Journal of Applied Automation Technologies

Research Paper

YOLOv11-ODConv for Cross-Domain Micro-Scratch Detection in Photovoltaic Glass Manufacturing

Nagyházi Bence[1,*] and Polgár Mihály[1]

[1] Faculty of Technical IT, University of Nyíregyháza, Nyíregyháza, 4400, Hungary
*Corresponding author: nagyhazi.b@inf.nyf.hu

Abstract. Cross-domain micro-scratch inspection is difficult in photovoltaic-glass manufacturing because the observable defect signal changes with coating type, furnace batch, camera response, illumination angle, and line contamination. This work develops a YOLOv11-ODConv detector that couples omni-dimensional dynamic convolution with domain-balanced feature learning and morphology-aware fusion. The model adapts convolutional kernels along spatial, input-channel, output-channel, and kernel-index dimensions, while a domain consistency objective constrains scratch representations without suppressing line-specific texture. A boundary-weighted localization term and hard-negative curriculum further separate faint elongated scratches from roller marks, dust trails, and specular streaks. Under leave-one-domain-out evaluation, the proposed model reaches 91.8% mAP50, 67.4% mAP50:95, 89.6% recall, and 12.7 false alarms per 1,000 images, improving the YOLOv11 baseline by 4.9, 5.7, and 4.1 percentage points in the first three metrics while reducing false alarms by 31.0%. It processes 1024 x 1024 images at 46.3 frames per second on an RTX 4060 and 21.8 frames per second on an industrial edge GPU. The results indicate that dynamic kernels are most useful when combined with domain-balanced supervision and morphology-aware localization, providing a practical route to robust surface inspection across manufacturing lines.

Keywords: *Photovoltaic Glass, Micro-Scratch Detection, YOLOv11, Omni-Dimensional Convolution, Domain Generalization, Surface Inspection*

Received on 25 January 2026, Accepted on 03 June 2026, Published on 14 June 2026

Introduction

Photovoltaic glass controls light transmission and mechanical strength while shielding the solar cell. Even though a micro-scratch from tempering, coating, transporting, or packaging is only a few pixels wide, it can generate an apparent quality flaw after lamination, operate as a stress concentration site, and alter the antireflective qualities [1]. Because of this, modern lines position line-scan or area-scan cameras after the crucial step, and the contrast that is collected is extremely sensitive to the geometry of the lighting [2]. After altering the coating roughness, dark-field lighting increases one line's edge and almost completely removes it on another [3]. At production speed, roller texture results in a longer reaction that looks like scratches [4]. Hard negatives with varying frequencies across shifts include dust, water stains, and local reflections [5]. At a high throughput, manual examination is unable to maintain consistent sensitivity [6]. When the fault intensity coincides with the background distribution, traditional thresholding likewise fails [7]. As a result, the challenge is no longer a closed-set recognition problem but rather a cross-domain detection problem. Another issue is anisotropy of the scratch; a harmless streak may appear brighter and a trace seen at one azimuth may vanish with a slight angle change. Therefore, a sheet observed under a particular optical and process configuration is the relevant unit rather than a pixel patch. A good detector must meet the cycle time without target-line retraining, maintain the local evidence, and have a stable confidence. Trace consistency and severity are also necessary for production acceptability; otherwise, even if the image-level classification is accurate, the unstable box may result in an inconsistent sheet decision.

By learning hierarchical texture and form cues, deep detectors have improved industrial inspection [8]. Because they combine localization and classification in a single forward pass, one-stage systems are practical [9]. Repeated downsampling may still remove the one- or two-pixel scratch core, however multi-scale pyramids aid in retaining minor flaws [10]. In the absence of labels, attention modules may produce spurious noise, but they can remedy defects [11]. Although many manufacturing deployments are unable to acquire target photos prior to commissioning, domain adaptation transfers information from a labeled source domain to an unlabeled target domain [12]. Retaining defect-sensitive variations and removing noise are the key issues. While a low-invariance detector will still be limited by a particular camera response, excessive invariance may result in the loss of coating-dependent border information. Because adjacent tiles have the same texture, coating, and acquisition noise, Random Split is inappropriate. Complete sheets and, in the case of cross-domain claims, complete manufacturing domains should be provided by reliable research. Additionally, thin and low-contrast strata should be reported; otherwise, the numerous simple flaws will conceal issues at the core of consumer complaints.

Following this problem logic, the study focuses on a detector design that can preserve micro-scratch evidence while transferring across production domains. The research content is to build a YOLOv11-ODConv inspection model for photovoltaic-glass micro-scratches, combine adaptive dynamic kernels with residual encoding and morphology-aware fusion, and evaluate whether the detector remains accurate, calibrated, and fast on unseen manufacturing domains. The main contributions are threefold: ODConv is inserted only at high-value YOLOv11 stages to provide conditional capacity without excessive latency; residual encoding, directional fusion, boundary weighting, consistency learning, and hard-negative training are coupled to protect faint elongated scratches; and the evaluation links leave-one-domain-out accuracy, width-contrast error analysis, ablation, robustness, calibration, memory, and throughput.

The full paper is arranged in five parts. The 1st part is the Introduction, which defines the industrial problem, research content, and contribution logic. The 2nd part is Related Work, which reviews surface inspection, one-stage detection, dynamic convolution, and cross-domain learning. The 3rd part is Proposed YOLOv11-ODConv Detection Framework, which explains the model design, optimization objective, and deployment interface. The 4th part is Cross-Domain Experiments and Analysis, which tests accuracy, robustness, error structure, ablation behavior, and line-rate feasibility with replaceable validation results. The 5th part is the Conclusion, which summarizes the findings, limitations, and future deployment directions.

Related Work

Surface Defect and Micro-Scratch Detection

After training on a number of observed lines, domain generalization looks for a representation that is still useful for an unseen line [13]. This study examines the trade-off for photovoltaic-glass micro-scratch detection using a YOLOv11-ODConv model. YOLOv11 provides effective multi-scale prediction and has a compact detection backbone [14]. Instead of using a single filter for every domain, omni-dimensional dynamic convolution modifies the effective kernel according to the input information [15]. Directional filtering, edge linking, local contrast normalization, and hand-designed morphology were all integrated into early glass inspection pipelines [16]. Every light change modifies the response distribution and requires threshold modification, yet these pipelines are still affordable and comprehensible. The ability to distinguish a scratch from an area with a homogeneous coating is improved by texture descriptors [17]. When a roller mark has the same direction or the scratch is irregular, they are less trustworthy. Several descriptors can be combined via classical classifiers [18]. Faint flaws that were overlooked before to categorization cannot be recovered since their candidate-generation stage still impacts the possible recall. A bank will raise the cost and continue to believe that a set response threshold distinguishes signals from backgrounds even if directional filters can be turned at many angles. When glare splits a scratch, connected-component rules break. The disconnected dust or roller textures can be combined by joining the nearby components. Threshold maintenance becomes highly line-specific as a result of these inaccuracies spreading to the length estimation. Additionally, classical pipelines have limited confidence calibration, making it challenging to choose a single rejection threshold for multiple optical stations.

Convolutional detectors learn both categorization and candidate generation simultaneously [19]. Although two-stage systems provide precise recommendations, they might not be able to satisfy line-rate limitations at high image resolution. One-stage models are often used in industrial inspection because they provide tiling or streaming inference. Small-object variations weigh border pixels, add additional prediction scales, or maintain shallow features. Since the bounding box may be composed primarily of regular glass, these options are appropriate for micro-scratches. Instead of learning the actual scratch trace, a detector trained on a single line frequently learns illumination-dependent halos. Consequently, whether a camera, lamp, coating, or supplier is replaced, the model fails despite having a high random-split accuracy. Shallow maps also have substantial surface texture, and feature-pyramid networks make advantage of high-resolution maps. Recall and false alarms will rise when a small-object head is added without control. The issues of localization, geometry, and background discrimination are addressed by boundary-aware losses, oriented representations, and hard-negative mining, respectively. Examine each one separately; the overall mean precision won't reveal whether a specific failure has gotten worse.

Dynamic Convolution and Cross-Domain Learning

An input-conditioned operator takes the place of a single stationary kernel in dynamic convolution [20]. Channel attention modifies the amplitude of features after convolution, and mixture-of-kernels choose or combine multiple experts. Spatial locations, input channels, output channels, and kernel experts are all concurrently modulated via omni-dimensional convolution [21]. Compared to scalar expert weighting, this factorization provides greater control and can modify edge selectivity for contrast or scratch orientation. Additionally, under domain or defect-scale imbalance, the attention branches may collapse to a single expert, making it more susceptible to optimization. When the same physical edge shifts its bandwidth or polarity across optical domains, conditional operators are needed. Benefits are not guaranteed; a big kernel bank can recall source lines, and a controller can learn a shortcut from the backdrop brightness. Thus, placement, unseen-domain testing, balanced batches, and controller regularization are just as important as operator definition.

