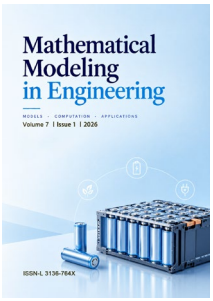
UNIVERSITY OF WEST ATTICA

eISSN: 3136-764X

Mathematical Modeling in Engineering

Vol. 7, No. 1, January, 2026



Diffusion Planner Model for Surface Crack Detection of Quick-Frozen Dumplings in Low Temperature Production Lines

Masato Taniguchi[a,*], Hiroshi Sato[a], Yuichi Yoshida[a]

[a] *Department of Computer Science, Keio University, Yokohama, 223-8522, Japan*

Received 13 December 2025; Revised 03 March 2026; Accepted 08 March 2026; Published 14 March 2026

Abstract

Surface cracks on quick-frozen dumplings are difficult to detect on low-temperature production lines because thin and discontinuous evidence is frequently obscured by frost, flour particles, packaging-film reflections, and nonuniform tunnel illumination. This study proposes a Diffusion Planner that treats inspection as constrained latent crack-path planning rather than direct texture classification or full-image generation. A temperature-normalized encoder first reduces appearance drift across freezer conditions, after which a contour-conditioned diffusion planner connects locally interrupted crack evidence and a compact decoder produces the final pixel-level mask. The planner is restricted to sparse task-relevant trajectories, while a confidence gate suppresses paths that conflict with normal pleat geometry. A production-line dataset containing 18,640 images and 42,310 annotated crack instances across four freezing-temperature bands was used for evaluation with product-sequence and acquisition-session separation. The proposed method achieved 95.6% precision, 94.2% recall, and a 94.9% F1 score, while reducing mean boundary error by 41.3% relative to the strongest conventional detector. At 1024 by 768 resolution, average inference latency was 24.8 ms on an industrial GPU, supporting a 1.2 m/s conveyor. Under heavy frost, diffusion planning increased recall by 3.7 percentage points, and temperature normalization reduced false alarms on legal folds by 28.6%. These results show that process-conditioned latent trajectory planning can provide accurate, interpretable, and deployment-oriented crack inspection for frozen food manufacturing.

Keywords: Quick Frozen Dumplings, Surface Crack Detection, Diffusion Planner, Low Temperature Vision, Industrial Inspection

1. Introduction

Quick-frozen dumplings are produced by means of forming, flouring, tunnel freezing, temporary buffering, weighing and packaging, and the surface condition of the product changes continuously during these processes [1]. A small crack in the wrapper may be negligible at the shaping station; however, after freezing and vibration, it can enlarge to a leakage defect that affects appearance, cooking stability and consumer acceptance [2]. Therefore, low-temperature inspection is not a simple extension of room-temperature food vision because frost crystals, flour residues, and cold-air turbulence alter the image statistics before the product reaches the camera [3]. Traditional sampling inspection cannot address the problem of connecting batch-level quality inspection with piece-level surface control in high-speed production, such as that over several hundred pieces per minute [4]. Machine vision has been applied to the production line of frozen food to reduce the amount of manual sorting, but identifying the crack target remains difficult due to its long, thin shape and local similarity to the normal pleats [5]. In fact, the defect boundary may be interrupted by frost beads or partially covered with powder; thus, a reasonable crack path needs to be inferred by the detector from incomplete evidence [6].

* Corresponding author: masato.tani@keio.jp

https://doi.org/10.68406/mme.2026.vol7iss1nm9:106-118

➢ 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.

The engineering demands are relatively high for real-time operation of the algorithm and stable decision-making under various conditions of shifts, freezers and recipes.

Although the current methods for food defect detection have achieved some results with the application of convolutional networks and transformer backbones and segmentation decoders, most of them assume that the defect texture is directly visible and the background distribution is relatively stable [7]. Quick-frozen dumplings violate both assumptions because the same product surface may have dry flour, condensed ice, a glossy frozen skin and mechanical folds in a single field of view [8]. Edge-based methods expand the cracks but are very sensitive to legal pleat ridges and sealing traces; region-based methods reduce isolated noise but may fail to detect hairline fracture segments that are only a few pixels wide [9]. Recently, generative models have shown that uncertainty in image structure evolution during denoising can be learned, and this characteristic is beneficial for reconstructing defects from discontinuous local cues [10]. Direct diffusion models are generally too expensive for production-line inspection, and if manufacturing data are not available, visually plausible but operationally irrelevant results may be generated [11]. Therefore, this paper introduces a planner-guided diffusion formulation for crack trajectory proposal through denoising and employs a compact deterministic decoder for the final segmentation mask.

This study addresses surface-crack detection in quick-frozen dumplings under low-temperature production-line imaging. The unresolved problem is not only weak local contrast, but also the need to connect discontinuous evidence without following normal pleat ridges. The proposed model therefore plans a sparse latent crack trajectory before mask classification. Its main contributions are a temperature-normalized encoder that accounts for freezer and frost variation, a contour-conditioned diffusion planner that enforces product geometry, and a confidence-aware decoder designed for real-time rejection control. The engineering evaluation combines pixel-level localization, product-level accuracy, calibration, latency, disturbance robustness, and continuous line-trial behavior.

To show the technical progress from the reasons for production to engineering verification, the rest of this paper is arranged as follows. Section 2 reviews relevant research on visual inspection of frozen food and generative defect modelling, focusing on the outstanding problem of fine crack localization and low-temperature robustness. Section 3 introduces the Diffusion Planner model and, in parallel, the production-line imaging workflow, network architecture and training objective. Section 4 is the experimental data set, implementation environment, quantitative analysis, robustness test and ablation results. Section 5 summarizes the research, discusses the limitations and future directions for application in intelligent frozen food production, etc.

2. Related Work

2.1. Visual Inspection of Frozen Food and Dough Products

Due to the demand for automatic inspection in food production, many issues with small-scale defects under varying light conditions, positions, and surface moisture have arisen. Early studies of frozen food used hand-designed color, texture and edge descriptors because these features could be realized by embedded controllers and interpreted by process engineers [12]. Although these methods are still suitable for the coarse filtering of foreign materials, their tolerance to local frost damage and deformation near the defect is relatively low compared with that of normal process marks. Dough products have a folded surface of the wrapper, flour clusters and sealing marks that appear as elongated dark lines; therefore, crack detection requires both local contrast sensitivity and structural discrimination. Some industrial vision systems have integrated ring lighting and threshold-based morphology to improve the visibility of edges; however, when changing the product formula or freezer set point, these pipelines need to be retuned [13]. Deep Convolutional Segmentation improves the feature representation of fine-grained defects by learning multi-scale texture patterns from annotated images. Supervised segmentation is also affected by distribution shift; for instance, a decrease in light intensity will alter the color temperature, specular reflection and size of frost grains suddenly.

Transformer-based inspection models have increased the receptive field and can distinguish between legal pleats and defect-like discontinuities; however, their performance is still affected by the density and diversity of labelled crack masks [14]. Annotation for quick-frozen dumplings is relatively expensive because the crack may be only one to three pixels wide in the raw image, and agreement needs to be reached between visual inspectors and process technicians. Although weakly supervised methods have reduced the cost of annotation, the generated blurry activation maps are not suitable for boundary-level sorting decisions [15]. The other problem is that the inspection model needs to deal with product overlap, conveyor vibration and partial occlusion by flour dust. The above conditions provide semantic context for the cracks but lack visual information; therefore, a model that can reason about the continuity of cracks can be built based on this.

