Alleviating Performance Disparity in Adversarial Spatiotemporal Graph Learning Under Zero-Inflated Distribution

Songran Bai[1,2], **Yuheng Ji**[1,2], **Yue Liu**[3], **Xingwei Zhang**[1,2], **Xiaolong Zheng**[1,2*], **Daniel Dajun Zeng**[1,2]

[1]Institute of Automation, Chinese Academy of Sciences
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences
[3]Institute of Data Science & School of Computing, National University of Singapore
songran.bai@mais.ia.ac.cn

Abstract

Spatiotemporal Graph Learning (SGL) under Zero-Inflated Distribution (ZID) is crucial for urban risk management tasks, including crime prediction and traffic accident profiling. However, SGL models are vulnerable to adversarial attacks, compromising their practical utility. While adversarial training (AT) has been widely used to bolster model robustness, our study finds that traditional AT exacerbates performance disparities between majority and minority classes under ZID, potentially leading to irreparable losses due to underreporting critical risk events. In this paper, we first demonstrate the smaller top-k gradients and lower separability of minority class are key factors contributing to this disparity. To address these issues, we propose MinGRE, a framework for Minority Class Gradients and Representations Enhancement. MinGRE employs a multi-dimensional attention mechanism to reweight spatiotemporal gradients, minimizing the gradient distribution discrepancies across classes. Additionally, we introduce an uncertainty-guided contrastive loss to improve the inter-class separability and intra-class compactness of minority representations with higher uncertainty. Extensive experiments demonstrate that the MinGRE framework not only significantly reduces the performance disparity across classes but also achieves enhanced robustness compared to existing baselines. These findings underscore the potential of our method in fostering the development of more equitable and robust models.

Introduction

Spatiotemporal Graph Neural Networks (STGNNs) have emerged as a vital component in modeling complex spatiotemporal dependencies within Spatiotemporal Graph Learning (SGL) under the Zero-Inflation Distribution (ZID) (Liu et al. 2023b,d; Zhao et al. 2023; Trirat, Yoon, and Lee 2023; Tang, Xia, and Huang 2023). The datasets that conforms to such distribution consist of a majority of zero observations and a minority of non-zero observations (Wilson et al. 2022; Lichman and Smyth 2018; Ghosh, Mukhopadhyay, and Lu 2006; Feng 2021). Effectively addressing ZID is pivotal for discerning sparse event patterns in urban crime analysis, traffic accident forecasting, and demand prediction (Zhuang et al. 2022; Wang et al. 2024; Liang et al. 2024; Jiang et al. 2024).

Nevertheless, recent studies have identified vulnerabilities within STGNNs, where adversaries could induce incorrect traffic predictions by slightly perturbing historical data (Zhu et al. 2024; Liu, Liu, and Jiang 2022; Li et al. 2022). Consequently, Adversarial Training (AT) has been introduced to bolster the robustness of these models (Liu, Zhang, and Liu 2023). This process generally encompasses three key stages: the selection of salient victim nodes, the generation of Adversarial Examples (AEs), and iterative optimization (Liu, Liu, and Jiang 2022; Liu, Zhang, and Liu 2023). However, the effectiveness of such spatiotemporal adversarial training has been evaluated primarily on dense datasets with normal distributions. Its effectiveness on sparse, zero-inflated datasets remains a significant and worthy area of exploration.

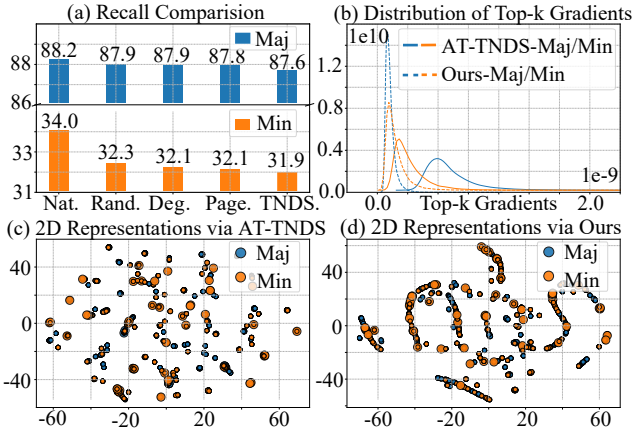

Figure 1: Impact analysis of spatiotemporal adversarial training on the ZID dataset NYC. (a) compares recall metrics between natural training and adversarial training approaches. (b) displays the distribution of top-K gradients for both majority and minority classes throughout the adversarial training. Panels (c) and (d) present two-dimensional projections of the learned features for majority and minority classes via AT-TNDS and our proposed method, respectively.

To this end, we initially investigate the performance of

*Corresponding author

existing spatiotemporal adversarial training methods in ZID scenarios, with a particular focus on the prediction performance and robustness of non-zero observations representing minority class, as this is crucial for addressing serious safety concerns such as incident underreporting (Yamamoto, Hashiji, and Shankar 2008) in real-world applications. Our empirical analysis of zero-inflated datasets revealed three findings as follows. 1) Conventional spatiotemporal adversarial training approaches tend to exacerbate the performance disparity between majority and minority classes, as illustrated in Figure 1(a), due to a more significant degradation of the minority class. 2) Our study further reveals that the top-k gradients of the minority class are generally weaker, leading to a dominance of majority class adversarial examples in training (see Figure 1(b)). 3) Furthermore, as illustrated in Figure 1(c), the separability of the minority representations deteriorates following adversarial training. Moreover, samples with high uncertainty exhibit greater prediction errors, as indicated by the size of the points in the figure. Thus, we posit that the smaller top-k gradients and lower separability of the minority class are two potential underlying causes of performance disparity.

To address current challenges, we propose the Minority Class Gradients and Representations Enhancement (MinGRE) framework. Our approach begins with a victim node selection strategy during adversarial training, crucial for generating fairer and more effective perturbations across classes. This strategy employs a cross-segment spatiotemporal encoder to capture complex inter-segment, intra-segment, and spatial dependencies. Additionally, we introduce a multi-dimensional attention-based gradient reweighting technique that adaptively adjusts spatiotemporal gradients throughout the training, reducing bias towards the majority class. Furthermore, inspired by Zha et al.'s work on maintaining continuity in representation space for regression tasks (Zha et al. 2024), we incorporate an uncertainty-guided contrastive learning loss. This loss function maximizes feature dissimilarity between classes, particularly in regions with high predictive uncertainty.

The main contributions of this paper are as follows:

- We analyze the adversarial robustness of SGL models under zero-inflated settings, identifying significant issues in performance disparity.
- We introduce a multi-dimensional attention-based gradient reweighting method to improve the selection of victim nodes in spatiotemporal adversarial training.
- We employ an uncertainty-guided contrastive loss to focus on representation learning in regression tasks, thereby reducing inter-class similarity and enhancing intra-class cohesion.
- Extensive experiments across various target models, attack methods, and datasets confirm the effectiveness of our proposed framework on both the robustness and disparity metrics.

