Knowledge-Guided Machine Learning for Automated Recognition of Irregular Patterns in Industrial Image Analysis

Latasha Morgan
Department of Computer Science, University of New Hampshire, Durham, NH, USA.
latashamorgan150@unh.edu

Abstract

Industrial image analysis increasingly depends on learned visual representations, yet irregular patterns such as fine scratches, porphyrin-like inclusions, weld discontinuities, textile faults, surface porosity, and deformation marks remain difficult to recognize reliably because their appearance is sparse, geometry is variable, contrast is weak, and nuisance factors often mimic defect evidence. Purely data-driven detectors can achieve strong fit to observed production conditions while failing when illumination, material batch, camera calibration, or process settings shift. This article develops a knowledge-guided machine learning framework for automated recognition of irregular industrial patterns. The central contribution is a constraint-coupled representation-and-decision architecture that separates pattern-bearing evidence from benign texture and sensor variation, injects process and imaging knowledge through differentiable consistency terms, scores irregularity against learned regular-pattern prototypes, and calibrates uncertainty before automated accept, review, or reject decisions. The formulation integrates convolutional representation learning, weak supervision, prior-consistency penalties, Mahalanobis-style irregularity scoring, ensemble-based uncertainty estimation, and cost-sensitive decision thresholds. The article analyzes why knowledge constraints can improve sample efficiency and shift robustness, how prior misspecification can induce systematic false negatives, and which failure modes require explicit review policies rather than additional model capacity. The resulting manuscript provides a technically grounded design logic for industrial vision systems that must recognize rare and ambiguous irregularities while maintaining auditable decision behavior. The framework is intended to support reliable inspection design, not to replace domain-specific validation.

Keywords

knowledge-guided learning; irregular pattern recognition; industrial image analysis; uncertainty calibration; anomaly detection; visual inspection

1. Introduction

Industrial image analysis has moved from handcrafted feature pipelines toward learned visual representations. Deep convolutional networks established that hierarchical filters can be trained end to end for image classification [1], and broader syntheses of deep learning clarified how representation learning reduces dependence on manually specified descriptors [2]. Residual connections made very deep networks trainable and improved gradient propagation [3], region-based detection methods enabled localization of objects and defects [4], and encoder-decoder segmentation architectures provided pixel-level outputs for irregular regions [5]. These developments are directly relevant to industrial inspection because

recognition tasks often require simultaneous classification, localization, boundary estimation, and severity assessment. A scratch may be only a few pixels wide, a weld irregularity may appear as a subtle intensity transition, and a surface inclusion may have no stable geometric form across materials. The visual evidence is therefore not simply a category label problem; it is a problem of separating pattern-bearing variation from benign process and imaging variation.

Automated anomaly and novelty detection provide an important complementary lineage. Isolation forests isolate sparse observations through random partitioning [6], while one-class and novelty-detection methods estimate a boundary around normal data [7]. Industrial anomaly benchmarks have made the recognition problem more concrete by emphasizing unsupervised or weakly supervised settings in which defective examples are rare or unavailable during training [8]. Reviews of shallow and deep anomaly detection show that reconstruction error, density estimation, one-class classification, and self-supervised representation learning each capture different assumptions about normality [9], and surveys of deep anomaly detection emphasize the difficulty of distinguishing semantic anomalies from nuisance variation [10]. These methods are valuable, but industrial recognition of irregular patterns is not identical to generic anomaly detection. A production image may contain a benign edge mark that is statistically rare but functionally acceptable, or a visually subtle irregularity that is rare and functionally critical. The decision boundary must therefore encode not only statistical deviation but also process knowledge and consequence.

