

Keep Your Friends Close, and the Right Neighbours Closer: Disaster-Conditioned Kernel-Regularized Graph Attention for Building Damage Classification

Fuad Hasan^{1*} and Chul Min Yeum¹

Computer Vision for Smart Structure (CViSS) Lab, University of Waterloo,
Waterloo, Ontario, Canada
fuad.hasan@uwaterloo.ca

Abstract. Disaster damage is spatial: buildings rarely fail in isolation. Yet using spatial context for damage classification remains surprisingly underexplored, and many pipelines still rely primarily on per-building appearance cues even when the dominant uncertainty is spatially structured. Complicating matters, the *right* neighbourhood is not the same across events. Floods, hurricanes, and wildfires can exhibit very different clustering behaviour, making spatial reasoning valuable but easy to misuse—naïve context aggregation can improve visual coherence while oversmoothing boundaries or propagating structured errors. We study this tension on xBD (the dataset used in the xView2 challenge) in a controlled post-localization, classification-only setup: each building is represented by a *pre/post combined* (PPC) patch cropped from the provided polygons, and spatial context is modelled with GPS-derived building graphs. Our approach keeps local evidence “close” by preserving strong spatial relationships in disaster damage patterns, while bringing only the *right* neighbours “closer” through a disaster-type-conditioned graph model that injects a learnable multi-scale spatial kernel prior into attention, allowing the effective neighbourhood scale to adapt across disaster types rather than being learned as a single global smoothing rule. To discourage coherence-by-smoothing, we add a residual de-correlation loss that penalizes positive Moran’s I in prediction residuals. We evaluate the method with (i) an xView2 holdout external-reference comparison under fixed building instances and (ii) zero-shot transfer evaluations that stress-test generalization under event and dataset shift: leave-one-event-out (LOEO) on xBD and cross-dataset transfer from xBD to Ida-BD. The model improves macro-F1 and substantially reduces residual spatial autocorrelation under zero-shot event shift, indicating better use of spatial context rather than naïve smoothing and enabling more reliable transfer to unseen events within known disaster types.

Keywords: Building damage assessment · xBD · kernel regularization · graph attention · geostatistics · zero-shot transfer

* Corresponding author.

Table 1: xBD spatial summary by disaster rollup. Moran’s I statistics are computed over valid train/test tiles.

Rollup	Tiles (Train/Test)	Moran’s I mean $\pm$ std (Train/Test)	Median Moran’s I (Train/Test)	Frac. $p < 0.05$ (Train/Test)	Buildings/tile (Train/Test)
earthquake	121 / 38	0.168 $\pm$ 0.222 / 0.356 $\pm$ 0.439	0.013 / 0.291	0.423 / 0.750	324.4 / 366.6
flood	898 / 80	0.364 $\pm$ 0.482 / 0.195 $\pm$ 0.522	0.339 / 0.071	0.453 / 0.090	75.6 / 46.7
hurricane	1219 / 387	0.306 $\pm$ 0.601 / 0.302 $\pm$ 0.589	0.226 / 0.203	0.233 / 0.260	72.5 / 75.0
tornado	719 / 93	0.437 $\pm$ 0.317 / 0.320 $\pm$ 0.533	0.313 / 0.249	0.376 / 0.433	93.2 / 89.5
tsunami	261 / 42	0.490 $\pm$ 0.407 / 0.484 $\pm$ 0.507	0.511 / 0.568	0.585 / 0.625	210.9 / 256.0
volcano	309 / 75	0.593 $\pm$ 0.369 / 0.398 $\pm$ 0.481	0.339 / 0.340	0.273 / 0.129	23.2 / 15.0
wildfire	5641 / 381	0.330 $\pm$ 0.625 / 0.568 $\pm$ 0.627	0.174 / 0.444	0.136 / 0.277	29.4 / 40.5

1 Introduction

Rapid post-disaster building damage assessment from remote sensing imagery is a core operational task for emergency response: it supports triage, resource allocation, and recovery planning when time matters most. The xBD dataset, released for the xView2 challenge, made this problem tractable at scale by providing paired pre/post satellite imagery, building polygons, and per-building damage labels across multiple disaster types and regions [8, 9, 27]. Yet two issues still limit real-world reliability: **cross-event generalization** and **spatially structured error**. Models that classify buildings independently from image patches often overfit event-specific appearance (roof materials, illumination, sensor differences), which gives rise to the errors or residuals sharing spatial clusters. Spatial context-aware models can improve local consistency but may blur boundaries and also propagate mistakes through neighbourhoods [2, 6, 22, 24, 25].

The key tension is that disaster damage is spatial—a behaviour famously coined as Tobler’s First Law: "Everything is related to everything else, but near things are more related than distant things" [23]; but, this spatial behaviour is not fixed [4]. In other words, *nearby* is not a universal parameter: floods, hurricanes, wildfires, and earthquakes produce different neighbourhood-scale damage patterns, so a single notion of “context” is usually too rigid. In practice, this makes spatial reasoning both valuable and risky. If the model brings in the wrong neighbours, it can become visually smooth but systematically wrong—coherent maps that hide structured mistakes.

A key observation behind our method is that the strength and scale of spatial clustering vary substantially across events and disaster types. Table 1 summarizes a compact event roll-up (our exact roll-up categorization is specified in Sec. 4.1) of ground-truth Moran’s I computed on xBD event tiles (using building severity labels mapped to $\{0, 1, 2, 3\}$) for rollups with enough valid buildings. Hurricanes and wildfires show substantially stronger average spatial autocorrelation than floods, motivating type-adaptive context: when correlation extends further (as is often the case in some disaster types), more distant neighbours can be informative; when it does not, they should not be “pulled closer” by the model.