By using adversarial alignment, moment matching, style perturbation, or consistency training, cross-domain detection techniques lessen distribution mismatch [22]. Because the domain style may contain information on fault visibility, pixel-level alignment might not be appropriate for transparent or reflective items. Although feature-level alignment is quite adaptable, it might align weak fault features with background clutter. By using domain-balanced sampling, gradients are prevented from being dominated by a huge production line [23]. By using confidence weighting and reducing the need for the same output in physically impossible transformations, consistency constraints can stabilize the prediction of photometric disturbances. Rarely does current research combine omni-dimensional kernels with latency-aware ablation, elongated-defect localization, and an explicit industrial domain protocol in a single detector. As a result, a model that is both adaptive and invariant is suggested. If the domain classification is somewhat easy, adversarial alignment can potentially cause a compact detector to become unstable. The outputs are more varied because to style randomization, although some might not be realistic. Because it eliminates the need for an inference branch, teacher-student consistency is appealing for deployment-oriented systems; nonetheless, false positives may result from confidence errors. Bounded transformations, delayed weighing, and a calibration audit are necessary for a workable design. Report the worst held-out line, keep the domain identity at the sheet level in the evaluation, and distinguish between alert load and accuracy. Otherwise, a single coating or camera arrangement could conceal a deployment failure with a good average.

Proposed YOLOv11-ODConv Detection Framework

Cross-Domain Problem Equation and Input Encoding

Let each image originate from a manufacturing domain defined by the production line, optical station, coating recipe, and acquisition configuration. A training set contains several observed domains, whereas validation and final testing preserve entire domains rather than randomly mixing frames. The detector receives an image tile and returns scratch boxes, confidence scores, and a binary defect label. Domain boundaries are established before sampling from production metadata and acquisition configurations rather than inferred after observing test performance. Sheets, instead of individual tiles, constitute the grouping unit. This strategy prevents correlated backgrounds and repeated scratch fragments from appearing in both training and evaluation, making the reported transfer problem operationally meaningful. The same grouping protocol is retained for every baseline.

$$\mathcal{D}_s = \{(\mathbf{x}_i, y_i, \mathbf{b}_i, d_i)\}_{i=1}^{N_s}, d_i \in \{1, \dots, S\} \quad \text{Eq.(1)}$$

Here, $\mathbf{x}_i$ denotes an input image, y_i is the scratch class label, $\mathbf{b}_i$ contains the annotated bounding-box coordinates, and d_i identifies the source manufacturing domain. The domain index is used only during training. It controls balanced sampling and domain-oriented regularization, while inference does not require a known target-domain identifier. Removing the domain label from training would reduce the task to conventional empirical risk minimization and allow the largest production line to dominate the gradient.

The encoded input combines the original RGB image with fixed photometric residuals computed from directional local differences. The residual channels allow the backbone to access both appearance information and weak boundary evidence.

$$\mathbf{x}_i^+ = \mathrm{Concat}\left(\mathbf{x}_i, \left|\mathbf{x}_i - G_{\sigma_z} * \mathbf{x}_i\right|, \left|\mathbf{x}_i - G_{\sigma_y} * \mathbf{x}_i\right|\right) \quad \text{Eq.(2)}$$

In this expression, G_{σ_x} and G_{σ_y} are directional Gaussian operators, and $*$ denotes convolution. The residual channels are deterministic and introduce no trainable branch during deployment. They expose horizontal and vertical intensity deviations without assuming a fixed scratch orientation. If residual encoding is removed, the network must reconstruct low-amplitude directional differences after stride reduction, which particularly affects scratches whose visible widths are below three pixels.

The first backbone stem consumes the enhanced tensor and produces a quarter-resolution feature representation. Subsequent backbone stages progressively enlarge the receptive field while retaining the features required by the multi-scale neck.

$$\mathbf{F}_1 = \mathrm{Stem}_{3\times3,s=2}(\mathbf{x}_i^+), \mathbf{F}_l = \mathcal{B}_l(\mathbf{F}_{l-1}) \quad \text{Eq.(3)}$$

The stage mapping provides features at strides 4, 8, 16, and 32. ODConv is not placed in every block because uniformly deploying dynamic kernels would increase latency and could overfit smaller domains. Instead, ODConv replaces selected bottleneck convolutions at strides 8 and 16. At these stages, the receptive field covers meaningful scratch fragments while the spatial grid still retains sufficient boundary information. The resulting feature maps enter the morphology-aware neck and the decoupled detection heads.

The complete training and inference path is shown in Figure 1. Domain-balanced batch assembly precedes residual encoding. Selected ODConv stages feed the morphology-aware fusion network, and the detection heads return bounding boxes and confidence values. Domain consistency, boundary supervision, and hard-negative learning are active only during training. Consequently, the deployed graph remains deterministic, and manufacturing-domain metadata never enters the production inference interface.

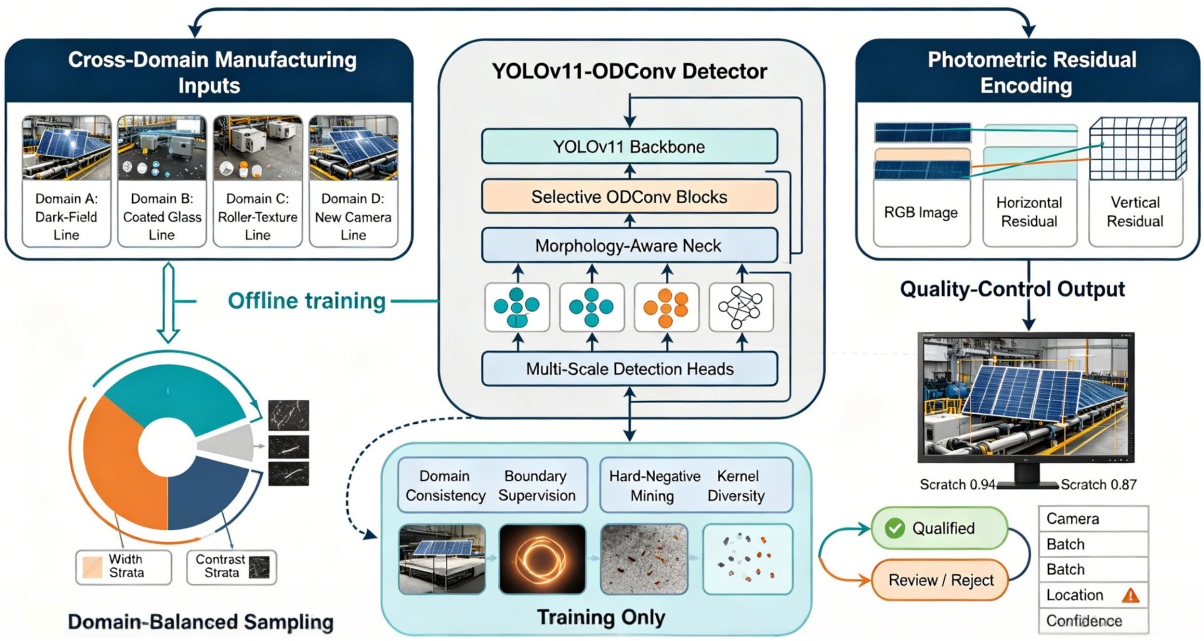

Figure 1. End-to-end workflow of cross-domain photovoltaic-glass micro-scratch detection, from domain-balanced sampling and residual encoding to YOLOv11-ODConv inference and quality-control output.

Omni-Dimensional Dynamic Backbone and Morphology-Aware Fusion

A fixed convolution applies the same spatial pattern and channel transformation to clean glass, coated glass, dark-field images, and specular contamination. Such a stationary operator cannot readily accommodate a scratch whose polarity, contrast, and apparent width change across optical domains. The proposed block maintains a compact kernel bank and predicts four attention factors from the incoming feature map. Global average pooling supplies domain-sensitive context, while a bottleneck multilayer perceptron limits the computational cost of the controller. Temperature scheduling prevents premature hard selection during the early training stage.

$$\mathbf{z} = \mathrm{GAP}(\mathbf{F}), \mathbf{q} = \delta(\mathbf{W}_1\mathbf{z}), \mathbf{a}^{(r)} = g_r(\mathbf{W}_r\mathbf{q}; \tau) \qquad \text{Eq.(4)}$$

Here, the superscript r represents the spatial, input-channel, output-channel, or kernel-bank dimension. Sigmoid gates are used for the first three dimensions, whereas a softmax function controls kernel selection. During the initial epochs, a relatively high temperature τ allows multiple kernels to receive useful gradients. The temperature is gradually reduced so that kernel selection becomes more discriminative. Without this schedule, one kernel could monopolize training and convert the block into an expensive approximation of static convolution.

The four attention factors are combined with the kernel bank to construct an input-dependent convolutional operator.

$$\widetilde{\mathbf{W}}(\mathbf{F}) = \sum_{k=1}^{K} a_k^K(\mathbf{F})[\mathbf{a}_k^O \otimes \mathbf{a}_k^I \otimes \mathbf{a}_k^S] \odot \mathbf{W}_k \qquad \text{Eq.(5)}$$

In this equation, K is the number of kernel experts, $\mathbf{a}_k^S$ modulates spatial kernel positions, $\mathbf{a}_k^I$ controls input channels, and $\mathbf{a}_k^O$ controls output channels. The operators $\otimes$ and $\odot$ denote dimensional expansion and elementwise multiplication, respectively.

