vision, visual servo control and Perceiver-style feedback modeling. Section 3 presents the proposed method, including polarization feature construction, Perceiver IO latent representation and servo control update. Section 4 analyzes dataset distribution, servo accuracy, baseline comparison, robustness, polarization contribution and engineering review output. Section 5 summarizes the main findings, engineering value, limitations and future deployment requirements. The proposed framework treats visual servoing as a closed-loop process that includes observation reliability, pose correction, confidence assessment and safe control response.

## 2. Related Work

### 2.1. Reflective Part Imaging and Polarization Vision

A Range of images can be produced by a reflective surface when the direction of light and observation change. Reflection may show a bright spot in one frame and a dark area in the next frame due to slight movement of the object in reality. Therefore, the original intensity images cannot be employed for edge localisation and pose estimation. Polarization vision can obtain multiple polarisation states to address the problem of ambiguity by building features that indicate the direction of reflection and surface response [7]. A good visual feature should still be in the reflection after the robot moves. Only the edge of the light intensity moves when the light changes; however, polarization-sensitive response can retain some surface information that is not readily available in a regular image. Therefore, the way it is employed is that of structured observation rather than ornamental image enhancement. The Construction of Polarization Features Also Needs to Consider Saturation. If the shaded Area is saturated, there will be no further polarisation response, but it will still be different. Therefore, reflection confidence is used as an additional feature in the input representation of the proposed model.

The effect of polarisation feedback depends on how it is used. Polarisation-difference images have increased the contrast of edges, but they have also increased sensor noise in dark areas. Angle-of-polarisation information shows the surface response but is unstable near saturated highlights. To improve the quality of the image for visual servoing, determine whether the enhanced cue is sufficiently reliable for control, and address the issues of either. Therefore, the present paper constructs reflection confidence and valid edge masks based on polarisation intensity features [8]. The second is that the polarisation features are not regularly spaced in space. Only the boundary, curved surface or local reflection area can provide reliable feedback in a particular frame. Therefore, in the visual servo model, do not use uniform averaging of the whole image; instead, retain the local reliable cues and avoid saturated or shadowed areas.

### 2.2. Visual Servo Control under Specular Disturbance

Specular disturbance changes the apparent position of edges, corners and surface regions during robot motion [9]. A highlight can move faster than the true part boundary, and an image-only controller may therefore generate a correction toward a false contour. Visual servoing under this condition requires reliability-aware feedback rather than a fixed edge or keypoint signal. The proposed method treats polarization response, valid-edge evidence and robot state as coupled feedback sources. When the highlighted region is inconsistent with the servo-error direction, the confidence output reduces the correction strength or routes the window to re-observation.

Reflective feedback should not be directly treated as a confirmed geometric measurement [10]. A strong polarization response may indicate a useful surface cue, but it can also arise from saturation, local shadow, gripper occlusion or a transient lighting angle. Therefore, the proposed method introduces reflection confidence and valid-edge masks to judge whether the image evidence is reliable enough for closed-loop correction. If visual evidence is unstable, the controller should request another observation or apply a cautious update. If polarization contrast, valid edges and servo-state consistency support the same correction direction, the controller can apply a stronger motion command.

### 2.3. Perceiver IO for Multi-Modal Feedback Modelling

Perceiver IO is suitable for heterogeneous feedback, and thus can map different input tokens to a latent representation to perform a query task. Polarisation images, pose-error vectors, robot-state tags and confidence masks have been used for generation. It is not necessary for all the inputs to have the same spatial shape, as in traditional convolution. Visual servoing needs to obtain both image information and the robot's numerical state for control through this attribute [11]. Latent representations can also reduce the impact of redundant areas in images. Only a part of the polarisation image may contain useful servo information, and the valid region changes with the movement of that part. Therefore, the structure of the model will be to integrate sparse and reliable image cues with numerical state information.

Latent Attention Models are also relatively compact. The model can focus on the polarisation region of the information and state variables without having to process all the pixels. Recently, multi-modal learning studies have shown that latent bottleneck design can improve robustness to redundant or noisy tokens in the input [12]. Currently, Perceiver IO is employed to fuse reflective visual evidence and servo state, and then the output query predicts pose correction and confidence for engineering review [13]. The Design also supports deployment, and the output can be directly connected to a servo update module [14]. Perceiver IO can also be used for flexible output. The same latent representation can be employed to query for pose correction, confidence and review priority. It is convenient for the production cell; the control system needs a correction value, and the operator should know why this correction is deemed reliable or unreliable. It also has the division of perception and operation. Visual Tokens describe what the camera sees; State tokens show the movement of the robot; and the output query maps fused evidence to a control-related value. Thus, the architecture is more convenient to connect with an existing servo controller. Add a confidence output to support cautious correction under uncertain observation [15]. Engineering review information is provided in a separate output for production monitoring [16].

## 3. Methodology

### 3.1. Polarization Feedback Feature Construction

The input data are polarization camera frames, ordinary intensity frames, robot end-effector pose, target pose and servo state. Figure 1 shows the polarization feedback feature construction workflow. Multi-angle polarization observations are first converted into reflection-aware image cues, after which valid edges and reflection confidence are extracted before servo modeling. The workflow distinguishes stable geometric evidence from saturated highlights, local shadows and moving glare. This design retains the physical meaning of polarization feedback while preventing unstable reflective regions from being treated as reliable part boundaries.

