OHATP: Graph Anomaly Detection with Orthogonal-Hyperspherical Augmentation and Topology Perception

Yihang Qiu[1] **Yi Zhang**[1] **Di Xiong**[1] **Ge Gao**[1] **Shuo Chen**[1,2]*
[1]School of Intelligence Science and Technology, Nanjing University, China
[2]Center for Advanced Intelligence Project, RIKEN, Japan
yihang.qiu@smail.nju.edu.cn shuo.chen@nju.edu.cn

Abstract

Contrastive Learning (CL) has been widely used for Graph Anomaly Detection (GAD). However, existing augmentation techniques in such CL frameworks usually rely on random masking or structural alterations, which obliterates the irregularities that define anomalies and leads to *anomaly distortion*. Even improved methods using *multi-scale sampling* usually suffer from *blind contextual extraction*, which risks disrupting critical topological structures or introducing irrelevant information in the learned representations, thereby hardly capturing complex anomalies. In this paper, we propose Orthogonal-Hyperspherical Augmentation and Topology Perception (**OHATP**), a novel method that leverages the synergies at both the *feature* and *structure* levels to mitigate anomaly distortion and *blind sampling defects*. Specifically, during *feature-level* augmentation, we adapt projections and noise based on node *message-passing* influence to strengthen *feature diversity*, while utilizing projection orthogonality and hyperspherical noise constraints to preserve *feature discriminability*. To complement this feature-level augmentation, we design a *structure-level* topology perception module that leverages the high-quality representations learned from augmentation to screen abnormal dense substructures. Extensive experiments on benchmark datasets demonstrate that OHATP substantially outperforms state-of-the-art methods, achieving improvements of over 6% in AUROC and 29% in AUPRC. Code is available at https://github.com/tyl419/OHATP.

1 Introduction

Graph Anomaly Detection (GAD) [1, 2, 3, 4, 5, 6, 7, 8], which aims to identify instances deviating from normal data [9, 10, 11], is widely applied in graph data mining [12, 13, 14, 15, 16, 17, 18]. Unlike the traditional domain-specific data [19, 20, 21], graphs include node feature and topological structure, where their inconsistency induces *attribute* (feature-neighborhood mismatch) and *topological* (dissimilar but densely connected nodes) anomalies [22].

Research efforts have been made to address the GAD task. Traditional methods rely on feature engineering or pattern recognition techniques, failing to capture deep nonlinear graph information effectively [23, 24, 25]. With the advancement of deep learning [26], Graph Neural Networks (GNNs) have been widely applied to GAD [27, 28, 29]. Among them, Graph Contrastive Learning (GCL) has achieved success by maximizing positive-pair agreement while pushing apart negative pairs [30, 31, 32, 33, 34]. The pioneering approach CoLA [13] identifies anomalies by contrasting target nodes with their corresponding local subgraphs. This framework has been advanced via multi-view

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

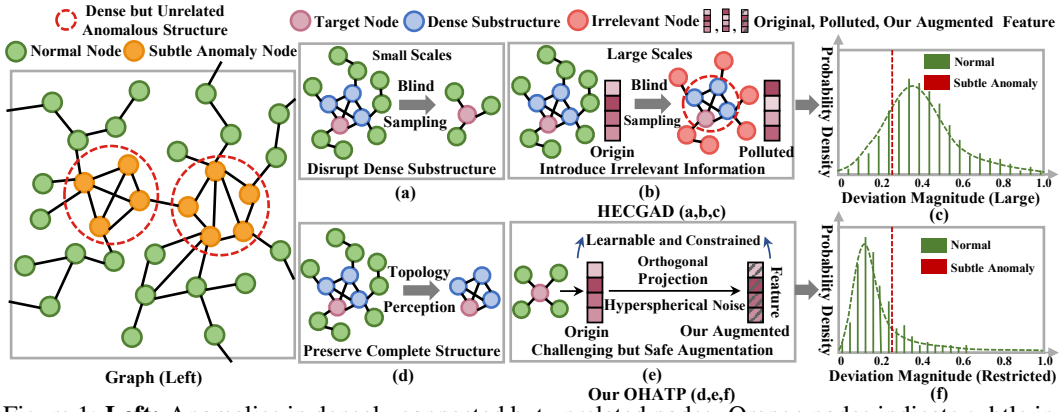

Figure 1: **Left:** Anomalies in densely connected but unrelated nodes. Orange nodes indicate subtle individual anomalies, and red circles denote anomalous combinations. **Right:** Blind-sampling methods like HECGAD [40] can disrupt key topological structures (a) or introduce irrelevant information (b), causing large embedding deviations and masking subtle anomaly signals (c). By contrast, our OHATP retains these structures via topology perception (d), while orthogonal-hyperspherical augmentation further restricts deviations (e), preserving anomaly signals (f).

contrastive objectives [11, 35, 36, 37], and refined by GADMCLG [38] through global and local information incorporation. Generally, the success of these methods hinges on generating effective contrasting views via strategies, especially *graph augmentations*, to expose *anomaly signals* [39].

However, simply transferring standard augmentations to GAD is usually problematic. These methods assume semantic invariance to perturbations, treating structural or attribute variations as noise. Although this enforces robust representation learning, such indiscriminate perturbations usually obliterate the irregularities that define anomalies, causing anomaly distortion. For instance, edge perturbation alters topological connections, masking topological anomalies; random masking degrades discriminability between normal and anomalous nodes, affecting attribute anomaly detection. While approaches like HECGAD [40] use multi-scale subgraph sampling to avoid this, they usually suffer from blind contextual extraction. This is detrimental for topological anomalies formed by densely connected but unrelated nodes with subtle individual anomaly signals (Fig. 1left). Specifically, small sampling scales disrupt critical topological structures (Fig. 1a), while large scales introduce irrelevant information (Fig. 1b), causing large *embedding deviations* that mask these subtle signals (Fig. 1c). Therefore, an ideal scheme should introduce necessary diversity without distorting anomalies, preserve critical dense substructures and restrict embedding deviations.

