Graph-Attention Network for Spatially-Aware Post-Hurricane Building Damage Assessment from UAV Imagery

Fuad Hasan[1,2], Chul Min Yeum[1,2], Ali Lesani[1,2], Rodrigo Costa[2]

[1]Computer Vision for Smart Structures (CViSS), Waterloo, Canada
[2]University of Waterloo, Waterloo, Canada

Keywords: Damage Assessment, Deep Learning, Graph Attention Network, Remote Sensing, UAV, Spatial Autocorrelation

Abstract

In the immediate aftermath of a hurricane, the rapid, accurate assessment of building damage is paramount for effective emergency response and the allocation of resources. Traditional methods of damage assessment, which rely on ground-based surveys, are often slow, hazardous, and subjective. While the advent of remote sensing (RS), through Unmanned Aerial Vehicles (UAVs) and the application of Convolutional Neural Networks (CNNs), has significantly advanced the automation of this process, these models operate on a pixel-level or object-level basis, failing to capture the inherent spatial relationships and contextual information within a disaster zone. Damage patterns are not spatially random; they exhibit strong spatial autocorrelation, a principle encapsulated by Tobler’s First Law of Geography. This paper introduces a novel approach that leverages Graph Attention Networks (GATs) to explicitly model spatial dependencies when evaluating building damage. By representing damaged buildings and their surroundings as nodes and edges in a graph, our model can learn and weigh the influence of neighboring structures and the local environment when assessing their damage level. This spatially-aware methodology moves beyond simple image classification to a more holistic scene understanding. We evaluate the method on DoriaNET, a geo-referenced UAV dataset collected after Hurricane Dorian (2019) that provides masked building patches, GPS centroids, structural metadata, and ordinal FEMA/HAZUS-style damage labels. By incorporating spatial context via a graph-based framework, our GAT model achieves superior performance in building damage classification compared to state-of-the-art CNN-based approaches, producing more coherent and accurate damage maps better suited to real-world disaster management scenarios.

1. INTRODUCTION

In the critical hours and days following a catastrophic hurricane, timely and accurate information on the extent and severity of damage is the cornerstone of an effective emergency response (Preliminary Damage Assessment (PDA) Guide, 2023). This information is not merely data; it is the critical input that guides life-saving decisions, from prioritizing search and rescue operations to allocating limited resources like medical supplies, food, and shelter (Al Shafian and Hu, 2024, NOAA National Centers for Coastal Ocean Science, 2024). The speed at which this assessment is conducted has profound humanitarian implications. Delays in understanding the on-the-ground reality can lead to disorganized and inefficient response efforts, amplifying public skepticism and, more critically, directly impacting the well-being of the affected population (Mitchell, 2020, Goldmann and Galea, 2014). Thus, the imperative for rapid damage assessment is driven by both logistical necessity and a fundamental need to mitigate human trauma (Goldmann and Galea, 2014).

Historically, this crucial task has been performed through traditional ground-based methods, such as windshield surveys and in-situ inspections conducted by first responders and volunteers (Preliminary Damage Assessment (PDA) Guide, 2023). However, these methods have severe limitations that render them increasingly inadequate for the scale and urgency of modern disasters. They are slow, labor-intensive, and costly, often taking days or weeks to cover a large affected area (NOAA National Centers for Coastal Ocean Science, 2024, Su et al., 2020). Furthermore, the data collected is prone to significant inconsistencies and subjectivity, as assessments can vary widely between different inspectors and may be influenced by numerous biases, including geographical, temporal, and even personal factors (Howarth and Butler, 2020). This lack of standardization can lead to inaccurate aid allocation, where assistance is not directed to those most in need (Benson et al., 2021). Most alarmingly, these ground-based surveys expose response personnel to significant physical and psychological harm. The combination of inefficiency, data unreliability, and severe risk to personnel creates a compelling case for a paradigm shift in assessment methodologies.

The convergence of remote sensing (RS) technologies and Artificial Intelligence (AI) has catalyzed this paradigm shift, offering a safer, faster, and more scalable alternative (van Westen, 2002, Wu et al., 2024). High-resolution imagery from satellites and, more recently, Unmanned Aerial Vehicles (UAVs), provides a synoptic, bird’s-eye view of the disaster zone without endangering personnel (Al Shafian and Hu, 2024, NOAA National Centers for Coastal Ocean Science, 2024). AI, and specifically deep learning models, has emerged as the key to unlocking the vast potential of this imagery, automating the laborious task of visual interpretation (Zhu et al., 2017). Among these models, Convolutional Neural Networks (CNNs) have become the state-of-the-art for image-based classification tasks, demonstrating remarkable success in automatically identifying and categorizing building damage from aerial and satellite imagery with high accuracy (Gupta et al., 2019, Kerle et al., 2020, Cheng et al., 2021b). These data-driven approaches are revolutionizing disaster management by providing near-real-time situational awareness that was previously unattainable (Microsoft Research, 2022, Gupta et al., 2019).

Despite their success, CNNs have a fundamental architectural limitation that curtails their effectiveness in geospatial analysis:

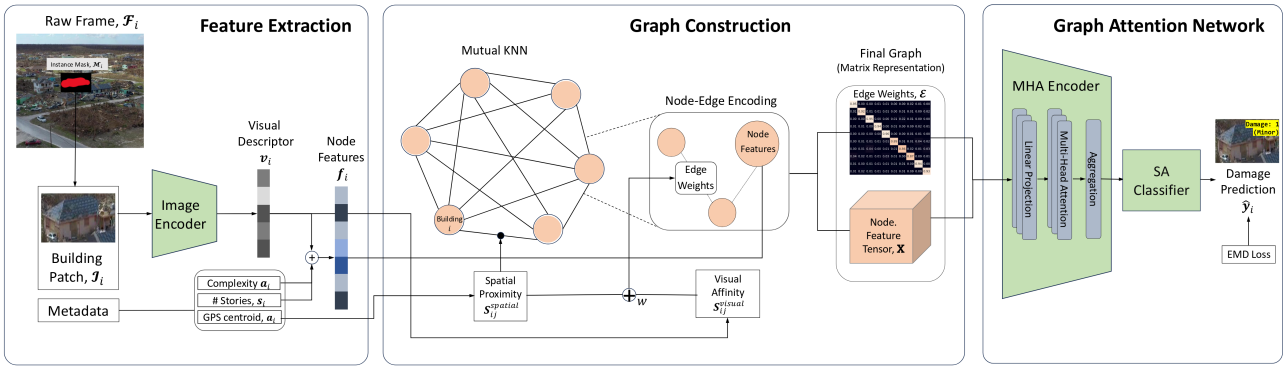


Figure 1. Overview of the proposed three-stage pipeline: (a) feature extraction with BiT encoder and metadata fusion; (b) kNN spatial graph with combined spatial/visual edge weights; (c) multi-head Graph Attention encoder and single-head classifier for ordinal damage prediction.

they are inherently geospatially naive. A CNN processes an image by applying filters to small, local patches of pixels (receptive fields), making it exceptionally good at recognizing textures and local patterns, such as a damaged roof or collapsed wall (Gupta et al., 2019). However, in doing so, they largely disregard the broader geospatial context and the relationships between objects in a scene. The pooling layers that give CNNs their translational invariance—the ability to recognize an object regardless of its position—achieve this by progressively discarding precise spatial information (Goodfellow et al., 2016). This design, optimized for general computer vision tasks, is a significant drawback when analyzing geographic phenomena, where the location and spatial arrangement of features are of paramount importance. Consequently, a CNN may correctly classify an individual damaged building but remains blind to the crucial contextual information that this building is part of a larger, contiguous cluster of destruction (Zhu et al., 2017).

This architectural flaw is particularly salient because post-disaster damage is not a spatially random process. It is governed by what geographers call Tobler’s First Law of Geography, which posits that ”everything is related to everything else, but near things are more related than distant things” (Tobler, 1970). This principle manifests as spatial autocorrelation: the tendency of values of a variable at nearby locations to be more similar than those at distant locations (Moraga, 2019). It is inherently a spatially correlated process due to the path of the hurricane and spatial characteristics of the exposure. Empirical studies of hurricane impacts have repeatedly confirmed this, showing that damage patterns, post-disaster psychological distress, and household relocation decisions exhibit statistically significant spatial clustering (Cao and Abad, 2022). This clustering can be quantified using statistical measures like Moran’s I (Esri Inc., 2023). The existence of this strong spatial dependency represents a rich source of information that is systematically ignored by standard CNN architectures. The failure to leverage this context results in less coherent damage maps and overlooks a key predictive signal present in the data.

To address this critical gap, we move beyond per-building, pixel-grid CNNs to a relational formulation grounded in Graph Neural Networks (GNNs). GNNs operate directly on graph-structured data comprised of nodes and edges (Scarselli et al., 2009), a natural fit for geospatial analysis that involves complex spatial dependencies encompassing both Euclidean proximity and non-Euclidean relational structures (Cui et al., 2021). Concretely, we fine-tune a high-capacity image encoder on DoriaNET to evaluate building damage and then freeze it; we fuse the resulting visual embeddings with structural metadata (e.g., number of stories and an annotation-effort proxy) to form node features; and we construct a spatial K-Nearest Neighbor (kNN) graph from GPS centroids (based on geodesic/Euclidean proximity) whose edges are weighted by a blend of geographic proximity and appearance affinity. On this graph, we apply a two-layer Graph Attention Network (GATv2) (Veličković et al., 2018) (a member in the family of GNNs): *(i)* a multi-head layer that update node representation by learning the weights according to relevance, and *(ii)* a single-head layer that produces logits for ordinal damage states. Training uses a weighted Earth Mover’s Distance (EMD) loss to respect class ordering. Unlike prior DoriaNET/SPDA-style CNN pipelines that classify each building largely in isolation, and prior geospatial GNN studies that are not specifically formulated for UAV-based post-hurricane ordinal damage assessment, our method constructs a building-level graph from GPS centroids and visual similarity and learns target-specific neighborhood influence through graph attention. By explicitly encoding spatial autocorrelation and neighborhood context, the method yields more coherent and context-aware damage assessments; such improvements enhance the reliability of damage intelligence and support fairer allocation of disaster aid (Benson et al., 2021).

In summary, our contributions are as follows:

- We introduce a *spatially aware* post-hurricane damage pipeline that integrates fine-tuned visual embeddings, structural metadata, and a proximity/appearance-weighted kNN graph within a GNN framework (GAT);
- We demonstrate consistent gains over strong per-building CNN/SPDA baselines on DoriaNET, with tighter confusion structure and lower ordinal error, and provide neighborhood-consistency analyses that empirically support the spatial-correlation premise underlying our approach.

2. RELATED WORK

This section expands on the themes introduced in Section 1 by first consolidating *dataset-specific baselines on DoriaNET*, and then situating them within the evolution of damage assessment methods, recent deep learning advances, and emerging graph-based approaches for spatial reasoning.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-101-2026 | © Author(s) 2026. CC BY 4.0 License.

