# Supplementary Material
# Evidence–Mechanism Routing for Latent Mechanism-Guided Zero-Transfer Industrial Anomaly Recognition

Changsong Du[1], Hongyu Zhou[1],YiLin Zhou[1],Changhui Yang[1,2,*]
[1]College of Mechanical Engineering, Chongqing University of Technology, Chongqing 400054, China
[2]Key Laboratory of Advanced Manufacturing Technology for Automobile Parts,
Ministry of Education, Chongqing 400054, China
*Corresponding author: yangchanghui@cqut.edu.cn

This supplementary material provides additional route-level analyses for Evidence–Mechanism Routing (EMR). The main manuscript establishes the core zero-transfer evidence–mechanism route through path activation, direct appearance-to-class collapse under cross-part oracle crops, S3 cross-part evidence discovery, and the S1–S5 diagnostic scenario matrix. To avoid confusion with the S1–S5 diagnostic scenarios in the main manuscript, supplementary modules are denoted as SM-1–SM-7.

The following modules are organized according to their main-manuscript usage: fixed operational configuration, parameter-level localization behavior, training-set latent readouts, evidence-supported localization visualizations, Random-GT-size control, photometric perturbation probes, and route-level implications.

These materials do not introduce a separate method. They document how intermediate route variables behave before the final class explanation or localization output, and clarify the fixed operational scope of the reported experiments.

## SM-1 Reproducibility Scope, Fixed Configuration, and Training Objectives

This module summarizes the fixed operational scope and training-objective implementation of the reported EMR experiments. All reported results are produced by one instantiated route unless an ablation setting is explicitly named. The purpose of this module is to clarify both the fixed inference chain used in S1–S5 and the operational training objectives used to instantiate the evidence–mechanism routing route.

**SM-1A. Fixed operational scope.** The fixed inference chain is

$$\begin{aligned} O \to \hat{F} \to R \to \mathcal{R}_{\text{cand}} = \{r_i\}_{i=1}^{K}, \\ r_i \to E_i \to M_i \to S_{\text{loc}}(r_i) \to B_i^{\text{ref}}, \\ \{B_i^{\text{ref}}\} \to G_m \to (B_m^{\text{inst}}, \mathcal{B}_m^{\text{sub}}, C_m, M_m). \end{aligned}$$

Here, $O$ denotes the observed image, $\hat{F}$ the estimated pseudo-background, $R = |O - \hat{F}|$ the self-residual, $\mathcal{R}_{\text{cand}}$ the candidate set, $E_i$ the evidence representation, $M_i$ the mechanism response, $S_{\text{loc}}(r_i)$ the localization-retention score, $B_i^{\text{ref}}$ the refined evidence box, $G_m$ the grouped anomaly instance, $B_m^{\text{inst}}$ the instance-level box, $\mathcal{B}_m^{\text{sub}}$ the evidence sub-box set, $C_m$ the group-level class explanation, and $M_m$ the group-level mechanism response.

**Table 1:** Fixed configuration and reproducibility scope of the reported EMR experiments.

| **Item** | **Fixed setting in the reported experiments** |
|---|---|
| Training resource | ID-SGF factual–counterfactual transition pairs with six fine-grained abnormal-transition types mapped to defect, oil, and artifact paths. |
| Test-time input | Observed image only. No paired factual image, manual mask, or target-domain annotation is used during inference. |
| Observed-only change modeling | A pseudo-background is estimated from the observed image, and $R = \lvert O - \hat{F} \rvert$ approximates local abnormal change. |
| Candidate proposal | A high-recall candidate proposal stage is fixed and is not tuned separately for S2–S5. |
| Path activation and rejection | Defect, oil, and artifact are independent target paths. The *other* state is target-path non-activation $[0, 0, 0]$, with $\tau = 0.625$. |
| Localization / classification decoupling | Candidate retention depends on proposal support, abnormal evidence, target-path activation, and background-consistency explanation, rather than final class probability alone. |
| Evidence shrinking and grouping | Retained candidates are refined by concentrated evidence components and grouped by bridge evidence, distance, compactness, and mechanism compatibility. |
| External scenarios | No image from S2–S5 is used for training, validation, threshold selection, target-domain finetuning, or scenario-specific tuning. |
| Evaluation protocol | Hit@0.1, Hit@0.3, Hit@0.5, mAP@0.3, Mean IoU, Mean Cover, Pred/Image, Unmatched Pred/Image, and Random-GT-size control are computed under fixed evaluation scripts. |

The reported results correspond to this fixed instantiated route. Individual modules can be improved in future work, but no module is replaced or re-tuned for the reported S1–S5 evaluations.