To this end, we propose Orthogonal-Hyperspherical Augmentation and Topology Perception (**OHATP**), aiming at leveraging the synergies at both the feature and structure levels to mitigate anomaly distortion and blind sampling defects. Specifically, during feature-level augmentation, we adapt *orthogonal projections* and uniform noise based on each node's message-passing influence, applying stronger perturbations to nodes with low connectivity and influence, to strengthen feature diversity required for contrastive learning. To reduce embedding deviations and preserve feature discriminability, the *projection orthogonality* maintains feature norms to avoid unintended loss or amplification of feature information, while *hyperspherical constraints* prevent excessive noise from polluting neighbors during graph propagation. This synergy introduces challenging yet safe perturbations without distorting anomalies (Fig. 1e) to learn more robust representations, preserving anomaly signals (Fig. 1f). Then, we leverage the augmented views to jointly optimize multi-view contrastive learning and node attribute reconstruction. Meanwhile, to further complement this feature-level augmentation, we design a structure-level topology perception module that extracts dense substructures (Fig. 1d) and detects anomalies based on the high-quality representations learned via our augmentation strategy. By maximally preserving these dense substructures, it effectively mitigates blind sampling defects. Finally, thanks to fusing representation learning with orthogonal-hyperspherical augmentation and topology perception, OHATP can effectively capture complex anomalies.

Accordingly, the major contributions of this paper can be summarized as follows:

- To avoid anomaly distortion caused by existing graph augmentations, we propose the orthogonal-hyperspherical augmentation that applies learnable perturbations with geometric constraints to generate contrasting views, preserving anomaly discriminability.

- To mitigate blind sampling defects and complement our new augmentation technique, we propose a topology perception module that extracts dense substructures via topology screening and detects anomalies based on the learned high-quality representations.
- Extensive experimental results on six datasets demonstrate that our approach OHATP substantially outperforms state-of-the-art GAD methods, achieving over 6% and 29% improvements in AUROC and AUPRC respectively compared with the best baselines.

2 Related Work and Preliminaries

2.1 Graph Contrastive Learning

Graph Contrastive Learning (GCL) drives the model to learn high-quality representations by pulling positive pairs closer and pushing negative pairs away. Existing methods mainly fall into two categories: negative-based and negative-free. Negative-based approaches maximize mutual information [41, 42, 43] or optimize node embeddings via node-subgraph contrast [44, 45] or hard negative sampling [46, 47, 48]. To reduce computational costs, negative-free methods abandon negative sampling, utilizing asymmetric architectures [49, 50] or specialized loss functions [51, 52] for robust learning.

2.2 Graph Anomaly Detection

GAD aims to identify anomalous instances deviating from normal data. Existing approaches fall into traditional and deep learning-based methods. Traditional techniques [24, 25, 53, 54, 55] rely on manual feature engineering or expert-designed statistical models to identify anomalous instances. However, these methods are labor-intensive and fail to capture deep nonlinear information.

With the advancement of deep learning, GNNs have become a widely adopted paradigm for GAD [56] by extracting both structural and attribute information from graphs to detect anomalies. Beyond generative methods relying on reconstruction errors [57, 58, 59, 60], Graph Contrastive Learning (GCL) has become another mainstream paradigm in this field. Following the pioneering node-subgraph contrast in CoLA [13], subsequent methods enhance representations through multi-scale signals [35, 36, 61], hybrid strategies [11, 27, 28, 29, 62], or adaptive message passing [63]. However, a core challenge in GCL is constructing effective contrastive views. Many approaches [37, 38, 64] rely on traditional graph augmentations that rely on random masking or structural alterations, leading to anomaly distortion. To avoid this, some works draw inspiration from semi-supervised paradigms to generate more reliable training signals, such as by creating pseudo-anomalies [65, 66] or enforcing consistency training via learnable perturbations [67], which rely on labeled data or heuristics, limiting their capability in unsupervised settings. Another approach attempts to replace destructive augmentations with multi-scale subgraph sampling [40, 68]. These strategies are usually blind, which disrupt critical structures or introduce irrelevant information in the learned representation. By contrast, our method integrates both feature and structure levels to address these issues, introducing learnable feature augmentation with geometric constraints to preserve anomaly signals and optimize representation learning, and preserving dense substructures via the topology perception module.

2.3 Task Definition

We focus on graph anomaly detection within an undirected attributed graph denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, where $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ denotes the set of N nodes and $\mathcal{E} = \{e_{ij}\}_{N\times N}$ is the set of edges. $e_{ij} = (v_i, v_j) \in \mathcal{E}$ indicates an edge between node v_i and v_j. The adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ represents the graph structure, where $A_{ij} = 1$ if and only if $e_{ij} \in \mathcal{E}$, otherwise $A_{ij} = 0$. The node attributes are defined as $\mathbf{X} \in \mathbb{R}^{N\times D}$, where D denotes the dimension of the node attributes, and $x_i \in \mathbb{R}^D$ is the attribute vector of node v_i.

In the context of graph anomaly detection, the node set $\mathcal{V}$ consists of a majority set of normal nodes $\mathcal{V}_{\text{norm}}$ that strictly follow underlying attribute-structural distribution patterns, and a small set of anomalous nodes $\mathcal{V}_{\text{anom}}$ ($|\mathcal{V}_{\text{anom}}| \ll |\mathcal{V}_{\text{norm}}|$) that deviate from these patterns. The goal of GAD is to learn an anomaly scoring function $\mathcal{F}(\cdot) : \mathcal{V} \to \mathbb{R}$. This function quantifies the degree of deviation from the normal patterns, satisfying the ranking inequality $\mathcal{F}(v_i) > \mathcal{F}(v_j)$ for any $v_i \in \mathcal{V}_{\text{anom}}$ and $v_j \in \mathcal{V}_{\text{norm}}$. A higher value of $\mathcal{F}(v_i)$ indicates a significant violation of the learned consistency patterns, implying a higher probability of being anomalous.