2.1 Baselines on DoriaNET

DoriaNET is a UAV-centric, building-level post-hurricane visual dataset from Hurricane Dorian (2019) with FEMA/HAZUS-style ordinal (damage-grade) labels and geo-referenced metadata (Cheng et al., 2022a, Cheng et al., 2022b). It aggregates *three* 5-minute UAV videos at 1280×720 and 30, FPS over Marsh Harbour, Bahamas; every 10th frame was annotated, yielding **271** frames that cover **147** buildings and **2,409** cropped building patches with pixel-accurate masks. For each building, the dataset provides GPS lat/long centroids, per-frame masks, number of stories, and an *annotation-effort* score (0: easy, 1: moderate, 2: difficult), recorded by human annotators; damage labels follow HAZUS wind-induced states 0–4 with an added state 5 for under-construction/destroyed cases, and all object/scene attributes are organized in a JSON manifest to facilitate downstream modeling and graph construction (Cheng et al., 2022b). The published baseline is a stacked preliminary damage assessment (SPDA) pipeline that (i) localizes building instances and (ii) classifies their *ordinal* damage level using a squared EMD loss; reported results on a pilot subset include ∼90% within ±1 class on seen data and ∼61% exact accuracy, establishing a widely cited benchmark for this dataset.(Cheng et al., 2021a, Cheng et al., 2022c).

Subsequent work using the same data family explored *uncertainty-aware* extensions (Monte Carlo dropout/Bayesian inference) to accompany point predictions with calibrated confidence estimates, improving decision support under ambiguous visual evidence (Cheng et al., 2022d, Cheng et al., 2022e). More recently, the generalizability of the stacked CNN across different hurricane events and scenes has been examined, highlighting performance drift under geographic and distributional shift and motivating models that exploit richer context (Cheng et al., 2022c).In addition to these, student reports and course implementations have reproduced DoriaNET-style pipelines and metrics, further reinforcing the baseline architecture and dataset protocols in the community (Amoyo, 2018).

2.2 Deep Learning for Automated Damage Classification

Building-damage mapping evolved from windshield surveys and manual inspections codified in FEMA's PDA Guide (Preliminary Damage Assessment (PDA) Guide, 2023) to RS/GIS workflows using aerial and satellite imagery (van Westen, 2002), and more recently to ultra-high-resolution imagery with automated vision pipelines (Singh et al., 2023, Wu et al., 2024).

Deep learning has become the default engine for evaluating building damage from RS. The xBD dataset standardized building-level pre/post image pairs and labels, catalyzing model development and benchmark evaluation (Gupta et al., 2019). Strong CNN baselines (e.g., U-Net/ResNet variants, attention modules) and two-stage "detect→classify" designs (e.g., BDANet) achieved competitive results on xBD and related corpora (Su et al., 2020, Shen et al., 2021, Kerle et al., 2020, Zhu et al., 2017). Within UAV imagery, the SPDA baseline on DoriaNET is notable for explicitly adopting an ordinal loss (EMD) in classification (Cheng et al., 2021a), a design choice later echoed in broader civil infrastructure damage studies (Tsai et al., 2024). At the same time, several works demonstrated that model confidence/uncertainty and cross-event robustness remain open challenges for operational deployment (Cheng et al., 2022d, Cheng et al., 2022e, Cheng et al., 2022c).

2.3 Graph Neural Networks in Geospatial Analysis

While CNNs excel at local texture/shape cues, they underutilize geospatial dependencies that are intrinsic to disaster footprints. In other words, with these algorithms, the presence of damaged buildings does not influence the detection of nearby building damage. On the other hand, GNNs naturally encode non-Euclidean structure and neighbourhood interactions (Scarselli et al., 2009). In the subject area of RS, knowledge-guided graph convolution and self-constructing graph attention have improved scene understanding and semantic segmentation by propagating context across spatially related objects (Cui et al., 2021, Zi et al., 2021). Our work builds on these insights by modeling *buildings as nodes* connected by *spatial proximity* and *visual affinity/similarity*, and by employing graph attention to learn neighbourhood influence weights, with the aim of complementing DoriaNET baselines that treat buildings largely in isolation (Veličković et al., 2018).

Building on the motivation that hurricane damage exhibits strong spatial dependence and that per-building CNN classifiers overlook neighbourhood context (Section 1) and the related work on graph-based spatial reasoning, we implement a three-stage pipeline (Figure 1): *(i)* feature extraction that fuses high-fidelity visual descriptors with structural metadata, *(ii)* spatial graph construction with spatially- and visually-aware edge weights, and *(iii)* graph attention for neighbourhood-aware, ordinal damage prediction.

3. PROPOSED METHODOLOGY

3.1 Stage I: Feature Extraction

Image Encoding Each building i is associated with a building patch $\mathcal{I}_i \in \mathbb{R}^{H\times W\times 3}$ obtained by cropping and resizing (e.g., 224×224) a raw frame $\mathcal{F}_i$ using its instance mask $\mathcal{M}_i$. The patch is encoded by an image encoder $T(\cdot)$ (Big Transfer, BiT) (Scheele et al., 2024, MIT Lincoln Laboratory, 2024) to produce a final convolutional feature map

$$T(\mathcal{I}_i) \;\rightarrow\; \mathbf{U}_i \in \mathbb{R}^{h\times w\times d}, \tag{1}$$

which we aggregate to a single d-dimensional visual descriptor by global average pooling over spatial locations:

$$\mathbf{v}_i \;=\; \frac{1}{hw}\sum_{p=1}^{h}\sum_{q=1}^{w}\mathbf{U}_i[p,q,:] \;\in\; \mathbb{R}^d. \tag{2}$$

Node Features To inject structural/contextual cues, we concatenate the visual descriptor with two encoded scalar metadata values: the number of stories s_i and complexity (a *redefinition* of the annotation-effort score in Sec. 2.1) a_i, yielding the *concatenated* node feature

$$\mathbf{f}_i \;=\; \begin{bmatrix}\mathbf{v}_i\\ s_i\\ a_i\end{bmatrix} \in \mathbb{R}^{d+2}. \tag{3}$$

3.2 Stage II: Graph Construction

Mutual kNN Let $\mathbf{c}_i = [\,\ell_i, \lambda_i\,]^\top \in \mathbb{R}^2$ denote the GPS centroid (latitude, longitude) of building i. We compute

$\mathbf{p}_i = [x_i, y_i]^\top$ via projection from WGS84 to a metric CRS (UTM 18N; meters). We then compute pairwise Euclidean distances in geographic space

$$d_{ij} = \left\|\mathbf{p}_i - \mathbf{p}_j\right\|_2, \tag{4}$$

and connect each node to its k nearest neighbors to form the adjacency list $\mathcal{N}(i)$ (kNN graph).