Related Work

Spatiotemporal Graph Learning Under ZID

Spatiotemporal graph learning has garnered substantial interest, particularly in domains with sparse or zero-inflated data (Li et al. 2024). Models like GMAT-DU (Zhao et al. 2023) and RiskSeq (Zhou et al. 2022) underscore the value of granular spatiotemporal data in data-scarce environments (Chen et al. 2024). The recent trend of employing graph neural networks (GNNs) with dynamic and multi-view approaches, exemplified by MADGCN (Wu et al. 2023) and MG-TAR (Trirat, Yoon, and Lee 2023), demonstrates the synergy between spatiotemporal dynamics and attention mechanisms to improve prediction accuracy. Furthermore, the integration of uncertainty quantification in STGNNs (Gao et al. 2023; Zhou et al. 2024; Zhuang et al. 2024; Gao et al. 2024) underscores the necessity for robust models capable of handling sparse data and providing dependable predictions.

Adversarial Robustness of Spatiotemporal Graph Learning

Adversarial attacks are crucial for assessing model robustness (Zhang, Zheng, and Mao 2021), especially in spatiotemporal contexts (Liu et al. 2024). Designing such attacks involves dynamically selecting victim nodes and generating time-dependent perturbations while ensuring the attacks remain imperceptible. (Zhu et al. 2024) proposed a query-based black-box attack using SPSA (Uesato et al. 2018) for gradient estimation and a knapsack greedy algorithm for node selection. (Liu, Liu, and Jiang 2022) introduced STPGD, an iterative method suitable for both white-box and gray-box scenarios. ADVERSPARSE (Li et al. 2022), on the other hand, targets graph structures by sparsifying them to disrupt spatial dependencies and increase prediction errors. Adversarial training has also shown promise in enhancing robustness (Jiang et al. 2023a), with AT-TNDS integrating spatiotemporal perturbations into the training process (Liu, Liu, and Jiang 2022). (Liu, Zhang, and Liu 2023) leveraged reinforcement learning for dynamic node selection, alongside knowledge distillation to stabilize the policy network. (Zhang et al. 2023) further strengthened spatiotemporal representations using contrastive loss within a self-supervised learning framework. However, current research primarily addresses dense, continuous data, often overlooking the discrete and sparse nature of critical spatiotemporal data (Wölker et al. 2023; Wang et al. 2021b).

Adversarial Training in Imbalanced Settings

Recent studies have underscored the critical impact of data imbalance on the effectiveness of adversarial training (Xiong et al. 2024; Yue et al. 2024; Dobriban et al. 2023). In such conditions, adversarial training can amplify the imbalance, thereby diminishing the model's performance on underrepresented classes (Wang et al. 2022). To address these challenges, approaches such as scale-invariant classifiers and two-stage rebalancing frameworks have been proposed (Wu et al. 2021). Furthermore, meta-learning-based sample-aware re-weighting has demonstrated potential in enhanc-

ing adversarial robustness within imbalanced datasets (Hou, Han, and Li 2023). These methods aim to balance class representation during training, with strategies like margin engineering and re-weighting showing promise in enhancing adversarial robustness under imbalanced settings (Qaraei and Babbar 2022).

Preliminaries

Spatiotemporal Prediction Under ZID

Let $\mathcal{G}_t^k = (\mathcal{V}, \mathcal{E}^k, \mathcal{A}^k, \mathcal{X}_t)$ denote multi-view undirected graphs at step t, where $\mathcal{V}$ is the set of N nodes that is time-invariant. $\mathcal{E}^k$ denotes the edge set of k^{th} view and $\mathcal{A}^k$ denotes the adjacency matrix of k^{th} view. Then $\mathcal{X}_t \in \mathbb{R}^{N*D}$ denotes the D-dimensional node features at time t. The prediction model aims to estimate future node states $\mathcal{Y}_{t+1:t+\Delta}$ as follows:

$$\hat{\mathcal{Y}}_{t+1:t+\Delta} = f_\theta\left(\mathcal{X}_{t-\mathcal{T}+1:t}, \mathcal{A}\right) \tag{1}$$

Here, $\mathcal{Y}_{t+1:t+\Delta}$ exhibits the characteristic of zero-inflation distribution, which means that non-zero labels are sparsely distributed in both temporal and spatial dimensions. For simplicity, we will use $\mathcal{X}_t^{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}*N*D}$ to represent node features from time $t-\mathcal{T}+1$ to time t in the following content. And we use $\mathcal{Y}_t^{\Delta}$, $\hat{\mathcal{Y}}_t^{\Delta} \in \mathbb{R}^{\Delta*N}$ to represent the real and predicted node states from time $t+1$ to time $t+\Delta$. The widely used weighted RMSE loss function (Wang et al. 2023) can be defined as:

$$\mathcal{L}\left(\hat{\mathcal{Y}}_t^{\Delta}, \mathcal{Y}_t^{\Delta}\right) = \frac{1}{\Delta * N} \sum_{\delta,n} w_t^{(\delta,n)} \left(y_t^{(\delta,n)} - \hat{y}_t^{(\delta,n)}\right)^2 \tag{2}$$

where $y_t^{(\delta,n)}$, $\hat{y}_t^{(\delta,n)}$ and $w_t^{(\delta,n)}$ represent the real state, the predicted state and the loss weight of node $\mathcal{V}_n$ at time $t+\delta$, respectively.

Spatiotemporal Adversarial Attack

The objective of spatiotemporal graph adversarial attacks is to maximize prediction errors by perturbing the historical attributes of a minimal subset of node features. The optimal AEs can be defined as (Liu, Liu, and Jiang 2022):

$$\underset{(\mathcal{X}_t^{\mathcal{T}})' \in \mathcal{B}(\mathcal{X}_t^{\mathcal{T}})}{\operatorname{argmax}} \sum_{t \in T_{test}} \mathcal{L}\left(f_{\theta^*}(\cdot), \mathcal{Y}_t^{\Delta}\right) \tag{3}$$

$$s.t. \quad \left\| \left((\mathcal{X}_t^{\mathcal{T}})' - \mathcal{X}_t^{\mathcal{T}} \right) \circ \mathcal{P}_t \right\|_p \leq \epsilon, \ \|\mathcal{P}_t\|_0 \leq \eta * N \tag{4}$$

where ϵ and η denote the attack budget and the proportion of nodes being attacked, respectively. And $\mathcal{P}_t \in \mathbb{R}^{1*N*1}$ represents a three-dimensional matrix containing only 0 and 1. If the elements $\mathcal{P}_t(:,i,:)$ are 1, it indicates that the node $\mathcal{V}_i$ will be attacked. And $\mathcal{B}\left(\mathcal{X}_t^{\mathcal{T}}\right) = \left\{\mathcal{X}_t^{\mathcal{T}} + \Phi_t^{\mathcal{T}} \circ \mathcal{P}_t \mid \|\Phi_t^{\mathcal{T}} \circ \mathcal{P}_t\|_p \leq \epsilon\right\}$ represents the allowed perturbation set. To solve the above optimization problem, (Liu, Liu, and Jiang 2022) firstly calculate the gradient-based time-dependent non-negative node saliency within a batch:

$$\mathcal{S}_{T_{batch}} = \|\text{Relu}\left(\frac{1}{B} \sum_{t \in T_{batch}} \nabla \mathcal{L}(\cdot)\right)\|_2 \tag{5}$$

then the victim nodes can be represented by $\mathcal{P}_t(:,i,:) = \mathbf{1}_{\mathcal{V}_i \in \text{top}_k\left(\mathcal{S}_{\mathcal{T}_{batch}}\right)}$, where $\mathcal{P}_t(:,i,:)$ will be 1 if $\mathcal{V}_i$ is the $top-k$ salient node within a batch $\mathcal{T}_{batch}$. Based on the victim nodes, the iteration process of Spatiotemporal Projected Gradient Desent (STPGD) can be defined as:

$$\left(\mathcal{X}_t^{\mathcal{T}}\right)'^{(i)} = clip_\epsilon\left(\left(\mathcal{X}_t^{\mathcal{T}}\right)'^{(i-1)} + \alpha \text{sign}\left(\nabla \mathcal{L}(\cdot) \circ \mathcal{P}_t\right)\right) \tag{6}$$

where $\text{clip}_\epsilon(\cdot)$ is the operation to bound the perturbation in a ϵ ball. And $\left(\mathcal{X}_t^{\mathcal{T}}\right)'^{(i)}$ represents the adversarial features of i^{th} iteration.

