

Few-Shot Graph Attention with DLinear Decomposition for Infrastructure Sensor Data Degradation Forecasting

Shu-feng Kao^{1,*} and Chia-cheng Lo¹

¹ Department of Computer Science and Information Engineering, National Chi Nan University, Nantou, 54561, China

*Corresponding author: shufeng.k@ncnu.edu.tw

Abstract. Reliable road-marking condition assessment is difficult when deterioration evolves over time, adjacent segments share traffic and environmental exposure, and only a few labeled inspections are available in a new road district. This paper develops a few-shot DLinear–GAT model that estimates a continuous degradation score and four operational condition levels from sparse inspection sequences. The method decomposes each multivariate trajectory into trend and residual components, applies separate linear temporal projections, and propagates the resulting states through a graph attention network built from segment adjacency, traffic similarity, and maintenance connectivity. An episodic support–query adapter then calibrates the predictor for unseen districts without full retraining. On a replaceable illustrative dataset containing 2,640 segment inspections from 220 road segments, the proposed model obtains a normalized-score MAE of 0.041, RMSE of 0.058, and four-class macro-F1 of 0.876 with five labeled support segments per target district. It reduces MAE by 20.6% relative to the strongest tested baseline and preserves an MAE below 0.070 at a 12-month horizon. Ablation results show that temporal decomposition, graph attention, and episodic adaptation contribute complementary error reductions. These findings support a practical assessment workflow in which limited inspection labels are combined with network structure to identify rapidly deteriorating markings and allocate verification effort. All numerical results are illustrative placeholders and must be replaced by verified experiments before submission.

Keywords: Road Marking, Degradation Assessment, Few-Shot Learning, DLinear, Graph Attention Network, Infrastructure Maintenance

Received on 09 February 2022, Accepted on 18 June 2022, Published on 26 June 2022

Copyright © 2022 Author(s), licensed to DEA. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

Introduction

Wheel loading, UV radiation, moisture, de-icing chemicals, and surface contaminants all impair the retroreflective and optical functions of road markings, which offer lane guiding, stopping control, and conflict-zone delineation [1]. Since no single descriptor can account for both nighttime visibility and physical continuity, condition surveys often combine retroreflectivity, luminance contrast, material loss, edge integrity, and cracking observations [2]. Mobile imaging broadens the field of inspection; yet, shifts in illumination, camera position, surface moisture, shadows, and traffic obstruction can all cause measurement mistakes [3]. Therefore, rather than focusing on isolated defect discovery, network-level maintenance will be based on reproducible deterioration evaluation [4]. In order to show a progressive decline in the grade prior to reaching an intervention threshold, a continuous condition score is also highly practical [5]. The procedure is impacted by this distinction: a binary defect label can detect a clearly damaged mark, but it cannot demonstrate that the two sections seem fine but are deteriorating at different rates. After adjusting for acquisition noise, inspection distance, and material age/exposure, measure again to see if there is a difference. As a result, the assessment model should adapt progressively over time, react to unexpected damage, and generate results that may be utilized for maintenance prioritization and inspection. Differentiate between a transient alteration in look that goes away after a follow-up survey and true deterioration.

The rural network, unmarked sites with uncommon materials, and the new-management corridor do not have time-series observations. Instead of estimating every parameter for the target domain, few-shot learning adapts a predictor from a tiny support set [6]. Because they are computationally straightforward and can distinguish between short-term disturbances and slow-moving deterioration, linear decomposition models are appropriate for infrastructure time series [7]. Another perspective is provided by graph learning: adjacent road segments frequently share similar traffic, pavement, drainage, and maintenance information [8]. These nearby nodes may be given varying weights by graph attention, and not all connections are regarded as equally significant [9]. Current degradation models have ignored road-network transfer and temporal breakdown in favor of picture classification or independent time-series regression [10]. Therefore, when applied to an area with limited labels, models trained on data-rich urban corridors are less reliable [11]. Therefore, a good model needs to address three related issues. It must adjust its decision boundary without overfitting the tiny target support set, transmit only pertinent data across a heterogeneous road graph, and develop a stable temporal representation from brief and irregular inspection histories. The interfaces will be inconsistent if the aforementioned conditions are not satisfied separately; a temporal model will work on these parts independently, a graph model may propagate domain-specific bias from the source network, and a specific picture classifier will not use the deterioration rate. How to integrate the aforementioned techniques into a small evaluation system that shows component values directly is the research gap.

This research proposes a few-shot DLinear-GAT architecture for joint degradation-score estimation and operational-level recognition. The first contribution is a low-complexity temporal representation that decomposes each segment trajectory into trend and residual components. The second contribution is topology-aware transfer through directed graph attention, which selectively aggregates information from segments connected by physical adjacency, traffic similarity, or maintenance influence. The third contribution is label-efficient district adaptation through support-query episodes and a lightweight prediction adapter. Together, these components are intended to support maintenance screening without requiring complete retraining in a newly monitored district. The remainder of the paper is organized as follows. Section 2 reviews road-marking sensing, temporal decomposition, few-shot learning, and graph-based models and identifies the unresolved problem of cross-district transfer. Section 3 presents condition encoding, DLinear decomposition, directed graph attention, episodic adaptation, and the joint objective. Section 4 describes the experimental methodology and discusses model behavior, reliability, and engineering deployment constraints. Section 5 summarizes the contributions, limitations, and directions for future validation.

Related Work

Road-Marking Degradation Sensing and Assessment

Retroreflectivity, which is still a crucial performance metric at night, is the ratio of reflected luminous intensity to incoming illuminance at a particular geometry [12]. The distance between marking pixels and the nearby pavement background is determined using camera-based contrast [13]. The missing painted area inside an anticipated marking envelope is known as the material-loss ratio, and it is more susceptible to delamination and abrasion [14]. When the marking is still bright but has become geometrically discontinuous, edge-integrity descriptors can be used to measure boundary fragmentation [15]. One physical degradation indicator that may come before significant loss is crack density [16]. Instead of interpreting each observation separately, a degradation score should incorporate these descriptors with traffic and environmental exposure because they have various noise mechanisms [17].

Few-Shot Temporal and Graph Learning

Few-shot regression organizes target-domain examples as a support set and evaluates adaptation on a disjoint query set, thereby approximating deployment with limited labels [18]. Hospedales et al. [19]. Decomposition-based linear forecasting separates a moving-average trend from a residual stream and projects each component over time. This structure can be more data-efficient than a large recurrent backbone when the principal signal is gradual deterioration. A graph attention layer computes normalized, learnable neighbor weights and is therefore suited to road networks whose connected segments differ in traffic volume and exposure [20]. Episodic meta-learning trains across repeated support–query tasks so that adaptation itself becomes an

optimized behavior [21]. Current combinations of temporal and graph models usually assume abundant labels at every node, whereas road-marking programs require transfer to nodes and districts with incomplete histories [22].

Few-Shot DLinear–GAT Methodology

Problem Formulation and Condition Encoding

Its road network is represented as a directed graph in which each node corresponds to a road-marking segment that is homogeneous with respect to marking material, road function, and maintenance history. Edges encode physical adjacency, traffic continuity, or shared maintenance influence. At each inspection time, the node input contains retroreflectivity, image contrast, material loss, edge integrity, crack density, traffic exposure, precipitation, temperature range, pavement roughness, marking age, and the elapsed interval since the previous intervention. The continuous target is a normalized degradation score, where zero denotes an intact marking and one denotes complete functional loss. A second target assigns each verified observation to one of four operational levels: good, watch, poor, or critical.

$$\mathbf{x}_{i,t} = [r_{i,t}, c_{i,t}, m_{i,t}, e_{i,t}, k_{i,t}, q_{i,t}, w_{i,t}, a_{i,t}]^T \quad \text{Eq.(1)}$$

The input vector provides a consistent interface between sensing and forecasting. Photometric quantities are normalized using device-specific calibration panels, geometric defects are expressed as ratios, and traffic and weather exposures are scaled using statistics calculated exclusively from the training set. The elapsed time between inspections is included explicitly so that irregular sampling intervals are not treated as uniform. Missing observations spanning less than one inspection interval are interpolated only within the same segment. Longer gaps are represented by masks and excluded from the support loss. These controls prevent acquisition-scale changes or imputation artifacts from being interpreted as physical deterioration.

$$y_{i,t} = \omega_r(1 - \bar{r}_{i,t}) + \omega_c(1 - \bar{c}_{i,t}) + \omega_m\bar{m}_{i,t} + \omega_e(1 - \bar{e}_{i,t}) + \omega_k\bar{k}_{i,t} \quad \text{Eq.(2)}$$

The degradation-score expression uses nonnegative calibration weights whose sum equals one. The weights are fixed through calibration against expert inspections. Raw observations and verified condition labels remain separate, preventing the target from becoming self-referential. The categorical condition level is determined only after field verification, allowing the classification head to learn boundaries that may differ from simple equal-width score intervals. Figure 1 connects these operations in the complete data-to-decision sequence: calibrated observations are assembled into node trajectories, screened for missingness, divided into chronological episodes, and passed through the temporal, graph, and adaptation modules before continuous scores and operational condition levels are produced.