A second body of work concerns the integration of prior knowledge into learning systems. Informed machine learning provides a taxonomy for combining data with symbolic rules, physical constraints, simulation models, knowledge graphs, and logical dependencies [11]. Physics-informed machine learning demonstrates how governing equations and physical constraints can be embedded in losses, architectures, or optimization procedures [12]. Scientific-knowledge integration for engineering and environmental systems shows that such knowledge can improve generalization when data are sparse, but also that knowledge-data conflicts must be handled explicitly [13]. Uncertainty estimation is equally important because industrial inspection systems frequently encounter inputs outside their training distribution. Bayesian approximation through dropout provides a practical route to predictive uncertainty [14], deep ensembles offer a simple and scalable alternative [15], and weakly supervised learning addresses incomplete or coarse labels that are common when operators annotate only image-level outcomes [16]. These ingredients suggest a recognition framework that is neither a purely black-box classifier nor a brittle rule-based vision system.

The central contribution of this article is a knowledge-guided machine learning formulation for irregular pattern recognition that couples four elements in one decision logic. First, the formulation explicitly decomposes an industrial image into a pattern-bearing component, a benign variation component, and sensor noise, so that irregularity evidence is not conflated with illumination or material texture. Second, it defines differentiable knowledge-consistency terms that encode process geometry, reflectance behavior, spatial relations, and morphological expectations. Third, it scores irregularity against learned regular-pattern prototypes in a latent space rather than relying solely on pixel reconstruction error. Fourth, it uses calibrated uncertainty and cost-sensitive thresholds to separate automated decisions from review decisions. This article develops the method analytically, explains the role of each constraint, and assesses likely failure modes. It does not claim that an

experimental evaluation has been performed. Instead, it provides a rigorous methodological treatment that can guide implementation and empirical validation in industrial settings.

The argument proceeds as follows. Section 2 formulates the recognition problem and identifies the knowledge sources relevant to irregular patterns. Section 3 presents the knowledge-guided representation architecture. Section 4 develops irregularity scoring and uncertainty-aware decision making. Section 5 specifies the optimization strategy and training constraints. Section 6 provides an analytical assessment of behavior, sensitivity, and failure modes. Section 7 discusses implementation logic and evaluation protocol. Section 8 concludes.

2. Problem Formulation and Knowledge Sources

2.1 Image formation and pattern separation

An industrial image is not a direct measurement of a defect class. It is a composition of surface structure, material texture, illumination, camera response, and stochastic noise. For a pixel or local region indexed by u, a useful analytical abstraction is

$$I(u) = \mathrm{P}(u;\theta) + B(u) + \varepsilon(u)$$

where I(u) is the observed image intensity or multichannel response, P(u;θ) is the pattern-bearing surface field parameterized by θ, B(u) is benign variation generated by texture, illumination gradients, reflectance changes, or fixturing, and ε(u) is sensor and quantization noise. The decomposition is not directly identifiable from a single image without assumptions, but it clarifies the recognition objective. Irregular pattern recognition should estimate evidence that P contains a nonconforming structure rather than merely detect large values of I−P. A bright reflection on a machined surface belongs to B if it is consistent with the imaging geometry; a faint scratch belongs to P if it violates process-specific regularity. Knowledge enters by constraining the plausible forms of P and B and by specifying which deviations are functionally relevant.

This formulation has a direct consequence for model design. If the representation is trained only to minimize reconstruction error on normal images, it may devote capacity to reconstructing nuisance variation and then confuse rare but benign texture with irregularity. If the representation is trained only with supervised defect labels, it may learn shortcuts tied to lighting, camera position, or batch-specific appearance. Knowledge-guided learning attempts to reduce both risks by combining data-driven discrimination with explicit consistency requirements. The approach is particularly appropriate for irregular patterns because their intra-class variability is high and their boundary between acceptable and unacceptable variation is often defined by process specifications rather than by appearance alone.

2.2 Irregularity as structured deviation

An irregular pattern can be described as a structured deviation from a family of acceptable patterns under a set of transformations. The family may include translation, rotation, scale, elastic deformation, illumination change, and material texture. The deviation may be local, as in a crack or void, or relational, as in misalignment between two jointed components. This perspective suggests that recognition should not depend on a single template. Instead, the model should learn an equivariant or approximately invariant representation of acceptable structure and then measure deviations in a space where nuisance transformations have been suppressed. The Mahalanobis-style scoring developed later follows

this logic: regular patterns form clusters or manifolds in latent space, and irregularity is measured relative to the conditional spread of those clusters.