Spatiotemporal Adversarial Training

Adversarial training in the context of spatiotemporal graph learning can also be regarded as a min-max optimization process, which enhances the robustness of the model against adversarial attacks. This can be formulated as:

$$\min_\theta \max_{(\mathcal{X}_t^{\mathcal{T}})' \in \mathcal{B}(\mathcal{X}_t^{\mathcal{T}})} \sum_{t \in T_{train}} \mathcal{L}\left(f_\theta(\cdot), \mathcal{Y}_t^{\Delta}\right) \tag{7}$$

Since the above problem is most likely a non-convex bi-level optimization problem, many studies approximate it by alternating first-order optimization, that is, training f_θ on the adversarial perturbed spatiotemporal graph in each iteration.

Methodology

This section delineates the MinGRE framework through two key components: the Adversarial Examples Generation Module and the Uncertainty-guided Contrastive Loss Module, as illustrated in Figure 2. The implementation of our proposed method is presented in the Appendix.

Adversarial Examples Generation Module

The primary challenge in generating adversarial samples is effectively reweighting gradients to ensure a more balanced selection of victim nodes. For instance, considering the weighted RMSE Loss, we can simplify the expression for the gradient $\nabla \mathcal{L}(\cdot)$ of a sample i using the chain rule, as follows:

$$\nabla \mathcal{L}(\cdot) = \frac{\partial \mathcal{L}}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{x_i} = w_i \cdot \frac{\partial \hat{y}_i}{x_i} = w_i \cdot \mathcal{G}_i \tag{8}$$

It can be observed from the above equation that the predefined weight w_i and the variable $\mathcal{G}_i$ determine the final magnitude of the gradients. Notably, w_i is set based on expert knowledge, and $\mathcal{G}_i$ assumes that gradient flow across different temporal dimensions of x_i holds uniform importance (Chen et al. 2021). However, node selection strategies based on these assumptions are unsuitable for ZID scenarios, as they can result in biased gradient distributions between majority and minority classes. To address this issue, we propose a multi-dimensional gradient reweighting strategy that employs segment and spatial attention to focus on samples within specific segments and nodes. Additionally, we introduce temporal attention mechanisms to differentiate the importance of gradient flows from various temporal dimensions of the input feature $\mathcal{X}$. This approach is implemented

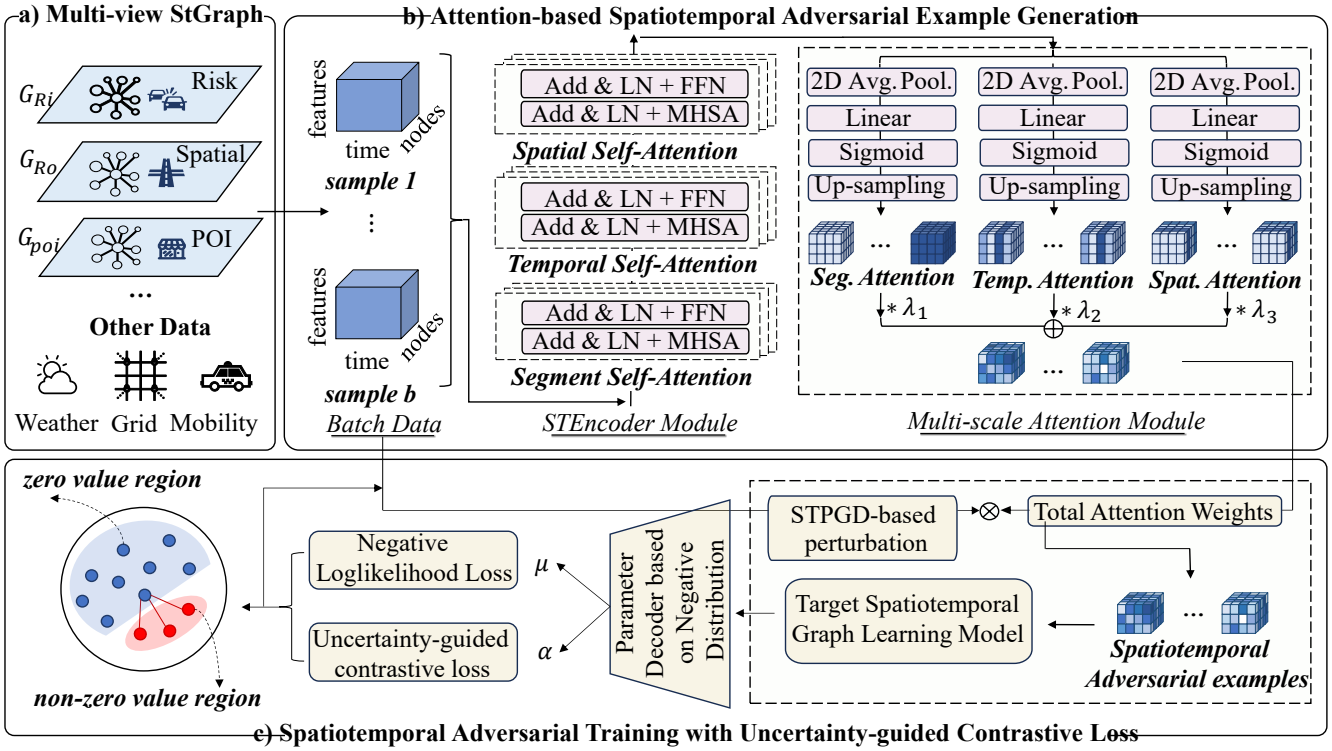

Figure 2: The overall framework of our proposed MinGRE.

through three key components: a Cross-Segment Spatiotemporal Encoder, Gradient Reweighting-based Adversarial Example Generation, and an Optimization Objective.