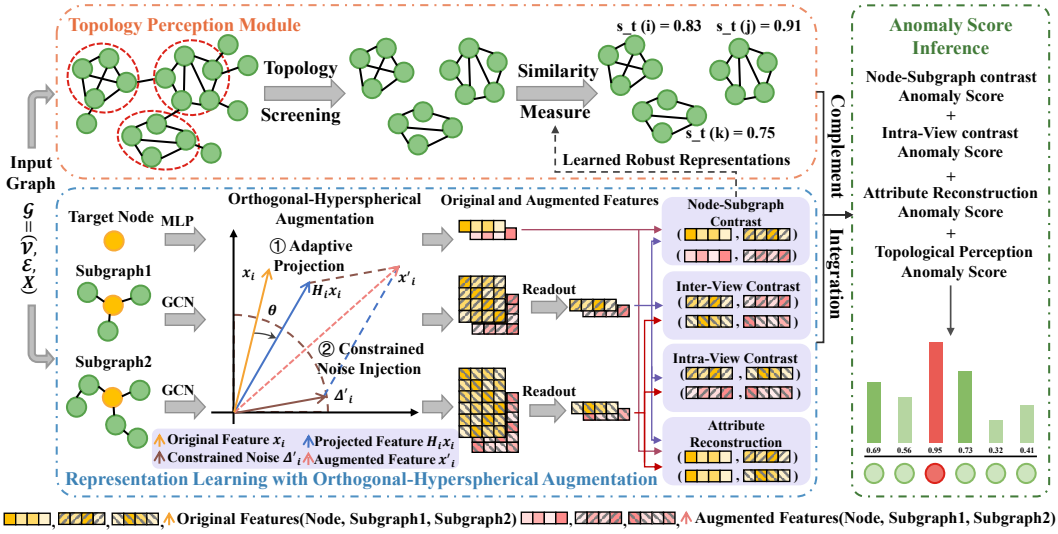

Figure 2: Overview of OHATP. Orthogonal-hyperspherical augmentation generates contrastive views to optimize multi-view contrastive learning and attribute reconstruction, while the topology perception module screens abnormal dense substructures. The final anomaly score fuses these metrics.

3 Methodology

In this section, we detail our OHATP (Fig. 2). First, via orthogonal-hyperspherical augmentation, we generate contrastive views to optimize multi-view contrastive learning and node reconstruction. Then, we adopt a topology perception module to isolate dense substructures, capturing topological anomalies. The final node anomaly score is derived by integrating these metrics.

3.1 Orthogonal-Hyperspherical Augmentation

To avoid anomaly distortion, we propose an orthogonal-hyperspherical augmentation synergizing adaptive projection with constrained noise. For node i with D-dimensional embedding $\mathbf{x}_i$, we adopt a learnable projection matrix $\mathbf{P}_i \in \mathbb{R}^{D\times D}$ and orthogonalize it via the Cayley transform [69]:

$$\mathbf{H}_i = \left(\mathbf{I} + \frac{\mathbf{P}_i - \mathbf{P}_i^\top}{2}\right)\left(\mathbf{I} - \frac{\mathbf{P}_i - \mathbf{P}_i^\top}{2} + \xi\mathbf{I}\right)^{-1}, \tag{1}$$

where $\mathbf{I}$ is the identity matrix, and $\xi \in (0,1)$ is a small positive regularization coefficient to ensure invertibility. The Cayley transform ensures the orthogonality of $\mathbf{H}_i$ [69] to maintain the feature vector norm, avoiding unintended loss or amplification of feature information. This adaptive projection can yield diverse node representations while preserving core information, enhancing the diversity of contrastive views. Initially, $\mathbf{P}_i$ has near-zero entries, making the projection negligible. During training, it is learned based on a node's message-passing influence to adaptively apply larger projection angles to nodes with low connectivity and influence. To further enhance robustness, we introduce *dimension-adaptive noise weights* to inject constrained random uniform noise:

$$\mathbf{x}'_i = \mathbf{H}_i\mathbf{x}_i + \mu\cdot\Delta'_i = \mathbf{H}_i\mathbf{x}_i + \mu\cdot\epsilon\cdot\frac{\omega\odot\operatorname{sign}(\mathbf{H}_i\mathbf{x}_i)}{\|\omega\odot\operatorname{sign}(\mathbf{H}_i\mathbf{x}_i)\|_2}, \quad \omega\sim U(0,1)^D \tag{2}$$

where $\mathbf{x}_i, \Delta'_i \in \mathbb{R}^D$, $\odot$ is the Hadamard product and $\operatorname{sign}(\cdot)$ is the sign function. The learnable noise intensity $\mu \in (0,1)$ is adjusted based on each dimension's message-passing contribution, applying smaller perturbations to critical dimensions. The injected noise vector Δ'_i is constrained by $\|\Delta'_i\|_2 = \epsilon$ with $\epsilon \in (0,1]$ to ensure that Δ'_i is numerically equivalent to a point on a hypersphere of radius ϵ, controlling the size of the perturbation. The direction constraint requires that $\mathbf{H}_i\mathbf{x}_i$ and Δ'_i reside within the same *hyper-octant* to preserve the sign structure of the projected representation. These constraints reduce embedding deviations that mask anomaly signals or adversely affect neighborhood message passing, preventing the misclassification of the node and its neighbors.