The distinction between statistical rarity and functional irregularity must be maintained. A rare lighting artifact can be statistically unusual but acceptable; a small crack can be statistically subtle but unacceptable. The knowledge layer therefore needs to encode functional relevance, such as whether a discontinuity crosses a load-bearing region, whether a scratch exceeds a specified width, or whether porosity occurs in a sealing surface. Such information can be represented as constraints, region masks, relational predicates, or cost-sensitive labels. It should not be reduced to a generic anomaly score without domain interpretation.

2.3 Knowledge sources for industrial irregular patterns

Four knowledge sources are especially useful. Imaging knowledge describes how illumination, camera response, lens distortion, and viewing geometry generate observed intensity. It supports constraints on shading consistency, specularity, and edge formation. Process knowledge describes manufacturing operations and their characteristic signatures. For example, machining marks have directional regularity, welded joints have bead geometry and heat-affected zones, and molded parts have parting lines and gate marks. Morphological knowledge describes the expected shape, width, orientation, and termination of defects. A scratch often has an elongated skeleton and tapered ends; porosity often appears as compact dark regions; a crack may exhibit branching and low contrast. Relational knowledge describes spatial dependencies among components, such as alignment between joint members, expected distances between features, and region-specific acceptance criteria.

These knowledge sources can be encoded in differentiable form. A shading-consistency term can compare predicted surface normals with observed intensity gradients. A directionality term can penalize latent features that respond strongly to expected machining marks unless they deviate from the learned orientation distribution. A relational term can compare detected feature positions with geometric constraints. A morphological term can use soft skeleton length, compactness, or boundary smoothness as auxiliary targets. The purpose is not to replace learning with handcrafted rules. The purpose is to bias learning toward representations that respect stable physical and process structure, so that scarce irregular examples are used more efficiently.

3. Knowledge-Guided Representation Architecture

3.1 Encoder and multi-scale latent representation

The recognition architecture begins with an encoder that maps an image region or full inspection image into a latent representation. For image I, the encoder is written as

$$z = f_{\psi}(I)$$

where f_ψ is a convolutional or transformer-based encoder with parameters ψ, and z is a latent vector or spatial latent map. In industrial inspection, a multi-scale representation is desirable because irregular patterns span different spatial extents. A fine scratch may be visible only at high resolution, while a deformation or misalignment may require global context. The encoder can therefore produce features at several resolutions and aggregate them through lateral connections or attention. The latent representation is not assumed to be a complete description of the image. It is a task-oriented summary that should retain pattern evidence while discarding irrelevant variation.

A second design decision concerns spatial preservation. Global pooling is convenient for classification but can erase small irregularities. The architecture should preserve spatial latent maps for localization and use region-level pooling only for image-level decisions. When pixel-level masks are unavailable, weak labels can be used to train a multiple-instance or attention-based head that identifies which local features support the image-level label [16]. This is important because irregular patterns are often sparse; a model that averages over the whole image can dilute their evidence.

3.2 Differentiable knowledge constraints

Knowledge constraints are introduced as auxiliary objectives on the latent representation or on intermediate predictions. A general form is

$$L_k = \sum_\{j = 1\}\text{^}K\ w_j\ ||g_j(z) - h_j(I)||_2\text{^}2$$

where K is the number of knowledge terms, w_j is a nonnegative weight, g_j(z) is a predicted quantity derived from the latent representation, and h_j(I) is a reference or target quantity derived from image measurements or process specifications. The term L_k is a consistency loss, not a claim that the knowledge is exact. For a shading constraint, g_j might be a predicted surface-normal gradient and h_j an image-derived shading gradient. For a directionality constraint, g_j might be a predicted orientation histogram and h_j a target histogram representing acceptable machining marks. For a relational constraint, g_j might be a predicted offset between component features and h_j a geometric tolerance. The weights express relative confidence in each constraint and can be adjusted during training to avoid domination by a single term.