Cross-Segment Spatiotemporal Encoder Building on the work of (Kossen et al. 2021), we have incorporated the Attention Between Datapoints (ABD) mechanism to capture pairwise interactions across different segments within a batch. Consider a batch of spatiotemporal segments denoted as $\mathcal{X} = \left\{\mathcal{X}_t^{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}*N*D} \mid t = t_1, \ldots, t_B\right\}$. The ABD layer processes these samples as follows:

$$\mathcal{O}_{sg}\left(\mathcal{X}\right) = \mathrm{LN}\left(\mathcal{R}\left(\mathcal{X}_{sg}\right) + \mathrm{FFN}\left(\mathcal{R}\left(\mathcal{X}_{sg}\right)\right)\right), \quad (9)$$

where $\mathcal{X}_{sg} = \pi_{\sigma(sg)}\left(\mathcal{X}\right)$ reshapes the input tensor $\mathcal{X}$ to conform to the dimensions $(\mathcal{T}, N, B, D)$. The function $\mathcal{R}\left(\mathcal{X}_{sg}\right)$, defined by

$$\mathcal{R}(\mathcal{X}_{sg}) = \mathrm{LN}\left(\mathcal{M}\left(\mathcal{X}_{sg}\right) + \mathcal{X}_{sg}\right), \quad (10)$$

represents the residual output of the ABD module. Here, $\mathcal{M}\left(\mathcal{X}_{sg}\right)$ computes the output from the multi-head self-attention mechanism, which is expressed as:

$$\mathcal{M}\left(\mathcal{X}_{sg}\right) = \mathrm{concat}\left(\mathcal{M}_{sg}^1, \ldots, \mathcal{M}_{sg}^k\right) \mathcal{W}_{sg}^{\mathcal{M}}, \quad (11)$$

where each $\mathcal{M}_{sg}^j$ is obtained by

$$\mathcal{M}_{sg}^j = \mathrm{softmax}\left(\frac{\mathcal{Q}_{sg}^j \left(\mathcal{K}_{sg}^j\right)^T}{\sqrt{d}}\right) \mathcal{V}_{sg}^j, \quad (12)$$

and embedding matrices $\left(\mathcal{Q}_{sg}^j, \mathcal{K}_{sg}^j, \mathcal{V}_{sg}^j\right)$ are computed as $\left(\mathcal{X}_{sg}\mathcal{W}_{sg}^{\mathcal{Q}_j}, \mathcal{X}_{sg}\mathcal{W}_{sg}^{\mathcal{K}_j}, \mathcal{X}_{sg}\mathcal{W}_{sg}^{\mathcal{V}_j}\right)$. This formulation effectively models complex interactions between the segments.

Similarly, in order to sequentially encode the temporal and spatial dependencies (Liu et al. 2023a), we connect the temporal self-attention and spatial self-attention mechanisms in series after the ABD module. The computation process is the same as equation (9), and the final output of the encoder is:

$$\mathcal{O}\left(\mathcal{X}\right) = \mathcal{O}_{sp}\left(\pi_{\sigma(sp)}\left(\mathcal{O}_{te}\left(\pi_{\sigma(te)}\left(\mathcal{O}_{sg}\left(\mathcal{X}\right)\right)\right)\right)\right), \quad (13)$$

where $\pi_{\sigma(te)}\left(\cdot\right)$ reshapes the input tensor to conform to the dimensions $(B, N, \mathcal{T}, D_h)$ and $\pi_{\sigma(sp)}\left(\cdot\right)$ reshapes the input tensor to conform to the dimensions $(B, \mathcal{T}, N, D_h)$.

Gradient Reweighting-Based Adversarial Example Generation We aim to increase the selection probability of non-zero regions in adversarial sample generation. A potential solution is to re-weight the spatiotemporal gradients $grad = \left\{\nabla\mathcal{L}\left(\cdot\right) \in \mathbb{R}^{\mathcal{T}\times N\times D} \mid t = t_1, \ldots, t_B\right\}$ during the iterative process, skewing the gradient distribution towards non-zero regions. This ensures the top-k node selection strategy targets nodes with more non-zero observations. We propose a learning-based re-weighting method using a multi-dimensional attention mechanism, integrating segment, temporal, and spatial attention matrices. The segment attention weight matrix Att_{sa}, inspired by channel attention mechanisms (Hu, Shen, and Sun 2018), is computed as follows:

$$Att_{sg} = \mathcal{C}\left(\sigma\left(g_3^{sg}\left(g_2^{sg}\left(g_1^{sg}\left(\mathrm{Pool}_{\mathcal{T},N}\left(\mathcal{O}\left(\mathcal{X}\right)\right)\right)\right)\right)\right)\right) \quad (14)$$

Here we first perform two-dimensional pooling compression $\text{Pool}_{\mathcal{T},N}$ in the temporal and spatial dimensions to obtain a matrix of shape $(B, 1, 1, D_h)$. Subsequently, we derive a weight matrix of shape $(B, 1, 1, 1)$ based on a three-layer perceptron $(g_1^{sg}(\cdot), g_2^{sg}(\cdot), g_3^{sg}(\cdot))$ and a Sigmoid layer $\sigma(\cdot)$, which represents the significance of different segments. Ultimately, the elements of this weight matrix are replicated and expanded by $\mathcal{C}(\cdot)$ to form a weight matrix of shape $(B, \mathcal{T}, N, D)$, reflecting the weight distribution across the original spatiotemporal gradients. Temporal and spatial attention are also similar to (14). So the final gradients after reweighting can be denoted as:

$$\hat{grad} = Att_1 \circ grad \circ Att_{te} \tag{15}$$

where $Att_1 = Att_{sg} + Att_{sp}$ can be used to correct w_i in (8), and Att_{te} can be used to reweight $\mathcal{G}_i$ in (8). Thus, the new attention-guided spatiotemporal graph adversarial sample generation process can be described as follows:

$$\mathcal{P}(:,:,i,:) = \mathbf{1}_{\mathcal{V}_i \in \text{top}_k\left(|\text{Relu}\left(\hat{grad}\right)\|_2\right)} \tag{16}$$

$$\mathcal{X}'^{(i)} = \text{clip}_\epsilon\left(\mathcal{X}'^{(i-1)} + \alpha \text{sign}\left(\hat{grad} \circ \mathcal{P}\right)\right) \tag{17}$$

Optimization Objective In the context of the performance disparity problem studied in this paper, we hope to strengthen the gradient of the minority samples, so we designed a specific optimization loss to guide the reweighting network. Given the coupled nature of adversarial attacks and the optimization of the reweighting network, we adopt a two-stage iterative strategy for learning. In the first stage, the optimization objective of an adversarial attack is:

$$\underset{\left(\mathcal{X}_t^{\mathcal{T}}\right)'_{\psi^*} \in \mathcal{B}\left(\mathcal{X}_t^{\mathcal{T}}\right)}{\text{argmax}} \sum_{t \in T_{train}} \mathcal{L}\left(f_{\theta^*}\left(\left(\mathcal{X}_t^{\mathcal{T}}\right)'_{\psi^*}\right), \mathcal{Y}_t^{\Delta}\right) \tag{18}$$

In the second stage, the optimization objective of reweighting the network is:

$$\begin{aligned} \underset{\psi}{\text{argmin}} \sum_{t \in T_{train}} & \lambda_1 \mathcal{L}\left(f_{\theta^*}\left(\left(\mathcal{X}_t^{\mathcal{T}}\right)^*_{\psi}\right), \mathcal{Y}_t^{\Delta}\right) \\ & + \lambda_2 MAE\left(\left(\hat{grad}_t^{\mathcal{T}}\right)_+, \left(\hat{grad}_t^{\mathcal{T}}\right)_-\right) \\ & + \lambda_3 \| \left(Att_{1t}^{\mathcal{T}}\right)'_+ \|_2 + \lambda_4 \| \left(Att_{1t}^{\mathcal{T}}\right)_- \|_2 \end{aligned} \tag{19}$$

where $\left(\hat{grad}_t^{\mathcal{T}}\right)_+$ and $\left(\hat{grad}_t^{\mathcal{T}}\right)_-$ respectively represent the spatiotemporal gradients of minority class and majority class. Similarly, $\left(Att_{1t}^{\mathcal{T}}\right)_+$ and $\left(Att_{1t}^{\mathcal{T}}\right)_-$ respectively represent the weight matrices of majority class and minority class.