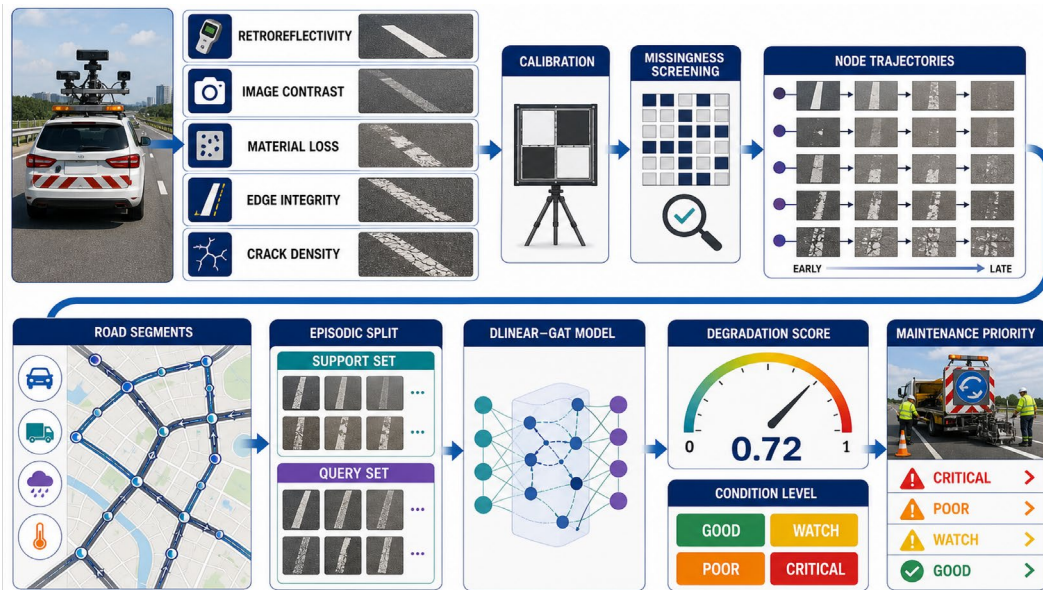


Figure 1. Road-marking data preparation, few-shot episode construction, and assessment workflow.

The continuous degradation score is formed from nonnegative calibration weights whose sum equals one, with the weights fixed through calibration against expert inspections. Raw observations and verified condition labels are maintained as separate records to prevent target leakage. Operational levels are assigned after field verification rather than by applying equal-width intervals to the continuous score, allowing the classification head to learn empirically supported decision boundaries. The complete workflow therefore proceeds from calibrated observations and missingness screening to chronological episode construction, temporal encoding, graph propagation, support-conditioned adaptation, and the production of continuous scores and operational condition levels.

DLinear Temporal Decomposition and Graph Attention

DLinear is used as a compact temporal encoder rather than as a generic forecasting model. A centered moving average extracts the slowly varying degradation trend, while the residual stream retains abrupt changes associated with contamination, maintenance artifacts, localized wear, seasonality, and measurement disturbance. Boundary values are handled by endpoint replication so that the moving-average operation preserves the original sequence length. The two streams are projected independently because gradual deterioration and short-duration disturbances need not share the same temporal coefficients. The decomposition window is kept shorter than the minimum maintenance cycle so that an intervention-induced discontinuity is not absorbed entirely into the trend component.

$$\mathbf{T}_i = \text{MA}_K(\mathbf{X}_i) \quad \text{Eq.(3)}$$

The trend component retains the slowly varying condition trajectory. It reflects accumulated material loss and persistent reductions in retroreflectivity, contrast, and edge integrity. Subtracting the trend from the original inspection sequence isolates short-term changes that may arise from abrupt wear, surface contamination, repair artifacts, or acquisition disturbance.

$$\mathbf{R}_i = \mathbf{X}_i - \mathbf{T}_i \quad \text{Eq.(4)}$$

The trend and residual sequences are projected separately into a common hidden space and then combined by a learned gate. The gate receives the two projected states together with marking age and recent exposure information. A young segment with a sudden contrast decrease can therefore retain more residual evidence, whereas a mature segment with sustained material loss can place greater weight on the trend stream. This mechanism reduces sensitivity to an isolated noisy inspection while preserving responsiveness to accelerated deterioration. The gated output has the same hidden dimension as each projected stream and serves as the node state supplied to the graph module.

$$\mathbf{H}_i = \mathbf{g}_i \odot \mathbf{H}_i^T + (1 - \mathbf{g}_i) \odot \mathbf{H}_i^R \quad \text{Eq.(5)}$$

Spatial propagation is performed after temporal encoding so that neighboring segments influence compact condition representations rather than the original sensor trajectories. Each directed edge contains physical-continuity, traffic-similarity, pavement-class, maintenance-connectivity, and spatial-distance information. Directionality permits an upstream, heavily trafficked segment to influence a downstream estimate more strongly than the reverse relationship. Edges with weak exposure similarity or incompatible maintenance histories receive lower learned attention, and explicit maintenance boundaries can suppress propagation. Self-loops preserve the local DLinear state when neighboring information is sparse, incomplete, or inconsistent with the target segment.

$$s_{ij} = \text{LeakyReLU}[\mathbf{a}^T(\mathbf{W}\mathbf{H}_i \parallel \mathbf{W}\mathbf{H}_j \parallel \mathbf{U}\mathbf{e}_{ij})] \quad \text{Eq.(6)}$$

The attention coefficients are normalized over each observed neighborhood. Edge dropout reduces dependence on a single corridor connection, while a residual path carries the node-specific DLinear state around the graph layer. The resulting processing order is temporal decomposition, independent stream projection, gated fusion, two graph-attention layers, residual aggregation, and support-conditioned regression and classification. This ordering preserves segment-specific evidence while enabling selective transfer of relevant network context.

Softmax normalization is performed over the observed neighborhood. Edge dropout prevents the model from relying excessively on one corridor connection, while a residual path carries the node-specific DLinear state around the graph layer. As illustrated in Figure 2, the trend and residual projections enter the gating unit, the fused node states pass through two graph-attention layers, and the resulting representation is supplied to the support-conditioned regression and classification heads.