This formulation is deliberately modular. It allows domain knowledge to be added without redesigning the entire network, and it allows constraints to be relaxed when they conflict with data. The latter point is essential. A rigid rule that is slightly wrong can create systematic false negatives. A soft penalty can express a preference while allowing the model to learn exceptions from evidence. The next section makes this trade-off explicit through the combined objective.

3.3 Regular-pattern prototypes and class-conditional structure

For irregular pattern recognition, it is useful to represent acceptable appearance as a set of prototypes or clusters rather than as a single normal distribution. Let μ_c denote the centroid of the c-th regular-pattern cluster in latent space, and let Σ_c denote its covariance or a shared covariance estimate. These quantities can be computed from training data or learned as parameters. They provide an interpretable reference for scoring. A new latent vector z is compared with the nearest regular-pattern cluster, and the distance is converted into an irregularity score. This approach is preferable to a single reconstruction threshold when normal appearance is multimodal, as it often is across product variants, lighting conditions, and orientations.

The prototype representation also supports process knowledge. Different clusters can correspond to known product families or machining conditions, and constraints can be applied conditionally. For example, a directional constraint might be active only for a turned surface, while a relational constraint might be active only for an assembled joint. This conditional structure reduces the risk that a constraint appropriate for one product family penalizes another. The architecture therefore combines a shared encoder with conditional knowledge heads and cluster-specific scoring.

4. Irregular Pattern Scoring and Uncertainty-Aware Decision Making

4.1 Mahalanobis-style irregularity score

Given a latent representation z and regular-pattern prototypes, a natural irregularity score is the minimum Mahalanobis distance to a regular cluster:

$$s(I) = \min_c (z - \mu_c)^T \Sigma^{-1} (z - \mu_c)$$

where s(I) is the irregularity score for image I, c indexes regular-pattern clusters, μ_c is the centroid of cluster c, and Σ is a covariance matrix, either shared across clusters or cluster-specific. The score is large when the latent representation lies far from all acceptable clusters relative to their expected spread. The covariance normalization prevents directions with high natural variability from dominating the score. This matters in industrial images because illumination gradients and texture variations often create large but benign latent variance. Without covariance normalization, the detector would over-penalize those directions and generate false alarms.

The score should be interpreted as evidence, not as a probability. A threshold on s(I) defines a statistical decision boundary, but the mapping from distance to operational risk depends on the cost of false negatives, the cost of manual review, and the prevalence of irregularity. The score can be calibrated against labeled validation data when such data are available, but it should not be assumed to be calibrated by construction. The next subsection introduces a separate uncertainty pathway so that distance and confidence are not conflated.

4.2 Uncertainty decomposition and review gating

Industrial recognition systems must distinguish between an image that is clearly acceptable, an image that is clearly irregular, and an image for which the model lacks evidence. Predictive uncertainty can be decomposed into aleatoric and epistemic components. For model parameters ω and target y, a standard decomposition is

$$\mathrm{Var}(y|I) = E_\omega[\mathrm{Var}(y|I,\omega)] + \mathrm{Var}_\omega(\mathrm{E}[y|I,\omega])$$

where the first term represents irreducible observation noise and the second term represents model uncertainty due to limited data or ambiguous inputs. In an industrial setting, aleatoric uncertainty may arise from sensor noise or genuinely ambiguous visual evidence near an acceptance boundary. Epistemic uncertainty may arise from unfamiliar materials, new lighting conditions, or rare defect morphologies. Deep ensembles provide a practical approximation to the second term by training multiple models and measuring disagreement [15], while Bayesian dropout provides an alternative approximation [14]. The distinction matters because aleatoric uncertainty may justify tighter imaging or improved resolution, whereas epistemic uncertainty justifies manual review or additional training examples.