Fig. 3 illustrates our augmentation in $\mathbb{R}^2$. Specifically, to enhance feature diversity, projections and noise adaptively apply perturbations based on nodes' message-passing influence. To restrict deviations, projection orthogonality preserves feature information, while hyperspherical constraints prevent neighbor pollution. This synergy introduces necessary diversity while confining it to a valid semantic neighborhood, enabling robust representation learning without distorting anomalies. We adopt uniform noise over Gaussian for comprehensive hyper-octant exploration. Node information is then propagated via a GCN with layer-wise noise injection, where the perturbed representation is computed as:

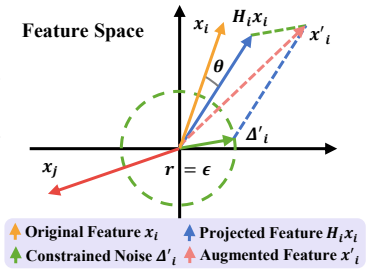

Figure 3: Illustration of orthogonal-hyperspherical augmentation in $\mathbb{R}^2$.

$$\tilde{\mathbf{X}}_i^{(l-1)} = \mathbf{H}_i \mathbf{X}_i^{(l-1)} + \mu \cdot \Delta_i^{(l-1)}, \tag{3}$$

where $\Delta_i^{(l)}$ is the layer-l random perturbation, and $\mathbf{X}_i^{(0)}$ is the initial node feature. These augmented features are propagated via standard GCN message-passing to produce updated node embeddings:

$$\mathbf{X}_i^{(l)} = \sigma\left(\hat{\mathbf{A}}_i \tilde{\mathbf{X}}_i^{(l-1)} \mathbf{W}^{(l-1)}\right). \tag{4}$$

Here, $\hat{\mathbf{A}}_i$ represents the normalized adjacency matrix with self-loops. $\mathbf{W}^{(l-1)}$ refers to the model parameter weight matrix, and $\sigma(\cdot)$ is the ReLU activation function.

3.2 Contrastive Learning Network and Attribute Reconstruction

This section aims to learn robust node representations. For node v_i, we generate subgraphs $\mathcal{G}_i^{(1)}$ and $\mathcal{G}_i^{(2)}$ by sampling its 1-hop and 2-hop neighborhoods with different scales ($K_1 \neq K_2$). Based on these views, we conduct contrastive learning and attribute reconstruction.

3.2.1 Contrastive Learning Network

Node-Subgraph Contrast We treat node v_i and its subgraphs as positives, and other subgraphs as negatives. With node features $\mathbf{h}_i$ and subgraph embedding $\mathbf{z}_i$, the similarities of positives (s_i^+) and negatives (s_i^-) are computed as: $s_i^+ = \text{sigmoid}\left(\mathbf{z}_i^T \mathbf{W} \mathbf{h}_i\right)$, and optimized via the BCE loss [41]:

$$\mathcal{L}_{NS}^{(w)}(v_i) = -\frac{1}{2}\left(\log\left(s_i^{(w)+}\right) + \log\left(1 - s_i^{(w)-}\right)\right), \tag{5}$$

where w denotes the sampled subgraphs for each node. The final node-subgraph loss combines both original ($\mathcal{O}$) and augmented ($\mathcal{A}$) views, balanced by a hyperparameter $\alpha \in (0, 1)$:

$$\mathcal{L}_{NS} = \frac{1}{2}\left(\alpha \sum_{w=1}^{2} \mathcal{L}_{NS}^{(w),\mathcal{O}} + (1-\alpha) \sum_{w=1}^{2} \mathcal{L}_{NS}^{(w),\mathcal{A}}\right). \tag{6}$$

During inference, the node-subgraph anomaly score $score_{\text{NS}}(v_i)$ is defined as the similarities margin: $s^-(v_i) - s^+(v_i)$. Following the loss design, the final score fuses two views, balanced by α.

Inter-View Contrast. To ensure noise robustness, inter-view contrast maximizes agreement between a node's subgraphs across original and augmented views, while pushing away subgraphs from different nodes as negative pairs. We optimize this objective using the NCE loss [70]:

$$\mathcal{L}_{NCE}^{(w)} = -\sum_{i=1}^{N} \log \frac{\exp\left(\langle \boldsymbol{z}_i^{\mathcal{O}}, \boldsymbol{z}_i^{\mathcal{A}} \rangle\right)}{\sum_{j=1, j\neq i}^{N} \left(\exp\left(\langle \boldsymbol{z}_i^{\mathcal{O}}, \boldsymbol{z}_j^{\mathcal{O}} \rangle\right) + \exp\left(\langle \boldsymbol{z}_i^{\mathcal{O}}, \boldsymbol{z}_j^{\mathcal{A}} \rangle\right)\right)}. \tag{7}$$

Here, $\mathbf{z}_i^{\mathcal{O}}$ and $\mathbf{z}_i^{\mathcal{A}}$ are the original and augmented subgraph representations for node v_i. N is the number of nodes. The total inter-view loss is the sum of NCE losses from the sampled subgraphs.

Intra-View Contrast. Intra-view contrast aims to align subgraphs $\mathcal{G}_i^{(1)}$ and $\mathcal{G}_i^{(2)}$ via *graph-EMD*. We first define a cost matrix combining attribute and topological distance, with entry $\mathbf{M}_{pq}$ defined as:

$$\mathbf{M}_{pq} = \left(1 - \frac{\boldsymbol{h}_p^T \boldsymbol{h}_q}{\|\boldsymbol{h}_p\|_2 \|\boldsymbol{h}_q\|_2}\right) \cdot \frac{1}{1 + \exp\left(-\Theta_{pq}/\delta\right)}, \tag{8}$$

where $\boldsymbol{h}_p$ and $\boldsymbol{h}_q$ denote the p-th node representation in $\tilde{\mathcal{G}}_i^{(1)}$ and the q-th node representation in $\tilde{\mathcal{G}}_i^{(2)}$, and Θ is their smallest hop count modulated by the temperature $\delta \geq 1$. To reduce computational complexity, we approximate the EMD via the Sinkhorn algorithm [71] based on cost matrix $\mathbf{M}$:

$$\text{g-EMD}\big(\tilde{\mathcal{G}}^{(1)}, \tilde{\mathcal{G}}^{(2)}\big) = \phi\big(\inf_{\mathbf{\Gamma}\in\mathbf{\Pi}} \langle \mathbf{\Gamma}, \mathbf{M}\rangle_{\mathrm{F}} + \frac{1}{\lambda}\sum_{i,j} \mathbf{\Gamma}_{ij}\big(\log \mathbf{\Gamma}_{ij} - 1\big)\big), \tag{9}$$

where $\langle \cdot, \cdot \rangle_F$ is the Frobenius inner product and λ is a regularization parameter. $\mathbf{\Gamma} \in \mathbf{\Pi}$ is a transport matrix over valid distributions $\mathbf{\Pi}$ and ϕ normalizes g-EMD to $[0, 1]$. To reduce overhead, we approximate $\mathbf{\Gamma}$ and g-EMD via Sinkhorn, yielding the similarity $s_{ii}^{(12)} = \text{sim}\left(\tilde{\mathcal{G}}_i^{(1)}, \tilde{\mathcal{G}}_i^{(2)}\right) = 1 - \text{g-EMD}\left(\tilde{\mathcal{G}}_i^{(1)}, \tilde{\mathcal{G}}_i^{(2)}\right)$. We then define the intra-view contrast loss via InfoNCE [72, 73]:

$$\mathcal{L}_{SS}^{(1)}\left(\tilde{\mathcal{G}}_i^{(1)}, \tilde{\mathcal{G}}_i^{(2)}\right) = -\log\left(\frac{\exp\left(s_{ii}^{(12)}/\tau\right)}{\sum_{j=1}^{N}\exp\left(s_{ij}^{(12)}/\tau\right) + \sum_{j=1}^{N}\mathbf{1}_{[j\neq i]}\exp\left(s_{ij}^{(11)}/\tau\right)}\right), \tag{10}$$

where $\tau > 0$ is the temperature, $\mathbf{1}_{[j\neq i]}$ is the indicator, and N is the node count. $\mathcal{L}_{SS}^{(2)}\left(\tilde{\mathcal{G}}_i^{(2)}, \tilde{\mathcal{G}}_i^{(1)}\right)$ can be computed similarly. To reduce false negatives, we sample top-k pairs with the largest g-EMD as hard negatives. Following Eq. (6), the overall intra-view loss $\mathcal{L}_{SS}$ combines original and augmented views, balanced by $\alpha \in (0, 1)$. At inference, the inter-view anomaly score is defined as the negative similarity of positives: $score_{\text{SS}}(v_i) = -\frac{1}{2}\left(s_{ii}^{(12)} + s_{ii}^{(21)}\right)$, with the final score aggregated as in $\mathcal{L}_{SS}$.

3.2.2 Attribute Reconstruction Learning

To further boost performance, we incorporate a node attribute reconstruction task. Unlike normal nodes, which align with their neighborhoods and are easily reconstructed, anomalies deviate and incur higher *reconstruction errors*. Specifically, we employ an MLP decoder to reconstruct the central node v_i from its neighbors' latent embeddings, with the loss defined as:

$$\mathcal{L}_{\text{recon}}^{(w)} = \frac{1}{N}\sum_{i=1}^{N}\left\|\text{MLP}^{(w)}\left(\boldsymbol{g}_i^{(w)}\right) - \mathbf{X}[i,:]\right\|_2^2. \tag{11}$$

Here, $\boldsymbol{g}_i^{(w)}$ concatenates unmasked neighbor representations for subgraph w and $\mathbf{X}[i,:]$ is v_i's original feature. The objective averages the loss over all subgraphs. At inference, the reconstruction score $score_{\text{recon}}(v_i)$ is computed by averaging the errors from Eq. (11) across subgraphs. Finally, we jointly optimize the overall loss with $\beta, \gamma \in (0, 1)$ to balance the contrastive and reconstruction modules:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{NS}} + \beta(\mathcal{L}_{\text{SS}} + \mathcal{L}_{\text{NCE}}) + \gamma\mathcal{L}_{\text{recon}}. \tag{12}$$

3.3 Topology Perception Module

Dense Substructure Isolation and Anomaly Evaluation. Topological anomalies manifest as dense clusters of *semantically unrelated nodes*. Though exhibiting subtle anomalies individually, they collectively form highly anomalous structures that blind sampling strategies overlook. As shown in Fig. 2, to mitigate blind sampling defects and complement our feature-level augmentation, we propose a structure-level topology perception module, whose workflow is illustrated in Fig. 4. Specifically, we leverage k-*core* decomposition [74, 75] as the topology screening algorithm to isolate dense substructures.

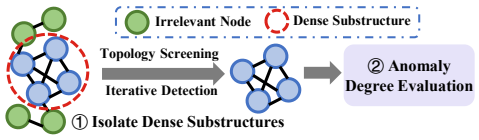

Figure 4: Illustration of topology perception.

However, normal nodes can also establish dense connections in certain scenarios. To distinguish them, we evaluate each substructure's anomaly level by measuring its internal feature consensus. Under the *homophily* assumption, normal cliques exhibit consistent semantics, while anomalous cliques display feature divergence. Thus, we measure this using the average cosine similarity of internal node embeddings, where lower similarity implies a higher anomaly likelihood. The similarity is defined as:

$$s_t^{(w)} = \frac{1}{|C|}\sum_{n=1}^{|C|}\cos(z_i, z_j)^{(n)}. \tag{13}$$

In this case, $|C|$ denotes the total number of node pairs within the identified substructure and w represents the sampled subgraphs for each node.