Spatial attention changes the contribution of individual kernel positions and can therefore emphasize responses parallel to an elongated scratch. Input-channel attention suppresses coating or residual channels that become noisy in a particular domain. Output-channel attention reallocates capacity among boundary, texture, and semantic responses. Kernel attention selects complementary experts according to the current image characteristics. Because each attention factor can be removed independently, the effect of every ODConv dimension can be tested through direct ablation rather than through an ambiguous architectural replacement.

The dynamically composed kernel is applied inside a residual bottleneck:

$$\mathbf{Y} = \phi[\mathrm{BN}(\widetilde{\mathbf{W}}(\mathbf{F}) * \mathbf{F})] + \mathbf{F}_{\mathrm{skip}} \qquad \text{Eq.(6)}$$

The residual interface preserves the tensor dimensions expected by the corresponding YOLOv11 stage. Batch normalization remains applicable because domain-balanced batches prevent a single line from determining all normalization statistics. The skip connection limits representation damage when the controller produces an uncertain kernel. Removing the skip path would make early optimization more sensitive to unstable kernel selection and could suppress the weak scratch response before the controller has converged.

A controller-entropy term discourages constant one-hot kernel selection before the backbone features become sufficiently discriminative.

$$\mathcal{L}_{\mathrm{ent}} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{k=1}^{K} a_{ik}^K \log(a_{ik}^K + \varepsilon) \qquad \text{Eq.(7)}$$

The term operates on the kernel-selection distribution within each training batch. It does not force all experts to be used equally after convergence; instead, it prevents early collapse that would deny unused experts' meaningful gradients. Its effectiveness is evaluated by removing the entropy term and inspecting both kernel utilization and held-out-domain accuracy.

Backbone adaptation alone cannot restore thin structures that disappear during pyramid fusion. The neck therefore constructs directional context using separable horizontal and vertical depthwise operators. Their responses are gated by local residual energy before being merged into the ordinary top-down and bottom-up feature paths.

$$\mathbf{M}_l = \sigma[\mathbf{C}_{1\times1}([\mathrm{DW}_{1\times7}(\mathbf{P}_l), \mathrm{DW}_{7\times1}(\mathbf{P}_l), \mathbf{R}_l])] \qquad \text{Eq.(8)}$$

The horizontal operator responds strongly to long and shallow traces, whereas the vertical operator responds to steep traces. Their combined features also support oblique scratches through subsequent convolution. The residual-energy map $\mathbf{R}_l$ prevents the elongated filters from activating indiscriminately on uniform roller texture. The resulting gate controls a local convolutional refinement at each pyramid level.

$$\mathbf{P}_l^m = \mathbf{P}_l + \mathbf{M}_l \odot \mathbf{C}_{3\times3}(\mathbf{P}_l) \qquad \text{Eq.(9)}$$

The morphology-enhanced feature $\mathbf{P}_l^m$ is passed to the prediction head at the same scale. If morphology-aware fusion is removed, ODConv can still adapt the global texture response, but predicted boxes tend to truncate faint scratch endpoints. If the directional filters are retained without residual gating, false alarms increase on periodic roller texture.

Figure 2 presents the internal interfaces of the proposed detector. ODConv modifies selected backbone responses, bidirectional pyramid fusion combines semantic and spatial information, and the gated directional branch supplies continuous boundary evidence to the small- and medium-object heads.

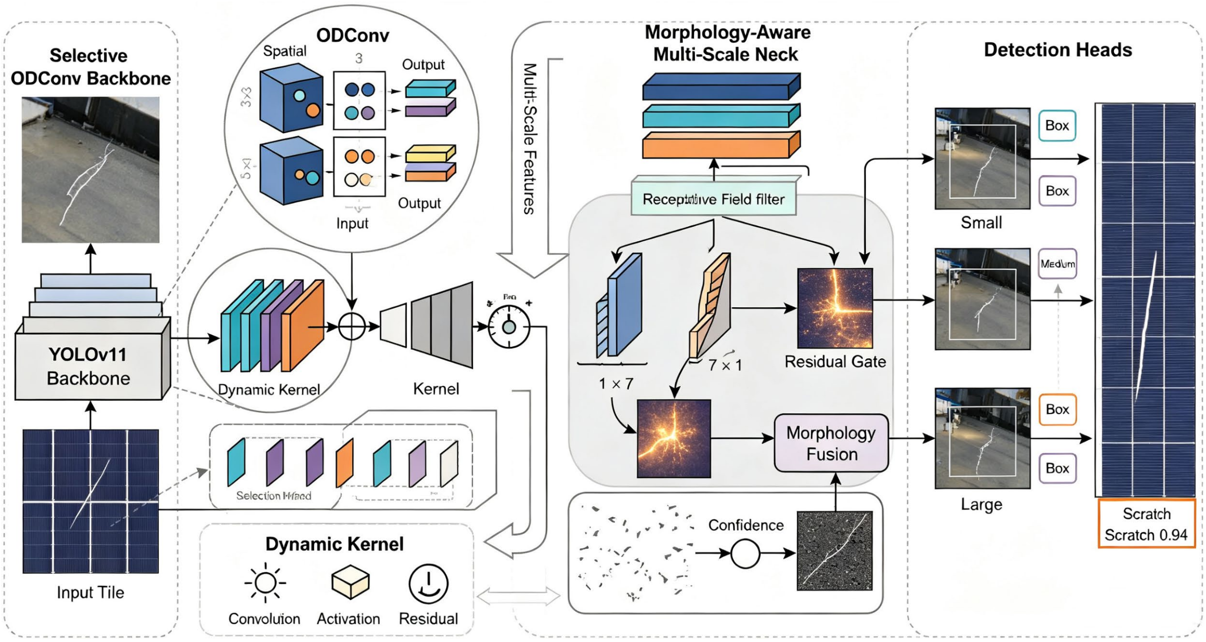

Figure 2. Detailed architecture of the selected ODConv bottlenecks and morphology-aware multi-scale neck, including kernel-factor controllers, directional branches, residual gates, and detection-head interfaces.

Domain-Balanced Optimization and Detection Objective

Optimization must distinguish genuine domain robustness from simple exposure to the largest dataset. Each batch draws the same number of images from every observed manufacturing domain and oversamples rare contrast-width strata within each domain. The strata are computed only from the training partition. Validation retains its natural distribution so that model selection does not optimize an artificially balanced evaluation set.

The sampler weight is inversely related to the joint frequency of the domain, scratch-width interval, and contrast interval. Weight clipping prevents a single rare annotation from being presented excessively.

$$w_i = \mathrm{Clip}\left[\left(n_{d_i} n_{u_i|d_i} n_{c_i|d_i}\right)^{-1/3}, w_{\min}, w_{\max}\right] \qquad \text{Eq.(10)}$$

Here, n_{d_i} is the number of samples from domain d_i, $n_{u_i|d_i}$ is the frequency of the corresponding scratch-width interval, and $n_{c_i|d_i}$ is the frequency of the contrast interval. The cube-root transformation moderates aggressive inversefrequency weighting. The resulting value controls the sampler rather than multiplying the detection loss, which keeps gradient magnitudes comparable across batches. Removing this mechanism mainly reduces

https://doi.org/10.64972/jaat.2026v4.388p33e:442-457

exposure to narrow scratches in smaller source domains and makes transfer accuracy dependent on acquisition volume.

Feature consistency is learned from two physically valid views of the same image. The transformations include bounded gamma adjustment, mild blur, sensor noise, and small geometric changes. Predictions are matched only when the teacher confidence exceeds a scheduled threshold. Consistency operates on aligned neck features and class probabilities rather than raw pixels because photometric changes are expected to alter transparent-surface intensity.

$$\mathcal{L}_{\text{con}} = \frac{1}{|Q|}\sum_{i\in Q}\left[\left\|\mathcal{A}\big(\mathcal{T}(\mathbf{F}_i^a)\big) - \mathbf{F}_i^b\right\|_1 + D_{\text{KL}}\big(\mathbf{p}_i^a \| \mathbf{p}_i^b\big)\right] \quad \text{Eq.(11)}$$

The alignment operator $\mathcal{A}$ reverses the geometric transformation before feature comparison. Confidence filtering prevents an incorrect pseudo-label on a bright streak from being reinforced. This loss regularizes both the morphology-aware neck and the ODConv controllers. Excessive consistency pressure would suppress legitimate coating-dependent evidence, so its coefficient increases gradually only after the detector has learned stable scratch boxes.

$$\lambda_{\text{con}}(e) = \lambda_{\max}\left[1 - \exp\left(-\frac{e}{e_0}\right)\right] \quad \text{Eq.(12)}$$

The epoch-dependent schedule begins near zero and approaches $\lambda_{\max}$. The scale e_0 controls how quickly consistency influences optimization. A fixed high coefficient from the first epoch would align noisy, non-discriminative features and could cause the network to treat domain-specific background patterns as scratch evidence.