The uncertainty estimate should be used as a gate rather than as a substitute for the irregularity score. A review policy can be written as an indicator decision:

$$D(I) = 1[s(I) \le \tau \wedge u(I) \le q]$$

where D(I)=1 denotes automated acceptance, s(I) is the irregularity score, u(I) is a scalar uncertainty estimate derived from ensemble disagreement or predictive variance, and τ and q are thresholds selected according to operational costs. The complement of D(I) triggers review or rejection according to the plant policy. This formulation allows a system to be conservative without forcing a single global threshold on all images. If the cost of false acceptance is high, q can be lowered so that more uncertain cases are reviewed. If review

capacity is limited, τ and q can be adjusted to maintain a controlled false-alarm rate. The thresholds are policy parameters, not learned constants, unless a cost-sensitive learning procedure explicitly optimizes them.

4.3 Coupling recognition with process decisions

Recognition outputs become useful only when they are connected to process decisions. Industrial inspection frequently involves accept, rework, quarantine, or reject actions, and the optimal action depends on both the probability of irregularity and the cost of each outcome. Real-time defect recognition for structural timber jointing illustrates how recognition can be coupled with optimized decision making rather than treated as an isolated classification task [17]. The lesson generalizes: a model score should be translated into an action through a decision rule that reflects production consequences. This does not mean that every recognition system requires a full optimization model. It means that threshold selection, review gating, and escalation logic should be designed with the downstream decision in view.

The decision layer should also preserve traceability. For each automated decision, the system should record the irregularity score, uncertainty estimate, active knowledge constraints, and localization evidence. This record supports auditing and helps distinguish model error from specification error. If an operator overrides a decision, the override becomes a valuable label for future calibration. The architecture therefore treats human review not as a failure of automation but as an explicit component of a risk-controlled inspection process.

5. Optimization Strategy and Training Constraints

5.1 Combined objective

The representation, knowledge constraints, and task head are trained with a combined objective. A general form is

where ψ denotes encoder parameters, φ denotes task-head parameters, L_{task} is the supervised or weakly supervised recognition loss, L_k is the knowledge-consistency loss, $R(\psi,\varphi)$ is a regularizer, and λ and β are nonnegative trade-off parameters. The task loss may be cross-entropy for labeled defects, multiple-instance loss for image-level labels, or a contrastive loss that pulls augmented views of the same regular pattern together. The regularizer may include weight decay, spectral normalization, or a penalty on excessive sensitivity to nuisance transformations. The objective is not assumed to have a unique optimum. Its purpose is to make the trade-off between data fit and knowledge consistency explicit and adjustable.

The choice of λ is consequential. If λ is too small, knowledge terms have little effect and the model reduces to a data-driven detector. If λ is too large, the model may satisfy constraints at the expense of recognition accuracy, especially when the constraints are approximate. A practical strategy is to begin with moderate λ values, monitor constraint residuals separately, and increase weights only for constraints that are stable across product families. The weights should not be tuned solely on overall accuracy, because irregular patterns are rare and overall accuracy can be dominated by acceptable images. Constraint-wise validation and false-negative analysis are more informative.

5.2 Weak supervision and label noise

Industrial labels are often weak. Image-level labels may indicate that a batch contains a defect without specifying its location. Operator annotations may be inconsistent near

acceptance boundaries. Weakly supervised learning provides a principled way to use coarse labels through multiple-instance learning, label noise modeling, or attention-based localization [16]. In the present framework, weak supervision interacts with knowledge constraints. A weakly supervised localization head can propose regions that support an image-level irregularity label, while a knowledge constraint can penalize proposals that violate expected morphology or spatial relations. This combination can reduce the risk that the model localizes incidental background features.

Label noise should be modeled rather than ignored. If a small fraction of labels is incorrect, a robust loss or a noise-transition model can reduce overfitting to erroneous annotations. The knowledge layer can also act as a regularizer: a proposal that strongly violates a stable process constraint is less likely to be a true irregularity even if the image-level label is noisy. This is not a substitute for careful annotation, but it improves resilience when labels are imperfect.

5.3 Calibration and threshold selection

Calibration is the process of aligning predicted scores with observed frequencies or costs. For irregularity detection, calibration should be assessed separately for the score $s(I)$ and the uncertainty $u(I)$. A model may rank images well while producing poorly calibrated probabilities. Temperature scaling, isotonic regression, or ensemble calibration can be applied when validation data are sufficient. Threshold selection should then account for prevalence and cost. In a production line with very low irregularity prevalence, even a small false-positive rate can generate a large review burden. The decision rule in Section 4.2 makes this trade-off explicit by separating the irregularity threshold τ from the uncertainty threshold q.