Uncertainty-Guided Adversarial Contrastive Loss

Previous studies show that feature separability helps mitigate performance degradation in minority classes during adversarial training in imbalanced classification tasks (Wang et al. 2022). In regression tasks, (Zha et al. 2024) highlighted the significance of continuous embeddings consistent with labels for enhancing model robustness and generalization. Moreover, mining hard negative and hard positive samples can effectively enhance the model's discriminative ability for these samples (Liu et al. 2023c). Building on this, we introduce an uncertainty-guided supervised contrastive learning approach. Given the abundance of zero-value regions, we prioritize hard-to-distinguish examples using uncertainty quantification based on parameter decoding (Pu et al. 2016). For the zero-inflated spatiotemporal data, the negative binomial distribution (Jiang et al. 2023b; Zhuang et al. 2022) is a more appropriate fit than the Gaussian assumption implied by RMSE, with its probability mass function defined as:

$$\mathcal{P}_{NB}(x_k; n, p) = \binom{x_k + n - 1}{n - 1} (1-p)^{x_k} p^n \tag{20}$$

where x_k and $n = \frac{\mu\alpha}{1-\alpha}$ are the number of failures and successes respectively, and $p = \frac{1}{1+\mu\alpha}$ is the probability of a single success.

The parameter decoding process based on the negative binomial distribution is as follows:

$$\left(\hat{\mu}_t^{\Delta}, \hat{\alpha}_t^{\Delta}\right) = f_{decoder}\left(h_{target}\left(\mathcal{X}_t^{\mathcal{T}}, \mathcal{A}\right)\right) \tag{21}$$

where $h_{target}\left(\mathcal{X}_t^{\mathcal{T}}, \mathcal{A}\right) = \hat{\mathcal{H}}_t^{\mathcal{T}}$ is the hidden feature embedding calculated by the target model before the output layer. And $\hat{\mu}_t^{\Delta}$ is the mean parameter of the distribution predicted by the decoder network, and $\hat{\alpha}_t^{\Delta}$ is the predicted dispersion parameter. We use the variance parameter $\hat{\alpha}_t^{\Delta}$ predicted by the decoder as an indicator of the difficulty of the region, and combine it with the supervised contrastive loss (Khosla et al. 2020) as a weight value. The final form of the adversarial training loss used in this paper is:

$$\begin{aligned} \mathcal{L}_{adv} = & \beta_1 \sum_{t \in T_{train}} \mathcal{L}_{nb}\left(\hat{\mu}_t^{\Delta}, \hat{\alpha}_t^{\Delta}, \mathcal{Y}_t^{\Delta}\right) \\ & + \beta_2 \sum_{t \in T_{train}} u_t \mathcal{L}_{scl}\left(\hat{\mathcal{H}}_t^{\mathcal{T}}\right) \end{aligned} \tag{22}$$

where $\mathcal{L}_{nb}$ represents the negative log-likelihood loss function based on the negative binomial distribution. And $u_t = \frac{2}{1+e^{-\hat{\alpha}_t^{\Delta}/\gamma}} - 1$ represents the normalized weights based on the uncertainty represented by the predicted variance. And $\mathcal{L}_{scl}\left(\hat{\mathcal{H}}_t^{\mathcal{T}}\right)$ represents the supervised contrastive learning loss function based on the feature embedding (Zhu et al. 2022).

Experiments

Datasets and Baselines

To evaluate the effectiveness of our proposed MinGRE, we conduct experiments on two benchmark datasets, including **NYC** and **Chicago**. The NYC and Chicago datasets contain finely-grained and sparse urban accident data, making them particularly well-suited for studying SGL models under ZID (Wang et al. 2021a). The detailed information on datasets is summarized in Table 1 of the Appendix.

We evaluated the adversarial robustness of our model by comparing it with various attack strategies: **STPGD-Random**, **STPGD-Degree**, **STPGD-Pagerank**, and the state-of-the-art **STPGD-TNDS** from Liu et al. (Liu, Liu,

Dataset	Attacks	Clean		STPGD-TNDS		Clean		STPGD-TNDS	
	Metrics	Rec-maj	Rec-min	Rec-maj	Rec-min	MAP-maj	MAP-min	MAP-maj	MAP-min
NYC	NT-WRMSE	88.182	33.956	87.012	27.416	0.7847	0.1869	0.7580	0.1467
	AT-Random	87.888	32.308	87.543	30.381	0.7808	0.1817	0.7642	0.1591
	AT-Degree	87.857	32.138	87.602	30.710	0.7801	0.1824	0.7683	0.1628
	AT-TNDS	87.586	31.893	87.856	30.974	0.7813	0.1782	0.7701	0.1458
	Ours	**88.189**	**33.992**	**88.191**	**34.004**	**0.7890**	**0.1924**	**0.7891**	**0.1924**
Chicago	NT-WRMSE	94.132	19.261	93.906	16.160	**0.8928**	0.0747	0.8661	0.0618
	AT-Random	94.071	18.426	93.954	16.816	0.8897	0.0890	0.8803	0.0840
	AT-Degree	94.054	18.187	93.989	17.293	0.8895	0.0887	0.8817	0.0868
	AT-TNDS	94.028	17.829	94.006	17.531	0.8898	0.0618	0.8854	0.0566
	Ours	**94.231**	**20.632**	**94.231**	**20.632**	0.8908	**0.0980**	**0.8908**	**0.0981**

Table 1: Evaluation of the robustness of spatiotemporal graph adversarial training techniques based on GSNet. The table provides a detailed analysis of natural and robust performance, with robustness assessed against the STPGD-TNDS attack. The evaluation metrics include Rec-maj, Rec-min, MAP-maj, and MAP-min, with the best results highlighted in bold and the second-best results underlined.

and Jiang 2022). Our method was benchmarked against spatiotemporal adversarial training methods: **AT-Random**, **AT-Degree**, **AT-Pagerank**, and **AT-TNDS** (Liu, Zhang, and Liu 2023). We also examined the effectiveness of different loss functions—**WRMSE** (Wang et al. 2021a), **NBL** (Jiang et al. 2023b), and **BMSE** (Ren et al. 2022)—on target models **GSNet** (Wang et al. 2021a) and **Graph WaveNet** (Wu et al. 2019).

Evaluations

Building on (Wang et al. 2021a), we evaluate model performance from a ranking perspective by calculating recall and precision for majority and minority classes under ZID. We use **Rec-maj**, **Rec-min** to quantify the overlap between predicted and actual zero, non-zero observations. Ranking quality is further assessed using Mean Average Precision (MAP) for the top-k matches (**MAP-maj**, **MAP-min**). Performance disparity is represented by the difference between zero and non-zero observations (**Rec-D**, **MAP-D**). These metrics are commonly employed to gauge accuracy and robustness disparity (Hu et al. 2023; Xu et al. 2021; Ma, Wang, and Liu 2022).

Main Results

We conduct a comprehensive analysis from three perspectives: robustness, performance disparity, and the effectiveness of sub-modules.