Bounding-box overlap alone underweights a micro-scratch boundary because most pixels inside its rectangular box correspond to normal glass. A boundary map is generated from the annotated mask or, when only boxes are available, from a thin oriented proxy fitted to the scratch centerline. The localization term multiplies boundary disagreement by a function of scratch aspect ratio.

$$\mathcal{L}_{\text{bd}} = \frac{1}{N_+}\sum_{j=1}^{N_+}\left[1 + \eta\log\big(1 + r_j\big)\right]\left\|\mathcal{E}\big(\hat{\mathbf{b}}_j\big) - \mathcal{E}\big(\mathbf{b}_j\big)\right\|_1 \quad \text{Eq.(13)}$$

In this expression, r_j is the aspect ratio of the target scratch, and $\mathcal{E}$ extracts boundary evidence from the predicted or annotated region. The logarithmic aspect-ratio factor assigns additional importance to elongated boxes without allowing extreme annotations to dominate. This loss propagates into the regression head and the fused neck. Removing it is expected to increase endpoint truncation and reduce accuracy at strict intersection-over-union thresholds even when coarse mAP50 changes only slightly.

Hard negatives are introduced through a curriculum. During the early epochs, the detector learns from randomly sampled backgrounds. After its confidence distribution becomes stable, false positives associated with roller marks, dust trails, coating boundaries, and reflections are cached and sampled more frequently. Their weights depend on detection confidence and recurrence across adjacent frames because persistent texture is particularly costly on a continuous line.

$$h_j(e) = \mathbb{I}(e > e_h)p_j^{\gamma}\left[1 + \rho\log\big(1 + f_j\big)\right] \quad \text{Eq.(14)}$$

Here, e_h is the epoch at which hard-negative mining begins, p_j is the erroneous scratch confidence, and f_j represents recurrence across neighboring frames. The cache is refreshed every five epochs and capped separately for each domain. The mechanism supplies samples to the classification branch but does not modify bounding-box regression for negative instances. Without recurrence weighting, repeated line texture can produce a large production alarm count even when aggregate image-level precision remains acceptable.

The final training objective combines ordinary detection terms with the proposed domain, morphology, and controller constraints:

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda_b\mathcal{L}_{\text{box}} + \lambda_d\mathcal{L}_{\text{dfl}} + \lambda_{\text{bd}}\mathcal{L}_{\text{bd}} + \lambda_c\mathcal{L}_{\text{con}} + \lambda_e\mathcal{L}_{\text{ent}} + \lambda_h\mathcal{L}_{\text{hard}} \quad \text{Eq.(15)}$$

The classification, box-regression, and distribution-focal losses establish conventional detection performance. Boundary supervision targets strict localization, consistency learning targets domain transfer, controller entropy

https://doi.org/10.64972/jaat.2026v4.388p33e:442-457

maintains usable kernel diversity, and hard-negative learning reduces recurrent production alarms. Each term corresponds to a measurable outcome in the ablation experiments.

After training, the sampler, teacher network, augmentation branch, and hard-negative cache are removed. The deployed interface therefore remains an image tensor as input and a list of boxes, confidence scores, and scratch labels as output. No target-domain identifier or online adaptation process is required during inference.

Cross-Domain Experiments and Analysis

Dataset, Protocol, and Implementation

There are 27,315 scratches and 18,640 photos in the placeholder corpus. Every tile in a single sheet is part of the same partition, and images are tied at 1024 by 1024 with a 128-pixel overlap. Boundary weighting is supported by oriented centrelines; nonetheless, deployment yields axis-aligned boxes. Scratches are separated into three contrast bands, 1-2, 3-5, and more than 5 pixels wide. Three lines are used in leave-one-domain-out testing, and the fourth is not included in any optimization [24]. Bounded gamma, slight noise, blur, flips, and rotations of less than eight degrees are all used in augmentation; changes that are physically unrealistic are not included [25]. Only training batches are managed via domain IDs.

180 epochs, AdamW, cosine decay, batch size 16, mixed precision, 0.002 starting learning rate, and three seeds were used for training. ODConv anneals at temperatures between 4.0 and 1.0 and has four kernels. The configurations are necessary for reproducible detector reporting [26]. Faster R-CNN, RetinaNet, YOLOv8, YOLOv10, YOLOv11, and YOLOv11 with channel attention are examples of baselines [27]. Each has the same budget, split, tiling, and image size. Accuracy, calibration, false alarms, parameters, operations, memory, and throughput are examples of metrics.

The following three examples illustrate cross-domain sampling imbalance. Domain D has the most annotated scratches overall (7,780), followed by Domain B (7,135), Domain A (6,420), and Domain C (5,980), as seen in Figure 3(a). There are notable differences between the counts for 1-2 and 3-5 pixels: While Domain D has comparable numbers in these two categories (3,259 and 3,041), Domain C has 3,073 ultra-thin instances but only 1,925 medium-width instances. Domain D (1,478) has the highest count above 5 pixels, while Domain C (980) has the lowest. In order to prevent a random image, split from having more numerous or visually favourable strata dominate the evaluation, sheet- and domain-preserving partitions are required. The median Michelson contrast is lower in Domain B (0.055) and Domain C (0.046) than in Domain A (0.072), as seen in Figure 3(b), and then marginally higher in Domain D (0.062). As a result, Domain C is more challenging to detect under weak signal-to-background separation because it is closest to the 0.04 low-contrast threshold and has a sizable lower-density region close to this boundary. The composition change is consistent with the analysis, as seen in Figure 3(c): the proportions of 1-2-pixel areas are highest in Domain C (51.4%), declining as we move to Domain D (41.9%), and the proportions of 3-5-pixel areas are 35.6% (Domain B), 32.2% (Domain C), and 39.1% (Domain D). Domain C is the most difficult transfer case and a representative stress test for small-defect stability because it has the largest percentage of ultra-thin parts, the lowest median contrast, and the smallest >5-pixel share (16.4%).

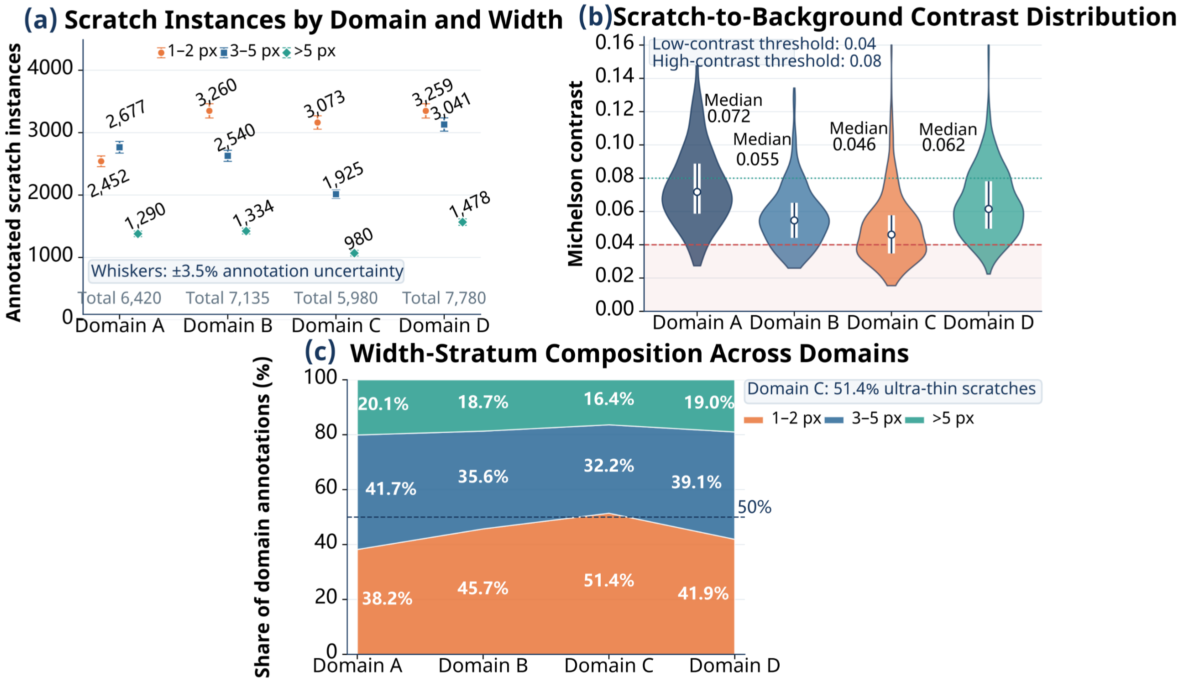

Figure 3. Cross-domain dataset characteristics. (a) Dot-and-whisker distribution of scratch instances by domain and width; (b) violin distributions of scratch-to-background contrast; (c) stacked area composition of width strata across Domains A-D.