2.2. Generative and Diffusion Models for Defect Detection

Generative models have been extensively applied to industrial anomaly detection by learning the characteristics of normal behavior and identifying deviations based on reconstruction errors [16]. Autoencoders and adversarial methods

are more suitable for problems with a small number of defective samples, but they may reconstruct small cracks too accurately or trigger false alarms due to modifications of harmless textures. Diffusion models are adapted to iteratively denoise latent variables, and at the same time, uncertainty in local image structures can also be expressed [17]. Diffusion-based methods have been employed for defect inspection to generate rare defects, estimate anomaly probability and improve feature robustness. The main limitation of computational cost for the production line is that many denoising steps per frame are required.

Recently, some research has tried to accelerate diffusion inference via latent sampling, distillation and conditional guidance [18]. These ideas can be applied to the inspection of frozen dumplings, but as the desired output is not a photorealistic image, a reliable decision on a narrow crack path is required. A planner-guided Design can use diffusion only in the uncertain region, and then a deterministic decoder produces the final mask. It is not a general synthesis of anomalies because the planner is constrained by crack geometry, production conditions and temperature-normalized features. Constraints are necessary for cold-weather production because frost and flour artifacts may be highly noticeable to the eye but are not serious defects [19]. Based on the above, this paper proposes a compact diffusion planner specifically for the problem of quick-frozen dumpling crack detection and conducts evaluations using process-level data instead of isolated laboratory images.

3. Diffusion Planner Model for Low Temperature Crack Detection

3.1. Production-Line Imaging Workflow

The inspection station is positioned between the freezing tunnel and automatic weighing, where dumpling surfaces are sufficiently stable for imaging and defective products can still be rejected before packaging. Shallow oblique illumination enhances wrapper discontinuities while limiting frost overexposure. Each encoder pulse triggers image acquisition, product-centered cropping, and temperature normalization using the freezer setpoint, enclosure temperature, and image-level frost statistics. Figure 1 shows the workflow from cold-zone imaging to crack-mask prediction and rejection control. The acquisition and inference stages are linked because chamber temperature, flour dosage, and conveyor humidity jointly alter otherwise normal surface features.

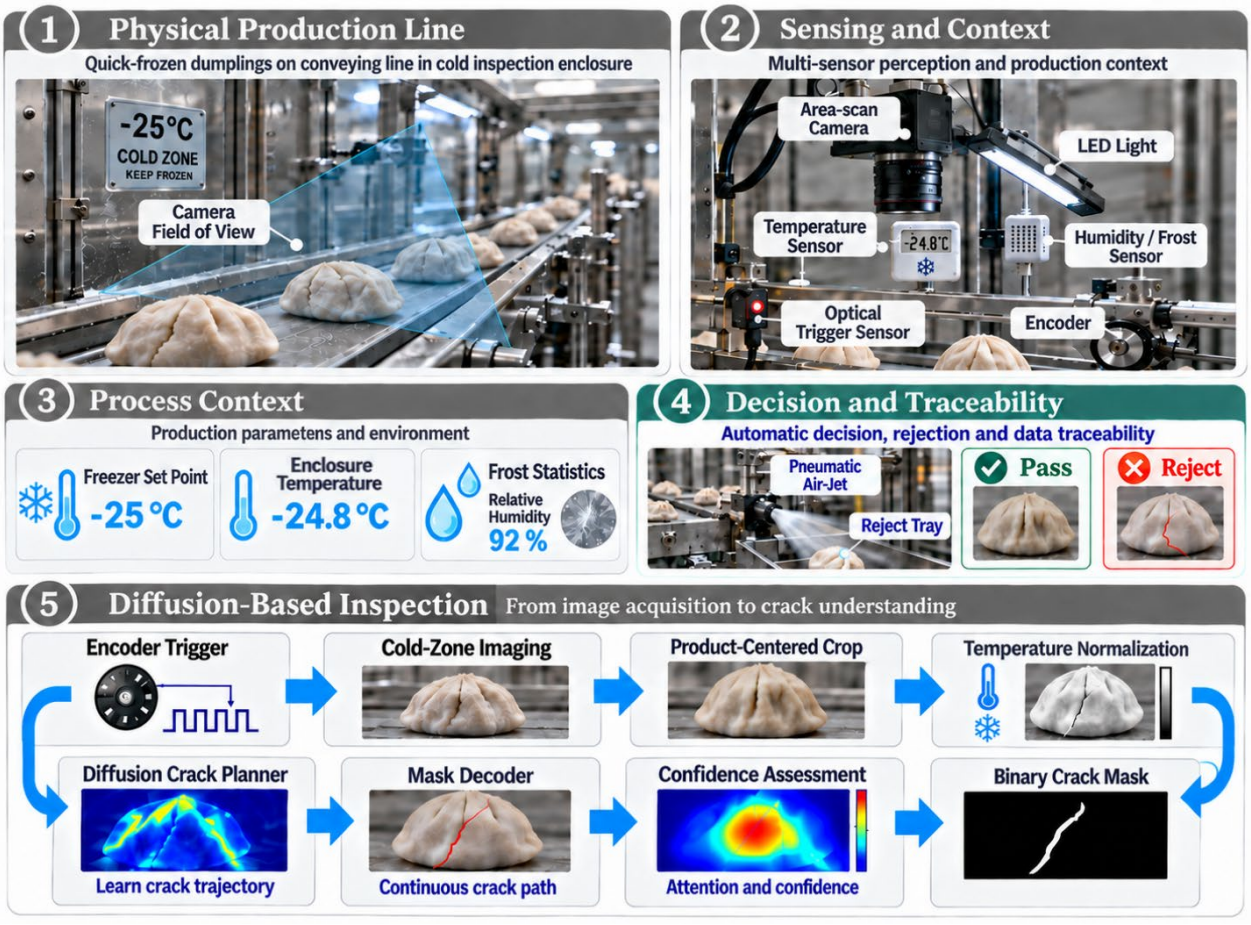


Figure 1. Low-Temperature Production Line Crack Detection Workflow

Let x denote the normalized dumpling image and τ the measured temperature descriptor. The encoder produces a multiscale feature tensor $E(x)$, while the image-level frost statistic s_f summarizes high-frequency crystal responses. The temperature-normalized representation F_τ is obtained by applying feature-wise affine calibration, whose scale and bias are estimated from τ and s_f. The correction is applied to encoder features rather than directly to raw pixels, allowing the model to preserve local crack evidence while compensating for process-dependent appearance drift.

$$F_\tau = \gamma(\tau, s_f)E(x) + \beta(\tau, s_f) \tag{1}$$

Here, E is the visual encoder, s_f is the frost statistic extracted from high-frequency image components, and γ and β are lightweight calibration functions. This formulation keeps the correction differentiable and avoids a hard separation between image preprocessing and feature learning. During deployment, the same module also reduces batch-to-batch bias when the production line changes from one filling recipe to another.

3.2. Conditional Diffusion Planner

The planner operates in a low-dimensional latent space rather than full-resolution pixel space. During training, each annotated crack mask is reduced to a one-pixel skeleton and encoded as a sparse crack-path variable. Forward diffusion corrupts this path, and the reverse process reconstructs it under visual-feature, timestep, and product-contour conditioning. The contour prior is supplied throughout denoising so that proposed trajectories remain within the dumpling boundary, while the image condition retains local evidence near weak or frost-interrupted segments.