Robustness Analysis Table 1 summarizes the natural and robust performance of various spatiotemporal adversarial training methods on the NYC and Chicago datasets. Key insights include: 1) Under the STPGD attack, the NT-WRMSE method shows significant declines in Rec-maj, Rec-min, MAP-maj, and MAP-min on the NYC dataset by approximately 1.3%, 19.3%, 3.4%, and 21.5%, respectively. This highlights the critical need to enhance SGL models' robustness across all classes under ZID scenarios. 2) Our method demonstrates superior robustness, particularly in minority classes, surpassing the state-of-the-art AT-TNDS by approximately 0.4%, 9.8%, 2.5%, and 31.9% in Rec-maj, Rec-min, MAP-maj, and MAP-min on the NYC dataset. While AT-TNDS achieves strong average robustness, it falls short in minority class protection, revealing the limitations of gradient-based victim selection strategies under zero-inflation contexts. The inherent gradient bias (Tan et al. 2021) leads to skewed adversarial examples, impeding uniform robustness enhancement.

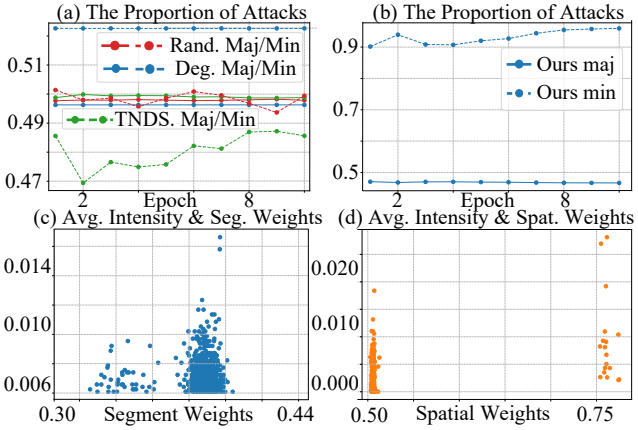

Figure 3: Effectiveness of sub-modules on NYC datasets.

Peformance Disparity Analysis In Table 2, we evaluate the performance disparity of various spatiotemporal adversarial training methods and zero-inflation distribution approaches on the NYC and Chicago datasets, yielding three main conclusions: 1) Spatiotemporal adversarial training, while boosting robustness, often increases the performance disparity between majority and minority classes. For example, AT-TNDS raises Rec-D and MAP-D by 3.7% and 2.1% on the NYC dataset, mainly due to the decline in minority class performance, highlighting the need to address this issue. 2) ZID methods reduce natural performance disparities, as seen in the comparison of NT-NBL and NT-BMSE,

Dataset	Attacks		Natural		STPGD-Random		STPGD-Degree		STPGD-Pagerank		STPGD-TNDS	
	Metrics		Rec-D	MAP-D	Rec-D	MAP-D	Rec-D	MAP-D	Rec-D	MAP-D	Rec-D	MAP-D
NYC	ZID	NT-WRMSE	54.23	0.5978	59.92	0.6188	59.78	0.6184	59.82	0.6182	59.60	0.6113
		NT-NBL	**53.37**	**0.5933**	**52.30**	0.6115	**53.83**	0.6043	54.38	0.6124	54.60	0.6136
		NT-BMSE	53.59	0.5959	54.21	0.5987	54.20	0.5995	54.26	0.5991	54.25	0.5989
	STAT	AT-Random	55.58	0.5991	56.88	0.6040	57.34	0.6010	57.40	0.6034	57.16	0.6051
		AT-Degree	55.72	0.5977	56.61	0.6033	56.93	0.6043	56.89	0.6029	56.89	0.6055
		AT-TNDS	56.21	0.6105	56.43	0.6354	56.32	0.6358	56.37	0.6337	56.43	0.6229
	Ours		54.20	0.5966	54.19	**0.5967**	54.19	**0.5966**	**54.19**	**0.5966**	**54.19**	**0.5967**
Chicago	ZID	NT-WRMSE	74.87	0.8181	77.86	0.8063	77.64	0.8060	77.75	0.8062	77.75	0.8043
		NT-NBL	74.43	0.7953	74.43	0.7953	74.43	0.7953	74.43	0.7953	74.43	0.7953
		NT-BMSE	**72.77**	**0.7920**	73.65	0.8056	**73.54**	0.8025	**73.54**	0.8021	73.71	0.8095
	STAT	AT-Random	75.65	0.8007	76.97	0.7973	76.81	0.7938	76.97	0.7945	77.14	0.7963
		AT-Degree	75.87	0.8008	76.59	0.7956	76.81	0.7947	76.70	0.7947	76.70	0.7949
		AT-TNDS	76.20	0.8280	76.53	0.8289	76.36	0.8287	76.42	0.8305	76.47	0.8288
	Ours		73.60	0.7928	**73.60**	**0.7927**	73.60	**0.7927**	73.60	**0.7928**	**73.60**	**0.7927**

Table 2: Evaluation of the performance disparity of spatiotemporal graph adversarial training techniques based on GSNet. The table provides a detailed analysis of natural and robust performance disparity under various attacks. The evaluation metrics include Rec-D and MAP-D, with the best results highlighted in bold and the second-best results underlined.

though they do not consistently achieve optimal robust performance. On the NYC dataset, these methods sometimes underperform compared to adversarial training in terms of robust disparity. 3) Our method achieves the lowest natural and robust performance disparities, reducing Rec-D and MAP-D by 3.6% and 2.3% on the clean NYC dataset, and by 4.0% and 4.2% on the perturbed NYC dataset, compared to AT-TNDS.

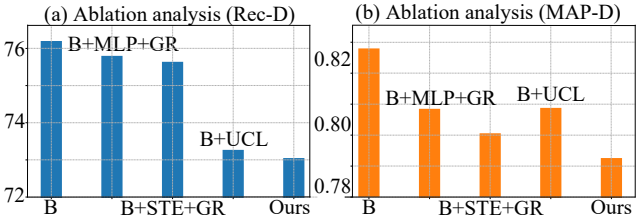

Figure 4: Ablation studies of the proposed Adversarial Examples Generation Module and Uncertainty-guided Contrastive Loss Module on Chicago datasets.

Effectiveness of Sub-Modules To assess the efficacy of our proposed module, we initially visualized the adversarial sample generation process, noting a reduced gradient disparity between minority and majority classes (see Figure 1). This recalibration introduces a greater number of minority samples into adversarial training, contrasting with the AT-TNDS method that included the fewest (see Figure 3). Furthermore, segment and spatial attention matrix analyses revealed that segments with frequent non-zero events of high intensity garnered higher weights (see Figure 3), indicating our mechanism's proficiency in capturing event frequency and intensity. Lastly, random visualizations of the feature space also showed improved separability (see Figure 1), with a slight increase in minority class contour coefficients (from -0.006 to 0.1), reflecting the inherent difficulty in differentiating the data.