## SM-2 Reproducibility Scope and Training Objectives

This module summarizes the fixed inference scope and the training objectives used in the reported EMR experiments. All reported results use the same instantiated route unless an ablation setting is explicitly named. The fixed route is

$$O \to \hat{F} \to R \to \mathcal{R}_{\text{cand}} \to E_i \to M_i \to S_{\text{loc}}(r_i) \to B_i^{\text{ref}} \to G_m \to (B_m^{\text{inst}}, \mathcal{B}_m^{\text{sub}}, C_m, M_m).$$

**Table 2:** Loss terms and inference rules used in EMR.

| **Module** | **Formula** |
|---|---|
| Observed / paired change | $\hat{F} = P(O), \quad R = \lvert O - \hat{F} \rvert, \quad \Delta X_i = \lvert \tilde{X}_i - F_i \rvert \odot A_i$ |
| Evidence-flux decomposition | $E_i = [E_i^{str}, E_i^{pho}, E_i^{shape}]$ |
| Evidence reconstruction | $\mathcal{L}_{rec} = \sum_{b \in \{str,pho,shape\}} \left\lVert \hat{\Delta X}_i^b - \Delta X_i \right\rVert_1$ |
| Evidence support | $\mathcal{L}_{sup} = -\frac{1}{HW} \sum_{u,v} \left[ A_i^{u,v} \log \hat{A}_i^{u,v} + (1 - A_i^{u,v}) \log(1 - \hat{A}_i^{u,v}) \right]$ |
| Latent mechanism routing | $M_i = \Phi_{\text{mech}}(E_i^{str}, E_i^{pho}, E_i^{shape})$ |
| Cause classification | $\mathcal{L}_{cls} = -\sum_{c=1}^{C} y_{i,c} \log P(C_i = c \mid E_i, M_i)$ |
| Mechanism–class soft exclusion | $\tilde{z}_{i,c} = z_{i,c} + \beta \sum_k M_{i,k} \log(\Pi_{k,c} + \epsilon)$ |
| Compatibility penalty | $\mathcal{L}_{comp} = -\sum_c y_{i,c} \sum_k M_{i,k} \log(\Pi_{k,c} + \epsilon)$ |
| Path activation | $\mathcal{L}_{path} = -\sum_{c \in \{defect,oil,artifact\}} [t_{i,c} \log S_{i,c} + (1 - t_{i,c}) \log(1 - S_{i,c})]$ |
| Other rejection | $\text{Other}(r_i) = \mathbb{I}\left[ \max_c S_{i,c} < \tau \right], \qquad \tau = 0.625$ |
| Localization retention | $S_{\text{loc}}(r_i) = \lambda_1 S_{\text{prop}}(r_i) + \lambda_2 S_{\text{evi}}(r_i) + \lambda_3 S_{\text{target}}(r_i) - \lambda_4 S_{\text{bg}}(r_i)$ |
| Target activation | $S_{\text{target}}(r_i) = \max_c S_{i,c}$ |
| Total objective | $\mathcal{L}_{total} = \lambda_{cls}\mathcal{L}_{cls} + \lambda_{sup}\mathcal{L}_{sup} + \lambda_{rec}\mathcal{L}_{rec} + \lambda_{path}\mathcal{L}_{path} + \lambda_{comp}\mathcal{L}_{comp}$ |

These definitions document the reproducible EMR inference-and-training configuration, rather than claiming full training-from-scratch reproducibility.