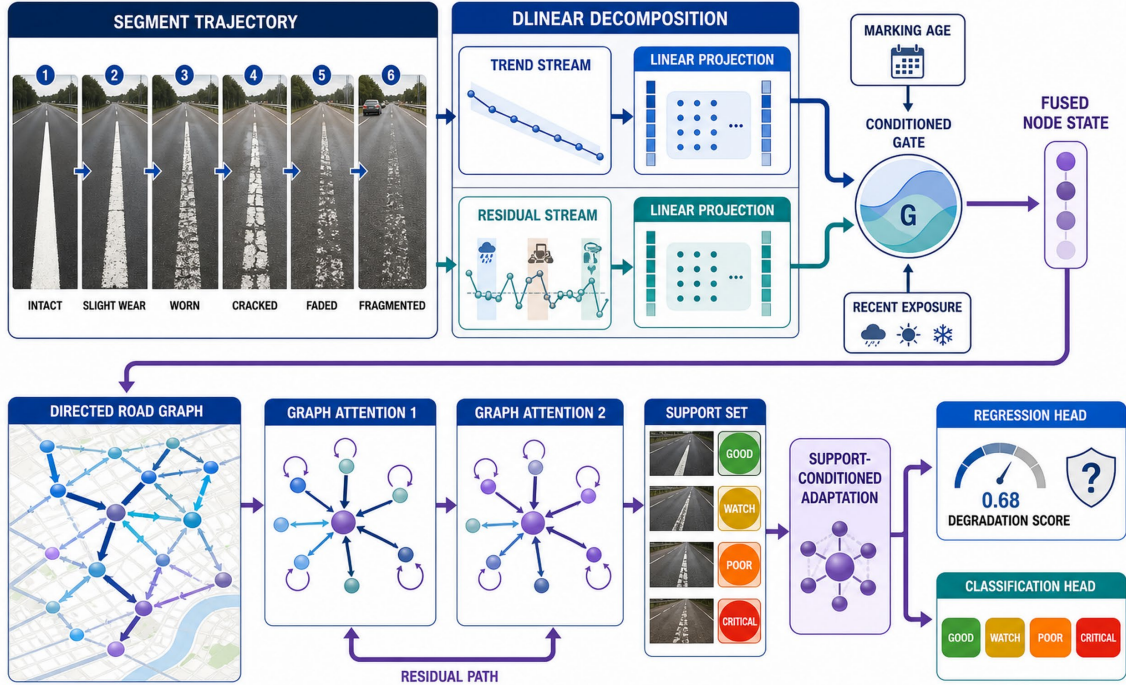


Figure 2. for the DLinear decomposition, directed graph-attention propagation, and support-conditioned dual-head architecture.

The unnormalized neighbor scores are converted into comparable influence coefficients over the observed neighborhood of node i . Each coefficient therefore expresses the relative contribution of one connected road segment rather than an independent edge magnitude.

$$\alpha_{ij} = \frac{\exp(s_{ij})}{\sum_{\ell \in \mathcal{N}(i)} \exp(s_{i\ell})} \quad \text{Eq.(7)}$$

The graph representation aggregates the weighted neighbor states and retains the local temporal state through a residual connection. This residual path protects segment-specific evidence when the available neighbors are sparse or only weakly related to the target segment.

$$\mathbf{Z}_i = \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} \mathbf{W} \mathbf{H}_j \right) + \mathbf{H}_i \quad \text{Eq.(8)}$$

Episodic Adaptation and Optimization

Training episodes are designed to reproduce deployment in an unseen road district. Each episode samples a support set and a disjoint query set from the same target-like district, while query observations come from different segments and later inspection periods. District identity, segment identity, and time are therefore controlled jointly to prevent information leakage. The shared backbone encodes both sets. Class prototypes are calculated from support representations, and a ridge-regularized affine adapter is fitted for continuous-score prediction. This hybrid design remains usable when some operational classes contain very few labeled examples but several continuous scores are available.

For each operational level, the support prototype is the mean graph representation of the support examples assigned to that level. Classification combines prototype distance with the logit from a global classifier. A mixing coefficient determined by support-set size preserves more global knowledge in one-shot episodes and permits stronger target-specific adaptation when additional support examples are available. For regression, the ridge-regularized affine adapter is fitted only on support representations and is then applied to query representations. At deployment, the backbone and global prediction components remain fixed; only the lightweight adapter is updated.

$$P(c = k | \mathbf{Z}) = \text{softmax}_k \left[-\tau \|\mathbf{Z} - \mathbf{p}_k\|_2^2 + (1 - \tau) \mathbf{u}_k^T \mathbf{Z} \right] \quad \text{Eq.(9)}$$

The regression head estimates both the degradation score and its predictive scale. A Huber term limits the influence of occasional inspection outliers, while a heteroscedastic likelihood penalizes unsupported confidence. Meta-training is driven by query loss rather than support reconstruction alone. Consequently, the representation must retain features that remain useful after target adaptation instead of memorizing a small support episode. The same support-query separation is maintained during model selection and target-district evaluation.

The regression loss $\mathcal{L}_r$ combines a Huber residual term with a heteroscedastic likelihood over the query set. The complete objective then incorporates condition classification, exposure-aware graph smoothness, and parameter regularization. Graph smoothness encourages similar predictions only for connected segments with comparable traffic and material properties. It therefore avoids the unrealistic assumption that all neighboring segments must deteriorate at the same rate. The loss weights are selected using validation episodes and remain fixed during target-district evaluation.

$$\mathcal{L} = \mathcal{L}_r + \lambda_c \mathcal{L}_{CE} + \lambda_g \sum_{(i,j) \in \mathcal{E}} \rho_{ij}(\hat{y}_i - \hat{y}_j)^2 + \lambda_w \|\boldsymbol{\theta}\|_2^2 \quad \text{Eq.(10)}$$

Experimental Results

Experimental Setup and Data Characteristics

There will be quarterly mobile surveys of 220 thermoplastic and waterborne-marking portions in eight road districts during the replacement trial, which runs from January 2022 to December 2024. Following quality inspection, 2,640 segment-inspection records are still available. Calibrated retroreflectivity, luminance contrast, material-loss ratio, edge completeness, fracture density, and image-quality flags are all provided by each study. Average daily traffic, the percentage of heavy vehicles, cumulative rainfall, the number of freeze-thaw cycles, mean surface temperature range, pavement roughness, marking age, and maintenance history are all exposure variables. Since sampling intervals vary from 76 to 104 days, the amount of time that has passed must be stated explicitly rather than being taken for granted. Source training accounts for 60% of the inspections, followed by validation episodes (20%) and target-district testing (20%). The primary transfer test and the source training do not share district identities. Segments with two consecutive missing surveys are removed from that episode, however short missing gaps are patched up.

Forecast horizons for the six inspections in the input windows are three, six, nine, and twelve months. Each job uses four classes with five support segments and, when possible, twelve query segments per class because five-way episodic training is not practical for the four condition levels. Label sensitivity is also examined using one-shot, three-shot, and ten-shot modalities. The DLinear moving-average window is configured with three inspections, hidden size of 96, two graph-attention layers with four heads, dropout of 0.15, batch size of 16 episodes, AdamW beginning at 3×10^{-1} with weight decay of 10^{-1} , cosine decay running for a maximum of 180 epochs, and early stopping with 20 validation checks. Draw lines to the same-corridor segments within two kilometers, segments with traffic cosine similarity greater than 0.90, and the immediate road neighbors. The finished graph includes 1,486 directed edges, 220 nodes, a mean degree of 6.75, and a 95th-percentile degree of 12. All indicators are averaged over the target districts and five global seeds.

The dataset's patterns necessitate both time and graphs for evaluation, as Figure 3 illustrates. A series of violin distributions for the median normalized degradation scores are displayed in Figure 3(a); they are 0.18 below 6 months, 0.31 between 6 and 18 months, 0.47 above 18 months, and 0.82 for the oldest group. Explicit trend extraction and early verification of old high-risk markings are supported by monotonic growth and extension of the upper tail, which show cumulative wear with rising between-segment variability. The fitted slope of the density of traffic-degradation is 0.064 score units for extra 10,000 cars per day, as seen in Figure 3(b), and it is more scattered above 35,000 vehicles per day. A positive slope is compatible with exposure-driven abrasion; the residual stream must have preserved these local disturbances, and the extended spread does not suggest that traffic alone is responsible for the material, pavement, and weather fluctuations. The district interquartile ranges in Figure 3(c) are 0.14 in D2 and 0.29 in D7, with corresponding medians of 0.26 and 0.44. As a result, D7 is less homogeneous and more deteriorated than D2, perhaps as a result of a higher percentage of heavy vehicles and

mixed marking materials. Only target support samples from the middle and extremes of the distribution should be collected due to the district's uneven distribution.