Ablation Study

We conduct ablation studies on the Chicago datasets to validate the proposed adversarial example generation and uncertainty-guided contrastive loss (UCL) modules. The baseline model (B) is a spatiotemporal adversarial training method (AT-TNDS) with weighted RMSE loss. STE, GR, and UCL represent the spatiotemporal encoder, gradient reweighting, and loss module, respectively. From Figure 4, we observe three conclusions as follows. 1) Gradient reweighting reduces performance disparity by more effectively selecting minority instances, while the spatiotemporal encoder enhances performance through the capture of cross-segment dependencies. 2) The "B+UCL" variant enhances feature separability, outperforming other methods on Rec-D. 3) Integrating gradient reweighting and UCL achieves the lowest performance disparity, confirming the effectiveness of the proposed modules.

Conclusion

In summary, our study highlights the critical need to address performance disparities in spatiotemporal graph learning under zero-inflated distributions for urban risk management (Jin et al. 2024). We show that traditional adversarial training worsens the performance gap between majority and minority classes, while our MinGRE framework reduces this disparity and improves model robustness. Visualizations and ablation studies confirm MinGRE's effectiveness in recalibrating gradients, enhancing inter-class separability, and accurately capturing non-zero events. These results emphasize MinGRE's potential to advance more equitable and robust models (Sun et al. 2023; Petrović et al. 2022) for urban risk management.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 72434005, Grant 72225011 and Grant 72293575.

References

Chen, C.; Zheng, S.; Chen, X.; Dong, E.; Liu, X. S.; Liu, H.; and Dou, D. 2021. Generalized Data Weighting via Class-Level Gradient Manipulation. In *Proc. of NeurIPS*.

Chen, M.; Yuan, H.; Jiang, N.; Bao, Z.; and Wang, S. 2024. Urban Traffic Accident Risk Prediction Revisited: Regionality, Proximity, Similarity and Sparsity. In *Proc. of CIKM*.

Dobriban, E.; Hassani, H.; Hong, D.; and Robey, A. 2023. Provable Tradeoffs in Adversarially Robust Classification. *IEEE Transactions on Information Theory*.

Feng, C. X. 2021. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *Journal of statistical distributions and applications*.

Gao, X.; Jiang, X.; Haworth, J.; Zhuang, D.; Wang, S.; Chen, H.; and Law, S. 2024. Uncertainty-aware probabilistic graph neural networks for road-level traffic crash prediction. *Accident Analysis & Prevention*.

Gao, X.; Jiang, X.; Zhuang, D.; Chen, H.; Wang, S.; and Haworth, J. 2023. Spatiotemporal Graph Neural Networks with Uncertainty Quantification for Traffic Incident Risk Prediction.

Ghosh, S. K.; Mukhopadhyay, P.; and Lu, J.-C. 2006. Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference*.

Hou, P.; Han, J.; and Li, X. 2023. Improving Adversarial Robustness with Self-Paced Hard-Class Pair Reweighting. In *Proc. of AAAI*.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-Excitation Networks. In *Proc. of CVPR*.

Hu, Y.; Wu, F.; Zhang, H.; and Zhao, H. 2023. Understanding the Impact of Adversarial Robustness on Accuracy Disparity. In *Proc. of ICML*.

Jiang, J.; Wu, B.; Chen, L.; Zhang, K.; and Kim, S. 2023a. Enhancing the Robustness via Adversarial Learning and Joint Spatial-Temporal Embeddings in Traffic Forecasting. In *Proc. of CIKM*.

Jiang, W.; Han, J.; Liu, H.; Tao, T.; Tan, N.; and Xiong, H. 2024. Interpretable Cascading Mixture-of-Experts for Urban Traffic Congestion Prediction. In *Proc. of KDD*.

Jiang, X.; Zhuang, D.; Zhang, X.; Chen, H.; Luo, J.; and Gao, X. 2023b. Uncertainty Quantification via Spatial-Temporal Tweedie Model for Zero-inflated and Long-tail Travel Demand Prediction. In *Proc. of CIKM*.

Jin, G.; Liang, Y.; Fang, Y.; Shao, Z.; Huang, J.; Zhang, J.; and Zheng, Y. 2024. Spatio-Temporal Graph Neural Networks for Predictive Learning in Urban Computing: A Survey. *IEEE Transactions on Knowledge and Data Engineering*.

Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised Contrastive Learning. In *Proc. of NeurIPS*.

Kossen, J.; Band, N.; Lyle, C.; Gomez, A. N.; Rainforth, T.; and Gal, Y. 2021. Self-Attention Between Datapoints: Going Beyond Individual Input-Output Pairs in Deep Learning. In *Proc. of NeurIPS*.

Li, C.; Zhang, B.; Wang, Z.; Yang, Y.; Zhou, X.; Pan, S.; and Yu, X. 2024. Interpretable Traffic Accident Prediction: Attention Spatial–Temporal Multi-Graph Traffic Stream Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*.

Li, J.; Zhang, T.; Jin, S.; Fardad, M.; and Zafarani, R. 2022. AdverSparse: An Adversarial Attack Framework for Deep Spatial-Temporal Graph Neural Networks. In *Proc. of ICASSP*.

Liang, K.; Zhou, S.; Liu, M.; Liu, Y.; Tu, W.; Zhang, Y.; Fang, L.; Liu, Z.; and Liu, X. 2024. Hawkes-Enhanced Spatial-Temporal Hypergraph Contrastive Learning Based on Criminal Correlations. *Proc. of AAAI*.

Lichman, M.; and Smyth, P. 2018. Prediction of Sparse User-Item Consumption Rates with Zero-Inflated Poisson Regression. In *Proc. of WWW*.

Liu, F.; Liu, H.; and Jiang, W. 2022. Practical Adversarial Attacks on Spatiotemporal Traffic Forecasting Models. In *Proc. of NeurIPS*.

Liu, F.; Tian, J.; Miranda-Moreno, L.; and Sun, L. 2024. Adversarial Danger Identification on Temporally Dynamic Graphs. *IEEE Transactions on Neural Networks and Learning Systems*.

Liu, F.; Zhang, W.; and Liu, H. 2023. Robust Spatiotemporal Traffic Forecasting with Reinforced Dynamic Adversarial Training. In *Proc. of KDD*.

Liu, H.; Dong, Z.; Jiang, R.; Deng, J.; Deng, J.; Chen, Q.; and Song, X. 2023a. Spatio-Temporal Adaptive Embedding Makes Vanilla Transformer SOTA for Traffic Forecasting. In *Proc. of CIKM*.

Liu, X.; Zhang, Z.; Lyu, L.; Zhang, Z.; Xiao, S.; Shen, C.; and Yu, P. S. 2023b. Traffic Anomaly Prediction Based on Joint Static-Dynamic Spatio-Temporal Evolutionary Learning. *IEEE Transactions on Knowledge and Data Engineering*.

Liu, Y.; Yang, X.; Zhou, S.; Liu, X.; Wang, Z.; Liang, K.; Tu, W.; Li, L.; Duan, J.; and Chen, C. 2023c. Hard Sample Aware Network for Contrastive Deep Graph Clustering. In *Proc. of AAAI*.

Liu, Z.; Chen, Y.; Xia, F.; Bian, J.; Zhu, B.; Shen, G.; and Kong, X. 2023d. TAP: Traffic Accident Profiling via Multi-Task Spatio-Temporal Graph Representation Learning. *ACM Transactions on Knowledge Discovery from Data*.

Ma, X.; Wang, Z.; and Liu, W. 2022. On the Tradeoff Between Robustness and Fairness. In *Proc. of NeurIPS*.