## SM-3 Parameter Visualization of Candidate Retention and Evidence Shrinking

This module visualizes how candidate-retention thresholding and support-map shrinking affect final boxes.

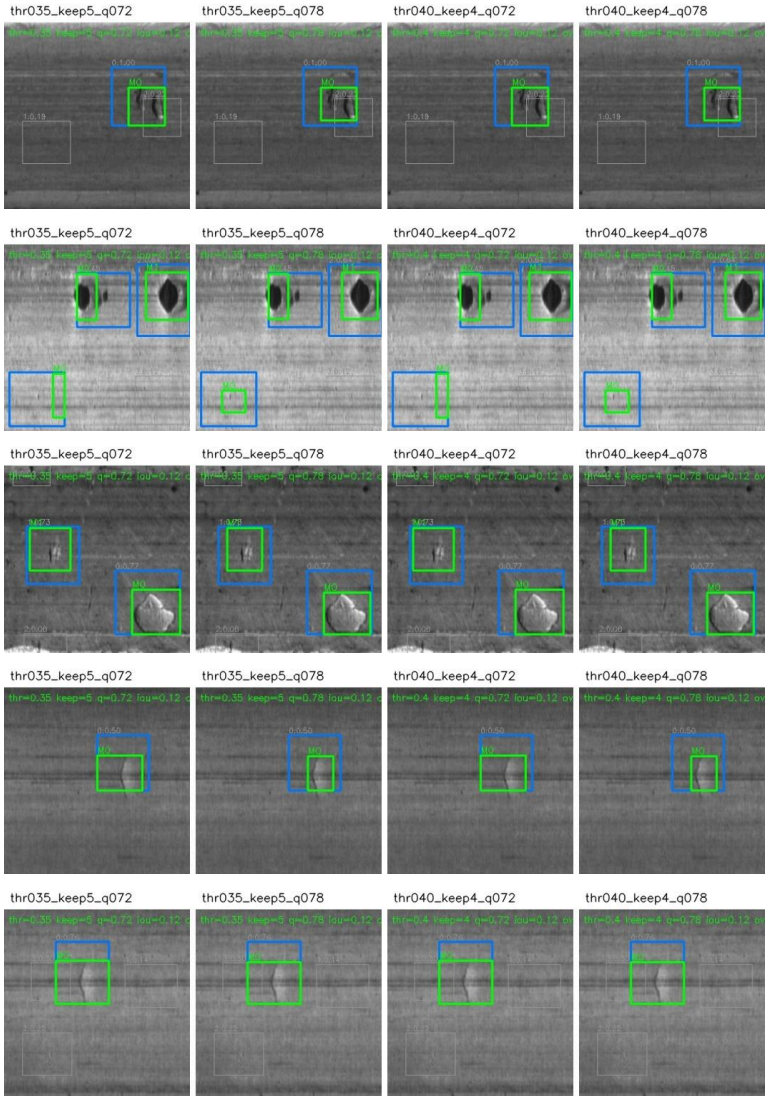


**Figure 1:** Parameter visualization of candidate retention and evidence shrinking. Columns correspond to different `select_thr` and `shrink_q` settings. White or light-gray boxes are raw proposals, blue boxes are retained proposals, and green boxes are merged final boxes.

The parameter `select_thr` controls whether a candidate is retained according to its localization-retention score. The parameter `shrink_q` controls the support-map quantile used during evidence shrinking; larger values keep only higher-response regions and produce more compact refined boxes, but may remove weak or diffuse evidence. This visualization shows how high-recall raw proposals are converted into compact evidence-supported outputs.

## SM-4 Training-Set Latent Mechanism Readout Profiles

This module reports latent readout profiles on EMR training candidates. The main manuscript reports the corresponding held-out in-domain test-set profiles. Together, the two views show whether the learned readout structure is preserved beyond the training samples.