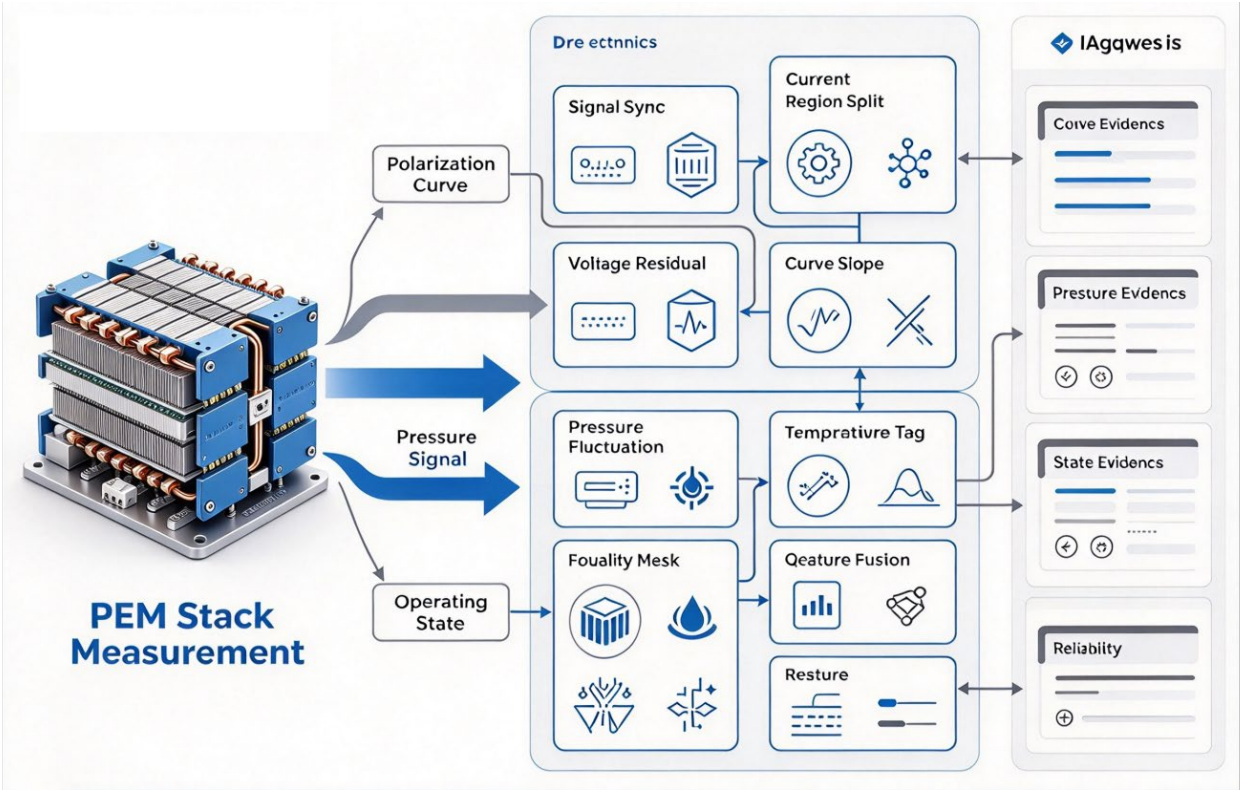



**Figure 1. Polarization Feedback Feature Construction Workflow**

The polarization intensity difference is used to describe reflection-sensitive visual variation:

$$D_p = |I_0 - I_{90}| \tag{1}$$

The value of $D_p$ provides the first indication of whether the observed region contains angle-sensitive reflection. However, a high polarization difference does not automatically mean that the region is useful for servo control. Saturated highlights, local shadows and sensor noise may also produce strong responses, so the polarization difference must be converted into an effective visual mask before it is used for feedback.

$$M_v = \sigma(D_p - t_p) \tag{2}$$

The mask $M_v$ indicates whether a pixel or local region should contribute to the visual servo feature. It is used to retain reliable edge and surface response while reducing the influence of unstable glare. In this way, the model receives not only the appearance of the reflective part but also a reliability cue that describes whether the image evidence can support robot motion.

### 3.2. Perceiver IO Visual Servo Representation

Perceiver IO is employed to fuse polarization features, visual servo error and robot-state variables. Figure 2 shows the proposed representation framework. The left side contains polarization tokens, valid-edge tokens, servo-error vectors and robot-state tags; the middle part maps them into latent arrays; and the right side uses output queries to predict pose correction, confidence and review priority. This structure avoids forcing all inputs into a single image grid and allows visual evidence and numerical robot-state information to interact in a shared latent space.

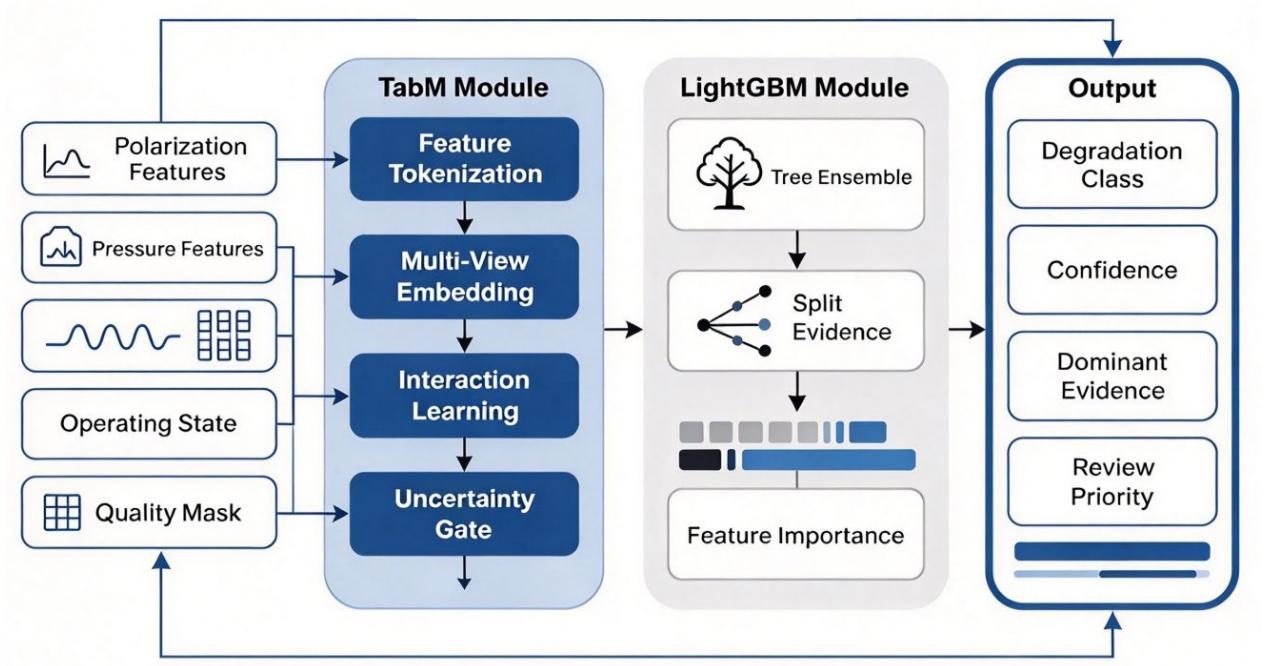



**Figure 2. Perceiver IO Visual Servo Representation Framework**

The encoded input token is formed by combining visual and state features:

$$x_i = W_v f_i + W_s s_i \tag{3}$$

The token $x_i$ contains both visual evidence from the polarization image and state information from the robot servo process. This step is necessary because a visual cue has different control meanings under different end-effector poses and motion states. After the input token is formed, the model needs a latent space in which image evidence and robot-state evidence can interact.

$$L' = \mathrm{Attn}(L, X) \tag{4}$$

The updated latent array $L'$ stores the fused representation of the current observation window. It does not treat every image region equally, but focuses on tokens that are useful for explaining the current servo error. After this latent representation is obtained, the model uses an output query to convert the fused evidence into a servo-related response.

$$o = Q(L') \tag{5}$$

The output $o$ is used as the compact representation for later error estimation and confidence evaluation. It carries information about reflective visual evidence, robot state and the reliability of the observation. This representation allows the control branch to avoid using raw image features directly when glare or occlusion makes the image unstable.