Node-Edge Encoding We first assign $\mathbf{f}_i$ as node features to the graph. To encode both geo-proximity and appearance, we assign each edge (i, j) a weight w_{ij} that blends a spatial proximity score and a visual affinity score. Spatial proximity is

$$S_{ij}^{\text{spatial}} \;=\; \frac{1}{1 + d_{ij}}, \tag{5}$$

while visual affinity uses cosine similarity of the visual descriptors, mapped to $[0, 1]$:

$$S_{ij}^{\text{visual}} \;=\; \frac{1}{2}\left(1 + \frac{\mathbf{v}_i^\top \mathbf{v}_j}{\|\mathbf{v}_i\|_2 \, \|\mathbf{v}_j\|_2}\right). \tag{6}$$

The combined edge weight is a convex blend

$$w_{ij} \;=\; \alpha\, S_{ij}^{\text{spatial}} \;+\; (1-\alpha)\, S_{ij}^{\text{visual}}, \quad \text{with } 0 \le \alpha \le 1. \tag{7}$$

The resulting graph is $\mathcal{G} = (\mathbf{X}, \mathcal{E})$, where $\mathbf{X} \in \mathbb{R}^{N \times (d+2)}$ stacks node features $\mathbf{f}_i$ and $\mathcal{E}$ stores edges (i, j) with weights w_{ij}.

3.3 Stage III: Graph Attention and Ordinal Prediction

We initialize node states with the concatenated features, $\mathbf{h}_i^{(0)} = \mathbf{f}_i$. A multi-head attentional encoder aggregates neighborhood context.

Multi-head attention (MHA) Encoder

(1) Linear Projection For each head $\ell \in \{1, \ldots, H\}$, apply a learned linear projection

$$\mathbf{t}_i^{(\ell)} \;=\; \mathbf{W}^{(\ell)}\, \mathbf{h}_i^{(0)} \;\in\; \mathbb{R}^{r'}. \tag{8}$$

(2) Attention computation For every $(i, j) \in \mathcal{E}$, compute unnormalised attention scores (GATv2 form)

$$e_{ij}^{(\ell)} \;=\; \mathrm{LeakyReLU}\Big(\mathbf{a}^{(\ell)\top}\big(\mathbf{t}_i^{(\ell)} + \mathbf{t}_j^{(\ell)}\big)\Big), \tag{9}$$

and normalise over the neighbourhood with a softmax:

$$\alpha_{ij}^{(\ell)} \;=\; \frac{\exp\Big(e_{ij}^{(\ell)}\Big)}{\sum_{k \in \mathcal{N}(i)} \exp\Big(e_{ik}^{(\ell)}\Big)}. \tag{10}$$

(3) Aggregation Each head outputs a weighted aggregation of neighbour projections,

$$\mathbf{h}_i^{(\ell)\prime} \;=\; \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(\ell)}\, \mathbf{t}_j^{(\ell)}, \tag{11}$$

and we concatenate head-wise outputs

$$\mathbf{h}_i^{(1)} \;=\; \Big\|_{\ell=1}^{H}\, \mathbf{h}_i^{(\ell)\prime}. \tag{12}$$

Self-attention (SA) Classifier Apply a pointwise ELU nonlinearity $\tilde{\mathbf{h}}_i^{(1)} = \mathrm{ELU}\big(\mathbf{h}_i^{(1)}\big)$, then a single-head graph-based self-attention classifier:

$$\mathbf{t}_i^{(2)} \;=\; \mathbf{W}^{(2)}\, \tilde{\mathbf{h}}_i^{(1)} \;\in\; \mathbb{R}^{r''}, \tag{13}$$

$$e_{ij}^{(2)} \;=\; \mathrm{LeakyReLU}\Big(\mathbf{a}^{(2)\top}\big(\mathbf{t}_i^{(2)} + \mathbf{t}_j^{(2)}\big)\Big), \tag{14}$$

$$\alpha_{ij}^{(2)} \;=\; \frac{\exp\Big(e_{ij}^{(2)}\Big)}{\sum_{k \in \mathcal{N}(i)} \exp\Big(e_{ik}^{(2)}\Big)}, \tag{15}$$

$$\mathbf{h}_i^{(2)} \;=\; \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(2)}\, \mathbf{t}_j^{(2)} \;\in\; \mathbb{R}^{r''}, \tag{16}$$

$$\mathbf{z}_i \;=\; \mathbf{U}\, \mathbf{h}_i^{(2)} \;\in\; \mathbb{R}^{C}. \tag{17}$$

The predicted ordinal damage level is $\hat{y}_i = \arg\max_c [\mathbf{z}_i]_c$. During training, we apply a softmax to $\mathbf{z}_i$ to obtain class probabilities and minimise a weighted Earth Mover's Distance (EMD) loss to respect the ordinal structure of damage states; inference uses the argmax over $\mathbf{z}_i$.