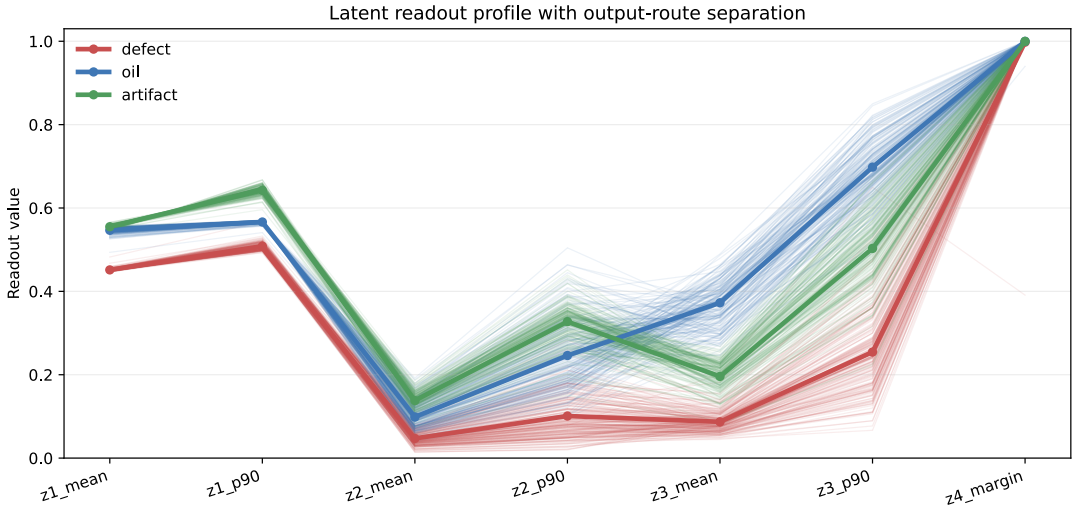


**Figure 2:** Training-set class-conditioned latent mechanism readout profiles. Thin lines denote individual training candidates and thick lines denote class-level averages. $z_1$–$z_3$ summarize latent readout groups, while $z_4$ denotes route-margin or path-level certainty.

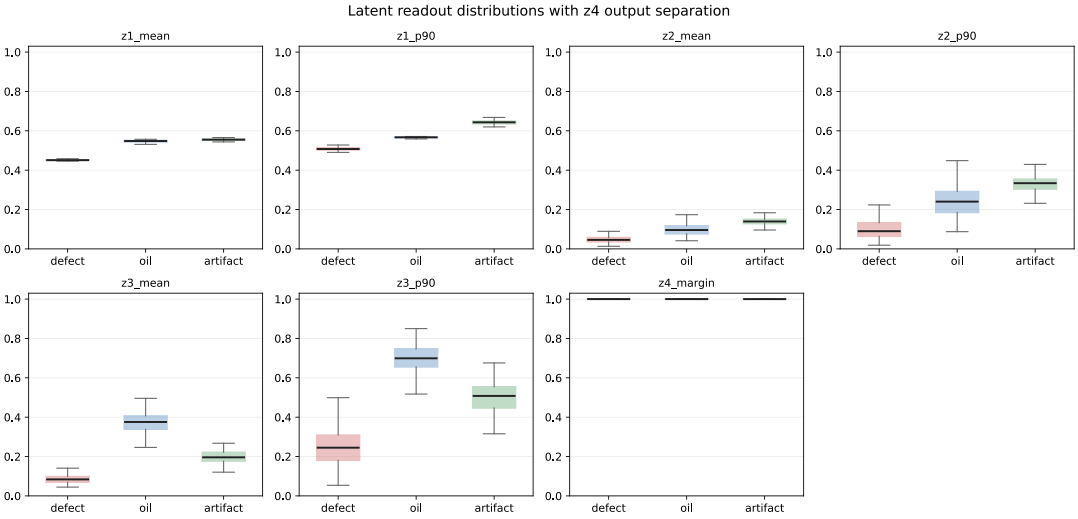


**Figure 3:** Training-set distribution of latent readouts across defect, oil, and artifact samples. The distribution plot complements the profile plot by showing spread and class-conditioned variation.

The training-set profiles show shared abnormal-evidence activation together with class-conditioned tendencies. The latent layer therefore does not reduce candidates to a single anomaly score, but organizes evidence through multiple readout dimensions. The readouts $z_1$–$z_3$ are unnamed latent variables and are interpreted from response patterns rather than manually assigned physical labels. In contrast, $z_4$ is a route-margin or path-level certainty readout derived from output behavior.