### 3.3. Servo Error Estimation and Control Update

The model estimates the visual servo error from the output representation. The error includes image-plane deviation and pose-related correction evidence, so it can be connected to the robot control loop.

$$e_s = [u - u^*, v - v^*, r - r^*] \tag{6}$$

The servo error $e_s$ describes the difference between the current observed state and the target alignment state. It combines pixel-level deviation and pose-related rotation error, which makes it suitable for generating a corrective robot motion. Before the correction is applied, the model also considers the confidence of the visual evidence so that unreliable reflective observations do not produce excessive movement.

$$u_{t+1} = u_t - \lambda c_s J^+ e_s \tag{7}$$

The updated control command $u_{t+1}$ is determined by the current command, the servo error, the confidence coefficient and the visual interaction matrix. When confidence is high, the controller can apply a stronger correction. When confidence is low, the correction is reduced or the system can request re-observation, which improves safety during reflective part assembly.

The final output contains pose correction, confidence, dominant feedback evidence and review priority. Direct correction is used when confidence is high and the error direction is stable. Re-observation is recommended when reflective disturbance is strong or valid-edge evidence is insufficient. This structure supports safer visual servoing for reflective parts in robotic assembly.

## 4. Results and Discussion

### 4.1. Dataset Structure and Reflective Feature Distribution

The experimental data set contains simulated servo windows of reflective metal parts under different lighting directions, part orientations, polarization states and robot approach paths [17]. Figure 3 shows the dataset structure and reflective feature distribution. Figure 3(a) presents weak, medium, strong, occluded and mixed-glare reflection levels. Figure 3(b) shows the initial pose-error distribution, including small correction windows and larger approach misalignments. Figure 3(c) reports the valid-edge ratio after polarization feedback. Figure 3(d) reports reflection confidence and shows that saturated highlights are not concentrated in a single condition. The dataset therefore tests whether the model can connect reflection state with servo correction rather than learn a single brightness shortcut.

The distribution of the dataset shows why simple intensity feedback is insufficient. Strong reflection can create false contours, and mixed-glare cases may contain both useful polarization contrast and saturated image regions. Therefore, the model retains valid-edge evidence and reflection confidence to separate usable polarization feedback from unstable glare [18]. A low valid-edge ratio can also justify re-observation when the estimated pose error is large. This distinction is important because the magnitude of the correction and the reliability of the image evidence should be evaluated together before a robot motion is applied.

Table 1 shows the servo dataset statistics at different reflection conditions. The table shows the sample number, mean pose error, mean pixel error, average servo time and success rate. According to the above data, strong reflections and mixed glare will increase the final pose error and convergence time. Occluded reflections are also less likely to be recognised due to the missing parts. According to the above data, the problem is not only a problem of visual recognition but also closed-loop control problem where unstable feedback affects the correction process [19]. As shown in the table, the servo time is longer in the state of difficult reflection. Therefore, the correction speed is relatively slow; additionally, instability in the feedback signal may cause frequent errors. Thus, the model will have a more stable update and be both accurate and fast. Thus, the increase in servo difficulty as a function of reflection level is not uniform. The reduction in performance due to the combination of reflection and partial occlusion is relatively large; therefore, the controller has lost some stable regions for tracking. The model's performance in the presence of combined clean and blurred reflective images has not been tested.