4. EXPERIMENTAL SETUP AND RESULTS

This section evaluates the proposed graph-attention pipeline on UAV imagery from *DoriaNET*, a Hurricane Dorian building-level dataset with georeferenced coordinates, FEMA-style ordinal labels, and per-structure metadata. We benchmark against per-building CNN baselines and the stacked preliminary damage assessment (SPDA) pipeline that optimizes an ordinal loss for classification (Cheng et al., 2021b). A public description of DoriaNET and its data protocol is available online. (Cheng et al., 2022b)

4.1 Dataset Description

DoriaNET aggregates UAV flights collected after Hurricane Dorian (2019) over multiple coastal and urban settings (See Section 2) and provides: (i) building instance masks and cropped patches, (ii) GPS centroids enabling spatial graph construction, (iii) structural metadata (e.g., number of stories), and (iv) ordinal building-damage labels consistent with FEMA categories.[1] For all experiments, we follow the split protocol accompanying the dataset, reserving held-out scenes for testing to assess generalization across neighborhoods. For context, we also refer to the SPDA baseline that reports building-level classification with an EMD loss on this data family (Cheng et al., 2021b).

4.2 Implementation Details

Backbone and feature extraction. We use the *LADIv2* classifier (Scheele et al., 2024, MIT Lincoln Laboratory, 2024) as the image encoder. The encoder is first *fine-tuned on DoriaNET* in a standard single-view classification setup using the dataset's ordinal labels. After convergence, the encoder is *frozen* and applied to masked building patches. As outlined in Section 3.1, we perform *mean pooling* over the final feature map to obtain a single descriptor per building, which is then concatenated with metadata to form node features.

Mutual kNN & graph construction. Nodes correspond to concatenated building features; edges are first connected by a *mutual* k-nearest-neighbor with k=1 and weighted with geodesic distances between GPS centroids after projecting to a metric CRS to obtain d_{ij} in meters and also visual affinity. To avoid the

kernel $1/(1+d)$ vanishing at urban scales, we normalize by a neighborhood length scale τ (the median distance to the k-th neighbor over the graph) and use $\delta_{ij} = d_{ij}/\tau$. The spatial term in Eq. (7) is then

$$S_{ij}^{\text{spatial}} = \frac{1}{1+\delta_{ij}}$$

Visual affinity maps cosine similarity to $[0, 1]$ (Refer to Eq. 6). The tuned edge weight, $\alpha = 0.369$, keeps proximity and appearance on compatible scales beyond the closest neighbors. We use mutual k=1 for sparsity and stability; ablations with $k \in \{2, 8\}$ increase connectivity with similar trends, but tends to dilute the message passing in the graph.

GNN architecture and training. We use a two-layer *GATv2* encoder. The first layer employs $H = 4$ attention heads, each with per-head width $r' = d_h/H = 64$, so the concatenated hidden state has $d_h = 256$ channels (cf. Sec. 3.3). This is followed by an ELU and a single-head classifier that maps $\tilde{\mathbf{h}}^{(1)} \in \mathbb{R}^{d_h}$ to logits. The graph module and classifier are trained end-to-end (with fine-tuned LADI-v2 features) using the *weighted EMD* loss to respect ordinality. We optimize with learning rate $\eta = 9.921\times10^{-4}$, a maximum of $E_{\max} = 40$ epochs, and early stopping with patience $p = 18$ based on validation MAE.

Hyperparameter search. Final hyperparameters were selected with Tree-structured Parzen Estimators algorithm (Bergstra et al., 2011, Bergstra et al., 2013), sweeping GAT width/heads, k, and the edge-blend coefficient α together with the learning rate and early-stopping settings.

4.3 Evaluation Metrics

We assess performance with three complementary metrics tailored to ordinal damage states:

Accuracy (overall correctness across classes) and **Precision** (macro-averaged across classes) quantify categorical agreement; **Mean Absolute Error (MAE)** penalizes ordinal mis-ordering by counting the absolute class distance between prediction and ground truth (lower is better). MAE directly encodes the operational cost of confusing, e.g., “minor” with “major,” which is more severe than confusing “minor” with “none.”

5. RESULTS & DISCUSSION

Spatial Behaviour. We first quantify spatial dependence in DoriaNET using Global/ Local Moran’s I under three weight matrices: geographic kNN (`geo`), the blended spatial–visual graph (`graph`), and the learned first-layer attention weights from the GATv2 encoder (`attn`). Table 1 reports a clear, positive Global Moran’s I in all cases, with a monotonic rise from `geo` $\rightarrow$ `graph` $\rightarrow$ `attn` ($0.182 \rightarrow 0.294 \rightarrow 0.341$; all $p<10^{-3}$). This progression is consistent with the intended design: (i) geographic proximity alone already captures that nearby buildings tend to have similar damage; (ii) adding visual affinity preferentially keeps neighbors that both are close and *look* alike while down-weighting nearby connections that do not carry informative signal about the label; and (iii) learned attention further concentrates weights on the most useful subset of edges, producing the highest measured autocorrelation. The Local Moran column supports the same picture: the share of nodes in statistically significant clusters increases from 16.0% (`geo`) to 25.4% (`attn`), indicating more—and more coherent—like-labeled neighborhoods.

Table 1. Spatial dependence diagnostics (Moran’s I) under three weighting schemes: geographic kNN (`geo`), blended spatial–visual graph (`graph`), and learned attention weights (`attn`).

Weight	Global Moran’s I	z-score / p-value	Local Moran ($p \leq 0.05$)
`geo`	0.182	3.71 / 2.1×10^{-4}	16.0%
`graph`	0.294	3.38 / 7.2×10^{-4}	22.7%
`attn`	0.341	3.95 / 7.9×10^{-5}	25.4%

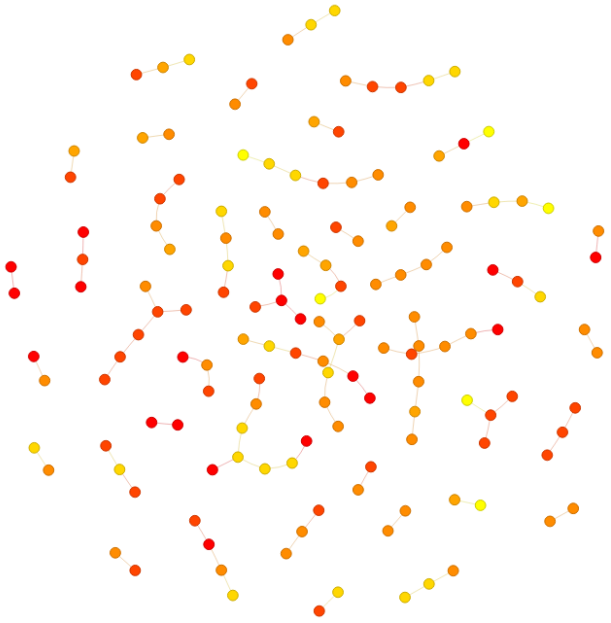


Figure 2. A snippet of the kNN spatial graph over building centroids. Nodes denote buildings; edges connect k nearest neighbors weighted by blended spatial–visual similarity. Colors indicate ordinal damage states.