## SM-5 Evidence-Supported Localization Visualization

This module visualizes how EMR processes candidate regions through observed-image responses, path-related localization maps, evidence or certainty readouts, and final evidence-supported regions.

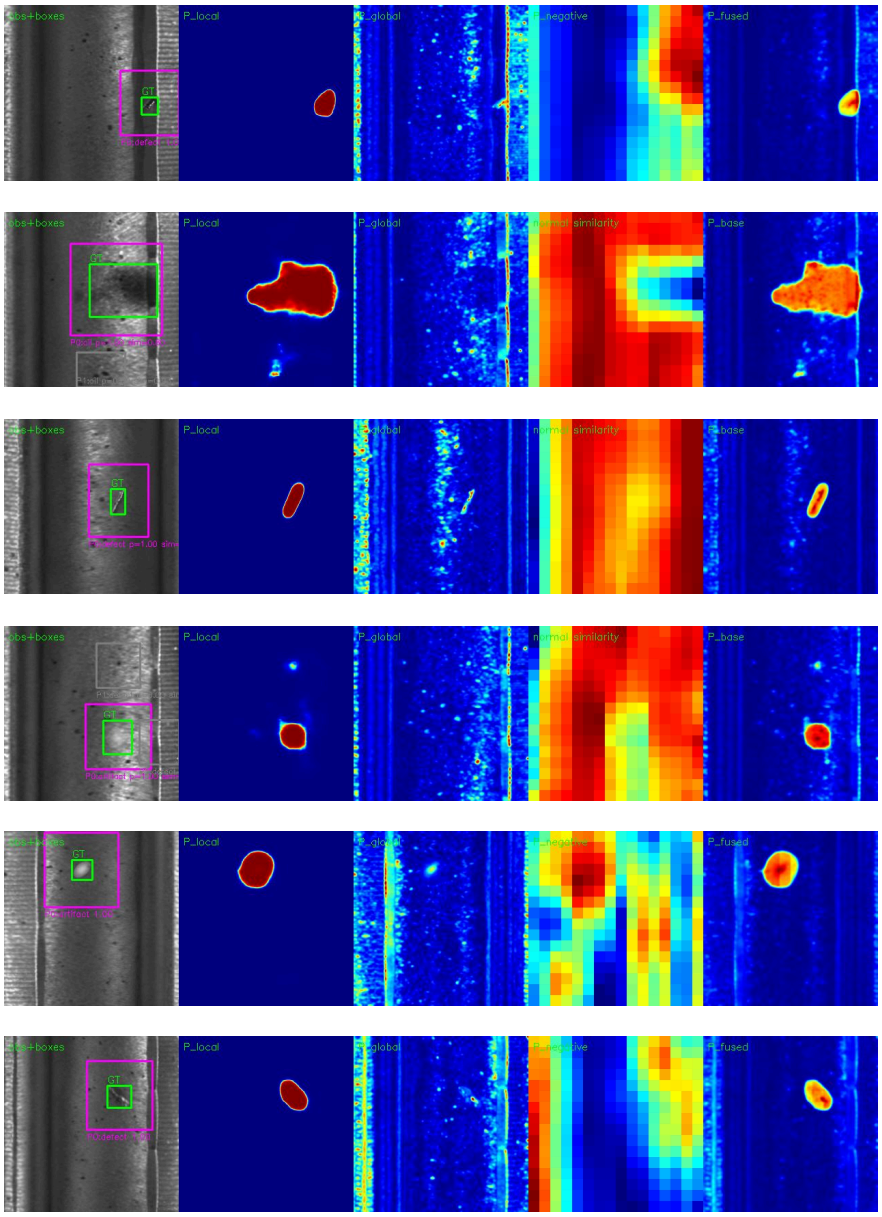


**Figure 4:** In-domain evidence-supported localization process. Each row shows one candidate-level example from observed region and predicted boxes to intermediate heatmaps and final evidence-supported localization.