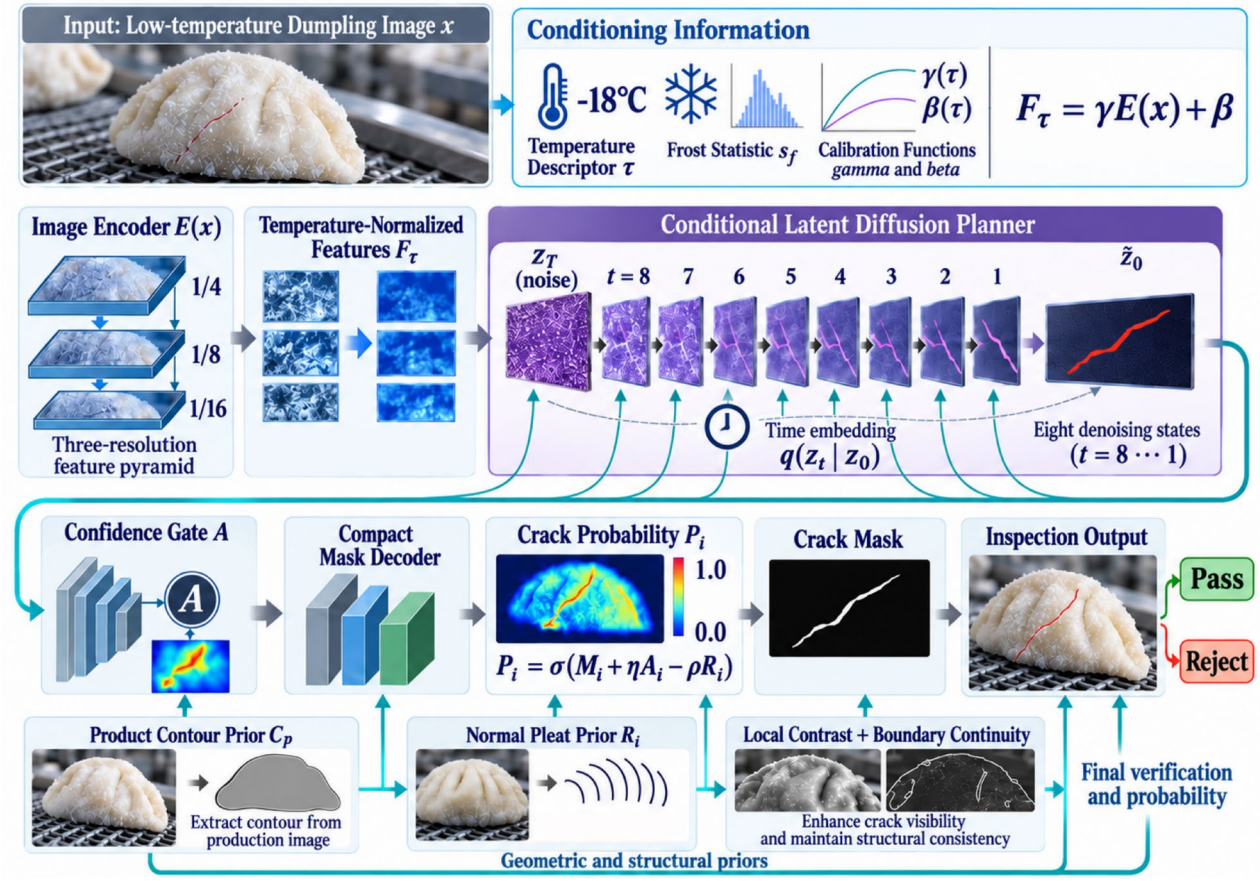


Figure 2. Diffusion Planner Architecture

The forward diffusion process is expressed as controlled corruption of the latent crack-path variable z_0.

$$q(z_t \mid z_0) = \mathcal{N}\left(\sqrt{\bar{\alpha}_t} z_0, (1 - \bar{\alpha}_t)I\right) \tag{2}$$

The reverse planner predicts the noise component under the given visual and process conditions. The denoising objective uses a moderate weighting term so that both early uncertain steps and late boundary-sensitive steps are learned.

$$\mathcal{L}_{diff} = \mathbb{E}_{t,z_0,\epsilon}\left[w_t \left\| \epsilon - \epsilon_\theta(z_t, t, F_\tau, C_p) \right\|_2^2\right] \tag{3}$$

In this equation, C_p is the product-contour prior and w_t is the timestep weight. The contour prior suppresses trajectories that leave the dumpling boundary, whereas the visual condition preserves the local evidence required to

distinguish cracks from pleats. A confidence gate subsequently converts the planned trajectory into an attention field for the decoder.

$$A = sigmoid\left(W_a\left[\hat{z}_0; F_\tau; C_p\right]\right) \tag{4}$$

3.3. Training Objective and Inference Strategy

The mask decoder combines visual features with the planned attention field. Training uses a compound objective containing segmentation, boundary, and planner-consistency terms to address class imbalance, thin-boundary localization, and agreement between the latent path and decoded mask. Loss weights, the confidence threshold, and the operating threshold are selected on the validation subset and remain fixed during final comparisons. This procedure prevents temperature-specific tuning on the held-out test data.

$$\mathcal{L} = \lambda_1 \mathcal{L}_{focal} + \lambda_2 \mathcal{L}_{dice} + \lambda_3 \mathcal{L}_{boundary} + \lambda_4 \mathcal{L}_{diff} \tag{5}$$

The inference strategy uses eight denoising steps after distillation. This number is sufficient because the latent planner searches for a crack path rather than reconstructing an image. Candidate masks with low planned confidence are not immediately discarded. They are passed to a compact verification layer that examines local contrast, boundary continuity, and overlap with normal pleat priors. The final crack probability at pixel i is computed by combining the decoder output with the planner confidence.

$$P_i = \sigma(M_i + \eta A_i - \rho R_i) \tag{6}$$

Here, M_i is the decoder logit, A_i is the planner attention value, and R_i is the normal-ridge response estimated from pleat templates. The parameters η and ρ are tuned on the validation set to balance missed cracks against false rejection. The resulting model is sufficiently compact for line-speed inspection because the computationally expensive denoising operation is restricted to the latent planner, while the high-resolution decoder remains deterministic.

4. Experimental Data and Analysis

4.1. Dataset Construction and Experimental Settings

Images were collected on a pilot quick-frozen dumpling line across four temperature zones, three conveyor speeds, and two camera angles. The acquisition protocol followed shift-based separation principles used in industrial evaluation [20]. To reduce temporal leakage, complete product sequences and acquisition sessions were assigned to only one of the training, validation, or test subsets; adjacent frames from the same product event were not split across subsets. The final dataset contained 18,640 images and 42,310 crack instances. Two trained inspectors reviewed pixel-level masks, and disagreements were resolved through joint reinspection of the original frame and its neighboring acquisition context. The protocol retained difficult frost, flour, reflection, and pleat cases rather than excluding them during quality control.

Table 1. Dataset split, crack instances, and low temperature disturbance conditions

Split	Images	Crack Instances	Temperature Range	Dominant Disturbance
Training	12,960	29,430	-34 to -18 C	Mixed frost and flour
Validation	2,840	6,410	-33 to -19 C	Film reflection
Test A	1,420	3,260	-31 to -20 C	Light frost
Test B	1,420	3,210	-34 to -18 C	Heavy frost
Line trial	5,000 frames	Manual audit	-32 to -21 C	Speed fluctuation

Low-temperature imaging studies show that defect contrast can decay faster than background texture contrast as frost crystals grow on the wrapper surface [21]. The comparison therefore included a U-Net variant, DeepLabV3+, SegFormer, a transformer detector, and an anomaly-reconstruction model. All baselines used the same data split, input resolution, and validation-based threshold selection. Augmentation was restricted to transformations compatible with conveyor imaging, and latency was measured under the same industrial-GPU configuration. These controls reduce the possibility that the reported differences arise from unequal preprocessing or test-time tuning.