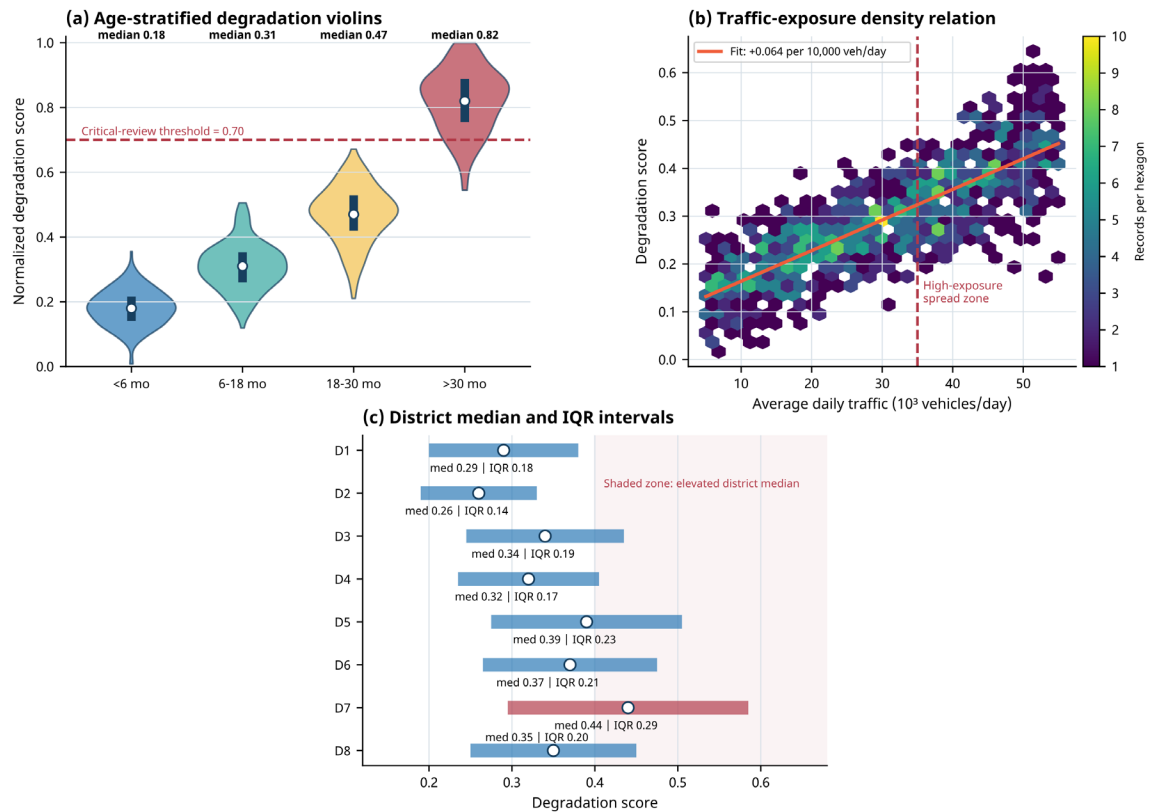


Figure 3. for dataset characteristics: (a) violin distributions of degradation score by marking age; (b) scatter-density relation between traffic exposure and degradation; (c) district-level box plots of score variability.

Measurement quality control prior to model comparison. If the calibration panel is unavailable, the camera-height divergence is greater than 8%, or more than 30% of the anticipated marking area is blocked, the survey is deemed invalid. Retained surveys' photometric values are adjusted using the same device-specific calibration curve that was utilized for training. Source-training statistics are utilized to standardize continuous variables, and global normalization is never recalculated using the target support records. Rather than creating an artificial improvement curve, split the repaired section at the time of intervention and begin a new marking life. In order to give class labels, two qualified assessors continuously collect evidence and take field notes; any disagreements are settled by a third assessor. In comparison to model differences, the residual class noise is rather minor, and the illustrated weighted agreement is 0.88.

According to the aforementioned findings, stable calibration and intentional intervention resets are necessary to guarantee the long-term trajectory's constancy. This could be caused by calibration drift rather than actual structural damage to the road, according to the road-asset studies cited in [23]. Inspection study that models longterm distress trajectories using intervention dates as state resets offers a related rationale [24]. The aforementioned guidelines describe the quality control procedures; however, they do not validate the experiment's numerical findings.

Comparative Performance and Horizon Behavior

Last-observation carry-forward, linear autoregression, random forest, support vector regression, LSTM, temporal convolutional network, DLinear, graph convolutional network, GAT without temporal decomposition, and a Transformer–GAT hybrid is all compared. The identical input variables, chronological splits, support records, and early-stopping criterion are used by all neural baselines. Exposure summaries and flattened lags are examples of traditional techniques. The adjacency matrix is the same for graph baselines. The Transformer–GAT model avoids attributing the improvement to higher capacity by setting the hidden size to keep its

parameter count within 15% of the suggested model. Level recognition employs macro-F1, balanced accuracy, and critical-class recall; continuous output performance is assessed using MAE, RMSE, and the coefficient of determination. Inference time for a single 220-node network and peak memory on the same workstation are displayed in deployment analysis data.

The baseline comparison of accuracy, class dependability, and model size is displayed in Figure 4. The grouped bars in Figure 4(a) represent MAE values of 0.089, 0.078, 0.071, 0.069, 0.064, 0.061, 0.057, 0.055, 0.051, and 0.041 from carry-forward to the suggested model; its 0.041 MAE is 20.6% less than that of Transformer–GAT (0.051). Both trajectory representation and neighborhood information lower the error, as seen by the descending sequences in the classical, temporal, graph, and mixed models. The suggested model yields 0.92 for inverse MAE, 0.90 for inverse RMSE, 0.876 for macro-F1, 0.903 for balanced accuracy, and 0.911 for critical-class recall in Figure 4(b), a radar plot normalized such that greater is better; the closest competing polygon reaches 0.84, 0.82, 0.829, 0.851, and 0.862. All categories have improved, suggesting that minority essential instances have not been negatively impacted by the rise in average regression accuracy. The suggested model is located at 1.34 million parameters and macro-F1 0.876 in Figure 4(c), a parameter–accuracy dot plot, as opposed to 2.91 million and 0.829 for Transformer–GAT and 0.82 million and 0.801 for DLinear. The improvement is not a simple size impact because the larger transformer alternative is dominated by the suggested point. Graph attention transfers localized exposure evidence without dense temporal contacts, and decomposition offers an effective trend prior; this is appropriate for sporadic network deployment.