Iterative Detection and Joint Evaluation. In practice, due to lacking prior knowledge of anomalous structures, we execute an *iterative detection mechanism*, progressively raising the core threshold k until no dense substructures are identified. Given the negative correlation between internal node-pair similarity and anomaly levels, we measure the abnormality of a substructure using the inverse of average similarity, ensuring that densely connected but semantically unrelated cliques, such as the red anomalous substructures depicted in Fig. 1left, receive higher anomaly scores. Furthermore, since large-scale dense substructures are rarer and more likely to be anomalous [76], we introduce the node count as an impact factor to weight the scores. During each iteration, nodes within the detected regions receive an anomaly estimation integrating average similarity with substructure size, whereas undetected nodes receive a score of zero. The overall anomaly score is the mean across all rounds:

$$score_t^{(w)}(v_i) = \frac{1}{\mathcal{R}^t} \sum_{r=1}^{\mathcal{R}^t} \frac{|C_t|}{s_t_r^{(w)}}. \tag{14}$$

Here, $|C_t|$ is the size of the substructure containing v_i, and $\mathcal{R}^t$ is the total number of detection rounds. Since different sampled subgraphs generate distinct embeddings, we calculate topological scores for each view and average them to derive the final score $score_t(v_i)$. Finally, the overall anomaly score for node v_i is computed by integrating this topology perception score with the contrastive and reconstruction scores from our representation learning, weighted by hyperparameters $\eta, \beta, \gamma \in (0, 1)$:

$$score(v_i) = (1-\eta) \cdot score_t(v_i) + \eta \cdot (score_{\text{NS}}(v_i) + \beta score_{\text{SS}}(v_i) + \gamma score_{\text{recon}}(v_i)). \tag{15}$$

During inference, each node's anomaly score is assessed via Eq. (15) across multiple rounds with distinct subgraphs, fully integrating neighborhood information for robust anomaly detection. The procedure of the proposed OHATP is summarized in Algorithm 1 in App. A.

4 Experiments

Datasets. To evaluate the performance of OHATP, we conduct experiments on six real-world benchmark datasets covering citation and social networks. Cora [77], Citeseer [77], Pubmed [77], ACM [78], and Citation [78] are citation networks, in which nodes represent academic papers and edges represent citation relationships between them. BlogCatalog [79], by contrast, is a social network in which nodes represent bloggers and edges capture their social connections. Including both network types allows us to evaluate OHATP under different graph structures and application settings. Detailed descriptions and statistics for all six datasets are provided in App. B.

Anomaly Injection Approach. Following previous work [7, 8, 13, 63, 80], we inject topology and attribute anomalies into the datasets. For topology anomalies, we randomly select m nodes and fully connect them, then repeat this operation p times. In practice, m is set to 15. The values of p are 5, 5, 10, 15, 20, and 20 for Cora, CiteSeer, BlogCatalog, Citation, ACM, and Pubmed, respectively. Attribute anomalies are generated by replacing a target node's features with those from the node that has the largest Euclidean attribute distance within 50 sampled candidates. To balance the ratio between the two anomaly types, we keep the number of attribute anomalies consistent with that of topological anomalies for each of the datasets.

Evaluation Metric. Following prior studies [37, 40, 68], we evaluate anomaly detection performance using two complementary metrics: the area under the receiver operating characteristic curve (AUROC) and the area under the precision–recall curve (AUPRC). AUROC measures how well a method distinguishes anomalous from normal instances across different decision thresholds, accounting for both the true-positive and false-positive rates. AUPRC summarizes the trade-off between precision and recall for the anomalous class and provides an additional perspective on the quality of detected anomalies. Higher values of either metric indicate better performance.

Experimental Settings. In OHATP, we use the Adam optimizer and project node and subgraph features into 64-dimensional latent space. All datasets are trained for 100 epochs with a batch size of 300. We set $\delta = 10$ in Eq. (8) and $\tau = 3$ in Eq. (10), while selecting the top-10 most dissimilar pairs

for hard negative sampling. The fusion parameter η is dynamically sampled from $\{0.2, 0.4, \ldots, 0.8\}$ during anomaly score calculation, and results are averaged over 256 rounds. Performance is evaluated using AUROC and AUPRC metrics. Additional configurations are detailed in App. E.1.

4.1 Comparison with State-of-the-Art Methods

We compare our proposed OHATP with two shallow methods, LOF [24] and ANOMALOUS [55], and ten deep-learning methods, CoLA [13], ANEMONE [35], Sub-CR [36], GRADATE [37], GADMCLG [38], SAMCL [68], HECGAD [40], UniGAD [81], ARC [80], and GADAM [63]. As shown in Table 1, OHATP outperforms all baselines, achieving up to 6% AUROC and 29% AUPRC improvement over the best competing methods on individual benchmark datasets across all settings.

Compared to shallow methods, deep learning approaches achieve substantially better results. As shown in Table 1, shallow mechanisms struggle severely on complex data—failing to surpass 0.18 AUPRC across all datasets—indicating their inability to effectively model high-dimensional node attributes and complex topology information. Deep methods consistently overcome these limitations.

Compared to methods adopting traditional graph augmentations, like GADMCLG, OHATP achieves significant performance improvements. While these methods contrast multiple views detect anomalies, their reliance on indiscriminate augmentations leads to anomaly distortion, which influences their performance on complex datasets like ACM, where GADMCLG only achieves 0.4947 in AUPRC. In contrast, our orthogonal-hyperspherical augmentation introduces learnable perturbations with geometric constraints, which strengthens feature diversity while preserving anomaly signals to avoid anomaly distortion. This enables OHATP to achieve significant performance.