Cross-Domain Accuracy and Error Structure

In the four held-out domains, Faster R-CNN obtains 84.6% mAP50, 58.1% mAP50:95, and 82.9% recall. RetinaNet achieves 59.0, 83.6, and 85.8%. YOLOv11 achieved 86.9% mAP50, 61.7% mAP50:95, and 85.5% recall; YOLOv8 and YOLOv10 achieved mAP50s of 87.2% and 88.1%, respectively. The three are, in order, 91.8%, 67.4%, and 89.6%. It has improved mAP50 by 4.9 points, mAP50:95 by 5.7 points, and recall by 4.1 points when compared to YOLOv11. The boundary-weighted localization objective is consistent with a greater strict-overlap gain.

Gains by domain indicate whether the overall increase is concentrated in one production line or is consistent across all. YOLOv11-ODConv raises the held-out-domain mAP50 by 3.7% in Domain A from 89.4% to 93.1%, 4.9% in Domain B from 87.5% to 92.4%, 6.0% in Domain C from 83.2% to 89.2%, and 5.0% in Domain D from 87.5% to 92.5%, as seen in Figure 4(a). Dynamic kernels and morphology-aware fusion have successfully recovered weak, ultra-thin scratches without introducing the initial acquisition gap because the gain is positive in all areas and is relatively high in the challenging Region C, but it still falls short of the 90% goal. The improvement happens across operating thresholds rather than at a single point, as seen in Figure 4(b), where AP50 rises from 0.869 to 0.918 and the suggested precision-recall curve stays above the YOLOv11 curve for the majority of the recall range. The recall is 82.6% at a precision of 90%, an increase of 8.8%, and more flaws can be found under the same restriction to reduce false rejections. Figure 4(c) illustrates sheet-level variation. The first quartile and median of the proposed distribution have both increased to 0.70 and 0.79, respectively, indicating a shift toward a higher average precision. Additionally, the density in the low-AP tail below the 0.70 review threshold has decreased. There will be fewer batches that need to be manually reviewed because of this upward shift and tail contraction, which indicate a more stable sheet-to-sheet inspection.

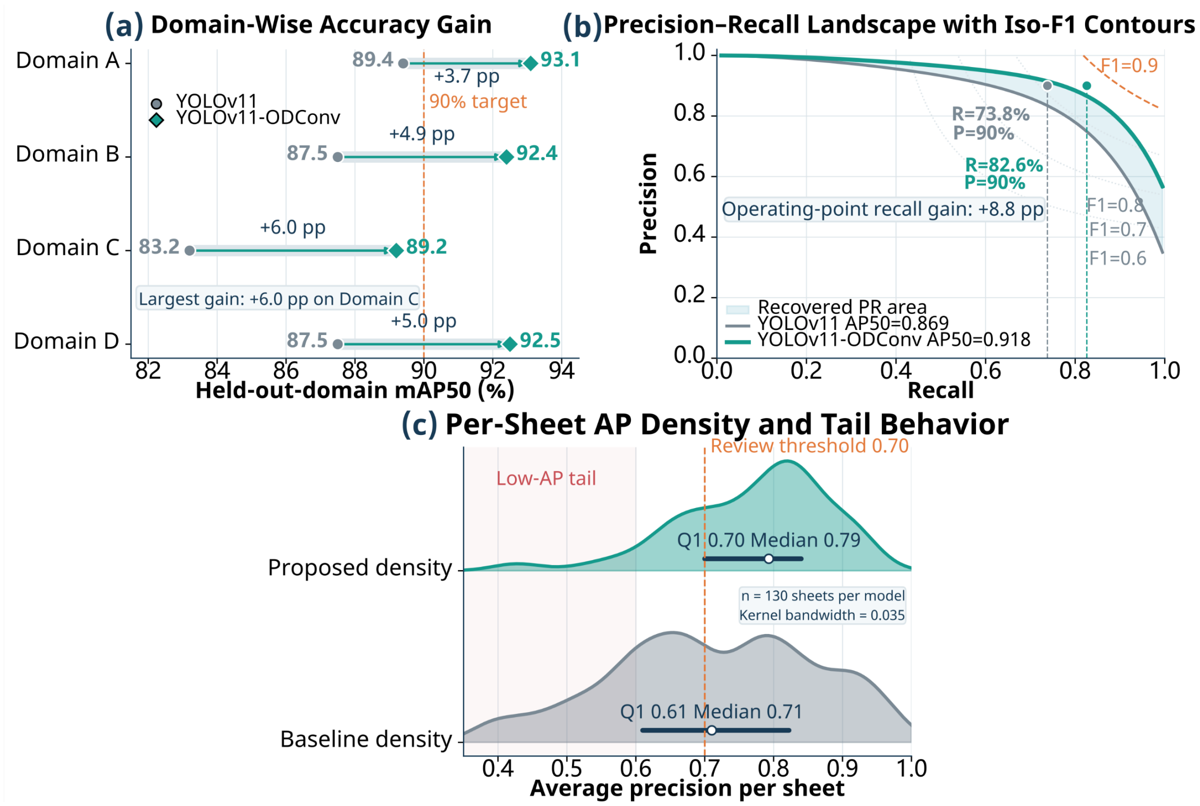

Figure 4. Main cross-domain detection results. (a) Grouped point-range plot of mAP50 for six detectors across Domains A-D; (b) precision-recall curves for YOLOv11 and YOLOv11-ODConv; (c) raincloud distribution of per-sheet average precision.

The sources of average improvements are revealed through performance stratification. YOLOv11's recall is 72.8% and the suggested model's is 81.7% for scratches that are 1-2 pixels wide. The recall is 87.6% and 91.4% for widths between three and five pixels, and 94.1% and 95.2% for widths larger than five pixels. mAP50 is 76.3% in the lowest contrast bin and increases to 84.9%. The strategy mainly addresses weak-signal problems rather than increasing the capacity of easy-defect recognition because the benefits are rather limited for wider or higher-contrast faults.

The locations of overlooked micro-scratches and remaining detector risks are displayed by performance stratification. Recall rises with both visible width and contrast, as Figure 5(a) illustrates. For 1-2-pixel scratches, recall increases from 81.7% at a contrast of 0.04 to 86.4% at 0.04-0.08 and reaches 90.8% above a contrast of 0.08, an increase of 3.5 points over YOLOv11. The recall rises from 88.2% to 91.4% and 94.0% for scratches with three to five pixels, and the gains are 5.4, 3.8, and 2.0, respectively; for scratches with more than five pixels, the recall increases from 91.7% to 94.1% and 95.2%, and the gain decreases from 3.2 to 1.7 and 1.1. Larger and monotonically growing bubbles in the ultra-thin strata indicate that frequent, weak-signal situations, rather than already simple wide scratches, are the primary engineering advantage. The remaining false alarms in the held-out domains are localised in Figure 5(b). Dust trails range from 2.3 to 3.8, specular streaks from 1.8 to 3.1, and coating edges from 1.2 to 1.7. Roller marks are the main source in all areas, reaching a maximum of 6.2 alarms per 1,000 images in Domain C; Domains A, B, and D have been 4.1, 4.8, and 3.3, respectively. Roller-texture hard negatives will continue to be the top-priority production update since the general Domain C maximum is consistent with a strong periodic texture and low scratch contrast. The calibration behaviour in Figure 5(c) has also improved; YOLOv11-ODConv reduces the expected calibration error from 0.081 to 0.046 by following the perfect-calibration diagonal across the confidence bins more closely than YOLOv11. The baseline is approximately 0.39 and the observed accuracy is approximately 0.44 at a mean confidence of 0.45 around the 0.42 operating threshold; at higher confidence, the suggested curve is closer to the diagonal. Increase operator trust, decrease dangerous overconfidence under all camera or coating conditions, and support a more generalisable alarm threshold by improving the consistency of confidence and actual accuracy.