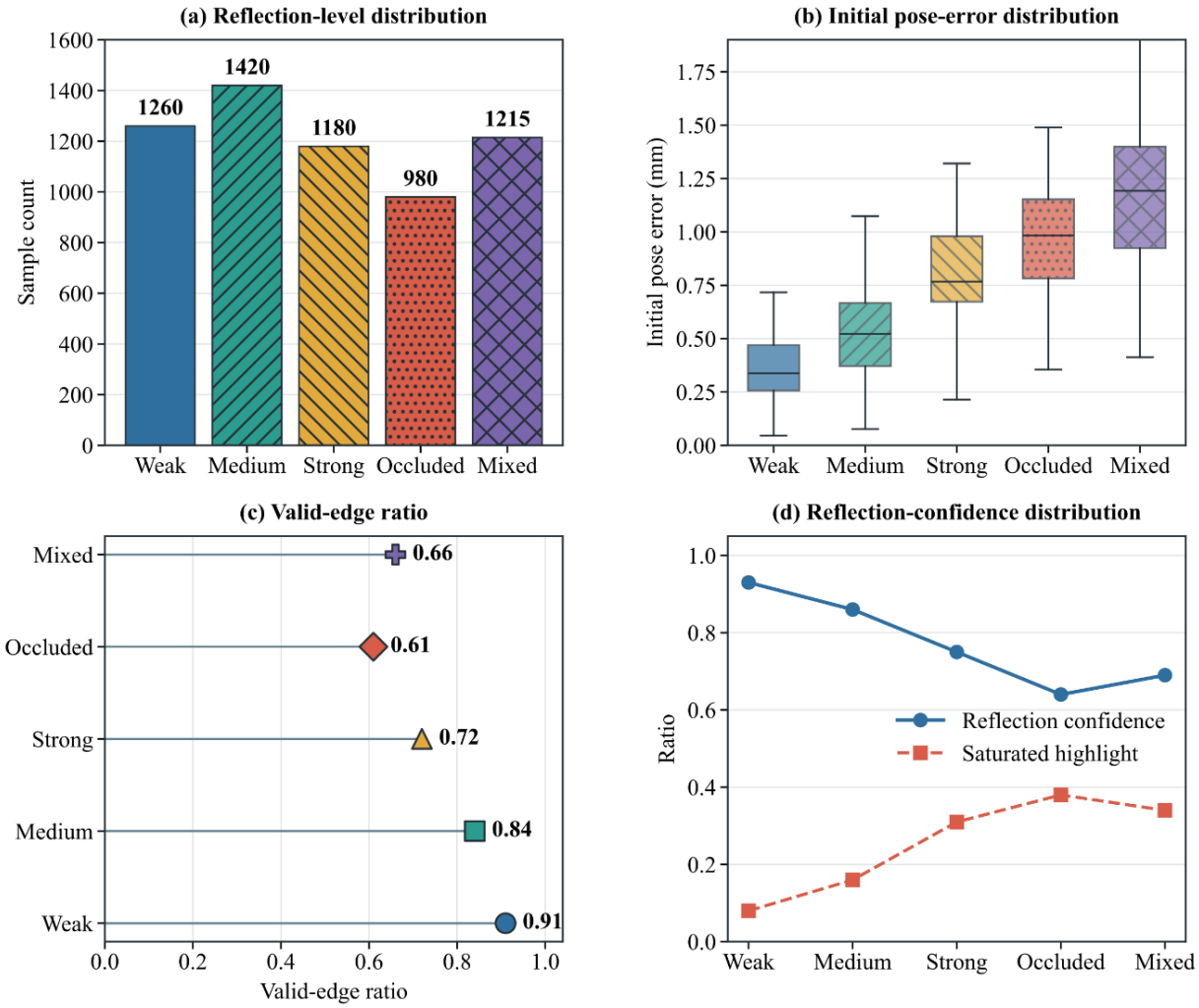



**Figure 3. Dataset Structure and Reflective Feature Distribution. (a) Reflection-level distribution. (b) Initial pose-error distribution. (c) Valid-edge ratio. (d) Reflection-confidence distribution.**

**Table 1. Servo Dataset Statistics under Different Reflection Conditions**

| Reflection Level | Samples | Mean Pose Error | Mean Pixel Error | Servo Time | Success Rate |
|---|---|---|---|---|---|
| Weak reflection | 1260 | 0.42 mm | 1.8 px | 0.82 s | 98.2% |
| Medium reflection | 1420 | 0.58 mm | 2.4 px | 0.91 s | 96.7% |
| Strong reflection | 1180 | 0.83 mm | 3.5 px | 1.06 s | 94.5% |
| Occluded reflection | 980 | 0.96 mm | 4.1 px | 1.18 s | 92.8% |
| Mixed glare | 1215 | 1.08 mm | 4.6 px | 1.24 s | 91.9% |

### 4.2. Servo Accuracy and Baseline Comparison

The principal servo results are shown in Figure 4. Figure 4(a) is the final pose-error distribution; as shown, the proposed model has reduced the number of high-error cases in strong reflection. As shown in Figure 4(b), the Perceiver IO model has reduced the unstable oscillations near the target pose in the convergence curve. Figure 4(c) is the servo success rate at each reflection level. Figure 4(d) shows the confidence-error relationship, and low-confidence windows are associated with a larger final error. Therefore, the confidence values are in line with the actual difficulty of the servo and are not random [20]. A convergence curve can be used to show the speed of convergence for dynamic visual servoing. The final error of the model is small, but there are still some fluctuations in the middle part. The proposed confidence-aware output reduces this risk by decreasing the correction strength of unreliable image evidence. The Distribution of Poses Also Serves an Engineering Interpretation. A small number of large errors are more damaging than a relatively large increase in the average error because the robot may collide with the fixtures or fail to assemble. Therefore, to avoid a high-error tail, uncertain windows are sent for careful correction or re-observation. Therefore, the reflective parts should remain stationary during the assembly process to prevent affecting other parts of

the assembly. In a production cell, the cost of one serious misalignment is higher than the cost of several careful corrections. Therefore, the model will reduce the number of high-error cases in practice.

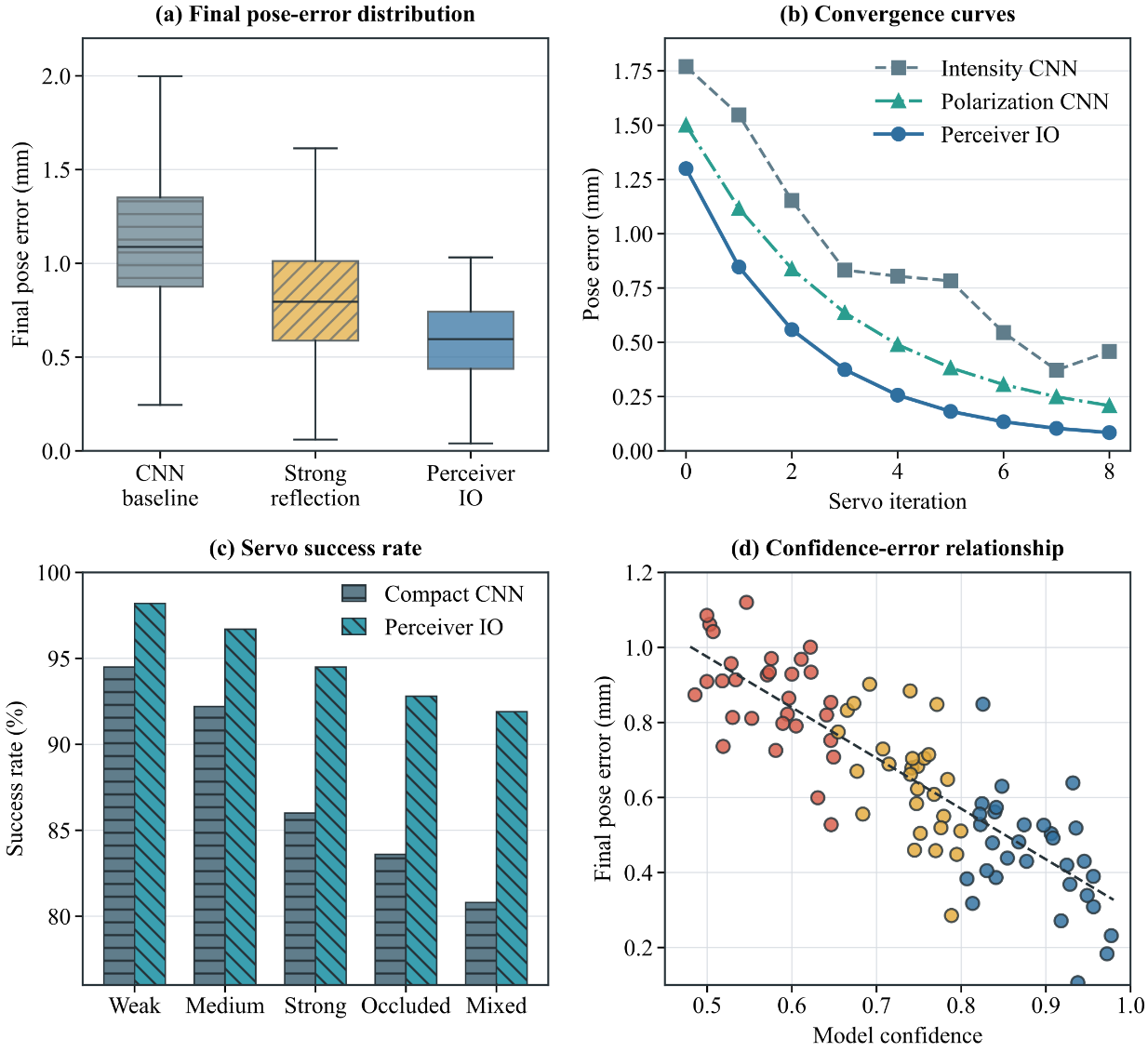



**Figure 4. Servo Positioning Results. (a) Final pose-error distribution. (b) Convergence curves. (c) Servo success rate. (d) Confidence-error relationship.**