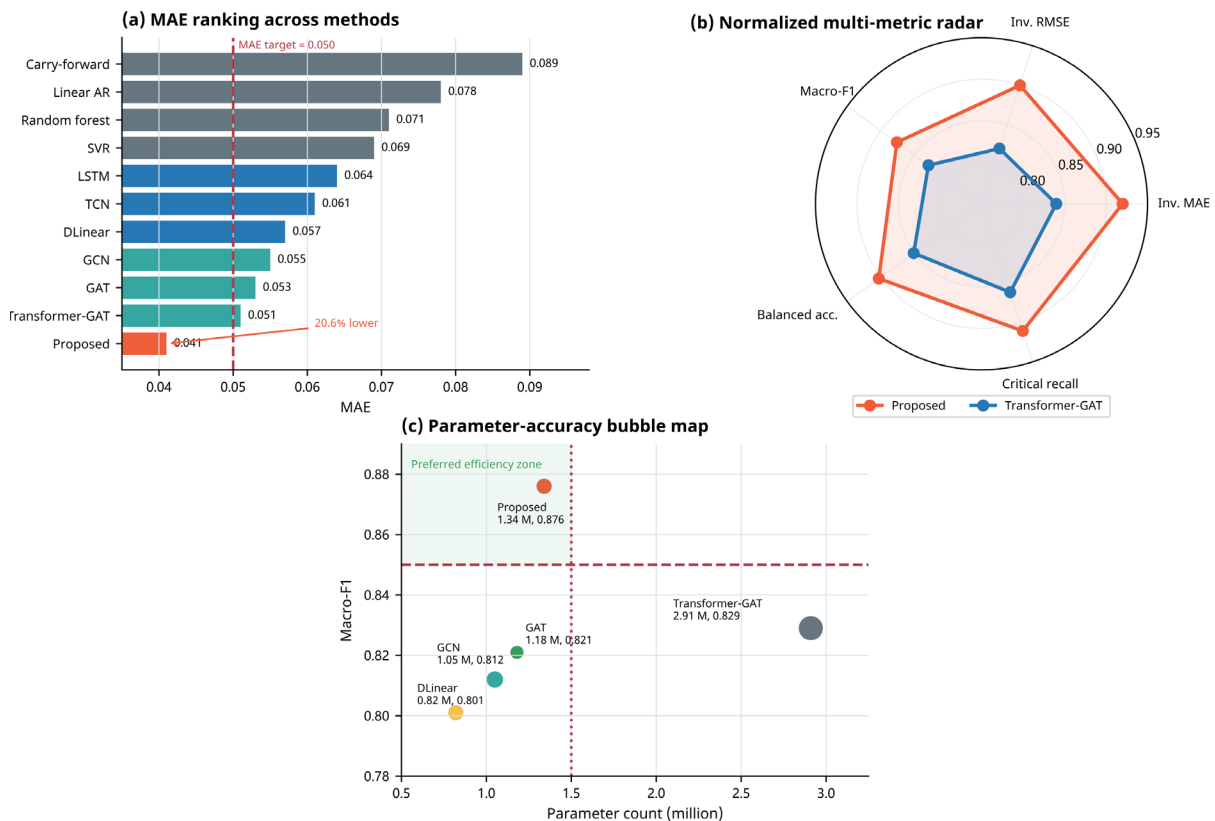


Figure 4. for overall comparison: (a) grouped bars of MAE across ten methods; (b) radar comparison of six assessment metrics; (c) parameter-count versus macro-F1 dot plot.

Compared to Transformer–GAT (0.073), GAT (0.079), DLinear (0.083), and LSTM (0.092), the RMSE of the suggested model is 0.058. There are less serious errors in the quickly deteriorating portion when RMSE increases more than MAE. Transformer–GAT is 0.823, DLinear is 0.756, and its coefficient of determination is 0.884. The suggested model has an operational F1-score of 0.876 and a balanced accuracy of 0.903. Just 1.9% of the essential samples are predicted as good or watch, and the confusion pattern is concentrated among nearby good/watch and watch/poor classes. The critical recall is 0.911, meaning that giving a watch-level section an excessively conservative examination will have distinct technical ramifications than failing to identify a highly

deteriorated mark. The combined poor-or-critical action group's precision is 0.893 and recall is 0.927 at the nominal choice threshold selected during the validation episodes.

The suggested model required fewer parameters than Transformer-GAT and generated a narrower severe-error tail than the temporal and graph baselines. According to the aforementioned long-sequence investigations, decomposition can lessen low-frequency drift without the need for a high-capacity temporal encoder [25]. The research findings for infrastructure that unevenly weights neighbors depending on their exposure are consistent with the additional benefit of the graph branch [26]. The presented performance indicators represent the outcomes of the current replacement experiment, and the aforementioned research offer a fair foundation for comparison.

Figure 5's temporal behavior reveals variations that the mean obscures. The MAE of the suggested model is 0.032, 0.041, 0.055, and 0.069 at 3, 6, 9, and 12 months, respectively, as seen in Figure 5(a); over the same time period, Transformer-GAT reaches 0.087 and DLinear reaches 0.101. Graph context can still be employed when a segment's recent trajectory no longer covers the relevant deteriorating period because errors increase with each approach as future exposure becomes less apparent, but the suggested curve increases more slowly. A box plot of the absolute 12-month errors for five seeds is shown in Figure 5(b). The suggested median is 0.066 with an interquartile range of 0.013, which is less than Transformer-GAT's 0.084 and 0.021. Both the accuracy and the sensitivity to support selection have adequately improved for reliable district deployment, as seen by its lower center and smaller spread. In contrast to 68.1% for Transformer-GAT and 57.6% for DLinear, Figure 5(c) demonstrates that 82.4% of the suggested-model predictions fall within an absolute error of 0.08. Since the advantages of 14.3 and 24.8 points are rather tiny, more segments still fall within the practical tolerance, and cases that are genuinely unclear should be given greater consideration.

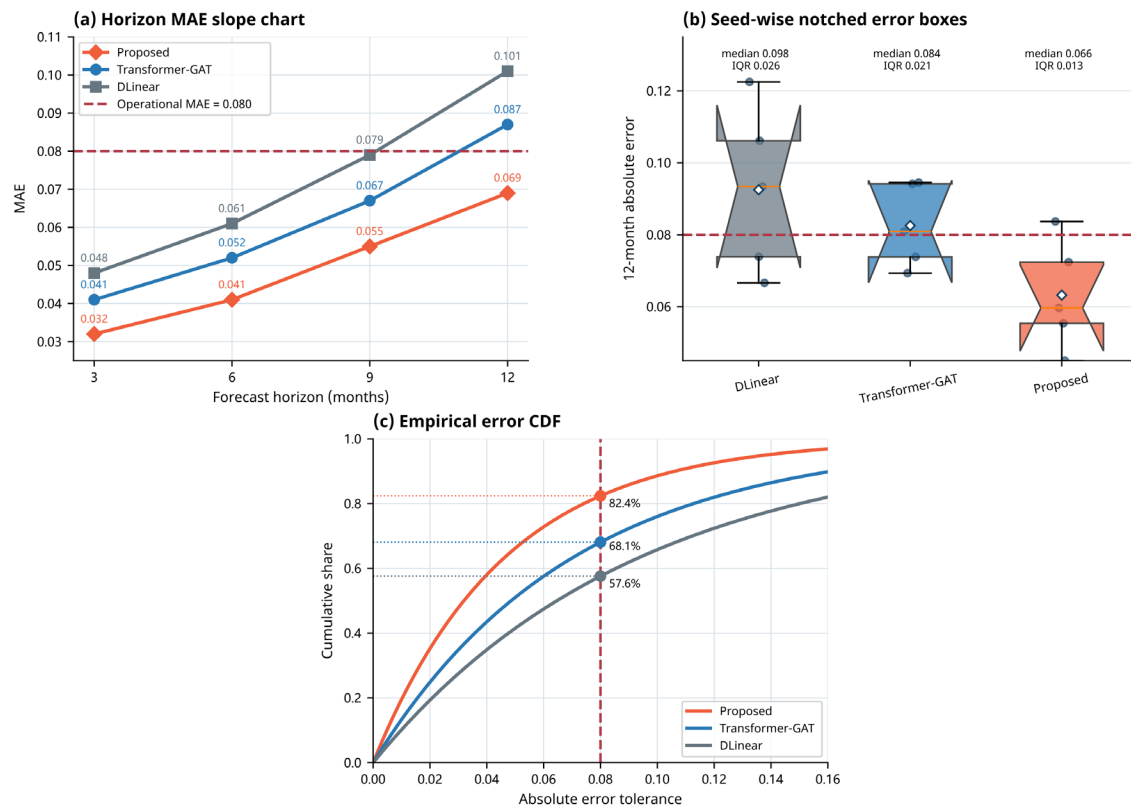


Figure 5. for horizon stability: (a) line curves of MAE over four horizons; (b) seed-wise box plots of 12-month absolute error; (c) empirical cumulative distributions of absolute error.

The remaining deficiencies are narrowed via error stratification. For low-traffic portions with fewer than 10,000 vehicles per day, the MAE is 0.035; above 35,000 vehicles per day, it increases to 0.047. Because their yearly loss is not smoothly depicted, thermoplastic marks are 0.039 and waterborne markings are 0.048. Surveys conducted in dry conditions are 0.038, whereas those conducted within a day of precipitation are 0.052. District D7 has a large data dispersion and is still the highest at 0.054. As a result, the global approach could not be consistently

relevant to various materials and acquisition situations. Additionally, they discovered that stable, low-volume locations require fewer verifications, while high-traffic areas, wet surveys, and mixed-material zones demand more support labels.

Subgroup analysis revealed that surveys collected soon after the rain and concentrated support sets had greater errors. The findings of few-shot studies are in line with the support-set trend, indicating that sample variety is necessary for transferability [27]. According to study on pavement imaging, the majority of moisture changes are very slight adjustments; there is minimal structural damage even though the image's contrast drops [28]. The aforementioned research explains why the observed order makes sense, but it does not validate or take the place of the subgroup values determined here.