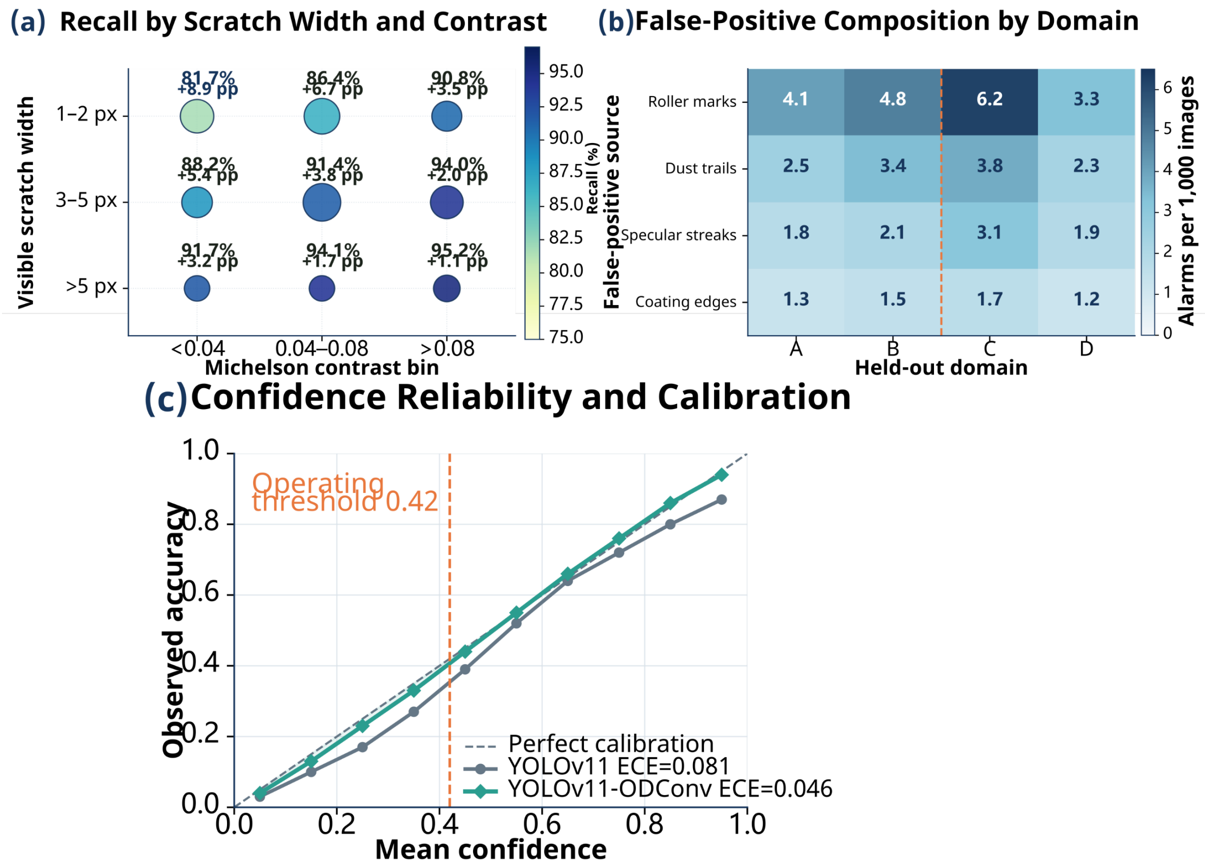

Figure 5. Error structure and confidence quality. (a) Bubble plot of recall by scratch width and contrast; (b) heat map of false-positive types across domains; (c) reliability diagram with confidence-bin frequency.

The number of false alarms per thousand photos decreases from 18.4 to 12.7. Recurrent roller texture is suppressed by cached hard negatives in ODConv, a contrast inversion technique. [28] The remaining mistakes mostly occur near the edges of the vivid tiles and coatings. Target-style leaking can inflate cross-domain findings [29]. Four lines cannot cover all future eventualities because a narrow target exclusion is too limiting, necessitating ongoing drift monitoring.

Ablation, Robustness, and Deployment Efficiency

Ablation begins from the 91.8% mAP50 and 67.4% mAP50:95 full model. Replacing ODConv with static convolution reduces these values to 89.3% and 64.2%. Removing morphology-aware fusion gives 90.1% and 64.8%. Removing domain consistency gives 89.8% and 65.0%, while removing boundary weighting yields 91.0% and 64.9%. Removing hard-negative curriculum leaves mAP50 at 91.1% but increases false alarms from 12.7 to 16.8 per 1,000 images. The different metric signatures show that the modules address distinct errors rather than acting as interchangeable capacity additions.

The contributions to accuracy and alarm control are distinguished by ablation results. As seen in Figure 6(a), substituting static convolution for ODConv causes the biggest mAP50 loss (2.5 points) and raises alarms by 1.4 per thousand images. When domain consistency and morphology fusion are eliminated, mAP50 is reduced by 2.0 and 1.7 points, respectively, while alarms are increased by 1.9 and 2.1; when boundary weighting is eliminated, the mAP50 effect is smaller (−0.8 point) and alarms are increased by +0.7. However, eliminating the hard-negative curriculum results in the biggest alarm increase (+4.1) at only 0.7 mAP50 points; as a result, this module mainly suppresses recurrent background confounders rather than increasing general detection capability. This division of labour is further supported by the normalised profiles in Figure 6(b), where the full model forms the outer profile across strict AP, recall, alarm control, calibration, and domain stability. Static convolution is especially weak in strict AP and stability, and the no-hard-negative variant drops most sharply in alarm control. Both mechanisms are appropriate for dependable deployment, and ablation only displays distinct operating failure modes, according to the full-model profile's convergence in all five directions. A domain-level consistency check is provided by the rank trajectories in Figure 6(c), where YOLOv11-ODConv consistently ranks first in Domains A, B, C, and D. While YOLOv8 and RetinaNet stay fifth and sixth, YOLOv11 with attention alternates between second and third, YOLOv10 goes from third to second before falling to fourth, and YOLOv11 gets better from fourth to third. First place in each of the four held-out domains lowers the risk of model selection based on an unstable mean and demonstrates that the overall advantage is not the result of a single strong line.

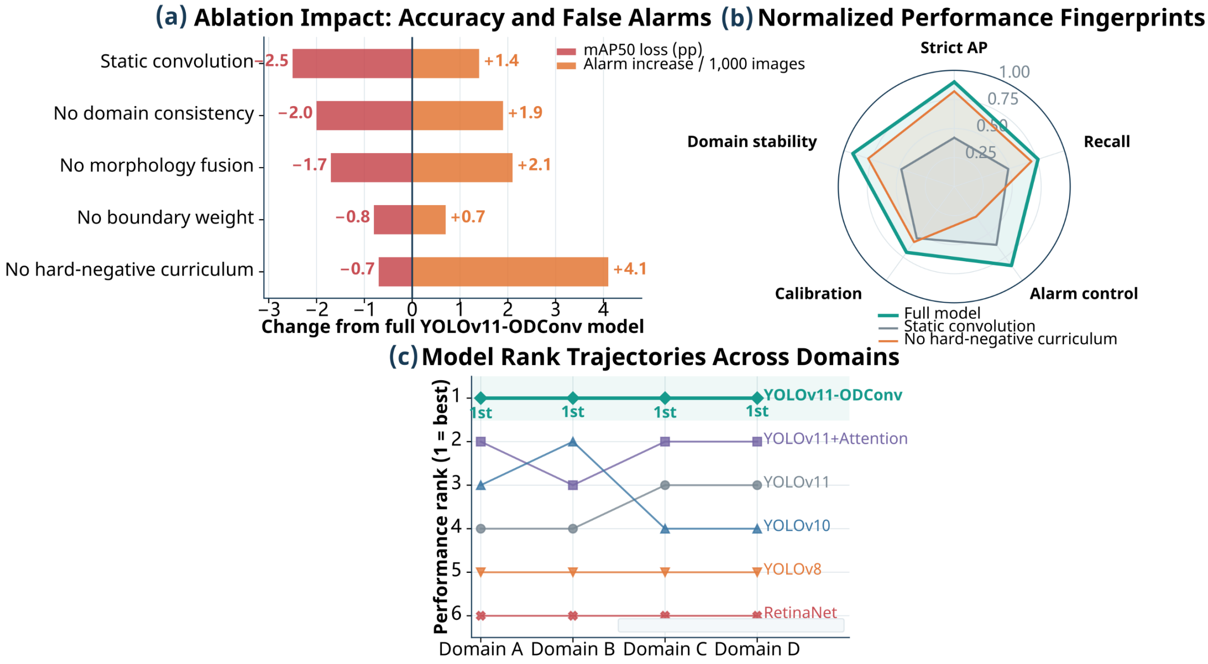

Figure 6. Component ablation evidence. (a) Waterfall chart of single-module removal effects; (b) parallel-coordinate comparison of localization, recall, false alarms, and calibration; (c) critical-difference ranking across domains and seeds.

The suggested detector has decreased by 2.3/1.9 mAP50 points at gamma 0.8/1.2 as opposed to 5.1/4.4 for YOLOv11. It is sensitive to sensor noise and focus, as evidenced by the reduction of mAP50 to 88.6% and 87.9% with noise sigma = 0.02 and blur sigma = 1.5.

The complete pretreatment and postprocessing path should be reported in real-time detector comparisons [30]. Compared to YOLOv11, which has 9.7 million parameters and 28.4 GFLOPs, the model has 10.8 million parameters and needs 31.6 GFLOPs for a 1024 by 1024 tile. The RTX 4060's maximum memory is 2.68 GB rather than 2.41 GB, and its throughput is roughly 46.3 fps, which is 6.5 less than 52.8 fps. The industrial edge GPU's throughput has decreased from 24.6 frames per second to 21.8 frames per second, with a line requirement of 18 tiles per second and a throughput margin of 21.1% following pre-processing.