Thresholds should be monitored after deployment because prevalence, imaging conditions, and product mix can change. A drift in the distribution of $s(I)$ or $u(I)$ may indicate a change in illumination, material batch, or camera alignment rather than a change in defect rate. Monitoring both score distributions and review rates helps distinguish model degradation from process change. This is a practical requirement for industrial vision systems, where the cost of silent false negatives can be high.

6. Analytical Assessment of Behavior and Failure Modes

6.1 Why knowledge constraints can improve sample efficiency

Irregular patterns are rare by definition. A purely supervised model must learn the boundary between acceptable and unacceptable variation from a small number of examples, and it may instead learn correlations with batch, lighting, or camera position. Knowledge constraints reduce the hypothesis space by requiring the representation to respect stable structure. For example, a shading-consistency constraint discourages the model from treating a benign illumination gradient as defect evidence, and a directionality constraint discourages it from treating expected machining marks as scratches. This narrowing of the hypothesis space can improve sample efficiency because fewer examples are needed to identify the relevant variation.

The benefit is not automatic. Knowledge constraints help only when they are approximately correct and when they are aligned with the recognition task. A constraint that is irrelevant to irregularity adds optimization burden without improving discrimination. A constraint that is systematically wrong can suppress true evidence. The framework therefore

treats knowledge as a soft bias rather than as a hard rule. The trade-off parameter λ controls how strongly the bias is enforced, and constraint residuals should be inspected during training.

6.2 Prior-data conflict and systematic false negatives

The most important failure mode is prior-data conflict. Suppose a morphological constraint encodes the expectation that cracks are elongated and low-contrast. If a new material produces cracks that are compact and bright, the constraint will penalize the correct evidence. The model may then assign a low irregularity score to a true defect. This failure is systematic rather than random, and it can be more dangerous than a generic false alarm because it is consistent across similar images. Mitigation requires constraint diagnostics, out-of-distribution detection, and periodic review of constraint residuals. When residuals are high for images that operators label as irregular, the constraint should be relaxed or made conditional on material or process context.

A related failure occurs when the knowledge layer encodes a specification that is outdated. If acceptance criteria change, the model may continue to enforce the old criteria. This is a governance issue as much as a modeling issue. The knowledge terms should be versioned and linked to process specifications so that changes can be propagated deliberately. The recognition system should not silently learn a new boundary without an explicit update.

6.3 Shortcut learning and nuisance sensitivity

Deep models are prone to shortcut learning when training data contain spurious correlations. In industrial images, common shortcuts include background color, fixture geometry, illumination direction, and image borders. A model may learn to classify a batch as defective because the defective images were captured under different lighting. Knowledge constraints can reduce this risk by forcing the representation to encode physically meaningful quantities, but they do not eliminate it. Data augmentation, adversarial training, and held-out testing across acquisition conditions are still necessary. The uncertainty pathway provides an additional safeguard because images with unfamiliar nuisance patterns may produce high ensemble disagreement.

Nuisance sensitivity is especially problematic for irregular patterns because the evidence is often low contrast. A small change in illumination can either reveal or obscure a scratch. The recognition system should therefore include acquisition consistency checks and, where possible, use multiple lighting directions or calibrated imaging. The model cannot recover evidence that is absent from the image, and an inspection system that ignores this limitation will produce misleading confidence.

6.4 Localization ambiguity and boundary effects

Irregular patterns often have ambiguous boundaries. A scratch may fade gradually, a porosity cluster may merge with surface texture, and a weld irregularity may transition smoothly into acceptable bead geometry. Pixel-level segmentation can therefore be unstable even when the image-level decision is correct. The architecture should preserve spatial evidence and report localization uncertainty rather than forcing a hard boundary. In evaluation, region-based metrics and boundary tolerance are more informative than exact pixel accuracy. The decision layer should also account for localization confidence when the downstream action depends on defect extent.