This paper is a controlled post-localization study: given candidate building instances, we isolate whether relational spatial context improves building-level damage classification under event shift. We avoid detection and segmentation in the experimental loop and use the xBD polygons directly: each node is a PPC

building patch (pre/post crop), and graph edges are formed from GPS-derived building centroids from their polygons. This keeps the question focused: *can we use neighbourhood context in a way that transfers across events—bringing in the "right" neighbours—without oversmoothing?*

Our answer is a **disaster-type-conditioned, multi-scale kernel prior** for graph attention, combined with a **residual de-correlation loss**. The kernel prior injects an interpretable, geostatistically motivated distance bias into attention and allows the effective neighbourhood scale to vary by disaster type, so the model can remain anchored to the per-building patch evidence while selectively amplifying informative neighbours. The residual term penalizes positive Moran’s I of prediction residuals [4, 20], explicitly discouraging clustered errors. Together, these components push the model toward using spatial context as a predictive signal rather than as a smoothing heuristic.

More generally, the method can be viewed as a *conditional context prior* for graph attention, consistent with relational-inductive-bias views of graph networks [1]: the effective neighbourhood scale is conditioned on a global/domain state rather than fixed globally. In this sense, the design is not an arbitrary distance embedding, but a monotone multi-scale log-prior attention mechanism paired with a residual spatial diagnostic; the same abstraction is relevant to structured vision problems where interaction strength varies with scene, geography, sensor, season, or event condition.

We evaluate the method in three complementary settings. First, the **xView2** holdout provides an external reference scale under the standard challenge setting [26]. Second, we use a harder **zero-shot leave-one-event-out (LOEO)** protocol on **xBD** to stress-test cross-event transfer and measure residual spatial structure, holding out one event at a time while providing its disaster type as metadata. Third, we test **zero-shot cross-dataset transfer** by training on xBD and evaluating directly on Ida-BD with a provided disaster-type token. Thus, xView2 contextualizes damage-classification scale, while xBD LOEO and xBD→Ida-BD are the primary stress tests for context modelling under event and dataset shift.

The contributions are:

- A **disaster-type-conditioned multi-scale (DCMS) kernel** parameterization that adapts correlation length scales across disaster types.
- A **kernel-regularized attention** layer for building damage classification that adds the disaster-type-conditioned spatial prior to GAT-style attention logits.
- A **residual spatial regularizer** based on Moran’s I, which explicitly discourages spatially clustered prediction errors.
- A **zero-shot transfer evaluation** that stress-tests context modelling under (i) xBD leave-one-event-out event shift and (ii) xBD→Ida-BD cross-dataset shift.

2 Related Work

Building damage assessment on xBD and xView2. The xBD dataset, introduced with the xView2 challenge, established large-scale, bitemporal satellite imagery for building damage assessment with polygon-level annotations and four damage categories [8,9]. Many strong methods use two-stage pipelines (localization + classification) or bitemporal fusion networks that explicitly compare pre/post imagery [6,22,25]. Recent dataset efforts such as Bright broaden building-damage assessment toward multimodal optical/SAR, all-weather EO benchmarks [3]. xFBD further studies focused building damage assessment and object-level separability, including settings where surrounding context may bias classification [19]. This is complementary to our motivation: spatial context should not be forced when labels are weakly clustered, and a conditional kernel prior can reduce neighbourhood influence in such regimes rather than smoothing indiscriminately. These methods are effective, but improvements can be entangled with detector quality, segmentation quality, and fusion design. Our setup removes those factors by using the provided polygons and focusing on building-level classification and contextual reasoning.

Graph-based context modelling. Graph neural networks offer a natural way to reason over nearby buildings. GraphSAGE [10], GCN [15], and GAT/GATv2 [2,24] provide increasingly expressive message passing, but standard attention is unconstrained and may learn unstable or over-smoothed spatial behaviour under event shifts. Prior work used vanilla graph attention (which we refer to as vanilla GAT) for spatially aware building damage assessment on UAV imagery and observed both benefits and oversmoothing-like failure modes in boundary-heavy scenes [11]. This motivates adding an explicit spatial prior and a residual spatial diagnostic to the learning objective.

Geostatistics and spatial diagnostics. The idea that nearby observations tend to be more related (Tobler’s first law) is foundational in spatial analysis [23]. Moran’s I and Geary’s C remain standard diagnostics for spatial autocorrelation [4,7,20]. We borrow this lens for model design: instead of using spatial statistics only for post-hoc analysis, we incorporate a learnable spatial kernel into attention and explicitly penalize residual spatial autocorrelation during training.

3 Method

Figure 1 gives the full pipeline. We first define a classification-only xBD task setup using PPC building patches and GPS-derived graphs, then encode patches, run graph reasoning with disaster-type-conditioned kernel-regularized attention, and finally optimize both classification loss and a residual Moran objective.

3.1 Task Setup and Graph Construction

We use the xBD dataset [8,9,27], released for the xView2 challenge, and formulate a controlled post-localization, classification-only building damage task.

Rather than treating xBD as a detection/segmentation task, we directly crop a PPC building patch for each annotated polygon and classify damage at the building level. Although experiments use xBD polygons to isolate context modelling, the graph module itself only requires candidate building instances and centroids; in deployment these could come from detector/segmenter outputs or existing footprints (e.g., OSM/cadastre), with localization noise affecting crop quality, node dropout, and centroid jitter.

We process xBD as a collection of *per-event* attributed graphs. Events are disjoint and we do not assume any persistent building identity across events; within each event e , we simply index the N_e building polygons as $i \in \{1, \dots, N_e\}$. For each building polygon, we compute a tight bounding box around the polygon vertices, add a small margin, crop the pre/post RGB regions, and resize to a fixed patch size. We refer to the resulting tensor as a **pre/post combined (PPC) building patch**. In implementation this is a stacked bitemporal tensor, but we use the term *patch* throughout for consistency.