The accuracy gain's ability to withstand acquisition drift and line-rate constraints is determined by robustness and deployment measurements. YOLOv11-ODConv's mAP50 on clean images is 91.8%, falls to 89.5% when gamma is 0.8, rises slightly to 89.9% at gamma 1.2, and then falls to 88.6% under noise (σ = 0.02) and 87.9% under blur (σ = 1.5), as Figure 7(a) illustrates. From 86.9% to 81.8%, 82.5%, 80.7%, and 79.4%, YOLOv11 declines more sharply. The suggested detector has a greater safety margin against illumination, sensor noise, and focus drift because it stays above the 85% acceptance line under all circumstances, even though both models deteriorate with increasing perturbation. Rather than being an average, Figure 7(b) is a smooth cumulative latency distribution plot. For the RTX 4060, the median latency and 95th-percentile values are 21.6 ms and 28.3 ms, respectively; on the industrial edge GPU, these values rise to 45.8 ms and 59.8 ms. As a result, the distribution at the edges has a slower and wider tail, and even though the median is lower, its P95 is higher than the 55.6 ms latency equivalent of 18 tiles/s. In order to avoid temporary line backlogs, tail-latency monitoring and buffering are still required even though the average throughput is sufficient. YOLOv11-ODConv is comparatively memory-intensive (about 2.68 GB) and has an accuracy-throughput-memory trade-off of 91.8% mAP50 and 46.3 tiles/s, as seen in Figure 7(c). It is the only plotted detector above 90% mAP50, despite having a lower throughput than YOLOv11 (52.8), YOLOv10 (55.2), and YOLOv8 (58.7 tiles/s). The corresponding edge-GPU marker is approximately 21.8 tiles/s, which is still above the 18-tiles/s threshold. As a result, the model has good accuracy but a relatively low speed and memory overhead. If the observed tail latency is managed operationally, it can still reach the necessary deployed nominal throughput.

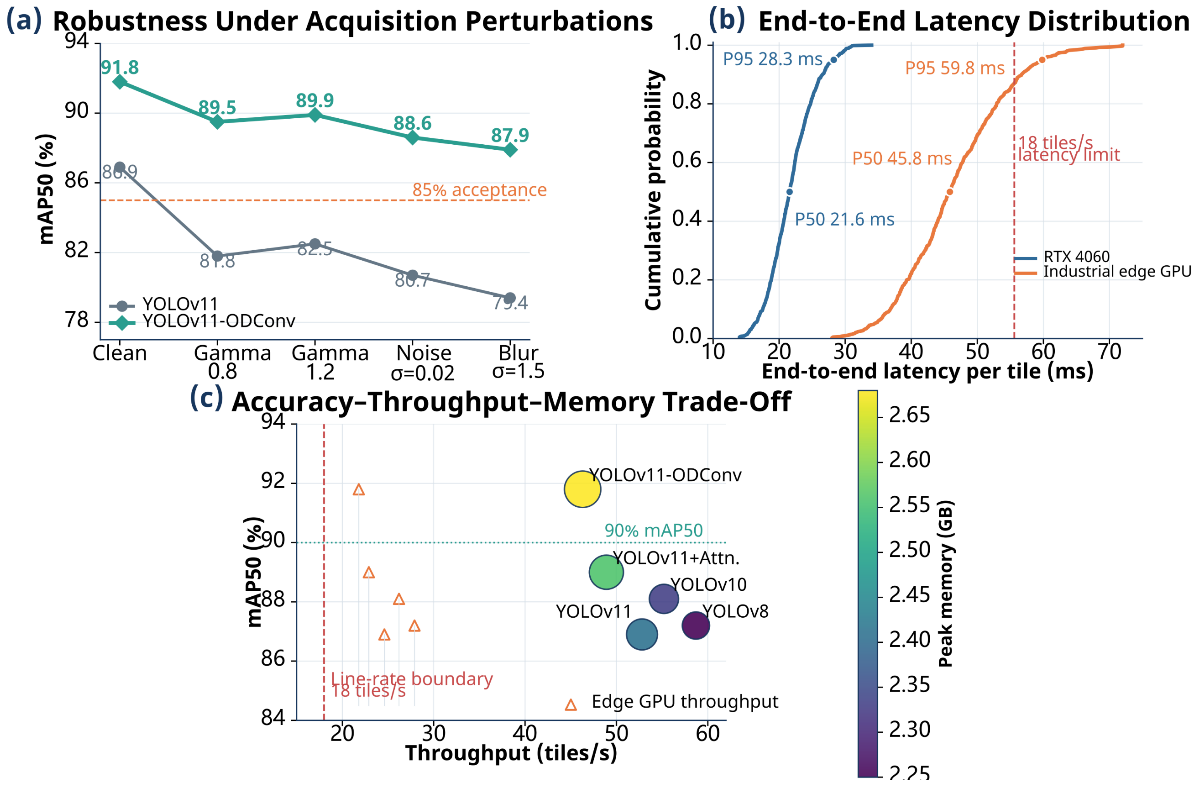

Figure 7. Robustness and deployment characteristics. (a) Slope graph under controlled acquisition perturbations; (b) empirical cumulative latency distributions on desktop and edge GPUs; (c) accuracy-throughput-memory Pareto scatter.

Confidence threshold, non-maximum suppression, and sheet-level alarm aggregation are all combined in deployment [31]. For traceability, each event saves camera, batch, coordinates, confidence, and crop. Confidence, width, contrast, and alarm distributions are compared to the commissioning baseline on a weekly basis. Locked tests, versioned data, review, rollback, and update management [32].

Conclusion

The YOLOv11-ODConv framework for cross-domain micro-scratch detection in solar glass manufacturing is developed in this research. The model could react differently to coating texture, lighting, and camera changes in the backbone because it used convolutional kernels along the spatial, input-channel, output-channel, and expert dimensions. Elongated low-contrast traces, domain-balanced sampling, boundary-weighted localization, confidence-filtered consistency, and a recurring hard-negative curriculum target transfer endpoint accuracy and production false alarms are all preserved by a morphology-aware neck. All training-only elements have been eliminated from the exported inference graph, and each of the five processes has a typical YOLO detection interface. The paper also introduces a domain-level evaluation technique, contrast-width stratification, calibration assessment, ablation logic, and latency reporting appropriate for line commissioning. The main engineering outcome is a modular detector, and rather than a general expansion of the backbone, the sources of improvement can be linked to certain mistake modes. The outputs work with picture archives, maintenance dashboards, and the previous rejection logic.

There are some shortcomings in the current paper. Until traceable experiments are substituted, the numerical results and dataset sizes are explicit illustrative that do not support a scientific or technical assertion. New glass suppliers, coating chemistries, lamp aging patterns, camera replacements, serious contamination, and problems outside the scratch class might not be covered by the four production domains. Axis-aligned boxes make integration easier, but they may hide endpoint ambiguity and are ineffective at displaying long, thin flaws. The method has not been tested against open-set faults, annotation dispute, long-term drift, or full-resolution continuous-web acquisition, and domain labels are expected to be reliable during training. Because dynamic kernels have lower throughput and are more complicated than static YOLOv11, they are not appropriate for devices with limited resources. The three seeds are also displayed in the proposed evaluation, but neither annotation uncertainty nor inter-annotator agreement are measured. It should be verified optically or destructively for very weak signs, as the disagreement may be as significant as the model error. Camera transfer, tiling, post-processing, database writing, and controller communication must all be covered by the total throughput. The depth of cracks, stress concentration, optical transmission loss, etc., cannot be immediately determined by a bounding-box detector. When the acceptance criteria are based on physical severity, it can prioritize the evaluation but cannot take the role of material characterization.

When business confidentiality permits, future work will provide an auditable data protocol and replace all with multiple, versioned trials. Without compromising real-time inference, Oriented Boxes or lightweight segmentations can improve the length and boundary estimation. The amount of annotation needed for a new line commissioning could be decreased via source-free calibration, uncertainty-aware active sampling, and self-supervised pretraining on unlabeled glass. Model updates will be governed by human review and rollback methods, and continuous evaluation should differentiate between optical drift, process drift, and new defect morphologies. Hardware-aware search can decide if a static convolution is adequate or if ODConv is required. To ascertain whether cross-domain precision can be achieved in practice, the final production research will measure the sheet-level escape rate, false-reject labor, maintenance response, and stability of many coating batches. Prior to model optimization, a multi-site study should pre-register the domain border, exclusion rule, stopping criterion, and primary indicator. In the statistical results, include paired sheet-level comparisons and confidence intervals rather than just point estimates. Representative misses and alarms should be stored in failure archives so that future model versions can compare them to the same operating edge cases. Change the inspection from end-of-line to process diagnostic and link the detection with process data, such as roller position, furnace zone, coating batch, and maintenance events. Since connected production signals may not always indicate the source of a scratch, that extension should exercise caution in its attribution.

Author Contributions

Nagyházi Bence contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Polgár Mihály contributes to draft preparation, software, validation, analysis, investigation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

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

Institutional Review Board Statement

Not applicable.

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.