The temperature-band results also show that the difficulty of the task is not directly related to the freezer set-point. At -30℃ and -26℃, all the models showed a slight increase of about 1-2 per cent because of reduced frost accumulation and stable wrapper contrast. However, at -34℃, the two performed differently: the drop for the proposed model from the warm-band value was only 1.6 per cent, while the edge-guided detector dropped by 5.5 per cent. Based on the manual inspection, among the 91 additional misses of the edge-guided detector, 64 occurred in areas where ice grains interrupted

the middle of a hairline crack. The planner did not have this error mode because it was aware of whether the two visible fragments were on a valid fracture path before the mask decoder assigned the final pixels.

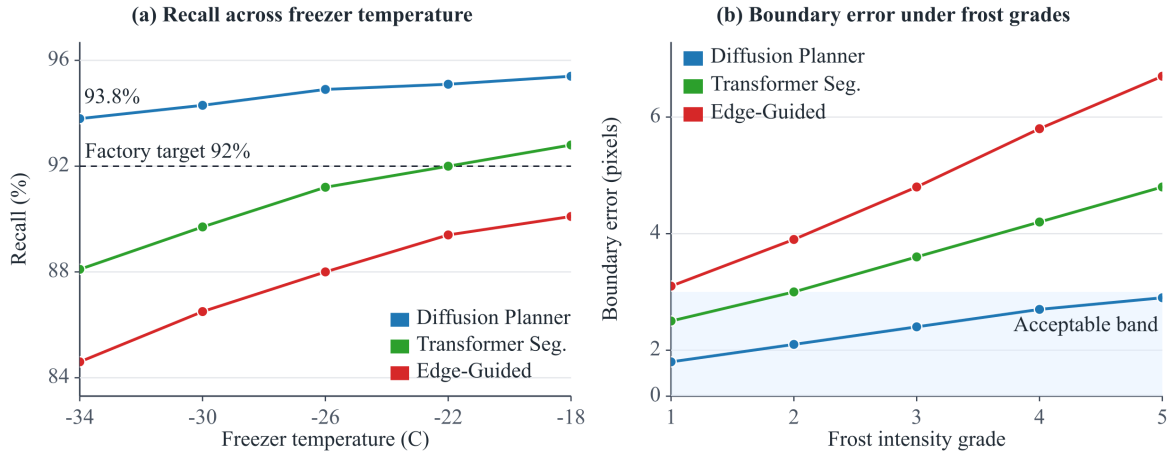


Figure 3. Temperature and frost robustness: (a) recall across freezer-temperature bands; (b) mean boundary error across frost-intensity grades

Wang et al. [22] reviewed machine-vision methods for industrial surface-defect detection and highlighted the engineering balance among detection accuracy, computational efficiency, and deployment conditions. The implementation used an industrial GPU with an 18 W power limit and an image acquisition rate of 40 frames per second. Figure 4(a) compares precision, recall, and F1 score across six models, where the proposed method reached 95.6%, 94.2%, and 94.9%, respectively. Figure 4(b) plots inference latency against input resolution and shows that the proposed model stayed below 25 ms at 1024 by 768. Figure 4(c) presents false rejection rate by dumpling orientation, with the largest reduction appearing for side-tilted samples where legal folds often mimic cracks. Figure 4(d) shows the distribution of crack confidence for accepted and rejected products, and the margin between the two distributions reached 0.31 on the normalized score scale.

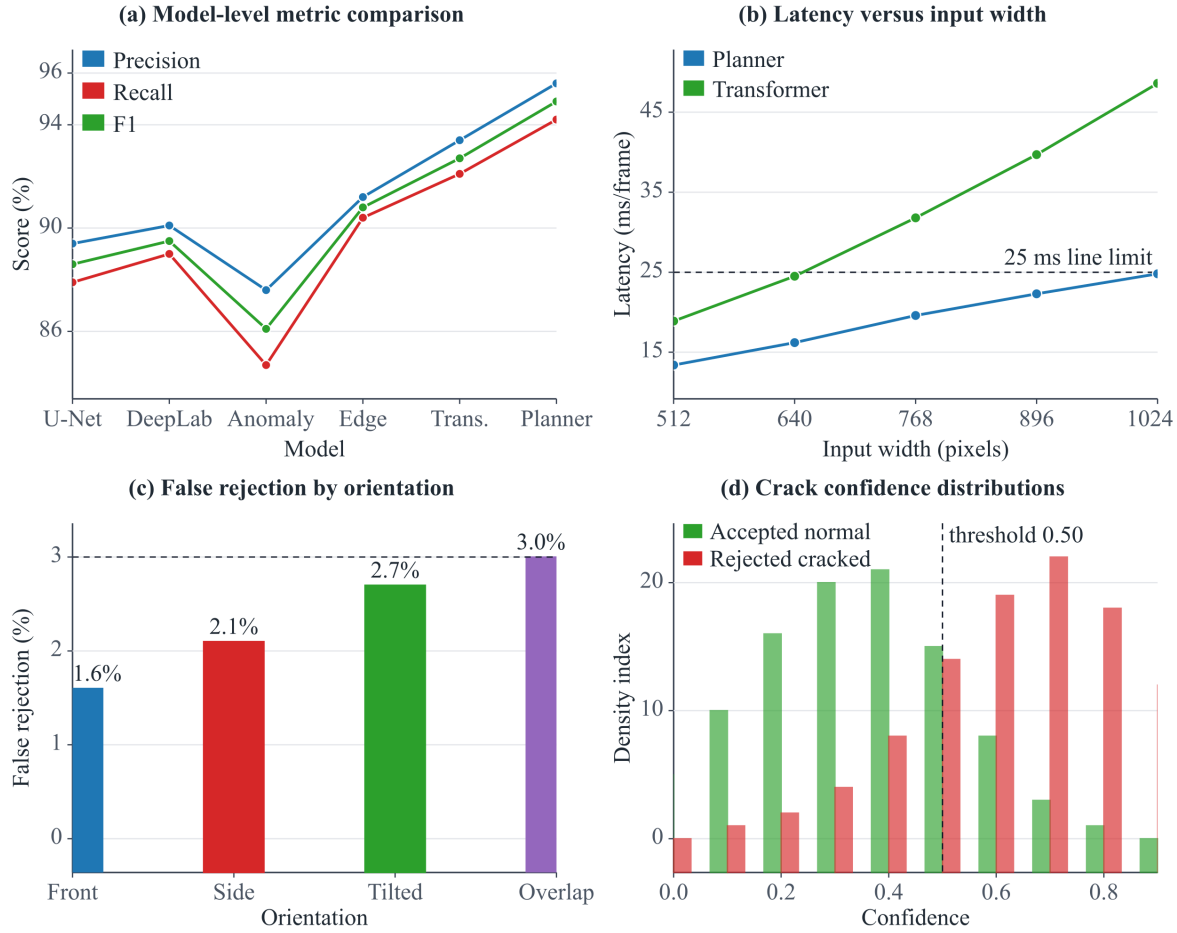


Figure 4. Accuracy and deployment efficiency: (a) precision, recall, and F1 comparison; (b) latency across input resolutions; (c) false rejection by product orientation; (d) crack-confidence distributions

The latency profile was examined at the module level because a model with high average speed can still fail on a line if the tail latency is unstable. Across 10,000 repeated inferences, the 95th percentile latency of the proposed method was 27.6 ms and the maximum observed latency was 31.2 ms. The transformer baseline had a similar mean accuracy but reached a 95th percentile latency of 55.8 ms at the same resolution. The rejection controller used a two-frame buffer, so the proposed model could tolerate short spikes without losing synchronization with the encoder pulse. The confidence distribution further indicated that the operating threshold did not sit in a dense overlap region; only 3.1% of normal samples and 4.4% of cracked samples fell within 0.04 of the thresholds.