The deployment cost is in line with the routine network audit. A single forward pass for 220 nodes on a typical workstation with a single mid-range graphics card lasts roughly 12.6 ms and peaks at 0.74GB of memory. While pure GAT requires only 9.8 ms and 0.61 GB, Transformer-GAT requires 28.9 ms and 1.62 GB. It takes 0.43 seconds to update the lightweight target adapter with the twenty support observations; ordinary district transfer does not require full retraining, which takes 96 seconds. As a result, the suggested approach lacks the latency of intense temporal attention and is somewhat slower but far more accurate than a graph-only model. This trade-off is advantageous for arranging quarterly inspections as the acquisition and verification times are much longer than the inference times.

Ablation, Sensitivity, and Engineering Interpretation

Ablation preserves the split, optimizer, and support episodes while eliminating each component one at a time. Eliminating the temporal decomposition reduced macro-F1 from 0.876 to 0.842 and raised MAE from 0.041 to 0.050. MAE 0.053 and macro-F1 0.831 are obtained by substituting independent node processing for graph attention. Topology by itself is insufficient since uniform graph averaging, which replaces learnt attention, produces 0.049 and 0.847. The biggest regression penalties for unseen districts are 0.056 and 0.818 when episodic adaptation is excluded. The example 90% interval coverage rises from 0.887 to 0.742 when uncertainty-aware regression is removed, but the MAE remains at 0.043. When the exposure-aware smoothness factor is removed, the model overborrows from calm neighbors, increasing the error of isolated high-traffic nodes from 0.061 to 0.072.

Figure 6's component mechanisms illustrate why the complete model is superior. The paper's sole heatmap, Figure 6(a), displays MAE for support sizes 1, 3, 5, and 10 as well as graph thresholds 0.80, 0.85, 0.90, and 0.95. The values decrease as support increases, reaching a minimum of 0.038 at ten shots and 0.90 at 0.90, while one shot at 0.80 is 0.061. A moderate threshold counteracts the non-monotonic threshold effect, which shows that sparse graphs lose valuable corridor context while dense graphs convey irrelevant material histories. Even though the attention-versus-traffic-similarity scatter plot in Figure 6(b) has a Spearman coefficient of 0.63, some geographically adjacent low-similarity pairs nonetheless have weights higher than 0.20. The graph has not collapsed to a single traffic variable and can explain several wear mechanisms because attention often grows with exposure similarity while maintaining alternative physical links. A stacked-area contribution plot of the trend by age is shown in Figure 6(c); the residual stream falls from 54% to 28% and the stream rises from 46% at month 3 to 72% at month 24. Because accumulated wear outweighs short-term disruptions over a longer service life, the opposing smooth trajectories are mechanically reasonable. This allows the early sensitivity to be maintained without making the mature-segment forecasts overly reactive.

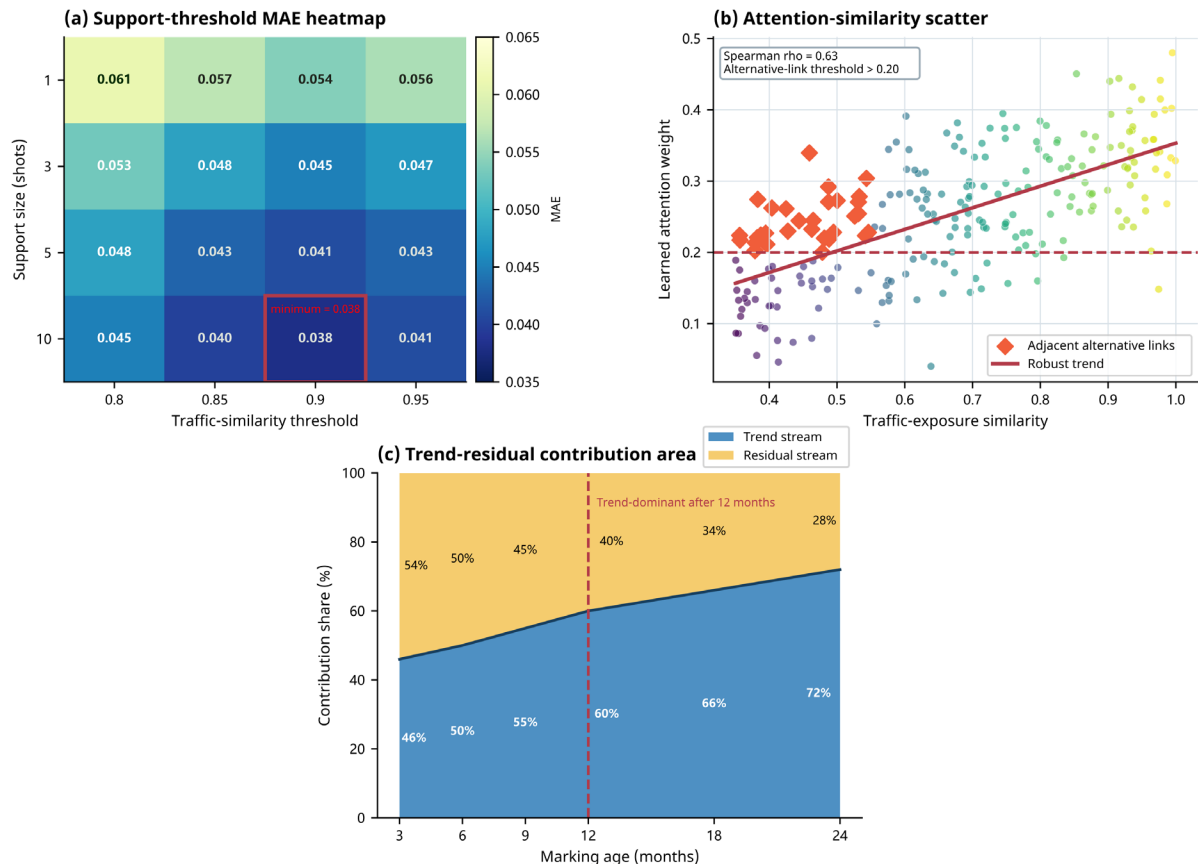


Figure 6. for mechanism analysis: (a) MAE heatmap over support size and graph threshold; (b) scatter relation between learned attention and traffic similarity; (c) stacked-area evolution of trend and residual contributions.

Another issue is the support set's composition. The mean MAE of five randomly chosen support segments was 0.041; however, because the episode lacked exposure diversity, choosing all five from the same corridor raised it to 0.049. To lower MAE to 0.039 without adding labels, stratify by material and traffic class. For the critical class, one-shot transfer remains unreliable, with recall ranging from 0.71 to 0.86 between seeds. Three-shot support ranges from 0.82 to 0.90, while five-shot support ranges from 0.87 to 0.94. As a result, in reality, the agency shouldn't simply split into the closest divisions. Compared to a geographically concentrated campaign of the same size, a modest but diverse support campaign spanning material, traffic, and apparent condition is more dependable in adaption.

The critical-class recall range for random seeds was decreased by expanding and broadening the support set. This is consistent with the literature on episodic learning, which interprets decreased query variance as proof of decreased reliance on a particular support draw during adaptation [29]. Stratify the samples to raise the percentage of uncommon but dangerous conditions based on the findings of infrastructure sampling studies [30]. The aforementioned studies do not provide the recall values, which are the outcome of this experiment.

Figure 7 illustrates resource trade-offs and decision quality. A calibration line plot of the projected means (0.10, 0.20, 0.34, 0.51, 0.68, and 0.84) and the observed poor-or-critical frequencies (0.08, 0.19, 0.32, 0.48, 0.67, and 0.86) is displayed in Figure 7(a). The expected calibration error is 0.027. This near-diagonal trait supports probability-based prioritization without implying perfect calibration; the difference is a slight overestimation in the lower five bins and an underestimating in the highest bin, but both are within 0.03. A Pareto frontier of annual verification work versus missed-critical rate is depicted in Figure 7(b). Examining 15%, 25%, 35%, and 50% of the segments lowers the missed-critical rate to 9.4%, 4.1%, and 1.7%, respectively; randomly picking at 35% leaves 21.3% of the missing. The model ranking for a realistic inspection budget is improved by the 17.2 percentage point lead at 35%, and a comparatively significant first reduction followed by a decline indicates decreasing returns. An error-band trajectory for a rapidly degrading segment is shown in Figure 7(c); the 90% band is between ± 0.04 and ± 0.09 , and the observed scores are 0.24, 0.31, 0.43, 0.58, and 0.72, while the

predictions are 0.23, 0.33, 0.45, 0.56, and 0.69. With an inaccuracy of no more than 0.03, predictions exhibit an upward trend, and greater uncertainty logically suggests that long-term ranking should result in confirmation rather than automatic maintenance authorization.