The xBD annotations also include geographic coordinates. For each polygon i in event e , we compute its centroid and project it to a local metric coordinate system (UTM per event/AOI), yielding a fixed position $p_i \in \mathbb{R}^2$. We then build a k -nearest-neighbour (kNN) graph with node set $V_e = \{1, \dots, N_e\}$ and directed edges $E_e = \{(i, j) : j \in \text{kNN}(i)\}$. Each edge is derived from the fixed geometry and stores fixed edge features u_{ij} (concatenated distance $d_{ij} = \|p_i - p_j\|_2$ and a relative direction encoding, θ_{ij} , which is a normalized displacement). We denote the resulting per-event attributed graph as

$$G_e = \left(V_e, E_e, \{v_i\}_{i \in V_e}, \{p_i\}_{i \in V_e}, \{u_{ij}\}_{(i,j) \in E_e} \right), \quad (1)$$

where p_i and u_{ij} define the *static* geometric scaffold, while $\{v_i\}$ are learned node representations updated by the network. Importantly, the kNN *topology* (V_e, E_e) and geometric attributes $\{p_i\}, \{u_{ij}\}$ are fixed once constructed; graph attention does not add/remove edges, but learns *edge weights* over this fixed neighbourhood during message passing.

3.2 Patch Encoder

For each building i in event e , a PPC patch x_i is fed to a pretrained ResNet-50 encoder [12] to produce an embedding $v_i^{(0)} \in \mathbb{R}^d$. This encoder family follows the xView2 first-place-derived bitemporal fusion baseline [25] and serves as the *patch-only* baseline when used without graph context in later parts of the paper. We also concatenate simple polygon metadata (area and perimeter) to the embedding before graph reasoning. In contrast to the fixed coordinates p_i , the node features are updated by message passing layers, yielding $v_i^{(\ell)} \rightarrow v_i^{(\ell+1)}$.

3.3 Disaster-Conditioned Multi-Scale (DCMS) Kernel Prior

Let $p_i \in \mathbb{R}^2$ be the projected centroid of building i , and $d_{ij} = \|p_i - p_j\|_2$ the edge distance. We define a disaster-type embedding vector z_d (learned from a

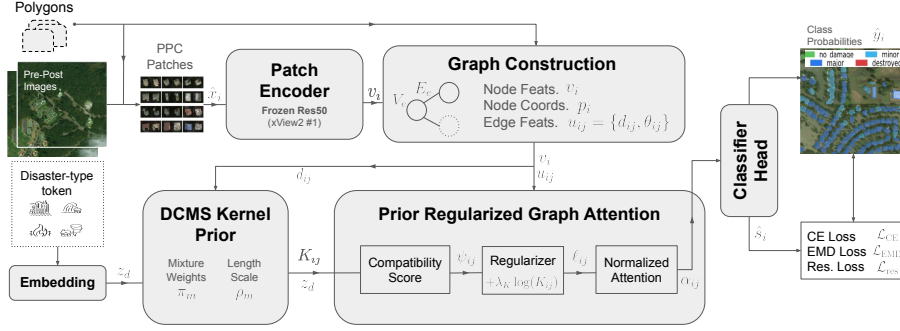


Fig. 1: Method overview. PPC patches form graphs over fixed building instances; a disaster-conditioned DCMS kernel regularizes graph-attention logits, and training combines classification/ordinal losses with residual Moran de-correlation.

disaster-type token parsed from the xBD filename for event e). From z_d , we predict the parameters of a multi-scale distance kernel:

$$\pi(e) = \text{softmax}(W_\pi z_d + b_\pi), \quad (2)$$

$$\rho_m(e) = \rho_{\min} + \text{softplus}(w_m^\top z_d + b_m), \quad m = 1, \dots, M, \quad (3)$$

where $\pi(e)$ are nonnegative mixture weights that sum to one and determine the disaster-type-specific emphasis across short/medium/long ranges, and $\rho_m(e)$ are disaster-type-specific length scales (shared across events of the same disaster type). Note that $\pi(e)$ and $\rho_m(e)$ are *event-level* quantities in the sense that they are shared across all edges within event e , while the resulting prior $K_{ij}(e)$ remains *edge-specific* through its dependence on d_{ij} .

We use a simple exponential family basis:

$$\kappa_m(d_{ij} | e) = \exp\left(-\frac{d_{ij}}{\rho_m(e) + \varepsilon}\right), \quad (4)$$

and define the DCMS spatial prior

$$K_{ij}(e) = \sum_{m=1}^M \pi_m(e) \kappa_m(d_{ij} | e), \quad (5)$$

where $M = 3$ captures *short*, *medium*, and *long* range context with disaster-type-dependent emphasis.

3.4 Kernel-Regularized Graph Attention

Given node features $v_i^{(\ell)}$, fixed edge features u_{ij} , and disaster-type embedding z_d , we compute a compatibility score

$$\psi_{ij}^{(\ell)} = a_\phi\left([W_q v_i^{(\ell)} \parallel W_k v_j^{(\ell)} \parallel u_{ij} \parallel z_d]\right), \quad (6)$$

where $a_\phi(\cdot)$ is a Multi-Layer Perceptron (MLP). We then inject the kernel prior (5) directly into the attention logit as a regularizer:

$$\ell_{ij}^{(\ell)} = \psi_{ij}^{(\ell)} + \lambda_K \log(K_{ij}(e) + \varepsilon). \quad (7)$$

Normalized attention is

$$\alpha_{ij}^{(\ell)} = \frac{\exp(\ell_{ij}^{(\ell)})}{\sum_{j' \in \mathcal{N}(i)} \exp(\ell_{ij'}^{(\ell)})}. \quad (8)$$

This keeps attention data-adaptive while anchoring it to an interpretable, disaster-type-aware spatial prior. While the neighbourhood $\mathcal{N}(i)$ is fixed by the kNN construction, the attention weights $\alpha_{ij}^{(\ell)}$ are recomputed at each layer, yielding a dynamic, layer-dependent *weighted* graph over the same edges. Equivalently, each layer induces weights on the fixed topology, i.e., $\tilde{G}_e^\ell = (V_e, E_e, \{\alpha_{ij}^{(\ell)}\})$.

The graph update is standard message passing:

$$v_i^{(\ell+1)} = v_i^{(\ell)} + \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(\ell)} W_v v_j^{(\ell)} \right), \quad (9)$$

followed by a classifier head producing class probabilities $\hat{y}_i \in [0, 1]^4$. Thus, during forward propagation the *only* quantities updated by graph reasoning are the node representations $\{v_i^{(\ell)}\}$ (and the induced attention weights $\{\alpha_{ij}^{(\ell)}\}$); the positions $\{p_i\}$, geometric edge features $\{u_{ij}\}$, and kNN topology (V_e, E_e) remain fixed.