4.2. Quantitative Comparison and Robustness Analysis

Calibration matters in rejection systems because an overconfident false alarm can remove acceptable products and disturb downstream weighing balance [23]. The strongest conventional segmentation baseline produced 91.8% F1 score, while the transformer baseline produced 92.7% F1 score. The proposed Diffusion Planner improved the score to 94.9%, mainly through recall gains on faint and interrupted cracks. Figure 5(a) gives the single heat map in this study, showing the interaction between temperature band and disturbance type; the lowest cell for the proposed model was 91.7% F1 under -34 C heavy frost, while the corresponding transformer value was 86.9%. Figure 5(b) gives the reliability curve, where the expected calibration error decreased from 0.071 for the transformer baseline to 0.034 for the proposed model.

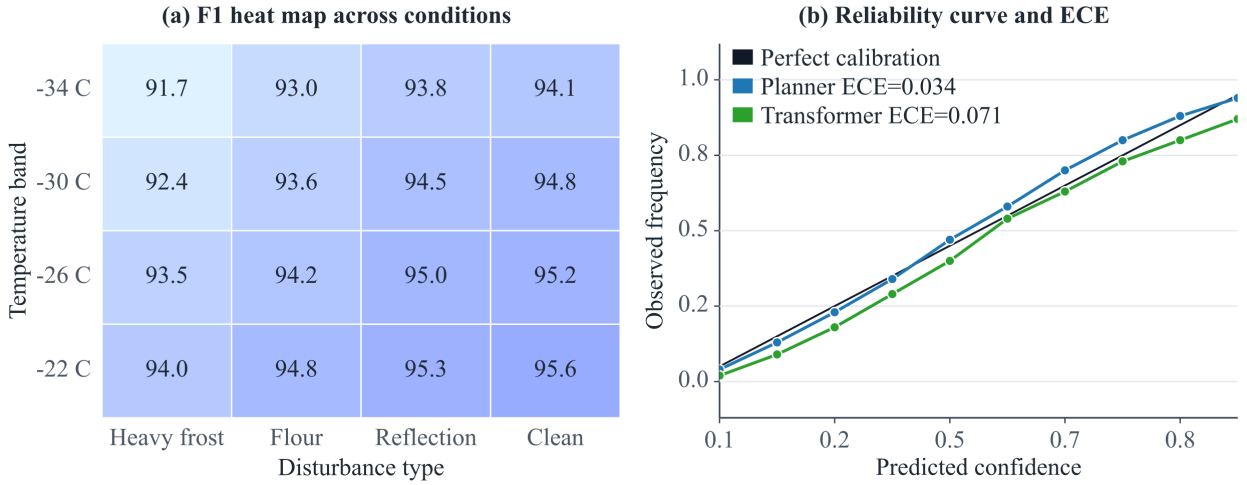


Figure 5. Robustness and calibration: (a) F1 heat map across temperature and disturbance conditions; (b) reliability curve and expected calibration error

Heatmap cells can be used to determine which disturbance needs to be addressed first in data collection. Heavy frost caused the most severe damage in all temperature ranges, but at -22°C, reflection was enhanced due to a smoother surface of the frozen wrapper and oblique light produced elongated highlights near the pleat valleys. Flour disturbance reduced the number of missed cracks but also increased false alarms, and dry flour formed narrow streaks after vibration on the conveyor. The proposed model outperformed the anomaly reconstruction baseline for flour by suppressing the connection of isolated powder streaks in defect paths through a contour prior and normal-ridge penalty. A second practical advantage of the calibration curve was that, at a predicted confidence of around 0.80, the observed crack frequency for the proposed method was 0.80, while that for the transformer baseline was 0.73; thus, the planner's output could be more conveniently adjusted by quality engineers.

Prior work on production inspection emphasizes that temporal stability is essential when the same product category is observed under cyclic illumination and conveyor vibration [24]. Boundary accuracy was analyzed because crack sorting depends not only on whether a defect exists but also on whether the defect crosses a severity threshold. Figure 6(a) shows the boundary error distribution, with the median error reduced from 3.9 pixels for the edge-guided detector to 2.1 pixels for the proposed method. Figure 6(b) compares severity-grade agreement and shows that the proposed model achieved 92.4% agreement for minor cracks and 96.8% agreement for severe cracks. Figure 6(c) plots crack length estimation error against visible crack length, where short cracks below 20 pixels remained the hardest group and still showed a 17.5% relative improvement. Figure 6(d) presents line-trial rejection stability over 5,000 consecutive frames, and the rejection rate varied within a 0.7 percentage point band after the first 400 frames.