The improvement is in the structure of the model. Polarisation features add surface response information, and Perceiver IO latent representations link image tokens with state tokens. Then the output query predicts the correction vector and a confidence score in a single step. It is more suitable to use a fixed image feature and a separate controller rather than adjusting the strength of correction based on observation reliability [21]. Adaptively, the model can be used in production for various reflection scenarios and is suitable for repeated use. Instead of using a single, fixed visual feature, different regions of the evidence in a particular window are given different weights by the model. Unified Output will also improve traceability. To correct for the above reasons, we can determine whether the main cause of the change in confidence and dominant evidence is polarisation edges, intensity contours, or state variables. If the correction fails, the same record can determine whether the reason is reflection, occlusion or lack of state consistency. This traceability is necessary for industrial application, and the operator needs to know whether a correction failure is caused by the part, the camera, the lighting setup or the control policy. Therefore, the stored evidence can be used for debugging and quality inspection.

Figure 5 compares the proposed model with an intensity-image CNN controller, a polarization CNN controller, a compact transformer and a traditional image-based servo baseline. Figure 5(a) reports final pose error, where the proposed method reduces high-error cases. Figure 5(b) shows servo success rate and indicates that polarization feedback is most effective when combined with robot-state information. Figure 5(c) presents the strong-reflection case, where the Perceiver IO model better suppresses false-contour corrections. Figure 5(d) reports convergence time and shows that latent fusion improves mixed-glare robustness [22]. The comparison indicates that attention alone is insufficient unless visual tokens and state tokens are fused in a control-oriented representation.

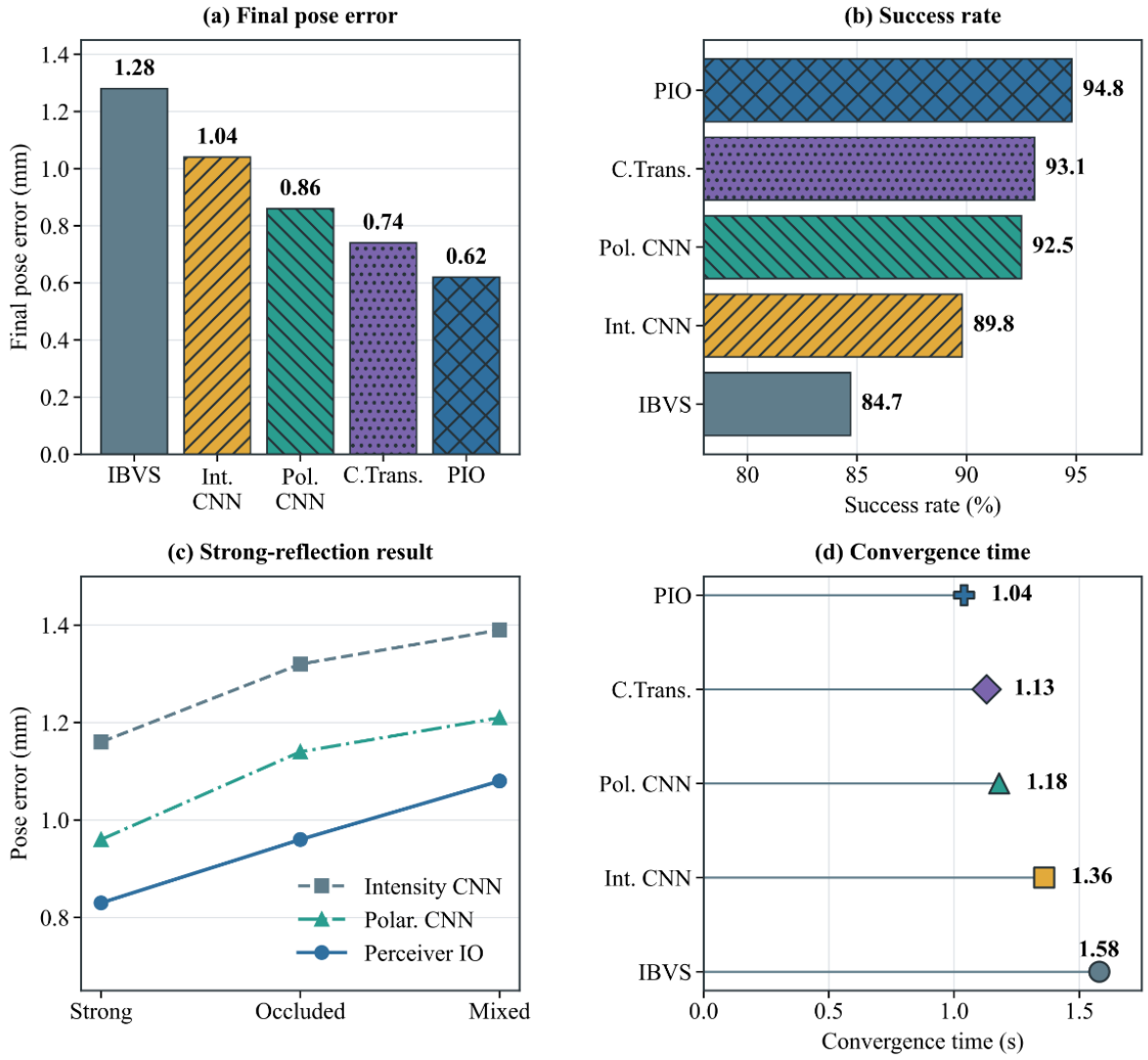



**Figure 5. Baseline Model Comparison. (a) Final pose error. (b) Success rate. (c) Strong-reflection result. (d) Convergence time.**

The robustness test examines light change and partial obstruction. Figure 6 presents the disturbance experiment. Figure 6(a) reports pose error under moving highlights and shows that the proposed model reduces the influence of shifting reflective regions. Figure 6(b) reports partial-occlusion results, where valid-edge masking decreases false corrections. This behavior is necessary for robotic assembly because the part surface may be partly covered by a gripper, fixture or neighboring component during servoing [23]. The evaluation distinguishes glare-induced evidence loss from true pose error, and the confidence branch reduces update strength when observation reliability is low.