3.5 Losses: Classification, Ordinal Consistency, and Residual De-correlation

We optimize a class-balanced classification loss with an ordinal auxiliary term and a residual Moran penalty:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda_{\text{emd}} \mathcal{L}_{\text{EMD}} + \lambda_{\text{res}} \mathcal{L}_{\text{res}}. \quad (10)$$

For class imbalance, we use effective-number weights $w_c = \frac{1-\beta}{1-\beta^{n_c}}$ (normalized so $\sum_c w_c = 4$) [5], giving

$$\mathcal{L}_{\text{CE}} = -\frac{1}{B} \sum_{i=1}^B w_{y_i} \log(\hat{y}_{i, y_i} + \varepsilon), \quad (11)$$

where $\hat{y}_i \in [0, 1]^4$ are predicted class probabilities [17].

To encourage ordinal consistency over $\{0, 1, 2, 3\}$, we use an EMD-style cumulative loss with $F_i^*(t) = \sum_{c \leq t} \mathbf{1}[y_i = c]$ and $\hat{F}_i(t) = \sum_{c \leq t} \hat{y}_{ic}$:

$$\mathcal{L}_{\text{EMD}} = \frac{1}{B} \sum_{i=1}^B \sum_{t=0}^3 (\hat{F}_i(t) - F_i^*(t))^2. \quad (12)$$

Let severities be $s \in \{0, 1, 2, 3\}$ and $\hat{s}_i = \sum_{c=0}^3 c \hat{y}_{ic}$. For nodes from event e , define residuals $r_i = s_i - \hat{s}_i$ and centered residuals $\tilde{r}_i = r_i - \bar{r}$. With a fixed row-normalized spatial weight matrix $W^{(e)}$ on (V_e, E_e) , we compute a differentiable Moran surrogate

$$I_r^{(e)} = \frac{n_e}{\sum_{ij} W_{ij}^{(e)}} \frac{\sum_{ij} W_{ij}^{(e)} \tilde{r}_i \tilde{r}_j}{\sum_i \tilde{r}_i^2 + \varepsilon}, \quad (13)$$

and penalize positive residual autocorrelation:

$$\mathcal{L}_{\text{res}} = \frac{1}{|\mathcal{E}_{\text{batch}}|} \sum_{e \in \mathcal{E}_{\text{batch}}} \max(0, I_r^{(e)}). \quad (14)$$

4 Experiments

4.1 Implementation Details

This paper reports the measured values shown in the accompanying tables and figures. In our implementation, we use the xBD dataset (released for the xView2 challenge) with paired pre/post disaster RGB tiles (typically 1024×1024 , ~ 0.8 m GSD), building polygons, and four building-level damage labels (no damage, minor, major, destroyed) [8, 9, 27]. We crop a PPC patch per polygon from the provided xy annotations (tight bounding box + small margin, then resize), compute centroids from the geographic annotations, project to local metric coordinates (UTM per event/AOI), and build a k -nearest-neighbour graph with $k = 16$ unless stated otherwise. The patch encoder is a ResNet-50 [12] with a bitemporal input stem, the kernel mixture uses $M = 3$ components, and optimization uses AdamW-style training [14, 18]. We freeze the patch encoder and train only the graph and classifier heads in all graph variants and the patch-only baseline, to isolate context modelling. We report macro-F1 and per-class F1 for all experiments, and compute ordinal MAE on the severity scale $\{0, 1, 2, 3\}$. The loss weights λ_{emd} and λ_{res} were tuned to be 0.25 and 0.1 respectively in our experiments. For spatial behaviour, we report Moran’s I of *residuals* (lower is better). For rollup-level reporting (e.g., Table 1 and rollup diagnostics in LOEO), we group xBD events into *disaster-type rollups* using the event token parsed from xBD filenames/metadata: *Hurricanes* (Harvey, Florence, Michael, Matthew), *Wildfires* (Santa Rosa, Carr, Woolsey, Pinery, Portugal), *Floods* (Midwest U.S. floods; monsoon in Nepal/India/Bangladesh), *Tornadoes* (Moore OK; Tuscaloosa AL; Joplin MO), *Tsunamis* (Indonesia; Sunda Strait), *Volcanoes* (Guatemala Fuego; Lower Puna), and *Earthquakes* (Mexico City). This rollup mapping is used for analysis/aggregation, as well as for training and inference which also operates on per-event graphs with a disaster-type token as described in Section 3.3.

Evaluation protocols. We use three protocols:

1. **xView2 official holdout external reference.** We reproduce the official xView2 rows from the xView2 challenge website [26] and report our patch-only and graph variants based on their classification scores.

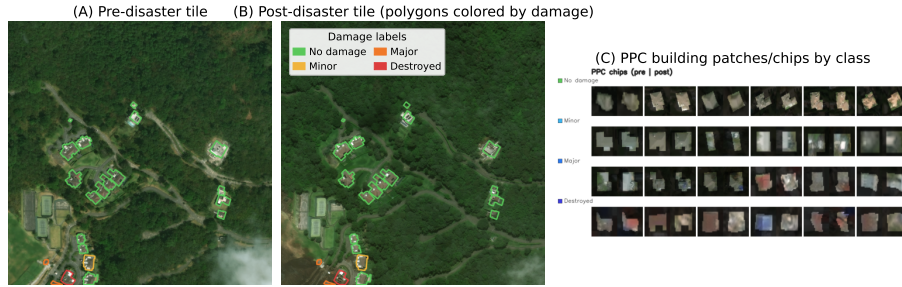


Fig. 2: PPC patch protocol: pre/post building-aligned crops are extracted from xBD polygons and resized for classification.

2. **Zero-shot cross-event transfer (Leave-One-Event-Out; LOEO).** We hold out one event at a time and train on the remaining events. This is a harder stress test for cross-event generalization and spatial behaviour; it is *not directly comparable* to the official holdout reference. In this protocol, the held-out event is evaluated with its associated *disaster type* provided as metadata (consistent with the disaster-type-conditioned design in Section 3.3). LOEO F1 scores are computed by pooling predictions over all held-out events and reporting global per-class F1 and macro-F1.
3. **Zero-shot cross-dataset transfer (xBD→Ida-BD).** We train on xBD and evaluate directly on Ida-BD [16] with no Ida-BD labels used for training, providing only a disaster-type token at inference (Ida-BD is a hurricane event).