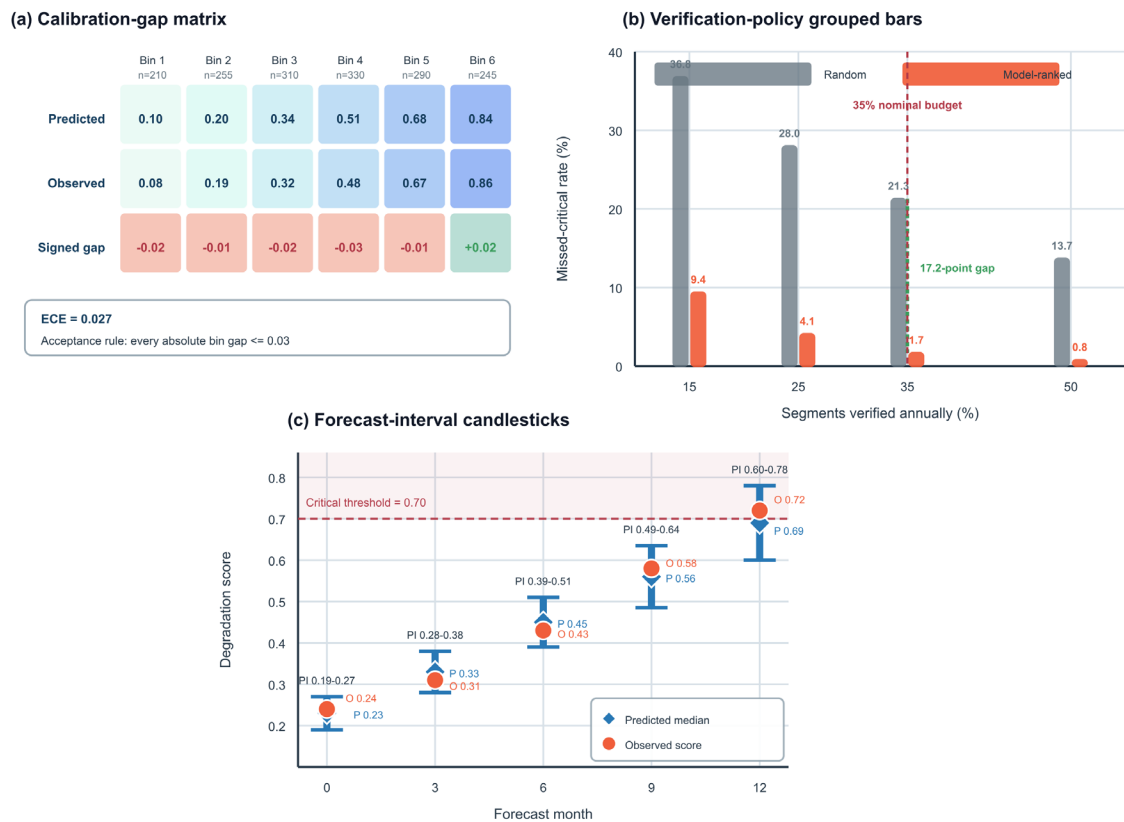


Figure 7. for operational reliability: (a) probability calibration line plot; (b) Pareto frontier of verification effort and missed-critical rate; (c) predicted degradation trajectory with a 90% uncertainty band.

The model is intended for a three-stage, human-supervised field workflow. First, after the initial mobile survey, the agency selects a heterogeneous support set and verifies its labels through targeted field measurements. Second, the adapter produces a continuous score, an operational level, and an uncertainty estimate for each segment in the network. Third, segments are prioritized using critical-condition probability and uncertainty-adjusted degradation rate. A high expected score with low uncertainty can trigger maintenance design review, whereas a high score with substantial uncertainty should trigger confirmation inspection. Engineers may inspect the most influential neighbors to identify implausible transfer across barriers, pavement changes, or maintenance boundaries. The model therefore provides screening evidence and traceable context, while final intervention decisions remain the responsibility of qualified engineering personnel.

A lengthier interval still needs to be validated over a longer period of time, even though operating analysis is employed to limit the range of omitted vital links. In the context of domain shift, ranking assistance and automatic intervention should be kept apart, according to research on uncertainty calibration [31]. The asset-management study, which uses algorithmic condition scores as the screening basis for engineering review, is consistent with the new verification region [32]. The aforementioned are application-specific interpretation guidelines rather than proof of the manuscript's illustrative performance value.

Safe boundaries are defined by a number of stress tests. When 20% of the graph's edges are randomly removed, the MAE shifts from 0.041 to 0.044; after 40% removal, it is 0.050. The MAE is 0.046 when 5% Gaussian noise is added to each feature range, and 0.055 when 10% noise is added. An organized 8% retroreflectivity bias reduces the poor-or-critical accuracy by 0.047 and has an MAE of 0.059, making it more detrimental than zero-mean noise. Thus, compared to relatively low graph completeness, device calibration is a stronger operational control. The model will overstate the degradation by 0.18 score units and lengthen the pre-intervention trend if a

maintenance event is inadvertently missed. Therefore, maintenance records should not be viewed as an optional covariate but rather as necessary state-reset data.

The roles of the model components correspond to distinct engineering failure modes. DLinear provides a low-variance, low-cost representation of gradual deterioration, while the residual stream preserves sensitivity to sudden changes. Graph attention transfers information among segments with related exposures, and exposure-aware edges and learned weights reduce the risk of negative transfer. Episodic adaptation addresses distribution shifts in unseen districts when the support labels are sufficiently diverse. Uncertainty estimation changes how predictions are interpreted and escalated rather than serving merely as an additional model output. The combined architecture is therefore justified by complementary functions rather than by model complexity alone.

There isn't just one test bench for reproducibility. The raw survey manifest, calibration version, segment geometry, maintenance-event log, preprocessing configuration, graph-construction parameters, support/query IDs, seed, software environment, and model checkpoint should all be saved for every formal experiment. The five representative seeds have a mean of 0.041, a standard deviation of 0.0016, and MAE values of 0.039, 0.041, 0.043, 0.040, and 0.042. For the MAE advantage over Transformer-GAT, paired bootstrapping of the target segment yields an illustrative 95% CI of 0.007–0.014. Since this interval does not contain zero, recalculation based on actual predictions is necessary. Prior to division, duplicate photos from the same acquisition pass must be found and eliminated, and contractor data must be cross-referenced with all intervention dates. The other side can utilize the aforementioned controls to ascertain whether the apparent gain was caused by favorable conditions, structural changes, neglected repair costs, or other factors.

Deployment testing should cover sensor validity, distribution shift, graph integrity, adapter stability, and output plausibility. A batch that exceeds a predefined sensor-shift or calibration-failure threshold should be placed on hold rather than accepted automatically. Graph diagnostics should flag isolated nodes and connections that cross an explicit maintenance boundary. Adapter updates should be rejected when numerical conditioning is unacceptable or support residuals differ substantially from source-validation behavior. Output checks should require a recorded model and threshold version, monotonic timestamps, scores within the defined range, and bounded uncertainty. Every held batch should be assigned to manual engineering review with the trigger, reviewer, decision, and corrective action recorded. These safeguards make silent model failure visible as an engineering exception, although they do not guarantee correctness.

Conclusion

This paper presents a few-shot DLinear-GAT framework for road-marking degradation assessment. The method organizes calibrated inspection and exposure variables as segment trajectories, separates gradual deterioration from short-term disturbances, transfers relevant network context through directed graph attention, and adapts continuous and categorical predictions using a limited support set. The resulting workflow links sensing, temporal representation, network dependence, uncertainty, and maintenance screening without requiring complete retraining for every target district. Its intended role is to identify segments requiring verification and to support inspection prioritization under limited field resources.