Boundary effects interact with knowledge constraints. A smoothness or compactness penalty may improve segmentation of compact defects but can truncate elongated scratches.

A directionality constraint may help identify scratches but may also merge them with machining marks. These trade-offs should be evaluated by defect type rather than by a single aggregate score. The framework therefore supports type-specific heads or conditional constraints, allowing different irregularity families to be modeled with different assumptions.

6.5 Explainability and auditability

Visual explanation methods can support operator trust by highlighting image regions that contribute to a decision. Gradient-weighted class activation mapping provides a practical way to localize evidence in convolutional networks [18]. However, explanation maps should not be treated as proof of causal reasoning. They can be unstable, and they may highlight correlated context rather than the defect itself. In the present framework, explanations are most useful when combined with knowledge residuals and uncertainty estimates. An explanation that highlights a region inconsistent with the active knowledge constraints should trigger review rather than automatic acceptance.

Auditability also requires model documentation. The system should record which knowledge constraints were active, which prototypes were nearest, and which uncertainty components contributed to the decision. This record supports root-cause analysis when a false negative occurs. It also allows operators to distinguish between a model that failed to see evidence and a model that saw evidence but applied an inappropriate acceptance criterion.

7. Implementation Logic and Evaluation Protocol

7.1 System integration and data flow

An implementation of the framework would acquire images under controlled illumination, preprocess them for calibration and region-of-interest extraction, compute multi-scale latent representations, apply conditional knowledge heads, score irregularity against regular-pattern prototypes, estimate uncertainty through an ensemble or Bayesian approximation, and apply the accept-review-reject policy. The system should maintain separate logs for image metadata, model version, knowledge-constraint version, score, uncertainty, and final action. This separation is important because it allows the plant to audit decisions without exposing raw model internals.

The data flow should also support active learning. Images that trigger review are the most informative for improving the model, especially when operators provide location annotations or acceptance rationale. These reviewed cases can be used to refine prototypes, recalibrate thresholds, and identify constraints that are too strict or too permissive. The recognition system is therefore part of a feedback loop rather than a one-time classifier. The loop should be governed by clear criteria for model updates so that changes are deliberate and traceable.

7.2 Evaluation metrics and validation design

Evaluation should reflect the operational cost structure. Image-level metrics such as area under the receiver operating characteristic curve and precision-recall curves are useful for ranking, but they do not capture localization quality or review burden. Region-level metrics such as intersection over union, boundary tolerance, and region-level precision and recall are appropriate when masks are available. For irregular patterns with ambiguous boundaries, tolerant boundary metrics and detection-at-overlap criteria are more informative than exact pixel accuracy. Calibration should be assessed with reliability diagrams or expected calibration error, and review efficiency should be reported as the proportion of images sent to manual inspection at a given false-negative rate.

Validation should include shifts in illumination, material batch, camera calibration, and product variant. A model that performs well only on the training acquisition conditions is not adequate for industrial deployment. Cross-condition validation is especially important for knowledge-guided models because constraints may be tuned to one condition and fail under another. The validation design should also include rare irregularity types, even if they must be oversampled or evaluated through targeted challenge sets. Aggregate accuracy is insufficient when the cost of missing a rare defect is high.

7.3 Relation to industrial inspection architectures

Industrial inspection has long relied on task-specific architectures and feature extraction strategies. Deep convolutional architectures have been designed for automated feature extraction in industrial inspection contexts, showing that domain-specific network design can outperform generic transfer in some settings [20]. The present framework is consistent with that lesson but adds explicit knowledge constraints and uncertainty-aware decision logic. It does not replace domain-specific architecture design; it provides a structure for integrating process knowledge into the representation and decision layers. The practical benefit is expected to be greatest when data are limited, when nuisance variation is strong, or when the cost of false acceptance is high.