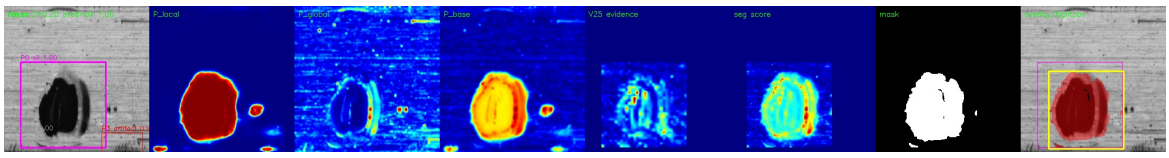


**Figure 5:** S3 cross-part evidence-supported localization. Candidate boxes are refined through concentrated abnormal evidence and path-level support under unseen-part and changed-imaging conditions.

The in-domain and S3 examples show that EMR does not simply output the largest high-recall proposal. Candidate boxes are refined through evidence concentration, target-path activation, and background-consistency filtering.

## SM-6 Complete Random-GT-Size Control

Random-GT-size control estimates how much box-level Hit can be caused by GT spatial occupancy and proposal budget alone. It uses no image content or model response, and samples random boxes according to GT counts and GT-size distributions:

$$N_{\text{rand}} = \alpha N_{\text{gt}}.$$

**Table 3:** GT-size random box control under different proposal budgets.

| Scenario | $\alpha$ | Pred/Image | Hit@0.1 | Hit@0.3 | Hit@0.5 | Mean Cover |
|---|---|---|---|---|---|---|
| S1 | 1 | 1.0000 | 0.0930 | 0.0158 | 0.0000 | 0.0818 |
| S1 | 2 | 2.0000 | 0.2000 | 0.0509 | 0.0105 | 0.1536 |
| S1 | 3 | 3.0000 | 0.2912 | 0.0667 | 0.0158 | 0.2246 |
| S2 | 1 | 1.2042 | 0.1839 | 0.0575 | 0.0115 | 0.1270 |
| S2 | 2 | 2.4083 | 0.2730 | 0.0718 | 0.0172 | 0.2110 |
| S2 | 3 | 3.6125 | 0.4023 | 0.1379 | 0.0345 | 0.2982 |
| S3 | 1 | 1.6154 | 0.0952 | 0.0476 | 0.0476 | 0.1515 |
| S3 | 2 | 3.2308 | 0.5714 | 0.1429 | 0.0476 | 0.3882 |
| S3 | 3 | 4.8462 | 0.4762 | 0.0952 | 0.0476 | 0.3003 |
| S4 | 1 | 0.1170 | 0.0662 | 0.0147 | 0.0000 | 0.0540 |
| S4 | 2 | 0.2340 | 0.1250 | 0.0184 | 0.0000 | 0.0848 |
| S4 | 3 | 0.3511 | 0.1397 | 0.0331 | 0.0037 | 0.1131 |
| S5 | 1 | 3.0667 | 0.5435 | 0.1957 | 0.0489 | 0.3015 |
| S5 | 2 | 6.1333 | 0.6685 | 0.3207 | 0.0870 | 0.4531 |
| S5 | 3 | 9.2000 | 0.8261 | 0.4837 | 0.1413 | 0.5620 |

The control shows that box-level Hit has different chance levels across diagnostic scenarios. In particular, the S5 diagnostic scenario has a high random-overlap level because of large GT regions and distributed defects. Therefore, S5 should be interpreted together with GT spatial occupancy, proposal budget, and the background mechanism prior.

## SM-7 Route Stability under Controlled Illumination Perturbations

This module provides a qualitative route-stability probe under controlled photometric changes, including the original image, global darkening, strong brightening, and left–right illumination imbalance.