Petrović, A.; Nikolić, M.; Radovanović, S.; Delibašić, B.; and Jovanović, M. 2022. FAIR: Fair adversarial instance reweighting. *Neurocomputing*.

Pu, Y.; Gan, Z.; Henao, R.; Yuan, X.; Li, C.; Stevens, A.; and Carin, L. 2016. Variational Autoencoder for Deep Learning of Images, Labels and Captions. In *Proc. of NeurIPS*.

Qaraei, M.; and Babbar, R. 2022. Adversarial Examples for Extreme Multilabel Text Classification. *Machine Learning*.

Ren, J.; Zhang, M.; Yu, C.; and Liu, Z. 2022. Balanced mse for imbalanced visual regression. In *Proc. of CVPR*.

Sun, C.; Xu, C.; Yao, C.; Liang, S.; Wu, Y.; Liang, D.; Liu, X.; and Liu, A. 2023. Improving Robust Fariness via Balance Adversarial Training. *Proc. of AAAI*.

Tan, J.; Lu, X.; Zhang, G.; Yin, C.; and Li, Q. 2021. Equalization Loss v2: A New Gradient Balance Approach for Long-Tailed Object Detection. In *Proc. of CVPR*.

Tang, J.; Xia, L.; and Huang, C. 2023. Explainable Spatio-Temporal Graph Neural Networks. In *Proc. of CIKM*.

Trirat, P.; Yoon, S.; and Lee, J.-G. 2023. MG-TAR: Multi-View Graph Convolutional Networks for Traffic Accident Risk Prediction. *IEEE Transactions on Intelligent Transportation Systems*.

Uesato, J.; O'Donoghue, B.; Kohli, P.; and van den Oord, A. 2018. Adversarial Risk and the Dangers of Evaluating Against Weak Attacks. In *Proc. of ICML*.

Wang, B.; Lin, Y.; Guo, S.; and Wan, H. 2021a. GSNet: Learning Spatial-Temporal Correlations from Geographical and Semantic Aspects for Traffic Accident Risk Forecasting. In *Proc. of AAAI*.

Wang, Q.; Wang, S.; Zhuang, D.; Koutsopoulos, H.; and Zhao, J. 2024. Uncertainty Quantification of Spatiotemporal Travel Demand With Probabilistic Graph Neural Networks. *IEEE Transactions on Intelligent Transportation Systems*.

Wang, S.; Zhang, J.; Li, J.; Miao, H.; and Cao, J. 2023. Traffic Accident Risk Prediction via Multi-View Multi-Task Spatio-Temporal Networks. *IEEE Transactions on Knowledge and Data Engineering*.

Wang, W.; Xu, H.; Liu, X.; Li, Y.; Thuraisingham, B.; and Tang, J. 2022. Imbalanced Adversarial Training with Reweighting. In *Proc. of ICDM*.

Wang, Z.; Jiang, R.; Cai, Z.; Fan, Z.; Liu, X.; Kim, K.-S.; Song, X.; and Shibasaki, R. 2021b. Spatio-Temporal-Categorical Graph Neural Networks for Fine-Grained Multi-Incident Co-Prediction. In *Proc. of CIKM*.

Wilson, T.; McDonald, A.; Galib, A. H.; Tan, P.-N.; and Luo, L. 2022. Beyond Point Prediction: Capturing Zero-Inflated & Heavy-Tailed Spatiotemporal Data with Deep Extreme Mixture Models. In *Proc. of KDD*.

Wölker, Y.; Beth, C.; Renz, M.; and Biastoch, A. 2023. SUSTeR: Sparse Unstructured Spatio Temporal Reconstruction on Traffic Prediction. In *Proc. of SIGSPATIAL*.

Wu, M.; Jia, H.; Luo, D.; Luo, H.; Zhao, F.; and Li, G. 2023. A multi-attention dynamic graph convolution network with cost-sensitive learning approach to road-level and minute-level traffic accident prediction. *IET Intelligent Transport Systems*.

Wu, T.; Liu, Z.; Huang, Q.; Wang, Y.; and Lin, D. 2021. Adversarial Robustness Under Long-Tailed Distribution. In *Proc. of CVPR*.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; and Zhang, C. 2019. Graph wavenet for deep spatial-temporal graph modeling. In *Proc. of IJCAI*.

Xiong, P.; Tegegn, M.; Sarin, J. S.; Pal, S.; and Rubin, J. 2024. It Is All about Data: A Survey on the Effects of Data on Adversarial Robustness. *ACM Computing Surveys*.

Xu, H.; Liu, X.; Li, Y.; Jain, A.; and Tang, J. 2021. To be Robust or to be Fair: Towards Fairness in Adversarial Training. In *Proc. of ICML*.

Yamamoto, T.; Hashiji, J.; and Shankar, V. N. 2008. Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. *Accident Analysis & Prevention*.

Yue, X.; Mou, N.; Wang, Q.; and Zhao, L. 2024. Revisiting Adversarial Training Under Long-Tailed Distributions. In *Proc. of CVPR*.

Zha, K.; Cao, P.; Son, J.; Yang, Y.; and Katabi, D. 2024. Rank-N-contrast: learning continuous representations for regression. In *Proc. of NeurIPS*.

Zhang, Q.; Huang, C.; Xia, L.; Wang, Z.; Yiu, S. M.; and Han, R. 2023. Spatial-Temporal Graph Learning with Adversarial Contrastive Adaptation. In *Proc. of ICML*.

Zhang, X.; Zheng, X.; and Mao, W. 2021. Adversarial Perturbation Defense on Deep Neural Networks. *ACM Computing Surveys*.

Zhao, S.; Zhao, D.; Liu, R.; Xia, Z.; Chen, B.; and Chen, J. 2023. GMAT-DU: Traffic Anomaly Prediction With Fine Spatiotemporal Granularity in Sparse Data. *IEEE Transactions on Intelligent Transportation Systems*.

Zhou, Z.; Shi, J.; Zhang, H.; Chen, Q.; Wang, X.; Chen, H.; and Wang, Y. 2024. CreST: A Credible Spatiotemporal Learning Framework for Uncertainty-aware Traffic Forecasting. In *Proc. of WSDM*.

Zhou, Z.; Wang, Y.; Xie, X.; Chen, L.; and Zhu, C. 2022. Foresee Urban Sparse Traffic Accidents: A Spatiotemporal Multi-Granularity Perspective. *IEEE Transactions on Knowledge and Data Engineering*.

Zhu, J.; Wang, Z.; Chen, J.; Chen, Y.-P. P.; and Jiang, Y.-G. 2022. Balanced Contrastive Learning for Long-Tailed Visual Recognition. In *Proc. of CVPR*.

Zhu, L.; Feng, K.; Pu, Z.; and Ma, W. 2024. Adversarial Diffusion Attacks on Graph-Based Traffic Prediction Models. *IEEE Internet of Things Journal*.

Zhuang, D.; Bu, Y.; Wang, G.; Wang, S.; and Zhao, J. 2024. SAUC: Sparsity-Aware Uncertainty Calibration for Spatiotemporal Prediction with Graph Neural Networks. In *Proc. of SIGSPATIAL*.

Zhuang, D.; Wang, S.; Koutsopoulos, H.; and Zhao, J. 2022. Uncertainty Quantification of Sparse Travel Demand Prediction with Spatial-Temporal Graph Neural Networks. In *Proc. of KDD*.