The framework also complements reconstruction-based and one-class methods. Those methods are useful for discovering unknown irregularities, but they can be sensitive to nuisance variation and can produce scores that are difficult to calibrate. The prototype-based score and knowledge constraints provide additional structure, while the uncertainty gate prevents overconfident automation in unfamiliar conditions. The combination is not a claim of universal superiority. It is a design pattern for industrial recognition problems in which irregularity is rare, appearance is variable, and decisions have consequences.

8. Conclusion

This article has developed a knowledge-guided machine learning framework for automated recognition of irregular patterns in industrial image analysis. The central problem is not simply classification accuracy but reliable separation of pattern-bearing evidence from benign process and imaging variation. The proposed formulation decomposes image formation, encodes differentiable knowledge constraints, represents acceptable appearance through regular-pattern prototypes, scores irregularity with a covariance-normalized distance, estimates predictive uncertainty, and gates automated decisions through a review policy. The method couples recognition with downstream process decisions and preserves traceability through score, uncertainty, and constraint logs.

The analytical assessment shows that knowledge constraints can improve sample efficiency and shift robustness when they are approximately correct, but they can also create systematic false negatives when they conflict with data or encode outdated specifications. Uncertainty estimation and review gating are therefore not optional additions; they are safeguards against overconfident automation. The framework is most appropriate for industrial settings where irregular patterns are rare, visual evidence is low contrast, nuisance variation is strong, and the cost of false acceptance is high. Its scientific significance lies in providing a coherent design logic that links representation learning, process knowledge, irregularity scoring, calibration, and decision making without conflating statistical rarity with functional nonconformance. Its practical significance lies in offering an auditable structure for

inspection systems that must operate under changing production conditions while maintaining controlled risk.

References

1. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), Advances in neural information processing systems (Vol. 25, pp. 1097–1105). Curran Associates.

2. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

3. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778). IEEE.

4. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), Advances in neural information processing systems (Vol. 28, pp. 91–99). Curran Associates.

5. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), Medical image computing and computer-assisted intervention – MICCAI 2015 (pp. 234–241). Springer.

6. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. In Proceedings of the 8th IEEE International Conference on Data Mining (pp. 413–422). IEEE.

7. Schölkopf, B., Williamson, R. C., Smola, A. J., Shawe-Taylor, J., & Platt, J. C. (2000). Support vector method for novelty detection. In S. A. Solla, T. K. Leen, & K.-R. Müller (Eds.), Advances in neural information processing systems (Vol. 12, pp. 582–588). MIT Press.

8. Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2019). MVTec AD: A comprehensive real-world dataset for unsupervised anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9592–9600). IEEE.

9. Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., Dietterich, T. G., & Müller, K. R. (2021). A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE, 109(5), 756–795.

10. Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2021). Deep learning for anomaly detection: A review. ACM Computing Surveys, 54(2), Article 38.

11. von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Pfrommer, J., Pick, A., Ramamurthy, R., Walczak, M., Machicao, J., Bauckhage, C., & Schuecker, J. (2023). Informed machine learning – A taxonomy and survey of integrating prior knowledge into learning systems. IEEE Transactions on Knowledge and Data Engineering, 35(1), 614–633.

12. Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3(6), 422–440.

13. Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. ACM Computing Surveys, 55(4), Article 82.

14. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the 33rd International Conference on Machine Learning (pp. 1050–1059). PMLR.

15. Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), Advances in neural information processing systems (Vol. 30, pp. 6402–6413). Curran Associates.

16. Zhou, Z. H. (2018). A brief introduction to weakly supervised learning. National Science Review, 5(1), 44–53.

17. Luo, D., Gattas, J. M., & Tan, P. S. S. (2020, July). Real-time defect recognition and optimized decision making for structural timber jointing. In The International Conference on Computational Design and Robotic Fabrication (pp. 36-45). Singapore: Springer Singapore.

18. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (pp. 618–626). IEEE.

19. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), Advances in neural information processing systems (Vol. 30, pp. 5574–5584). Curran Associates.

20. Weimer, D., Scholz-Reiter, B., & Shpitalni, M. (2016). Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. CIRP Annals, 65(1), 417–420.