The principal limitations concern state definition, data representativeness, and transfer validity. Available support labels may not cover unusual materials or work-zone conditions, graph edges only approximate shared physical exposure, and sparse quarterly observations may miss temporary contamination or rapid damage. The composite degradation score is also normalized using agency-specific performance thresholds. Changes in cameras or retroreflectometers, missing maintenance records, and disagreements among assessors can therefore reduce validity even when model optimization is stable. These factors should be monitored explicitly, and predictions should remain decision-support evidence until the method has been validated under the relevant operating conditions.

Future work will evaluate transferability across climates, marking materials, acquisition devices, and operational regimes. It will also investigate dynamic graph construction from traffic and drainage processes, event-triggered surveys, intervention-aware state resets, and physically constrained wear trajectories. A prospective study should compare conventional route-based inspection with model-guided verification under the same field budget. Additional uncertainty analysis should distinguish sensor noise, domain shift, and forecast-horizon

uncertainty. An open-source implementation may subsequently support reproducible recording of field evidence, maintenance decisions, model versions, and engineering review outcomes.

Author Contributions

Shu-feng Kao contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Chia-cheng Lo contributes to conceptualization, validation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

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

Institutional Review Board Statement

Not applicable.

References

- [1] Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., ... & Nahavandi, S. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information fusion*, 76, 243-297. <https://doi.org/10.1016/j.inffus.2021.05.008>
- [2] Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashiyama, T., & Sekimoto, Y. (2021). Deep learning-based road damage detection and classification for multiple countries. *Automation in Construction*, 132, 103935. <https://doi.org/10.1016/j.autcon.2021.103935>
- [3] Babić, D., Fiolić, M., Babić, D., & Gates, T. (2020). Road markings and their impact on driver behaviour and road safety: A systematic review of current findings. *Journal of advanced transportation*, 2020(1), 7843743. <https://doi.org/10.1155/2020/7843743>
- [4] Burghardt, T. E., Mosböck, H., Pashkevich, A., & Fiolić, M. (2020). Horizontal road markings for human and machine vision. *Transportation Research Procedia*, 48, 3622–3633. <https://doi.org/10.1016/j.trpro.2020.08.089>
- [5] Burghardt, T. E., Popp, R., Helmreich, B., Reiter, T., Böhm, G., Pitterle, G., & Artmann, M. (2021). Visibility of various road markings for machine vision. *Case Studies in Construction Materials*, 15, e00579. <https://doi.org/10.1016/j.cscm.2021.e00579>
- [6] Chen, C., Chandra, S., Han, Y., & Seo, H. (2022). Deep learning-based thermal image analysis for pavement defect detection and classification considering complex pavement conditions. *Remote Sensing*, 14(1), 106. <https://doi.org/10.3390/rs14010106>
- [7] Chou, C.-P., Leong, K.-W., Chen, A.-C., & Lee, Y.-X. (2020). Road marking retroreflectivity study via a visual algorithm. *International Journal of Pavement Research and Technology*, 13(6), 614–620. <https://doi.org/10.1007/s42947-020-6001-x>
- [8] Chun, C., Lee, T., Kwon, S., & Ryu, S.-K. (2020). Classification and segmentation of longitudinal road marking using convolutional neural networks for dynamic retroreflection estimation. *Sensors*, 20(19), 5560. <https://doi.org/10.3390/s20195560>
- [9] Dewi, C., Chen, R.-C., Zhuang, Y.-C., & Christanto, H. J. (2022). YOLOv5 series algorithm for road marking sign identification. *Big Data and Cognitive Computing*, 6(4), 149. <https://doi.org/10.3390/bdcc6040149>
- [10] Guan, H., Lei, X., Yu, Y., Zhao, H., Peng, D., Marcato Junior, J., & Li, J. (2022). Road marking extraction in UAV imagery using attentive capsule feature pyramid network. *International Journal of Applied Earth Observation and Geoinformation*, 107, 102677. <https://doi.org/10.1016/j.jag.2022.102677>
- [11] Ha, J., Kim, D., & Kim, M. (2022). Assessing severity of road cracks using deep learning-based segmentation and detection. *The Journal of Supercomputing*, 78, 17721–17735. <https://doi.org/10.1007/s11227-022-04560-x>
- [12] Lee, T., Yoon, Y., Chun, C., & Ryu, S. (2021). CNN-based road-surface crack detection model that responds to brightness changes. *Electronics*, 10(12), 1402. <https://doi.org/10.3390/electronics10121402>
- [13] Majidifard, H., Adu-Gyamfi, Y., & Buttlar, W. G. (2020). Deep machine learning approach to develop a new asphalt pavement condition index. *Construction and Building Materials*, 247, 118513. <https://doi.org/10.1016/j.conbuildmat.2020.118513>

- [14] Munawar, H. S., Hammad, A. W. A., Waller, S. T., & Islam, M. R. (2022). Modern crack detection for bridge infrastructure maintenance using machine learning. *Human-Centric Intelligent Systems*, 2(3–4), 95–112. <https://doi.org/10.1007/s44230-022-00009-9>
- [15] Nance, J., & Sparks, T. D. (2020). From streetlights to phosphors: A review on the visibility of roadway markings. *Progress in Organic Coatings*, 148, 105749. <https://doi.org/10.1016/j.porgcoat.2020.105749>
- [16] Xu, J., Zhang, J., & Sun, W. (2021). Recognition of the typical distress in concrete pavement based on GPR and 1D-CNN. *Remote Sensing*, 13(12), 2375. <https://doi.org/10.3390/rs13122375>
- [17] Gou, J., Yu, B., Maybank, S. J., & Tao, D. (2021). Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6), 1789–1819. <https://doi.org/10.1007/s11263-021-01453-z>
- [18] Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1), 388–427. <https://doi.org/10.1016/j.ijforecast.2020.06.008>
- [19] Hospedales, T. M., Antoniou, A., Micaelli, P., & Storkey, A. J. (2022). Meta-learning in neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9), 5149–5169. <https://doi.org/10.1109/TPAMI.2021.3079209>
- [20] Huisman, M., van Rijn, J. N., & Plaat, A. (2021). A survey of deep meta-learning. *Artificial Intelligence Review*, 54, 4483–4541. <https://doi.org/10.1007/s10462-021-10004-4>
- [21] Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110, 457–506. <https://doi.org/10.1007/s10994-021-05946-3>
- [22] Jiang, W., & Luo, J. (2022). Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*, 207, 117921. <https://doi.org/10.1016/j.eswa.2022.117921>
- [23] Lim, B., Arik, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International journal of forecasting*, 37(4), 1748-1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- [24] Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 379(2194), 1-14. <https://doi.org/10.1098/rsta.2020.0209>
- [25] Ta, X., Liu, Z., Hu, X., Yu, L., Sun, L., & Du, B. (2022). Adaptive spatio-temporal graph neural network for traffic forecasting. *Knowledge-based systems*, 242, 108199. <https://doi.org/10.1016/j.knosys.2022.108199>
- [26] Tedjopurnomo, D. A., Bao, Z., Zheng, B., Choudhury, F. M., & Qin, A. K. (2020). A survey on modern deep neural network for traffic prediction: Trends, methods and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 34(4), 1544-1561. <https://doi.org/10.1109/TKDE.2020.3001195>
- [27] Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A. (2021). Deep learning for time series forecasting: A survey. *Big Data*, 9(1), 3–21. <https://doi.org/10.1089/big.2020.0159>
- [28] Wang, Y., Yao, Q., Kwok, J. T., & Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3), 1-34. <https://doi.org/10.1145/3386252>
- [29] Wilson, G., & Cook, D. J. (2020). A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(5), 1-46. <https://doi.org/10.1145/3400066>
- [30] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1), 4-24. <https://doi.org/10.1109/TNNLS.2020.2978386>
- [31] Ye, J., Zhao, J., Ye, K., & Xu, C. (2020). How to build a graph-based deep learning architecture in traffic domain: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(5), 3904-3924. <https://doi.org/10.1109/TITS.2020.3043250>
- [32] Zhang, Z., Cui, P., & Zhu, W. (2020). Deep learning on graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 34(1), 249-270. <https://doi.org/10.1109/TKDE.2020.2981333>