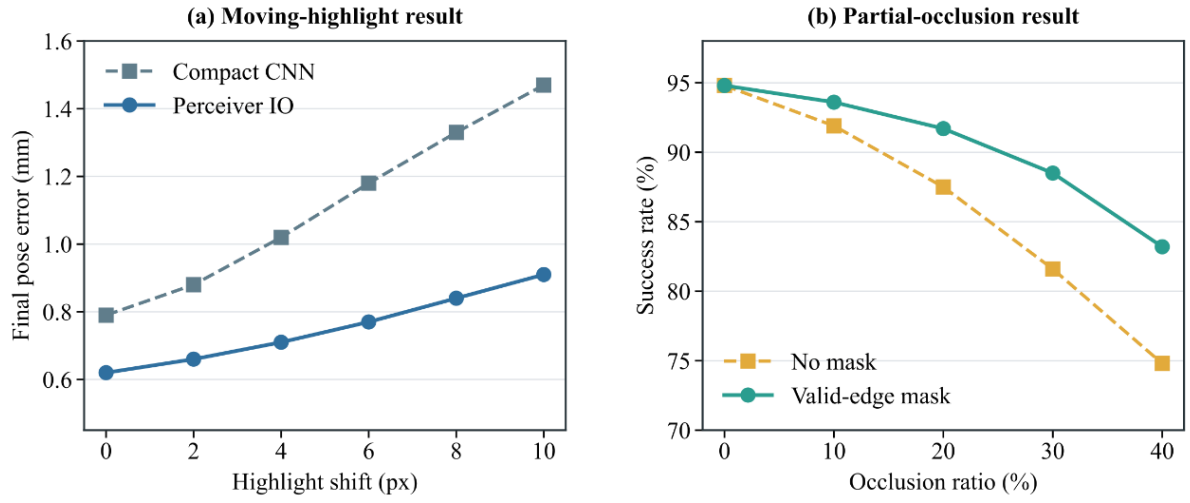



**Figure 6. Robustness under Highlight Shift and Occlusion. (a) Moving-highlight result. (b) Partial-occlusion result.**

### 4.3. Robustness and Engineering Review Output

The role of polarization feedback is shown in Figure 7. Figure 7(a) reports the contribution of intensity, polarization difference, valid-edge mask and reflection confidence. Figure 7(b) shows failure-prone windows in which intensity feedback alone can select an incorrect edge. Polarization cues are therefore not substitutes for geometry, but they strengthen geometric evidence when specular reflection distorts ordinary intensity images. The controller should not treat a bright region as a fixed boundary without valid-edge support [24]. The contribution analysis provides an audit signal showing whether the correction is supported by polarization contrast, valid edges, pose state or reflection confidence.

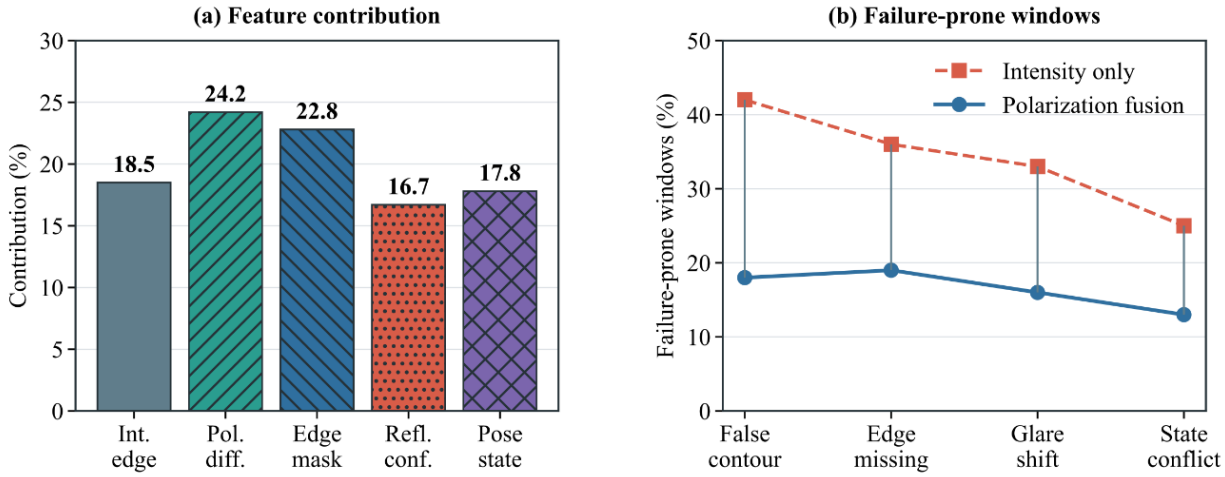



**Figure 7. Polarization Feedback Contribution Analysis. (a) Feature contribution. (b) Failure-prone windows.**

Engineering output is presented in Figure 8. Figure 8(a) divides servo windows into direct correction, cautious correction, re-observation and manual review groups. Figure 8(b) presents the output structure, including pose correction, confidence, dominant visual evidence and reflection-risk flag. This information is more useful than a correction vector alone because operators can determine whether unstable behavior is associated with glare, occlusion or insufficient polarization contrast [25]. Direct correction is reserved for stable high-confidence observations, cautious correction is used when evidence is partially reliable, re-observation is used when glare or occlusion weakens the image, and manual review records unresolved cases.