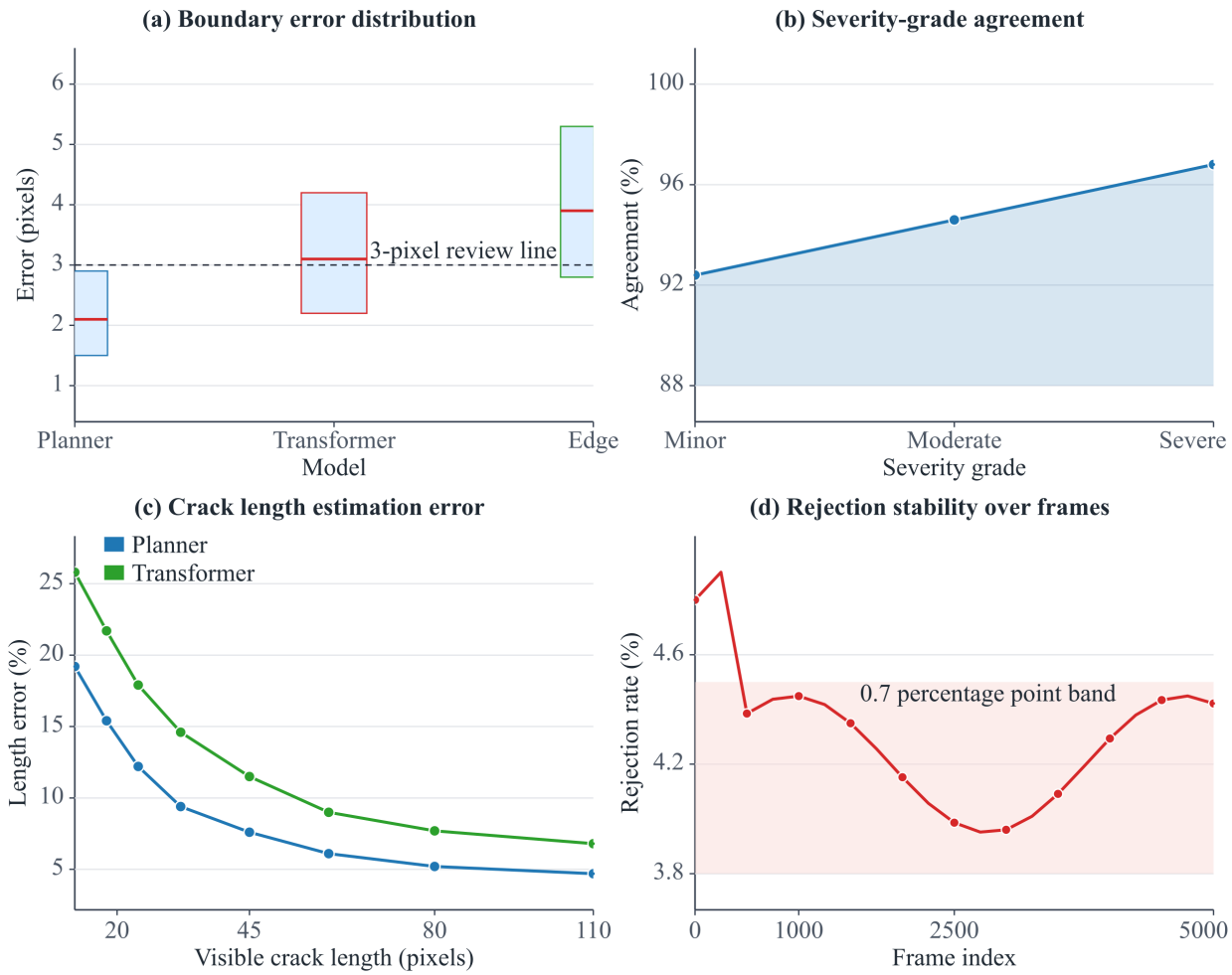


Figure 6. Boundary and line stability: (a) boundary-error distribution; (b) severity-grade agreement; (c) crack-length estimation error; (d) rejection stability during continuous production

Severity-grade agreement was calculated after converting the predicted mask into crack length, mean opening width, and location-weighted risk. The proposed model made 34 disagreements on minor cracks, 27 disagreements on moderate cracks, and 11 disagreements on severe cracks in the inspected subset. Most minor-crack disagreements were caused by fragmented masks that slightly underestimated length, while severe-crack disagreements were mainly caused by tail-region defects whose visual boundary merged with a pleat. The line-trial curve was recorded without stopping the conveyor or cleaning the camera window, so the early 400-frame drift reflected normal enclosure stabilization after the freezer door was opened. Once the enclosure temperature stabilized, the rejection rate remained close to the manual audit trend, and no frame queue overflow was recorded.

Manufacturing vision findings indicate that process-conditioned features are more reliable than image-only features when the acquisition environment is physically constrained [25]. The robustness test also included illumination deviation, camera height offset, and flour dosage variation. Under a 15% illumination reduction, the proposed model lost 1.4 percentage points of F1 score, whereas the anomaly reconstruction model lost 5.8 percentage points. With a 12 mm camera height offset, the proposed model retained 93.6% F1 score because the planner used contour-normalized coordinates rather than raw pixel distance. Figure 7(a) shows precision-recall curves under four disturbance settings, and the area under the curve remained above 0.973 in every setting. Figure 7(b) reports the ablation gains of temperature normalization, contour prior, diffusion planning, and confidence gating, where diffusion planning contributed the largest recall increase.

A precision-recall curve can be used to examine the effects of different strictness or leniency levels in the factory's rejection policy. Under clean imaging, the precision of the curve was still above 0.95 at a recall of 0.98, and therefore a high-throughput sorting detector could be used. In the case of an extreme low-temperature fluctuation, the rate-of-change drops below 0.91 and is no longer accurate. This behavior is better than a sudden drop because, when the quality of raw materials changes, the threshold can be gradually adjusted. As shown in the ablation plot, Temperature Normalization and Contour Prior were also employed simultaneously; the former reduced appearance drift, and the latter constrained geometrically impossible predictions. Removing both modules resulted in an increase in off-product false positives near the joint of the dumplings.

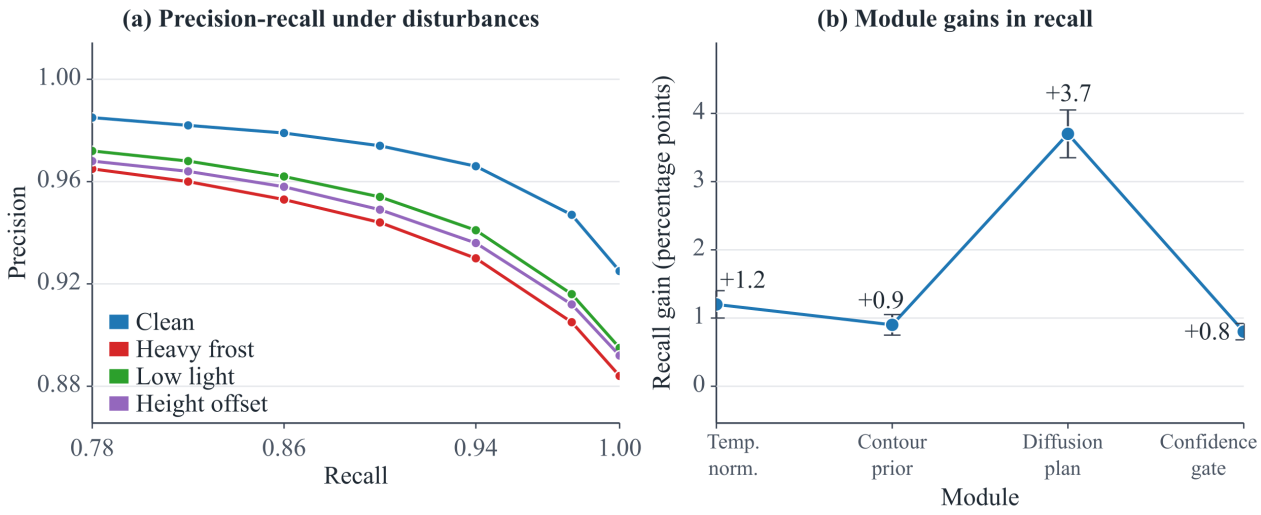


Figure 7. Disturbance robustness and ablation: (a) precision-recall curves under four disturbances; (b) incremental gains from the principal model modules

4.3. Ablation Study and Error Interpretation

Research on crack-like perception indicates that continuity cues are more informative than raw contrast when defect segments are partially occluded [26]. Starting from the encoder-decoder baseline, modules were added sequentially to expose their cumulative effects. Temperature normalization reduced false alarms on glossy frozen skin from 4.7% to 3.3%, while the contour prior reduced off-product responses by 61.2%. The diffusion planner increased recall for interrupted cracks from 90.5% to 94.1% by linking local evidence across frost-covered regions. The confidence gate primarily suppressed trajectories that conflicted with stable pleat templates rather than increasing interrupted-crack recall.