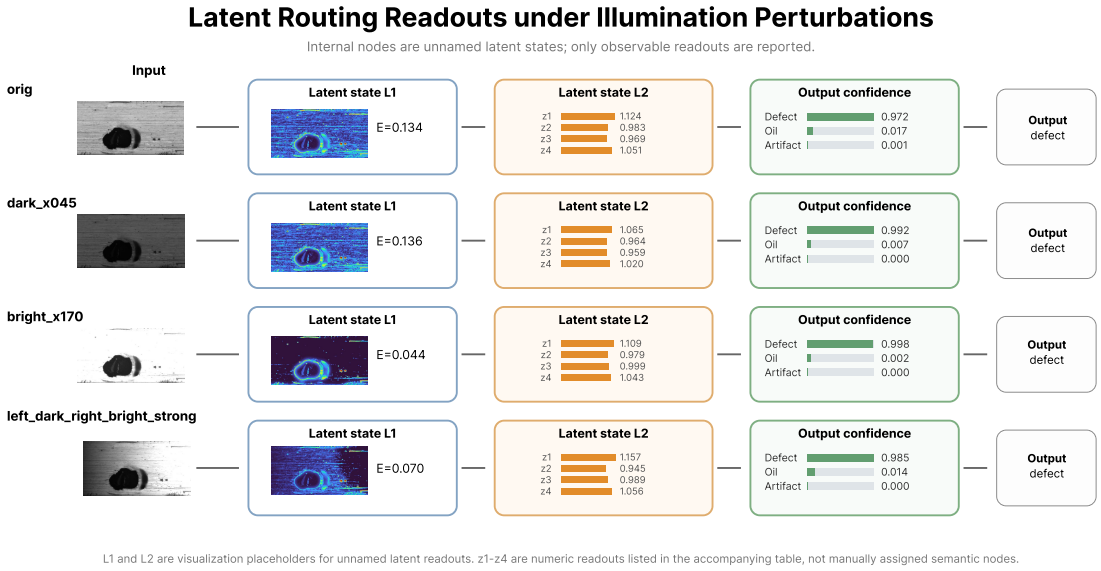


**Figure 6:** Latent routing readouts under illumination perturbations. Low-level evidence responses vary with photometric changes, while the final defect pathway remains consistently activated.

All tested variants are still routed to the defect output with high confidence. This suggests that, for this sample, the defect-related route is not determined only by brightness change. The result should be read as a qualitative route-stability probe rather than a formal intervention test.

## SM-8 Route-Level Implications of Supplementary Analyses

Table 4 summarizes how the supplementary modules support a mechanism-oriented reading of EMR.

**Table 4:** Route-level implications supported by the main experiments and supplementary analyses.

| Route-level property | Main evidence | Supplementary support | Implication |
|---|---|---|---|
| Change-driven evidence learning | EMR is trained from factual–counterfactual transitions and tested with observed residual evidence. | Fixed-configuration analysis and route visualizations show how event-induced changes enter evidence readouts and candidate decisions. | Event-based evidence |
| Beyond direct appearance mapping | Direct crop-to-class classifiers collapse under cross-part oracle crops, while EMR preserves abnormal evidence in S3. | Localization and perturbation visualizations show that appearance alone does not determine the final route. | Route necessity |
| Path activation and non-activation | Target-path activation improves candidate filtering; *other* is modeled as path non-activation. | Fixed configuration and training-set latent readouts expose intermediate responses before final class explanation. | Selective routing |
| Mechanism-compatible zero-transfer | S3 and the S2–S5 matrix show stronger transfer when unseen scenarios remain compatible with learned evidence routes. | S3 localization visualization and Random-GT-size control clarify behavior beyond dense prediction or random overlap. | Compatibility range |
| Observable intermediate responses | The route exposes evidence flux, target-path activation, background explanation, shrinking, and grouping stages. | Latent readouts, localization maps, and parameter visualizations show route variables beyond final class scores. | Route inter-pretability |
| Expandable prior and mechanism space | Main-manuscript S5 cases indicate the need for broader background priors and global-state evidence. | Random-GT-size control and fixed-configuration analysis clarify the current operating range. | Expandable range |

Overall, the supplementary analyses show that EMR exposes intermediate variables and decision stages rather than only final anomaly scores. These variables do not yet define a complete causal model, but they provide an experimental basis for future mechanism-aware and causal-oriented visual inspection.