Furthermore, we compare OHATP against recent multi-scale subgraph sampling methods, which avoid destructive augmentations by extracting neighbor contexts at varying scales. However, their performance is still lower than our OHATP by a relatively large margin. These methods risk disrupting critical topological structures or introducing irrelevant information in the learned representations, thereby hardly capturing complex anomalies. Our OHATP fuses both feature and structure levels, isolating these dense substructures through the structure-level topology perception module and optimizing representation learning via the feature-level orthogonal-hyperspherical augmentation. Consequently, OHATP effectively mitigates blind sampling defects, achieving state-of-the-art performance on six datasets and remaining highly competitive on the ACM and Pubmed datasets.

It is worth noting that OHATP achieves relatively weaker performance on the BlogCatalog and Citation datasets compared with the other datasets. We argue this is likely related to weaker separation between injected anomalies and normal graph structures, which makes dense normal communities harder to separate from injected clique-like anomalies. Specifically, for BlogCatalog, many naturally dense social communities have a relatively large number of nodes. Therefore, although the injected topological anomalies are 15-node cliques, their structural contrast is weaker than in sparser benchmarks, reducing the discriminative power and complicating the anomaly detection task. For Citation, sparse high-dimensional textual attributes and heterogeneous citation behaviors make the task more challenging. Interdisciplinary or survey papers may legitimately cite works from multiple topics, making their attributes differ from those of their neighbors and resemble injected anomalies. Despite these challenges, OHATP still achieves state-of-the-art performance on both datasets.

Table 1: Results on six datasets. The best result is boldfaced, with the second-best underlined.

Category	Methods	Dataset											
		AUROC						AUPRC					
		Cora	Citeseer	BlogCatalog	ACM	Pubmed	Citation	Cora	Citeseer	BlogCatalog	ACM	Pubmed	Citation
Shallow	LOF	0.3538	0.3484	0.6796	0.2843	0.3934	0.3059	0.0411	0.0339	0.0317	0.0321	0.0256	0.0360
	ANOMALOUS	0.3434	0.3556	0.7221	0.7280	0.5303	0.6356	0.1181	0.1084	0.1236	0.1752	0.1096	0.1413
Deep	CoLA	0.9065	0.8863	0.7825	0.8127	0.9493	0.7296	0.5065	0.4303	0.3278	0.3214	0.4169	0.1928
	ANEMONE	0.9122	0.9227	0.7866	0.8300	0.9552	0.8028	0.5320	0.6281	0.3309	0.4274	0.3539	0.2578
	Sub-CR	0.9133	0.9248	0.8135	0.8428	0.9594	0.7903	0.5922	0.6083	0.4038	0.3469	0.5854	0.2968
	GRADATE	0.8902	0.8513	0.7981	0.8632	0.9084	0.7072	0.5612	0.2473	0.2113	0.3258	0.4447	0.1539
	GADMCLG	0.9514	0.9397	0.8197	0.9241	0.9865	0.8445	0.7417	0.6523	0.4036	0.4947	0.7631	0.3482
	SAMCL	0.9504	0.9375	0.8154	0.9245	0.9893	0.8495	0.6367	0.5658	0.2985	0.3323	0.7936	0.3316
	HECGAD	0.9321	0.9382	0.8204	0.8790	0.9867	0.8448	0.6042	0.5712	0.4196	0.3924	0.7729	0.3561
	UniGAD	0.8736	0.9005	0.7967	0.8165	0.8323	0.8352	0.6875	0.5223	0.3198	0.4618	0.6145	0.3519
	ARC	0.8858	0.9154	0.7482	0.8016	0.8459	0.8031	0.5097	0.4702	0.3624	0.4072	0.3162	0.3996
	GADAM	0.9576	0.9394	0.8066	0.9256	0.9284	0.8345	0.7480	0.7178	0.2977	0.4467	0.4761	0.3879
	OHATP (Ours)	**0.9752** (+0.0176)	**0.9550** (+0.0153)	**0.8314** (+0.0110)	**0.9869** (+0.0613)	**0.9919** (+0.0026)	**0.8617** (+0.0122)	**0.8702** (+0.1222)	**0.7606** (+0.0428)	**0.4302** (+0.0106)	**0.7929** (+0.2982)	**0.8253** (+0.0317)	**0.4163** (+0.0167)

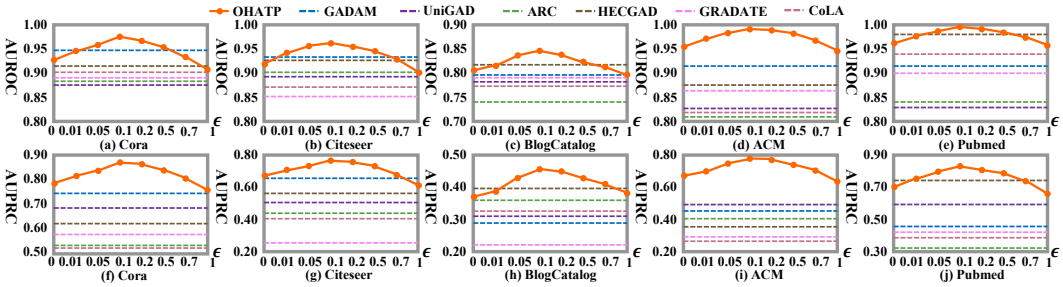

Figure 5: AUROC and AUPRC results of our and compared methods on five datasets with diverse noise magnitude levels.

4.2 Effectiveness of Orthogonal-Hyperspherical Augmentation

To validate our new augmentation technique, we evaluate OHATP against a variant without orthogonal-hyperspherical augmentation. As shown in Fig. 6, the full OHATP outperforms this variant across all benchmark datasets in AUROC and AUPRC. Without augmentation, relying only on the original graph restricts the model from exploring diverse contexts and learning robust representations. We also investigate the impact of noise magnitude, a key augmentation parameter, where OHATP outperforms the competing methods under most settings (see Fig. 5). These show that our learnable augmentation with geometric constraints can preserve feature discriminability while strengthening diversity required for contrastive learning, enabling robust representations for more accurate graph anomaly detection.