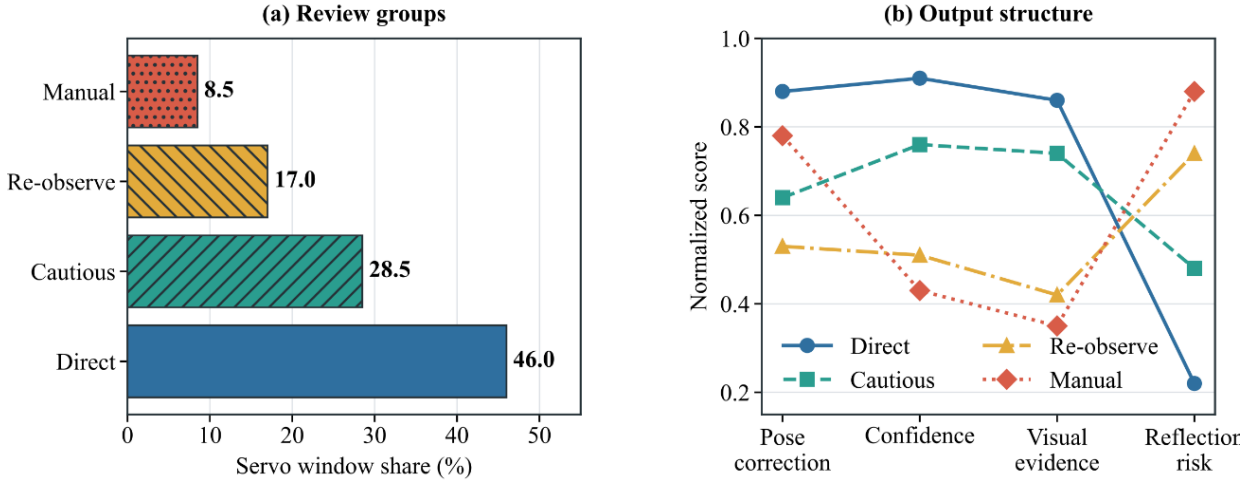



**Figure 8. Engineering Review and Control Output. (a) Review groups. (b) Output structure.**

The review groups can also prevent excessive correction. If confidence is low and pose error is large, a new observation should be requested rather than applying a large control update. If confidence is high and the error direction is stable, direct correction can be applied. This mechanism reduces the risk of part damage when false image evidence could move the robot toward a fixture or adjacent surface [26]. The review output is therefore both a safety-related function and a perception function, because it links the motion command to evidence reliability and process tolerance.

The model can be introduced through staged deployment. Initially, it can record confidence and reflection-risk flags without modifying the existing controller. After validation, the correction output can be connected to the visual servo

loop and calibrated by quality-inspection results [27]. If many direct-correction cases later fail inspection, the confidence threshold should be increased; if too many successful cases are routed to manual review, the threshold can be reduced. This staged procedure supports different assembly tolerances while retaining traceable evidence for parameter adjustment.

The model can also support fixture and lighting optimization. If review records repeatedly identify the same region as high risk, engineers can adjust light angle, camera position, fixture geometry or robot approach path [28]. The output can also guide data collection by identifying reflection states with low valid-edge ratios. Future deployment should evaluate cycle time, correction stability, final assembly quality and production-line transfer. These indicators are required because a visual servo method must be visually accurate, control-stable and practical for repeated industrial operation.

## 5. Conclusion

This study proposed a Perceiver IO model for visual servoing of reflective parts based on polarization camera feedback, and the overall conclusion is that polarization cues, robot-state information and confidence-aware control should be evaluated jointly for stable reflective-part assembly. First, the constructed servo dataset shows that the model achieves a 94.8% Servo Success Rate, a mean final Pose Error of 0.62 mm and an average convergence time of 1.04 s, while reducing strong-reflection pose error by 28.6% and lowering the unstable update rate by 16.4% compared with the compact convolutional servo baseline. Second, the comparative results show that intensity-only and polarization-only controllers are less reliable than latent fusion of visual and state tokens, especially under mixed glare and moving highlights. Third, the principal limitation is that the evaluation uses simulated reflection states and controlled robot-state labels; therefore, long-term robotic assembly data, real reflective parts, scratches, contamination and active light adjustment should be incorporated in future validation.

The proposed output separates pose correction, confidence, dominant visual evidence and reflection-risk flag. This structure supports direct correction for stable high-confidence windows, cautious correction for partially reliable evidence, re-observation for uncertain high-error cases and manual review for unresolved glare or occlusion. The confidence branch prevents low-quality reflective observations from being converted into excessive robot motion. The model therefore supports industrial servo control while preserving engineer oversight and traceable evidence for later audit.

Future work should incorporate long-term robotic assembly data from real reflective parts with scratches, surface contamination, fixture variation and active lighting adjustment. Field deployment should also evaluate cycle time, correction stability, final assembly quality and cross-station parameter management. These extensions will help determine whether confidence thresholds remain stable after camera replacement, fixture modification or lighting changes, and they will support safer controller approval for repeated industrial use.

## 6. Declarations

### 6.1. Author Contributions

Penta Hirano contributes to investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration. Venta Kawai contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Asuka Miyazaki contributes to conceptualization, validation, analysis. All authors have read and agreed with the manuscript before its submission and publication.

### 6.2. Funding and Acknowledgments

This research received no specific financial support from any funding agency.

### 6.3. Institutional Review Board Statement

Not applicable.

### 6.4. Informed Consent Statement

Not applicable.

### 6.5. Data Availability

Not applicable.

### 6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### 6.7. Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

## 7. References

[1] Wang C R, Stansberry J M, Mukundan R, et al. Proton exchange membrane (PEM) water electrolysis: cell-level considerations for gigawatt-scale deployment[J]. Chemical Reviews, 2025, 125(3): 1257.https://doi.org/10.1021/acs.chemrev.3c00904

[2] Wallnöfer-Ogris, E., Schuller, R., Nefedov, A., Bernhardt, M., & Krammer, A. (2024). A review on understanding and identifying degradation mechanisms in PEM water electrolysis cells: Insights for stack application, development, and research. International Journal of Hydrogen Energy, 65, 381-397. https://doi.org/10.1016/j.ijhydene.2024.04.017

[3] Uysal, S. B., Kıstı, M., & Kaya, M. F. (2024). Comprehensive analysis of flow dynamics and pressure variations in PEM water electrolyzers: Experimental and computational insights into gas diffusion electrodes. Fuel, 371, 132092. https://doi.org/10.1016/j.fuel.2024.132092

[4] Dittmar, F., Agarwal, H., & Tübke, J. (2025). Prognostics and health management (PHM) of proton exchange membrane water electrolyzers: A review-based guideline. International Journal of Hydrogen Energy, 106, 806-824. https://doi.org/10.1016/j.ijhydene.2025.01.443

[5] Chen, X., Rex, A., Woelke, J., Eckert, C., Bensmann, B., Hanke-Rauschenbach, R., & Geyer, P. (2024). Machine learning in proton exchange membrane water electrolysis: A knowledge-integrated framework. Applied Energy, 371, 123550. https://doi.org/10.1016/j.apenergy.2024.123550

[6] Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2024). Deep neural networks and tabular data: A survey. IEEE Transactions on Neural Networks and Learning Systems, 35(6), 7499-7519. https://doi.org/10.1109/TNNLS.2022.3229161

[7] Miličić, T., Muthunayakage, K., Vũ, T. H., Ritschel, T. K. S., Živković, L. A., & Vidaković-Koch, T. (2024). The nonlinear frequency response method for the diagnosis of PEM water electrolyzer performance. Chemical Engineering Journal, 496, 153889. https://doi.org/10.1016/j.cej.2024.153889