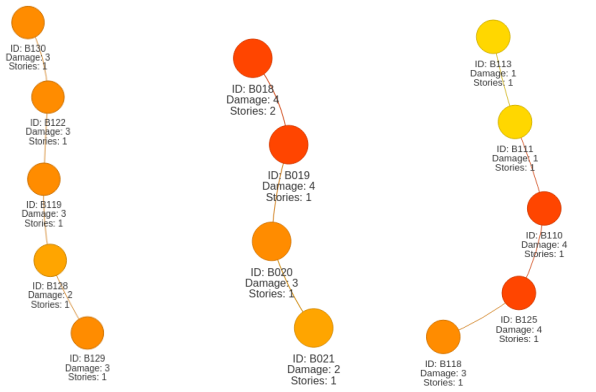
To interpret these results at the building scale, Table 2 summarizes agreement in kNN neighborhoods by ordinal tolerance. Nearly 70% of neighborhoods fall within ±1 class of the anchor and >90% within ±2, showing that most local clusters occupy a narrow ordinal band. This distribution is exactly what the positive Moran’s I reflects: as we tighten tolerance, the mass decays quickly if nearby labels are similar, while a broad tolerance (±3) captures almost all clusters. In short, the dataset exhibits the block-level consistency expected of wind/surge footprints and contiguous built forms.

Table 2. Neighborhood agreement by ordinal tolerance.

Tolerance	Clusters within tolerance
Within ±1 damage state	69.8%
Within ±2 damage states	91.1%
Within ±3 damage states	98.2%

Figure 2 sketches the kNN graph used for message passing, and Figure 3 shows representative neighborhoods at ±1, ±2, and ±3 tolerances. These visuals match the tables: many anchors sit in locally consistent pockets, while mixed triads occur at micro-boundaries (e.g., across a road or shoreline) or where construction types shift.

Attention Behaviour. Given the measurable autocorrelation above, the first-layer attention heatmap (Figure 4) explains why `attn` produces the strongest I. High weights generally follow the kNN scaffold but are not uniform: the model increases weights on neighbors (near diagonals) with the same (or near)



(a) Cluster within ± 1 (b) Cluster within ± 2 (c) Cluster within ± 3

Figure 3. Representative neighborhoods at increasing ordinal tolerance. Node colors encode ordinal damage labels.

damage state and decreases weights on conflicting neighbors (further from diagonals). In effect, attention learns a data-driven mask over the blended graph—keeping edges that help prediction and soft-removing those that add noise. Occasional longer links (off-diagonals) appear when distant buildings share strong visual cues (e.g., similar roof failures). These links are rare but helpful, and they account for improvements over `geo` and even over `graph`, which cannot adapt edge influence per target node.

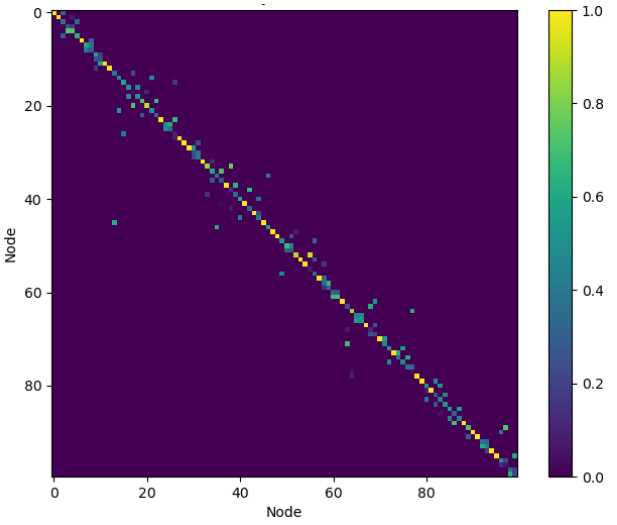


Figure 4. GATv2 first-layer attention. Weights concentrate on informative neighbors within the blended graph, with occasional visually driven long-range links.

Ordinal Fidelity. Table 3 summarizes aggregate performance on DoriaNET. Our two-layer GATv2 with BiT embeddings and metadata achieves **0.778** accuracy, **0.751** precision, and **0.283** MAE, outperforming per-building CNN baselines and an SPDA-style ordinal classifier (Cheng et al., 2021b). The *MAE* is the key: lower MAE means fewer large class jumps, which is what we expect when local consensus helps “pull” ambiguous cases toward neighborhood-consistent labels. The confusion matrix (Figure 5) aligns with this interpretation—a tight diagonal with most residuals within ± 1 class—mirroring the neighborhood tolerance statistics (Table 2) and the stronger clustering seen under `attn` in Table 1.

Table 3. Performance on DoriaNET (test). Baselines marked †.