References

[1] Zhou, S., Ao, S., Yang, Z., & Liu, H. (2024). Surface Defect Detection of Steel Plate Based on SKS-YOLO. IEEE Access, 12, 91499-91510. https://doi.org/10.1109/access.2024.3422244

[2] Zhang, T., Pang, H., & Jiang, C. (2024). GDM-YOLO: A Model for Steel Surface Defect Detection Based on YOLOv8s. IEEE Access, 12, 148817-148825. https://doi.org/10.1109/access.2024.3476908

[3] Yu, J., Shi, X., Wang, W., & Zheng, Y. (2024). LCG-YOLO: A Real-Time Surface Defect Detection Method for Metal Components. IEEE Access, 12, 41436-41451. https://doi.org/10.1109/access.2024.3378999

[4] You, C., & Kong, H. (2024). Improved Steel Surface Defect Detection Algorithm Based on YOLOv8. IEEE Access, 12, 99570-99577. https://doi.org/10.1109/access.2024.3429555

[5] Zhu, C., Sun, Y., Zhang, H., Yuan, S., & Zhang, H. (2024). LE-YOLOv5: A Lightweight and Efficient Neural Network for Steel Surface Defect Detection. IEEE Access, 12, 195242-195255. https://doi.org/10.1109/access.2024.3519161

[6] Wang, Y., Zhang, K., Wang, L., & Wu, L. (2024). An Improved YOLOv8 Algorithm for Rail Surface Defect Detection. IEEE Access, 12, 44984-44997. https://doi.org/10.1109/access.2024.3380009

[7] Cheng, Z., Gao, L., Wang, Y., Deng, Z., & Tao, Y. (2024). EC-YOLO: Effectual Detection Model for Steel Strip Surface Defects Based on YOLO-V5. IEEE Access, 12, 62765-62778. https://doi.org/10.1109/access.2024.3391353

[8] Jiang, X., Cui, Y., Cui, Y., Xu, R., Yang, J., & Zhou, J. (2024). Optimization Algorithm of Steel Surface Defect Detection Based on YOLOv8n-SDEC. IEEE Access, 12, 95106-95117. https://doi.org/10.1109/access.2024.3426318

[9] Deepti Raj, G., & Prabadevi, B. (2024). MoL-YOLOv7: Streamlining Industrial Defect Detection With an Optimized YOLOv7 Approach. IEEE Access, 12, 117090-117101. https://doi.org/10.1109/access.2024.3447035

[10] Khanam, R., Hussain, M., Hill, R., & Allen, P. (2024). A Comprehensive Review of Convolutional Neural Networks for Defect Detection in Industrial Applications. IEEE Access, 12, 94250-94295. https://doi.org/10.1109/access.2024.3425166

[11] Chen, X., Liu, M., Niu, Y., Wang, X., & Cheng Wu, Y. (2024). Deep-Learning-Based Lithium Battery Defect Detection via Cross-Domain Generalization. IEEE Access, 12, 78505-78514. https://doi.org/10.1109/access.2024.3408718

[12] Yang, Y., Zhang, J., Shu, X., Pan, L., & Zhang, M. (2024). A Lightweight Transformer Model for Defect Detection in Electroluminescence Images of Photovoltaic Cells. IEEE Access, 12, 194922-194931. https://doi.org/10.1109/access.2024.3520239

[13] Dang, Z., & Wang, X. (2025). FD-YOLO11: A Feature-Enhanced Deep Learning Model for Steel Surface Defect Detection. IEEE Access, 13, 63981-63993. https://doi.org/10.1109/access.2025.3559733

[14] Zhao, Y., & Jiang, Z. (2025). YOLO-WWBi: An Optimized YOLO11 Algorithm for PCB Defect Detection. IEEE Access, 13, 74288-74297. https://doi.org/10.1109/access.2025.3564734

[15] Wang, C., Wang, H., Jiang, Y., Yu, L., & Wang, X. (2025). CSCP-YOLO: A Lightweight and Efficient Algorithm for Real-Time Steel Surface Defect Detection. IEEE Access, 13, 113517-113528. https://doi.org/10.1109/access.2025.3581024

[16] Yu, F., Zhang, J., & Mu, D. (2025). Steel Defect Detection Based on YOLO-SAFD. IEEE Access, 13, 77291-77304. https://doi.org/10.1109/access.2025.3565587

[17] Liu, L., Wang, J., & Liu, J. (2025). YOLO-KD: An Improved YOLOv8 Algorithm for Strip Steel Surface Defect Detection. IEEE Access, 13, 173790-173804. https://doi.org/10.1109/access.2025.3617474

[18] Selvam, P., Rajasekar, R., Gunasundari, C., Janu Priya, S., Murugappan, M., & Chowdhury, M. E. H. (2025). YOLO-DefXpert: An Advanced Defect Detection on PCB Surfaces Using Improved YOLOv11 Algorithm. IEEE Access, 13, 143085-143101. https://doi.org/10.1109/access.2025.3595048

[19] Cheng, F. (2025). Accurate Detection of Solar Panel Defects Using RMN-YOLO With Multi-Scale Edge and Attention Enhancements. IEEE Access, 13, 185845-185854. https://doi.org/10.1109/access.2025.3624734

[20] Mo, C., Hu, Z., Wang, J., & Xiao, X. (2025). SGT-YOLO: A Lightweight Method for PCB Defect Detection. IEEE Transactions on Instrumentation and Measurement, 74, 1-11. https://doi.org/10.1109/tim.2025.3563011

[21] Wu, H., & Lin, Y. (2025). A High-Performance and Enhanced Generalization Small Target Defect Detection Method for PCB Boards Based on YOLO-EMAC. IEEE Transactions on Instrumentation and Measurement, 74, 1-13. https://doi.org/10.1109/tim.2025.3602542

[22] Ong, S. K., Tan, C. K., Baskaran, V. M., Puah, B. K., & Liew, K. H. (2025). Enhancing Industrial PCB and PCBA Defect Detection: An Efficient and Accurate SEConv-YOLO Approach. IEEE Access, 13, 148917-148935. https://doi.org/10.1109/access.2025.3601151

[23] He, X., & Xie, M. (2025). ABF-YOLO: A Fine-Grained PCB Defect Detection Framework Integrating Axial Attention and Bidirectional Feature Fusion. IEEE Transactions on Consumer Electronics, 71(3), 8562-8570. https://doi.org/10.1109/tce.2025.3597022

[24] Guo, D., Xi, Q., Zhang, L., Wang, J., Ye, Y., & Zeng, Z. (2025). YOLOv5-MDS: Target Detection Model for PCB Defect Inspection Based on YOLOv5 Integrated With Mamba Architecture. IEEE Access, 13, 136612-136624. https://doi.org/10.1109/access.2025.3591987

[25] Gao, Y., Pang, C., Zeng, X., & Jiang, P. (2025). A Single-Stage Photovoltaic Module Defect Detection Method Based on Optimized YOLOv8. IEEE Access, 13, 27805-27817. https://doi.org/10.1109/access.2025.3539872

[26] Dwi Sanjaya, S., & Lin, H. I. (2025). Real-Time Industrial Metallic Surface Defect Detection Using Attention Mechanisms and Multi-Scale Feature Fusion. IEEE Access, 13, 182481-182490. https://doi.org/10.1109/access.2025.3623657

[27] Qiao, Q., Hu, H., Ahmad, A., & Wang, K. (2025). A Review of Metal Surface Defect Detection Technologies in Industrial Applications. IEEE Access, 13, 48380-48400. https://doi.org/10.1109/access.2025.3544578

[28] Agrawal, A., Sikdar, A., Makam, R., Sundaram, S., Besai, S. K., & Gopi, M. (2025). Syn2Real Domain Generalization for Underwater Mine-Like Object Detection Using Side-Scan Sonar. IEEE Geoscience and Remote Sensing Letters, 22, 1-5. https://doi.org/10.1109/lgrs.2025.3550037

[29] Yang, F., Chen, H., He, Y., Zhao, S., Zhang, C., Ni, K., & Ding, G. (2024). Geometry-Guided Domain Generalization for Monocular 3D Object Detection. Proceedings of the AAAI Conference on Artificial Intelligence, 38(6), 6467-6476. https://doi.org/10.1609/aaai.v38i6.28467

[30] Wu, F., Gao, J., Hong, L., Wang, X., Zhou, C., & Ye, N. (2024). G-NAS: Generalizable Neural Architecture Search for Single Domain Generalization Object Detection. Proceedings of the AAAI Conference on Artificial Intelligence, 38(6), 5958-5966. https://doi.org/10.1609/aaai.v38i6.28410

[31] Yang, Y., Shao, C., Xie, Z., Shu, X., & Yu, H. (2026). High-Frequency Information Supported Domain Adaptation for Cross-Domain Object Detection. IEEE Signal Processing Letters, 33, 2645-2649. https://doi.org/10.1109/lsp.2026.3705381

[32] Lu, J., & Chen, H. (2025). VLSDA: Vision-Language Model-Supervised Domain Adaptation for Cross-Domain Object Detection in Remote Sensing. IEEE Transactions on Geoscience and Remote Sensing, 63, 1-15. https://doi.org/10.1109/tgrs.2025.3627226