4.2 Main Results: xView2 Holdout Comparison

Table 2 provides a compact xView2 holdout reference. The official xView2 evaluation scores localization and classification jointly, whereas our protocol fixes building instances using provided polygons and evaluates the classification/context module only. Therefore, the official row should be read as a broad external reference point, not a direct head-to-head comparison. The meaningful comparison here is among the patch-only encoder, vanilla GAT, and our kernel-regularized graph model under the same fixed building instances and patch protocol. Under this controlled setting, Vanilla GAT improves Macro-F1 to 0.84102, and the full model further increases it to 0.87288, with gains concentrated in the harder *Minor* and *Major* classes. The full xView2 external-reference table is included in the supplementary material for completeness.

4.3 Zero-shot Cross-Event Stress Test and Ablations (LOEO)

The LOEO setting is a stricter test and better reflects event shift. Table 3 shows three key patterns. First, the patch-only baseline is fragile under event shift

Table 2: xView2 holdout external reference. Macro-F1 is the arithmetic mean of class F1 over U/Mi/Ma/D; official solution #1 is shown only as a reference.

Method	Macro-F1	Per-class F1 (U, Mi, Ma, D)
xView2 official solution #1 [†]	0.80446	(0.92344, 0.64445, 0.78591, 0.86403)
Patch-only encoder	0.82205	(0.92880, 0.67320, 0.81520, 0.87100)
Vanilla GAT on patch encoder	0.84102	(0.93160, 0.70050, 0.84030, 0.89168)
Ours (kernel + disaster + residual)	0.87288	(0.93400, 0.76200, 0.89200, 0.90352)

[†] Official solution #1 is shown only as an external reference and is not directly comparable to our fixed building instances and classification-only protocol.

Table 3: LOEO ablation study. Residual Moran’s I is computed on severity residuals (lower is better).

Variant	Kernel prior	Disaster cond.	Residual loss	LOEO Macro-F1	Per-class F1 (No, Mi, Ma, D)	Residual Moran’s I ↓	Severity MAE ↓
Patch-only (xView2 1st-place-style Res50 backend)	×	×	×	0.433	(0.61, 0.37, 0.41, 0.34)	0.248	0.562
Vanilla GAT (kNN graph baseline)	×	×	×	0.453	(0.57, 0.40, 0.44, 0.40)	0.256	0.541
+ Kernel prior only	✓	×	×	0.473	(0.60, 0.42, 0.46, 0.41)	0.182	0.507
+ Disaster conditioning only	×	✓	×	0.460	(0.59, 0.41, 0.45, 0.39)	0.205	0.521
+ Residual de-correlation only	×	×	✓	0.450	(0.58, 0.39, 0.44, 0.39)	0.139	0.529
+ Kernel prior + Disaster conditioning	✓	✓	×	0.488	(0.61, 0.44, 0.48, 0.42)	0.116	0.484
+ Kernel prior + Residual loss	✓	×	✓	0.490	(0.60, 0.44, 0.49, 0.43)	0.096	0.474
Full (Kernel + Disaster + Residual)	✓	✓	✓	0.503	(0.62, 0.45, 0.50, 0.44)	0.079	0.458

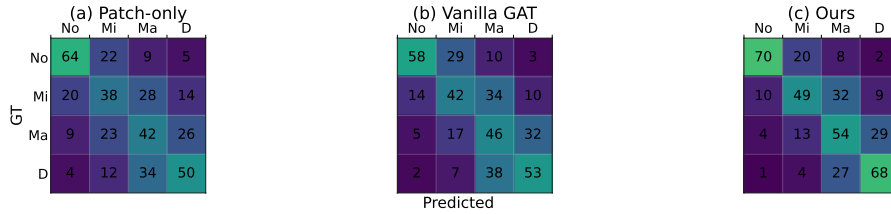
(macro-F1 0.433). Second, vanilla GAT improves macro-F1 slightly (0.453) but leaves high residual Moran’s I (0.256) and can oversmooth mixed neighbourhoods. Third, the full model reaches the best LOEO macro-F1 (0.503) while dramatically lowering residual Moran’s I to 0.079, which is the main empirical signature of the proposed objective.

Among single components, the kernel prior contributes the strongest standalone gain in both macro-F1 and spatial consistency. Disaster conditioning alone yields a smaller but consistent improvement, and the residual term alone mostly attacks spatially structured error (large Moran drop, smaller F1 gain). The pairwise ablations support the same narrative: *kernel + residual* is stronger than either component alone, and adding disaster conditioning produces the best final result. Table 4 adds two sensitivity checks: bitemporal PPC input improves the difficult Minor/Major distinction over post-only imagery, and the full model is stable around $k = 16$, which slightly minimizes residual Moran’s I among the tested graph connectivities.

Figure 3 makes the LOEO error modes more explicit. Patch-only produces broader off-diagonal mass under event shift, especially across adjacent severity levels and occasional severe confusions. Vanilla GAT improves some local consistency, but its confusion mass shifts toward adjacent-class transitions consistent with oversmoothing-like behaviour (e.g., stronger No→Minor and Destroyed→Major leakage in mixed neighbourhoods). The proposed model increases diagonal concentration across all classes while keeping the realistic hard confusions concentrated in *Minor↔Major* and *Major↔Destroyed*, which matches the qualitative boundary behaviour outlined later in Figure 5.