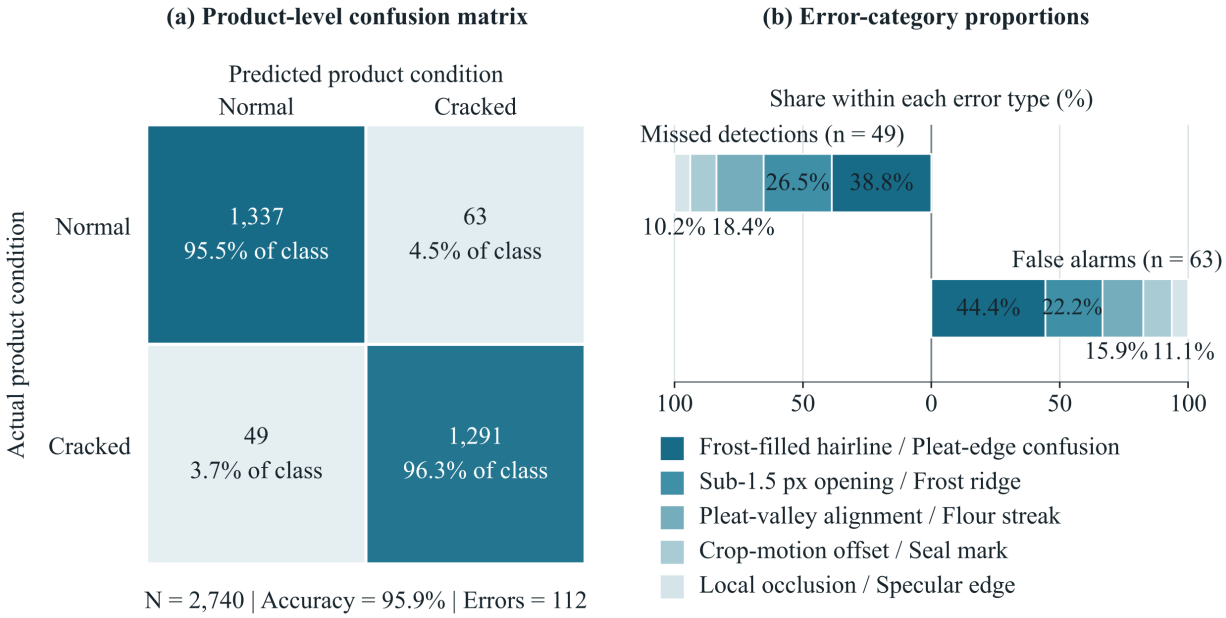


Figure 8. Error interpretation: (a) product-level confusion matrix; (b) category proportions for missed detections and false alarms

Fewer denoising steps were also taken by the model to determine if the planner could be further simplified. After four steps, the latency is 2.9 ms; however, the recall rate for interrupted cracks has decreased by 1.8%. Step Six increased the recall and was relatively stable in the neighborhood of the operating point. The eight steps offered a reasonable trade-off of recall, calibration and latency; the twelve steps increased the F1 score by only 0.2% and were over the time limit. Therefore, a small latent planner can be employed in place of a full iterative image generator. The planned trajectory maps were visually sparse, covered less than 6% of the product area, and were therefore suitable for interpretation in error analysis.

Industrial inspection literature indicates that borderline defects often require either higher resolution optics or complementary lighting, because a purely algorithmic solution cannot recover evidence that is absent from the sensor signal [27]. The product-level confusion matrix in Figure 8(a) records 1,337 normal samples, 1,291 cracked samples, 63 false alarms, and 49 missed cracks on the held-out test subset. The error-category analysis in Figure 8(b) shows that frost-filled hairline cracks accounted for 38.8% of missed cases, while pleat-edge confusion accounted for 44.4% of false alarms. The missed cases were concentrated in images where the crack opening was below 1.5 pixels and the local frost brightness exceeded the wrapper brightness by more than 18 gray levels.

Reliability analysis in quality-control systems favors smooth threshold responses because rejection policy can be adjusted without abrupt workload changes [28]. Reducing the threshold from 0.50 to 0.42 increased recall from 94.2% to 96.8% but raised false rejection from 2.4% to 3.9%; increasing it to 0.58 reduced false rejection to 1.6% while recall fell to 91.7%. The threshold of 0.50 was selected on the validation set and applied unchanged to the reported test conditions. Efficient generative inspection also benefits from restricting generation to task-relevant variables [29]. Profiling assigned 9.6 ms to the encoder, 8.7 ms to the eight-step planner, 4.8 ms to mask decoding, and 1.7 ms to post-processing. Peak inference memory remained below 2.1 GB, and latent-path planning reduced denoising computation by 72.4%, enabling 40.3 frames per second during the 5,000-frame trial.

The deployment simulation introduced trigger jitter, reduced product spacing, and asynchronous rejection commands. A 15 ms trigger offset increased crop misalignment by 3.6 pixels, although contour-guided localization corrected most of the displacement. When product spacing was reduced from 42 mm to 28 mm, overlapping cases increased by 18.2% and false rejection rose from 2.4% to 2.9%. Rejected images, masks, confidence scores, temperature descriptors, and encoder timestamps were retained for post-production audit and alignment with rejection events.

Reviewing rejected samples by planner confidence and estimated severity identified 183 confirmed defects among the first 200 cases, compared with 174 when ranking by decoder probability alone. This observation suggests that path confidence and defect location improved audit prioritization under the tested conditions. Approximately 24 products per thousand were rejected or routed for re-examination, and fewer than three per thousand required manual handling. These workload figures are specific to the pilot line, tested speed, product mix, and review policy and should not be generalized to other factories without local validation.

Food quality models increasingly relate defect location to downstream consequences instead of treating all surface marks equally [30]. The planner can still connect flour-interrupted pleats into false crack trajectories or suppress a true crack that follows a pleat valley. Location also changes practical severity, particularly between tail and central-wrapper defects. Recipe-specific pleat geometry therefore remains an unresolved source of uncertainty and requires product-family calibration.

5. Conclusion

This study developed a Diffusion Planner for surface-crack inspection of quick-frozen dumplings under low-temperature production disturbances. Temperature-aware normalization reduced appearance drift, the contour-conditioned planner connected discontinuous evidence across frost-covered regions, and confidence gating prevented uncertain trajectories from dominating the decoded mask. Sparse latent planning retained pixel-level localization while remaining compatible with production-line timing. The resulting framework links image formation, geometric reasoning, uncertainty estimation, and rejection control in a traceable inspection process.

Several limitations remain. The pilot-line dataset does not cover the full variation in wrapper recipes, dough rheology, flour application, equipment wear, optical-window contamination, freezer maintenance, and seasonal humidity encountered across factories. Visible-light imaging cannot reliably recover cracks whose evidence is completely hidden by frost or aligned with deep pleat valleys. Recipe-dependent pleat geometry may also weaken the contour and ridge priors for unfamiliar product families. Moreover, the study evaluates visible crack shape and location but does not establish a direct relationship with leakage, rupture, texture after cooking, or other functional quality outcomes. These constraints require line-specific validation before deployment outside the development environment.

Future work should evaluate the planner across factories, product geometries, recipes, and seasonal humidity ranges under a common protocol. Controlled oblique illumination, polarization, or lightweight near-infrared sensing could supplement insufficient visible contrast while preserving real-time operation. Domain adaptation and continual calibration should accommodate new pleat patterns without extensive mask annotation. Crack trajectories and confidence estimates could also be integrated with forming pressure, belt speed, freezing temperature, and ingredient measurements to distinguish isolated defects from systematic process drift. Linking predicted crack location and severity with standardized cooking tests would further support functional-quality assessment, selective rejection, and upstream process correction.

6. Declarations

6.1 Author Contributions

Masato Taniguchi contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Hiroshi Sato contributes to validation, analysis, investigation, data collection, draft preparation. Yuichi Yoshida 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] Alabi, K. P., Olalusi, A. P., Isa, J., & Jaiyeoba, K. F. (2024). Supercooling phenomenon of magnetic field-assisted freezing and its impacts on quality preservation of frozen fruits and vegetables. Food Biophysics, 19(4), 833-844. https://doi.org/10.1007/s11483-024-09873-3

[2] Ayon, S. T. K., Siraj, F. M., & Uddin, J. (2025). Steel surface defect detection using learnable memory vision transformer. Computers, Materials & Continua, 82(1), 499-520. https://doi.org/10.32604/cmc.2025.058361

[3] Block, S. B., da Silva, R. D., Lazzaretti, A. E., & Minetto, R. (2024). LoHi-WELD: A novel industrial dataset for weld defect detection and classification, a deep learning study, and future perspectives. IEEE Access, 12, 77442-77453. https://doi.org/10.1109/access.2024.3407019