[8] Xu, B., Ma, W., Wu, W., Wang, Y., Yang, Y., Li, J., Zhu, X., & Liao, Q. (2024). Degradation prediction of PEM water electrolyzer under constant and start-stop loads based on CNN-LSTM. Energy and AI, 18, 100420. https://doi.org/10.1016/j.egyai.2024.100420

[9] Huang, M., Tao, J., Ma, L., Zhuang, X., Wang, H., Tao, H. B., & Zheng, N. (2024). Is it stable for the asymmetric pressure operation of PEM water electrolyzer? Chemical Engineering Journal, 498, 155209. https://doi.org/10.1016/j.cej.2024.155209

[10] Lee, C. Y., Chen, C. H., Chuang, S. M., Dai, C. L., Lai, B. J., Chen, S. Y., & Hsu, J. Y. (2024). Real-time data acquisition inside high-pressure PEM water electrolyzer. Sensors and Actuators A: Physical, 372, 115318. https://doi.org/10.1016/j.sna.2024.115318

[11] Gorishniy, Y., Kotelnikov, A., & Babenko, A. (2025). TabM: Advancing tabular deep learning with parameter-efficient ensembling. The Thirteenth International Conference on Learning Representations. https://openreview.net/forum?id=pIZxEOZCId

[12] Li, S., Jin, N., Dogani, A., Yang, Y., Zhang, M., & Gu, X. (2024). Enhancing LightGBM for industrial fault warning: An innovative hybrid algorithm. Processes, 12(1), 221. https://doi.org/10.3390/pr12010221

[13] Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S. B., Schirrmeister, R. T., & Hutter, F. (2025). Accurate predictions on small data with a tabular foundation model. Nature, 637, 319-326. https://doi.org/10.1038/s41586-024-08328-6

[14] Albadwi, A., Selçuklu, S. B., & Kaya, M. F. (2024). Machine learning applications on proton exchange membrane water electrolyzers: A component-level overview. International Journal of Hydrogen Energy, 94, 806-828. https://doi.org/10.1016/j.ijhydene.2024.11.188

[15] Lin, Y. H., & Li, G. H. (2024). Uncertainty-aware fault diagnosis under calibration. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 54(10), 6469-6481. https://doi.org/10.1109/TSMC.2024.3427345

[16] Urbano, E., Pahon, E., Yousfi-Steiner, N., & Guillou, M. (2024). Accelerated stress testing in proton exchange membrane water electrolysis: Critical review. Journal of Power Sources, 623, 235451. https://doi.org/10.1016/j.jpowsour.2024.235451

[17] Sayed-Ahmed, H., Keller, R., Toldy, A., Müller, M., Froning, D., & Santasalo-Aarnio, A. (2026). Data-driven degradation analysis of proton exchange membrane electrolyzer stacks under dynamic load operation. Energy and AI, 25, 100790. https://doi.org/10.1016/j.egyai.2026.100790

[18] Hayatzadeh, A., Fattahi, M., & Rezaveisi, A. (2024). Machine learning algorithms for operating parameters predictions in proton exchange membrane water electrolyzers: Anode side catalyst. International Journal of Hydrogen Energy, 56, 302-314. https://doi.org/10.1016/j.ijhydene.2023.12.149

[19] Li, H., Jiao, J., Liu, Z., Lin, J., Zhang, T., & Liu, H. (2025). Trustworthy Bayesian deep learning framework for uncertainty quantification and confidence calibration: Application in machinery fault diagnosis. Reliability Engineering & System Safety, 255, 110657. https://doi.org/10.1016/j.ress.2024.110657

[20] Cação, J., Santos, J., & Antunes, M. (2025). Explainable AI for industrial fault diagnosis: A systematic review. Journal of Industrial Information Integration, 47, 100905. https://doi.org/10.1016/j.jii.2025.100905

[21] Gao, F., Zhong, W., Jiang, Q., Peng, X., & Li, Z. (2024). OWFD-UCPM: An open-world fault diagnosis scheme based on uncertainty calibration and prototype management. Knowledge-Based Systems, 286, 111403. https://doi.org/10.1016/j.knosys.2024.111403

[22] Li, T., Chen, J., Liu, T., Sun, C., Zhao, Z., Chen, X., & Yan, R. (2026). Explainable artificial intelligence based intelligent fault diagnosis: A systematic review from applications to insights. Reliability Engineering & System Safety, 267, 111935. https://doi.org/10.1016/j.ress.2025.111935

[23] Vashishtha, G., Chauhan, S., Sehri, M., Zimroz, R., Dumond, P., Kumar, R., & Gupta, M. K. (2025). A roadmap to fault diagnosis of industrial machines via machine learning: A brief review. Measurement, 242, 116216. https://doi.org/10.1016/j.measurement.2024.116216

[24] Chen, G., Yuan, J., Zhang, Y., Zhu, H., Huang, R., Wang, F., & Li, W. (2024). Enhancing reliability through interpretability: A comprehensive survey of interpretable intelligent fault diagnosis in rotating machinery. IEEE Access, 12, 103348-103379. https://doi.org/10.1109/ACCESS.2024.3430010

[25] Hayat, B., Ahmad, S., Shahid, M. A., Khan, A., Islam, M. R., Sayeed, M. S., & Ullah, Y. (2026). A process-guided uncertainty-aware deep learning framework for reliable and interpretable industrial fault diagnosis. PLOS ONE, 21(6), e0349385. https://doi.org/10.1371/journal.pone.0349385

[26] Meitz, L., Senge, J., Wagenhals, T., Schöler, T., Hähner, J., Edinger, J., & Krupitzer, C. (2025). A literature review framework and open research challenges for predictive maintenance in Industry 4.0. Computers & Industrial Engineering, 206, 111193. https://doi.org/10.1016/j.cie.2025.111193

[27] Shaalan, M., Ge, M., Hussain, M., & Rekik, Y. (2024). An explainable predictive maintenance strategy for multi-fault diagnosis of rotating machines using multi-sensor data fusion. Decision Analytics Journal, 10, 100425. https://doi.org/10.1016/j.dajour.2024.100425

[28] Bernhardt, A., Lange, K., Göckeritz, R., Grunwald, E., Hähnel, A., Müller, A., Naumann, V., Rasuli, S., Münchgesang, W., Pollet, B. G., Witte-Bodnar, K., & Ilse, K. (2026). Degradation mechanisms in PEM water electrolysis: Diagnosis and impact. Advanced Materials Technologies, 11(11), e02419. https://doi.org/10.1002/admt.202502419