Table 4: Sensitivity checks. Input sensitivity uses the fixed-instance holdout protocol; k -sensitivity uses LOEO. Mi/Ma averages Minor and Major F1.

k	Graph sensitivity		Input sensitivity		
	Macro-F1	Moran's I	Setting	Macro-F1	Mi/Ma F1
8	0.498	0.081	Post-only	0.828	0.762
16	0.503	0.079	Pre+Post (PPC)	0.873	0.827
32	0.500	0.084			

**Fig. 3: LOEO confusion matrices.** Row-normalized percentages under the LOEO protocol.

4.4 Zero-shot Cross-Dataset Transfer: xBD $\rightarrow$ Ida-BD

To probe generalization beyond xBD, we evaluate zero-shot transfer from xBD to Ida-BD, a high-resolution satellite dataset constructed for Hurricane Ida with paired pre/post tiles and polygon-level building annotations [16]. Table 5 reports published xBD $\rightarrow$ Ida-BD transfer baselines and our patch-based results when training only on xBD and evaluating directly on Ida-BD while providing a disaster-type token at inference. Prior work reports transfer performance in end-to-end localization+classification or segmentation-style pipelines (and may merge classes such as Destroyed into Major), which makes their reported scores not directly comparable to our polygon-patch classification protocol. We therefore include these reported baselines primarily to contextualize the difficulty of the dataset shift, and emphasize the relative gains among the patch-based baselines and our model under a consistent setup.

4.5 Qualitative and Spatial Diagnostics

Figure 4a visualizes the learned mean disaster-type-conditioned kernel priors aggregated by disaster rollup. Beyond decaying with distance, the key result is that the *mixture shape* changes: some rollups emphasize shorter ranges while others assign more weight to medium/long scales. This pattern shows qualitative consistency with the spatial statistics in Table 1: for rollups with slower spatial decay, more distant buildings can remain informative because damage labels are more spatially correlated. This supports our core motivation that “useful neighbourhood size” is disaster-type dependent.

Table 5: Zero-shot transfer from xBD to Ida-BD. Models train on xBD and evaluate directly on Ida-BD.

Method	Pipeline / notes	Token	Macro-F1	Per-class F1 (No, Mi, Ma, D)
Siam-UNet [13]	D merged into Ma [†]	–	0.458	(0.916, 0.208, 0.251, –)
xView2 winner [21]	per-class weighted F1	–	0.268	(0.667, 0.211, 0.154, 0.041)
Patch-only (ours)	patch cls (polygons)	✓	0.275	(0.670, 0.190, 0.160, 0.080)
Vanilla GAT (ours)	patch cls + graph	✓	0.295	(0.690, 0.220, 0.180, 0.090)
Ours (Kernel + Disaster + Residual)	patch cls + graph	✓	0.335	(0.710, 0.270, 0.230, 0.130)

[†] DAHiTra reports three damage classes on Ida-BD (Destroyed merged into Major); Macro-F1 is averaged over the reported damage classes.

Table 6: Rollup LOEO Macro-F1 / residual Moran’s I. Higher F1 and lower Moran’s I are better.

Method	earthquake	flood	hurricane	tornado	tsunami	volcano	wildfire
Patch-only	.470/.160	.397/.313	.438/.263	.424/.368	.415/.427	.382/.359	.505/.260
Vanilla GAT	.482/.174	.416/.335	.459/.283	.442/.384	.429/.437	.401/.364	.542/.283
Ours	.548/.134	.474/.212	.517/.198	.502/.294	.488/.301	.445/.320	.547/.204

Figure 4b compares how attention weights vary with inter-building distance for vanilla GAT and the proposed kernel-regularized GAT. The plotted medians are computed separately for edges linking buildings with the same damage label and edges linking buildings with different damage labels. In vanilla GAT, attention is less consistently structured with distance and shows weaker separation between same-label and different-label edges, reflecting noisier neighbourhood aggregation. In contrast, our kernel-regularized attention follows a more coherent distance-decay trend while preserving feature-driven variation, which reduces pathological long-range links and stabilizes local smoothing. Additional rollup-level plots and per-rollup kernel tables are provided in the supplementary material.

Figure 5 shows a representative LOEO evaluation tile (Hurricane-Harvey), showcased due to the tile’s rich spatial autocorrelation (high Moran’s I). The patch-only model, which was not trained on Hurricane Harvey, becomes sensitive to pre/post appearance shifts and produces mild spatially autocorrelated residuals because its errors are presumably driven by latent scene factors (hazard intensity, neighbourhood morphology, and imaging conditions) that are themselves spatially clustered, even without any explicit spatial message passing. Vanilla GAT improves local coherence but exhibits oversmoothing-like errors in mixed neighbourhoods. Our model is more stable along boundaries and maintains a cleaner damage pattern without collapsing mixed regions, even though some oversmoothing errors seem to be present.

Table 6 summarizes rollup-level LOEO Macro-F1 and residual Moran’s I in a compact F1/Moran format. The proposed model generally improves rollup-level accuracy while consistently lowering residual autocorrelation relative to both baselines, matching the intended effect of the residual de-correlation objective.

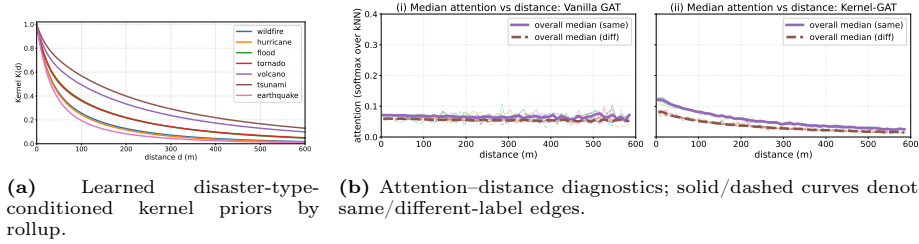


Fig. 4: Kernel and attention-distance diagnostics. Full rollup-specific grids and kernel parameter tables are provided in the supplementary material.