[4] Bu, T., Liu, X., Zhu, J., Ma, T., & Huang, X. (2025). Unsupervised pavement anomaly detection via conditional diffusion model and domain adaptive feature selection. Computer-Aided Civil and Infrastructure Engineering, 40(28), 5195-5211. https://doi.org/10.1111/mice.70099

[5] Cao, Y., Liu, J., Wang, L., & Lian, J. (2025). CTrans-SegDiff: CTransfomer-based diffusion model for ultrasound image segmentation. IEEE Signal Processing Letters, 32, 3505-3509. https://doi.org/10.1109/lsp.2025.3601987

[6] Cao, Y., Ni, Y., Zhou, Y., Li, H., Huang, Z., & Yao, E. (2024). An auto chip package surface defect detection based on deep learning. IEEE Transactions on Instrumentation and Measurement, 73, 1-15. https://doi.org/10.1109/tim.2023.3347799

[7] Chen, Q., Luo, H., Yao, H., Luo, W., Qu, Z., Lv, C., & Zhang, Z. (2025). Center-aware residual anomaly synthesis for multiclass industrial anomaly detection. IEEE Transactions on Industrial Informatics, 21(9), 7276-7286. https://doi.org/10.1109/tii.2025.3575122

[8] 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

[9] Genç, İ. Y., Gürfidan, R., & Açikgözoğlu, E. (2025). Quality determination of frozen-thawed shrimp using machine learning algorithms powered by explainable artificial intelligence. Food Analytical Methods, 18(6), 935-945. https://doi.org/10.1007/s12161-025-02768-0

[10] Lanko, V., & Makarov, I. (2024). Graph-attention diffusion for enhanced multivariate time-series anomaly detection. IEEE Open Journal of the Industrial Electronics Society, 5, 1353-1364. https://doi.org/10.1109/ojies.2024.3501014

[11] Li, B., Ge, H., Liu, Y., & Tang, G. (2025). Multiscale recovery diffusion model with unsupervised learning for video anomaly detection system. IEEE Transactions on Industrial Informatics, 21(3), 2104-2113. https://doi.org/10.1109/tii.2024.3493390

[12] Liu, X., Zhao, Y., Wang, S., & Wei, J. (2024). TransDiff: Medical image segmentation method based on Swin transformer with diffusion probabilistic model. Applied Intelligence, 54(8), 6543-6557. https://doi.org/10.1007/s10489-024-05496-w

[13] Liu, Y., Feng, Y., Cheng, J., Zhan, H., & Zhu, Z. (2025). MambaDiff: Mamba-enhanced diffusion model for 3D medical image segmentation. IEEE Transactions on Image Processing, 34, 5761-5775. https://doi.org/10.1109/tip.2025.3607615

[14] Mao, Y., Zhang, J., Xiang, M., Lv, Y., Li, D., Zhong, Y., & Dai, Y. (2025). Contrastive conditional latent diffusion for audio-visual segmentation. IEEE Transactions on Image Processing, 34, 4108-4119. https://doi.org/10.1109/tip.2025.3580269

[15] Meng, F., Xiang, X., Wu, D., & Gong, T. (2024). Defect detection on the surface of the bellow expansion joints based on machine vision. IEEE Access, 12, 131244-131254. https://doi.org/10.1109/access.2024.3431635

[16] Olaniyi, E. O., & Kucha, C. (2025). Advances in precision systems based on machine vision for meat quality detection. Food Engineering Reviews, 17(4), 809-834. https://doi.org/10.1007/s12393-025-09404-x

[17] Ren, F., Fei, J., Li, H., & Doma, B. T. (2024). Steel surface defect detection using improved deep learning algorithm: ECA-SimSPPF-SIoU-Yolov5. IEEE Access, 12, 32545-32553. https://doi.org/10.1109/access.2024.3371584

[18] Ruan, J., Wang, H., Zhao, J., Li, D., & Yang, H. (2024). Effect of magnetic field on frozen food quality characteristics. Food Engineering Reviews, 16(3), 396-421. https://doi.org/10.1007/s12393-024-09366-6

[19] Shaloo, M., Princz, G., Hörbe, R., & Erol, S. (2024). Flexible automation of quality inspection in parts assembly using CNN-based machine learning. Procedia Computer Science, 232, 2921-2932. https://doi.org/10.1016/j.procs.2024.02.108

[20] Sui, J., Yu, J., Song, Y., & Zhang, J. (2024). Anomaly detection for telemetry time series using a denoising diffusion probabilistic model. IEEE Sensors Journal, 24(10), 16429-16439. https://doi.org/10.1109/jsen.2024.3383416

[21] Wang, J., Zou, H., Zhou, M., & Su, R. (2025). An efficient deep neural network for surface defect detection in industrial edge sensing. IEEE Transactions on Industrial Informatics, 21(3), 2560-2569. https://doi.org/10.1109/tii.2024.3507954

[22] Wang, Q., Wang, M., Sun, J., Chen, D., & Shi, P. (2025). Review of surface-defect detection methods for industrial products based on machine vision. IEEE Access, 13, 90668-90697. https://doi.org/10.1109/access.2025.3571297

[23] Xu, J., Zhang, C., & Zeng, W. (2025). A conditional diffusion-based crack segmentation model: ConditionCrack segmentation. IEEE Transactions on Intelligent Transportation Systems, 26(10), 17041-17054. https://doi.org/10.1109/tits.2025.3577492

[24] Yang, C., Guo, Z., Fernandes Barbin, D., Dai, Z., Watson, N., Povey, M., & Zou, X. (2025). Hyperspectral imaging and deep learning for quality and safety inspection of fruits and vegetables: A review. Journal of Agricultural and Food Chemistry, 73(17), 10019-10035. https://doi.org/10.1021/acs.jafc.4c11492

[25] Yang, H., Hua, W., Xu, Z., & Sun, J. (2025). Domain-generalized discrete diffusion model for cross-domain medical image segmentation. IEEE Transactions on Medical Imaging, 44(11), 4131-4142. https://doi.org/10.1109/tmi.2025.3564474

[26] Yao, H., Luo, W., Lou, J., Yu, W., Zhang, X., Qiang, Z., & Shi, H. (2024). Scalable industrial visual anomaly detection with partial semantics aggregation vision transformer. IEEE Transactions on Instrumentation and Measurement, 73, 1-17. https://doi.org/10.1109/tim.2023.3343832

[27] Yao, H., Luo, W., Yu, W., Zhang, X., Qiang, Z., Luo, D., & Shi, H. (2024). Dual-attention transformer and discriminative flow for industrial visual anomaly detection. IEEE Transactions on Automation Science and Engineering, 21(4), 6126-6140. https://doi.org/10.1109/tase.2023.3322156

[28] You, J., Li, D., Wang, Z., Chen, Q., & Ouyang, Q. (2024). Prediction and visualization of moisture content in Tencha drying processes by computer vision and deep learning. Journal of the Science of Food and Agriculture, 104(9), 5486-5494. https://doi.org/10.1002/jsfa.13381

[29] Zhang, Y., Zhou, P., & Ma, E. (2024). Anomaly detection of industrial smelting furnace incorporated with accelerated sampling denoising diffusion probability model and Conv-Transformer. IEEE Transactions on Instrumentation and Measurement, 73, 1-11. https://doi.org/10.1109/tim.2024.3375410

[30] Zhang, Z., Yao, L., Wang, B., Jha, D., Durak, G., Keles, E., Medetalibeyoglu, A., & Bagci, U. (2025). DiffBoost: Enhancing medical image segmentation via text-guided diffusion model. IEEE Transactions on Medical Imaging, 44(9), 3670-3682.

https://doi.org/10.1109/tmi.2024.3519307