Model	Accuracy ↑	Precision ↑	MAE ↓
ResNet-50 per-building (Cross Entropy)†	0.741	0.702	0.397
SPDA-CNN (EMD)† (Cheng et al., 2021b)	0.754	0.718	0.372
Ours: GATv2 (BiT + meta + spatial)	**0.778**	**0.751**	**0.283**

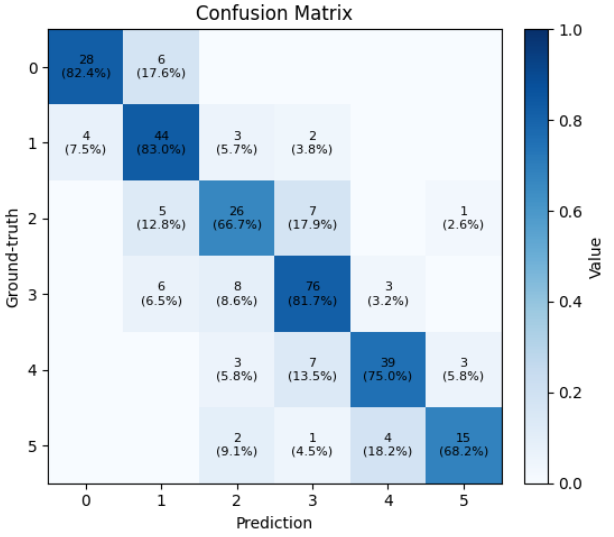


Figure 5. Confusion matrix of predicted vs. ground-truth ordinal damage states. Residuals lie predominantly within ± 1 class, matching the low MAE.

Synthesis and implications. Together, the spatial diagnostics (Table 1), neighborhood tolerance distributions (Table 2), attention patterns (Figure 4), and aggregate/error metrics (Table 3, Figure 5) tell a single story. Damage in DoriaNET is spatially clustered; blending proximity with visual similarity strengthens the useful part of that structure by reducing nearby links that are not predictive; attention then sharpens it further by emphasizing the most informative neighbors per target node. The net effect is better ordinal calibration (lower MAE) and cleaner error structure (residuals within ± 1), which translates into more coherent, actionable maps.

Limitations point to next steps. Mixed micro-environments (e.g., lightly and heavily damaged homes side-by-side) can still confuse message passing; centroid-based kNN may bridge across roads or water where exposure differs. A practical refinement is to introduce a *semivariogram-guided* graph: estimate empirical (directional) semivariograms on labels or residuals to recover the correlation length (range), nugget, and anisotropy; use the range to set k and to rescale distances; use the nugget to cap over-smoothing in noisy pockets; and gate or down-weight edges whose semivariance is anomalously high given their separation (barriers). Fitting a simple variogram model (exponential/Matérn) also yields a covariance kernel $C(h)$ that can replace the ad-hoc $1/(1+d)$ term in w_{ij} or serve as a lightweight prior for attention weights. Alongside barrier-/topology-aware edges, adaptive k, and adding environmental covariates (e.g., wind, surge) to guide edge strength, this semivariogram-guided construction is a practical way to make the graph more faithful to the spatial process while preserving the interpretability we observe here.

6. CONCLUSION AND FUTURE WORK

This work set out to test a simple premise: *post-hurricane damage is spatially correlated, and models that reason over neigh-*

borhoods should therefore make better, more actionable predictions than per-building classifiers. We operationalized this premise with a three–stage pipeline—(i) BiT-based visual feature extraction with structural metadata, (ii) kNN graph construction with blended spatial–visual affinities, and (iii) graph attention for neighborhood-aware ordinal inference—and evaluated it on DoriaNET. The resulting GATv2 model delivers consistent gains over CNN and SPDA-style baselines, reaching **0.778** accuracy, **0.751** precision, and an **MAE of 0.283**. Beyond the scores, the attention maps concentrate on proximate structures, the confusion matrix exhibits tight diagonal mass with most residuals within ± 1 class, and the kNN-cluster statistics align with Tobler's law—all converging evidence that fusing spatial context with visual cues yields more faithful damage maps.

From an operational standpoint, the improvement in MAE is particularly salient: preserving the *ordering* of damage severity reduces costly large hops (e.g., "minor" $\rightarrow$ "major") and produces contiguous swaths of consistent predictions that are easier to triage, route, and resource. The attention mechanism also provides a measure of interpretability by revealing which neighborhoods drive each decision, complementing the aggregate neighborhood-agreement analyses.

There remain meaningful challenges. Heterogeneous microenvironments—especially mixtures of lightly affected (0–1) and moderately/heavily affected (3–4) structures—can still mislead message passing, and centroid-based kNN can connect areas that are geographically close but separated by roads, water, or distinct exposure. Practical next steps include: (i) *semivariogram-guided neighborhoods* to set a data-driven range (neighborhood size), account for anisotropy, and down-weight edges whose semivariance is high for their separation; (ii) *topology-/barrier-aware edges* that respect road networks, parcels, and water boundaries; (iii) *adaptive k and heterophily-aware attention* to prevent over-smoothing when nearby labels disagree; and (iv) adding environmental covariates (wind, surge, elevation) as node/edge features so the graph encodes exposure, not just distance. A complementary axis is *spatio-temporal and multi-sensor fusion*—linking repeated UAV passes and pairing UAV with satellite/LiDAR—to capture both fine-scale cues and neighborhood evolution. Finally, *uncertainty quantification* (calibration, ensembling), *cross-event/domain generalization* (shift-robust training, test-time adaptation), and *model compression* (distillation, pruning/quantization) are practical enablers for field deployment.

In sum, explicitly encoding spatial correlation with graph attention moves damage assessment beyond isolated image cues toward coherent, neighborhood-aware intelligence. The empirical gains and the interpretability of the learned neighborhood weights suggest a promising path for scalable, equitable, and timely post-disaster mapping.

References

Al Shafian, A., Hu, X., 2024. Advances in Rapid Damage Identification Methods for Post-Disaster Buildings: A Review. *Buildings*, 14(4), 898.