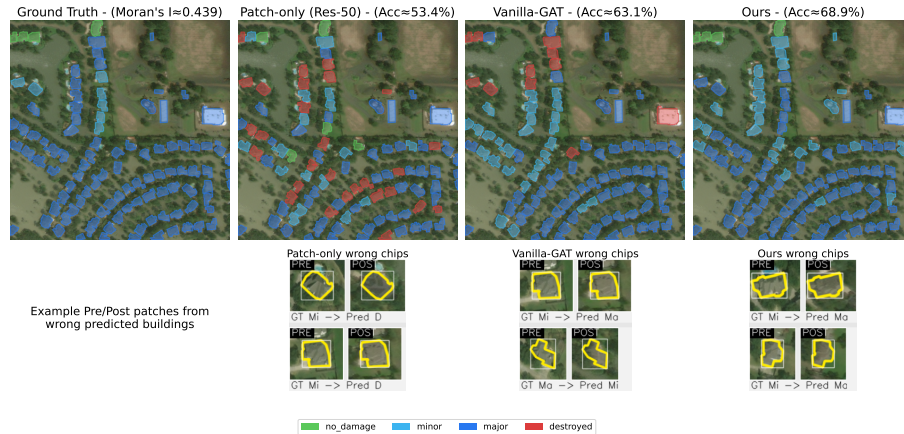


Fig. 5: LOEO qualitative map on a high-Moran Hurricane-Harvey tile. Left-to-right: ground truth, patch-only, vanilla GAT, and ours.

5 Discussion

The holdout and LOEO results serve different purposes. The xView2 holdout numbers provide an external reference scale for damage-classification performance, but the controlled comparisons are the patch-only, vanilla GAT, and proposed graph variants under fixed building instances. The post-only check confirms that bitemporal PPC input remains useful even in this building-level setting, especially for the Minor/Major distinction where pre-disaster appearance reduces ambiguity. LOEO is the more diagnostic setting for our claim: when the held-out event changes roof materials, imaging conditions, or disaster morphology, patch-only predictions leave spatially clustered residuals, indicating that appearance errors are not independent across space. This is also why we view the fixed-instance protocol as a feature rather than a shortcut: by holding localization constant, the experiments isolate whether spatial context itself improves damage classification under event shift.

Vanilla GAT confirms that graph context is useful but not sufficient. It improves over the patch-only encoder, yet unconstrained attention can still propagate mistakes across mixed neighbourhoods, especially near damage boundaries. The issue is not graph reasoning itself; it is the lack of a principled, disaster-type-specific spatial scale. The learned DCMS kernels address this by making the effective context radius conditional on disaster type, while the residual Moran penalty discourages coherent-but-wrong neighbourhoods; the compact (k)-sensitivity check shows this behaviour is stable around the chosen (k=16). More broadly, the results suggest that spatial context should be treated as a conditional cue rather than a universal prior: when labels are strongly clustered, neighbouring buildings can provide useful evidence, but when spatial clustering is weak, forcing neighbourhood agreement can become a source of bias.

The spatial diagnostics support this interpretation: the learned kernels do not collapse to a single default curve, and Table 6 shows the F1/Moran trade-off across disaster rollups rather than only in aggregate. The cross-dataset transfer results in Table 5 provide an additional stress test under xBD→Ida-BD dataset shift. Although published Ida-BD baselines use different pipelines and label handling [13, 21], the fixed building instances and patch protocol show the same trend: the proposed model improves over patch-only and vanilla GAT, with gains concentrated in the harder *Minor* and *Major* classes. This does not imply that instance-level graph reasoning replaces full-image models; rather, it addresses a complementary question. Full-image systems jointly solve localization, scene understanding, and classification, whereas our controlled setting asks whether, once candidate buildings are available, relational spatial structure can improve the robustness of building-level damage labels.

Some notable limitations remain. First, the method is not an end-to-end xView2 system: it assumes candidate building instances and centroids are available, and our experiments use the provided xBD polygons to avoid confounding localization quality with contextual reasoning. In deployment, these instances could come from a detector, segmenter, cadastral layer, or OSM-style footprint source, but errors in those inputs would affect crop quality, node availability, and graph geometry. Evaluating the method with detector-generated footprints and explicitly modelling localization uncertainty are therefore important next steps. Second, our GPS-only k -NN graphs ignore scene semantics (e.g., roads, parcels, barriers) and can connect buildings that are near in Euclidean space but weakly coupled in damage. Third, we assume the disaster-type token is known at inference; robustness to missing, noisy, or ambiguous event metadata is untested. Finally, we use optical PPC patches only and do not leverage modalities such as SAR that matter under cloud/smoke. Future work can address these by using topology-/uncertainty-aware edges, inferring conditioning when metadata is unavailable, integrating detector-generated instances, and extending to multimodal inputs while retaining residual spatial diagnostics.

6 Conclusion

We presented a disaster-type-conditioned, kernel-regularized graph attention model for xBD building damage classification in the controlled post-localization, classification-only xBD setting using a clean PPC building-patch protocol. The method keeps local evidence “close” by retaining strong per-building patch representations, but improves reliability by explicitly modelling spatial behaviour instead of relying on naive image-only recognition under shift. It then brings the *right* neighbours closer by injecting a disaster-type-adaptive, multi-scale geostatistical prior into attention, so neighbourhood influence is distance- and type-aware rather than an unconstrained tendency to smooth. A residual Moran objective further discourages coherent-but-wrong neighbourhood errors by penalizing spatial autocorrelation in severity residuals. Across an external-reference comparison and zero-shot transfer evaluations under event shift (xBD LOEO) and dataset shift (xBD→Ida-BD), the model improves macro-F1 and reduces residual Moran’s I, supporting the central claim: better event generalization starts with keeping your “friends” close (i.e., learning to utilize the rich spatial relationships in disaster damage patterns to make more informed predictions), and the *right* neighbours closer (i.e., also learning *which* spatial context to trust and at what scale, not by smoothing more aggressively).

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Alliance Program (ALLRP 570826-21), in partnership with Statistics Canada, Natural Resources Canada, Credit Valley Conservation, Toronto and Region Conservation Authority, and the City of Kitchener.

The authors acknowledge the use of OpenAI’s ChatGPT for manuscript preparation, including drafting, grammar and structure refinement, and assistance in identifying potentially relevant references. All scientific content, results, interpretations, and cited references were manually reviewed and verified by the authors.

References

1. Battaglia, P.W., Hamrick, J.B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gülgehre, Ç., Song, F., Ballard, A., Gilmer, J., Dahl, G.E., Vaswani, A., Allen, K., Nash, C., Langston, V., Dyer, C., Heess, N., Wierstra, D., Kohli, P., Botvinick, M., Vinyals, O., Li, Y., Pascanu, R.: Relational inductive biases, deep learning, and graph networks. arXiv preprint arXiv:1806.01261 (2018). <https://doi.org/10.48550/arXiv.1806.01261>, accessed: 2026-06-26
2. Brody, S., Alon, U., Yahav, E.: How attentive are graph attention networks? In: International Conference on Learning Representations (ICLR) (2022), <https://openreview.net/forum?id=F72ximsx7C1>, accessed: 2026-06-26

3. Chen, H., Song, J., Dietrich, O., Broni-Bediako, C., Xuan, W., Wang, J., Shao, X., Wei, Y., Xia, J., Lan, C., Schindler, K., Yokoya, N.: Bright: A globally distributed multimodal building damage assessment dataset with very-high-resolution for all-weather disaster response. *Earth System Science Data* **17**, 6217–6253 (2025). <https://doi.org/10.5194/essd-17-6217-2025>
4. Cressie, N.A.C.: *Statistics for Spatial Data*. Wiley, revised edn. (1993). <https://doi.org/10.1002/9781119115151>
5. Cui, Y., Jia, M., Lin, T.Y., Song, Y., Belongie, S.: Class-balanced loss based on effective number of samples. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 9268–9277 (2019). <https://doi.org/10.1109/CVPR.2019.00949>
6. Deng, L., Wang, Y.: Post-disaster building damage assessment based on improved U-Net. *Scientific Reports* **12**, 15862 (2022). <https://doi.org/10.1038/s41598-022-20114-w>
7. Geary, R.C.: The contiguity ratio and statistical mapping. *The Incorporated Statistician* **5**(3), 115–146 (1954). <https://doi.org/10.2307/2986645>
8. Gupta, R., Goodman, B., Patel, N., Hosfelt, R., Sajeew, S., Heim, E., Doshi, J., Lucas, K., Choset, H., Gaston, M.: Creating xBD: A dataset for assessing building damage from satellite imagery. KiltHub preprint (May 2019). <https://doi.org/10.1184/R1/8135576.v1>, accessed: 2026-06-26
9. Gupta, R., Hosfelt, R., Sajeew, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., Gaston, M.: xBD: A dataset for assessing building damage from satellite imagery. *arXiv preprint arXiv:1911.09296* (2019). <https://doi.org/10.48550/arXiv.1911.09296>
10. Hamilton, W.L., Ying, R., Leskovec, J.: Inductive representation learning on large graphs. In: *Advances in Neural Information Processing Systems (NeurIPS)*. vol. 30 (2017). <https://doi.org/https://doi.org/10.48550/arXiv.1706.02216>
11. Hasan, F., Lesani, A., Yeum, C.M., Costa, R.: Graph-attention network for spatially-aware post-hurricane building damage assessment from UAV imagery. *ISPRS Congress* (2026), accepted for presentation
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
13. Kaur, N., Lee, C., Mostafavi, A., Mahdavi-Amiri, A.: Large-scale building damage assessment using a novel hierarchical transformer architecture on satellite images. *Computer-Aided Civil and Infrastructure Engineering* **38**(15), 2072–2091 (2023). <https://doi.org/10.1111/mice.12981>
14. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: *International Conference on Learning Representations (ICLR)* (2015), <https://arxiv.org/abs/1412.6980>, accessed: 2026-06-26
15. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: *International Conference on Learning Representations (ICLR)* (2017), <https://openreview.net/forum?id=SJU4ayYgl>, accessed: 2026-06-26
16. Lee, C., Kaur, N., Mahdavi-Amiri, A., Mostafavi, A.: Ida-BD: Pre- and post-disaster high-resolution satellite imagery for building damage assessment from hurricane ida. *DesignSafe-CI dataset release* (2022), <https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-3563>, accessed: 2026-06-26
17. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. pp. 2980–2988 (2017). <https://doi.org/10.1109/ICCV.2017.324>

18. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (ICLR) (2019), <https://openreview.net/forum?id=Bkg6RiCqY7>, accessed: 2026-06-26
19. Melamed, D., Johnson, C., Zhao, C., Blue, R., Morrone, P., Hoogs, A., Clipp, B.: xFBD: Focused building damage dataset and analysis. arXiv preprint arXiv:2212.13876 (2023). <https://doi.org/10.48550/arXiv.2212.13876>
20. Moran, P.A.P.: Notes on continuous stochastic phenomena. *Biometrika* **37**(1/2), 17–23 (1950). <https://doi.org/10.2307/2332142>
21. Parupati, B.C.R., Kshirsagar, S., Bagai, R., Dutta, A.: Towards robust building damage detection: Leveraging augmentation and domain adaptation. In: 2025 IEEE Green Technologies Conference (GreenTech). pp. 163–167 (2025). <https://doi.org/10.1109/GreenTech62170.2025.10977720>
22. Shen, Y., Zhu, S., Yang, T., Chen, C., Pan, D., Chen, J., Xiao, L., Du, Q.: BDANet: Multiscale convolutional neural network with cross-directional attention for building damage assessment from satellite images. *IEEE Transactions on Geoscience and Remote Sensing* **60**, 1–14 (2022). <https://doi.org/10.1109/TGRS.2021.3080580>
23. Tobler, W.R.: A computer movie simulating urban growth in the detroit region. *Economic Geography* **46**(sup1), 234–240 (1970). <https://doi.org/10.2307/143141>
24. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: International Conference on Learning Representations (ICLR) (2018), <https://openreview.net/forum?id=rJXMpikCZ>, accessed: 2026-06-26
25. Weber, E., Kané, H.: Building disaster damage assessment in satellite imagery with multi-temporal fusion. arXiv preprint arXiv:2004.05525 (2020). <https://doi.org/10.48550/arXiv.2004.05525>
26. xView2 Challenge Organizers: xView2 challenge leaderboard. <https://xview2.org/challenge>, accessed: 2026-06-26
27. xView2 Challenge Organizers: xView2 dataset. <https://xview2.org/dataset>, accessed: 2026-06-26