Amoyo, A., 2018. Post-disaster building damage classification from aerial imagery (dorianet) – final report. Technical report, University of Central Florida, CRCV.

Benson, A. et al., 2021. Beyond damages: Social equity in allocating disaster assistance. Natural Hazards Center.

Bergstra, J. S., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems (NeurIPS)*, 24, 2546–2554.

Bergstra, J., Yamins, D., Cox, D. D., 2013. Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms. *Proceedings of the 12th Python in Science Conference (SciPy)*, 13–20.

Cao, E., Abad, A., 2022. Use of local spatial autocorrelation to assess typhoon-damaged buildings. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2022, 1077–1082.

Cheng, C.-S., Behzadan, A. H., Noshadravan, A., 2021a. Deep Learning for Post-Hurricane Aerial Damage Assessment of Buildings. *Computer-Aided Civil and Infrastructure Engineering*, 36(6), 695–710.

Cheng, C.-S., Behzadan, A. H., Noshadravan, A., 2022a. Dorianet: A visual dataset from hurricane dorian for post-disaster building damage assessment. DesignSafe-CI (Published Dataset).

Cheng, C.-S., Behzadan, A. H., Noshadravan, A., 2022b. Dorianet: A visual dataset from hurricane dorian for post-disaster building damage assessment. Technical report, Texas A&M University.

Cheng, C.-S., Behzadan, A. H., Noshadravan, A., 2022c. Exploring the generalizability of deep convolutional neural networks for post-hurricane damage assessment. *ASCE Lifelines Conference 2022: Long-Term Resilience*, ASCE.

Cheng, C.-S., Noshadravan, A., Behzadan, A. H., 2022d. Bayesian inference for uncertainty-aware post-disaster damage assessment using artificial intelligence. *International Conference on Computing in Civil Engineering (ICCCE)*. Preprint available.

Cheng, C.-S., Noshadravan, A., Behzadan, A. H., 2022e. Uncertainty-aware Convolutional Neural Network for Explainable AI-Assisted Disaster Damage Assessment. *Preprint*. https://www.researchgate.net/publication/361111923.

Cheng, M.-Y. et al., 2021b. Post-Disaster Preliminary Damage Assessment Using Unmanned Aerial Vehicle Imagery and Artificial Intelligence. *Computer-Aided Civil and Infrastructure Engineering*, 36(8), 998–1015.

Cui, Z., Li, F., Cao, Z., Wang, Y., Yang, Y., 2021. Knowledge-Guided Graph Convolutional Networks for Semantic Segmentation of Remote Sensing Images. *Remote Sensing*, 13(21), 4279.

Esri Inc., 2023. How spatial autocorrelation (global moran's i) works. ArcGIS Pro Documentation.

Goldmann, E., Galea, S., 2014. Mental Health Consequences of Disasters. *Annual Review of Public Health*, 35, 169–183.

Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT Press.

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., Gaston, M., 2019. xBD: A Dataset for Assessing Building Damage from Satellite Imagery. *arXiv preprint arXiv:1911.09296*. https://arxiv.org/abs/1911.09296.

Howarth, C., Butler, D., 2020. Problems with damage assessments can keep disaster victims from receiving the help they need. Urban Institute.

Kerle, N., Nex, F., Gerke, M., Duarte, D., Vetrivel, A., 2020. Automated Building Damage Classification from Remote Sensing Images: From Experiment to Operational Emergency Response. *Remote Sensing*, 12(17), 2839.

Microsoft Research, 2022. Ai for good: Disaster response (icdm 2022 workshop summary). Workshop summary.

MIT Lincoln Laboratory, 2024. Mitll/ladi-v2-classifier-large. Hugging Face model card.

Mitchell, S., 2020. Developing a rapid damage assessment procedure. Technical report, National Fire Academy.

Moraga, P., 2019. *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. Chapman and Hall/CRC.

NOAA National Centers for Coastal Ocean Science, 2024. New study advances detection of building damage after hurricanes. NOAA News.

Preliminary Damage Assessment (PDA) Guide, 2023. Technical report, Federal Emergency Management Agency (FEMA).

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., Monfardini, G., 2009. The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, 20(1), 61–80.

Scheele, S., Picchione, K., Liu, J., 2024. Ladi v2: Multi-label dataset and classifiers for low-altitude disaster imagery.

Shen, Y., Zhu, S., Yang, T., Chen, C., Pan, D., Chen, J., Xiao, L., Du, Q., 2021. Bdanet: Multiscale cnn with cross-directional attention for building damage assessment from satellite images.

Singh, D. K. et al., 2023. Post Disaster Damage Assessment Using Ultra-High-Resolution Imagery and Transformer Models. *Remote Sensing*, 15(20), 5001.

Su, J., Wang, Y., Li, M. et al., 2020. A perspective on xbd: Building damage assessment dataset and methods. arXiv preprint.

Tobler, W. R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46(sup1), 234–240.

Tsai, F.-J. et al., 2024. A Class Distance Penalty Deep Learning Method for Post-Disaster Building Damage Assessment. *KSCE Journal of Civil Engineering*.

van Westen, C. J., 2002. Remote sensing and geographic information systems for natural disaster management. *Remote Sensing and Digital Image Processing*, Springer, 407–440.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y., 2018. Graph attention networks. *International Conference on Learning Representations (ICLR)*.

Wu, J. et al., 2024. Post-Disaster Building Damage Assessment and Reconnaissance: A Bibliometric Review of Current Methodologies and Technologies. *Infrastructures*, 9(8), 2344.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and a List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36.

Zi, W., Xiong, W., Chen, H., Li, J., Jing, N., 2021. SGA-Net: Self-Constructing Graph Attention Neural Network for Semantic Segmentation of Remote Sensing Images. *Remote Sensing*, 13(21), 4201.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-101-2026 | © Author(s) 2026. CC BY 4.0 License.