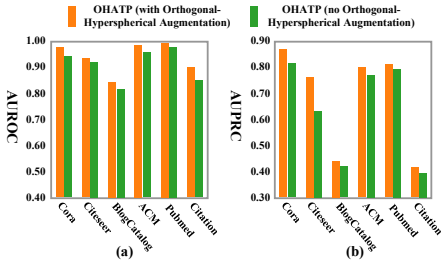

Figure 6: AUROC and AUPRC results comparison with and without orthogonal-hyperspherical augmentation.

4.3 Ablation Study

Various Module Combination Strategy. To investigate the contribution of the two core components in OHATP, we compare our full model with its ablation variants. As reported in Table 2, removing either component notably degrades performance, confirming their indispensability. Specifically, relying solely on Topology Perception (TP) without Representation Learning (RL) yields flawed embeddings, degrading the reliability of similarity evaluations in topological anomaly detection. Conversely, employing RL without TP restricts the model's capacity to identify anomalous substructures, missing topological anomalies. By unifying both modules, OHATP can learn more robust representations and capture critical structures, achieving reliable GAD performance.

Table 2: Ablation Study on module combination.

Metric	Component		Dataset					
	RL	TP	Cora	Citeseer	BlogCatalog	ACM	Pubmed	Citation
AUROC	✗	✓	0.9269	0.8973	0.7051	0.8579	0.9446	0.7781
	✓	✗	0.9346	0.9420	0.7631	0.8354	0.9731	0.8096
	✓	✓	**0.9752**	**0.9550**	**0.8314**	**0.9869**	**0.9919**	**0.8617**
AUPRC	✗	✓	0.7558	0.6850	0.2107	0.4612	0.7284	0.3948
	✓	✗	0.6973	0.7528	0.3655	0.3945	0.7096	0.2653
	✓	✓	**0.8702**	**0.7606**	**0.4302**	**0.7929**	**0.8253**	**0.4163**

Augmentation Strategy. To validate our orthogonal-hyperspherical augmentation, we compare it against five common baselines spanning three categories: attribute perturbations like Gaussian Noise Feature (GNF) and Feature Masking (FM), structural alterations including Graph Diffusion (GD) and Edge Modification (EM), and Multi-Scale Subgraph Sampling (MSP). As shown in Table 3, OHATP consistently achieves the best performance. We attribute this to the limitations of unconstrained augmentations, as random perturbations on attributes or structures degrade anomaly discriminability, leading to anomaly distortion. MSP avoids this but introduces irrelevant information, causing large embedding deviations and masking anomaly signals. Conversely, by introducing learnable perturbations with geometric con-

Table 3: Ablation study for graph augmentation.

Metric	Method	Dataset					
		Cora	Citeseer	BlogCatalog	ACM	Pubmed	Citation
AUROC	GNF	0.9574	0.9409	0.7621	0.9655	0.9718	0.8352
	FM	0.9525	0.9357	0.7835	0.9652	0.9808	0.8489
	GD	0.9564	0.9184	0.7253	0.9714	0.9837	0.8469
	EM	0.9458	0.9367	0.7425	0.9542	0.9680	0.8438
	MSP	0.9565	0.9410	0.8157	0.9693	0.9762	0.8378
	OHATP (Ours)	**0.9752**	**0.9550**	**0.8314**	**0.9869**	**0.9919**	**0.8617**
AUPRC	GNF	0.8116	0.7395	0.3620	0.7610	0.8053	0.3871
	FM	0.8303	0.7410	0.3914	0.7681	0.7964	0.3810
	GD	0.8176	0.6639	0.3455	0.7717	0.7768	0.3924
	EM	0.7846	0.6791	0.3716	0.7807	0.8069	0.3829
	MSP	0.8285	0.6348	0.4195	0.7630	0.7921	0.3906
	OHATP (Ours)	**0.8702**	**0.7606**	**0.4302**	**0.7929**	**0.8253**	**0.4163**

straints, our augmentation strengthens feature diversity while preserving discriminability, ensuring necessary variations without anomaly distortion or large deviations.

Topology Screening Methods. To evaluate the efficacy of the topology screening methods within our topology perception component, we conduct an ablation analysis against four classical community detection algorithms. Specifically, Louvain maximizes modularity through localized node clustering and community merging. Greedy Modularity iteratively adjusts node groupings to maximize the overall network modularity. Label Propagation operates by passing labels across the graph in a semi-synchronous manner, and PercoMCV relies on clique percolation followed by an evaluation of eigenvector centrality. As shown in Table 4, our adopted k-core strategy consistently yields superior results compared to these alternatives. This demonstrates its effectiveness in isolating suspicious subgraphs.

Table 4: Ablation study for topology screening.

Metric	Method	Dataset					
		Cora	Citeseer	BlogCatalog	ACM	Pubmed	Citation
AUROC	Louvain	0.9285	0.9376	0.7152	0.8116	0.9456	0.7459
	Greedy Modularity	0.9146	0.9336	0.7518	0.8427	0.9389	0.8010
	Label Propagation	0.9433	0.9440	0.7679	0.7830	0.9362	0.8374
	PercoMCV	0.9564	0.9467	0.7756	0.8259	0.9365	0.8326
	OHATP (Ours)	**0.9752**	**0.9550**	**0.8314**	**0.9869**	**0.9919**	**0.8617**
AUPRC	Louvain	0.6229	0.6019	0.3424	0.3588	0.5340	0.1664
	Greedy Modularity	0.4957	0.5785	0.3614	0.3856	0.4891	0.2130
	Label Propagation	0.6128	0.7126	0.3746	0.2463	0.4533	0.2517
	PercoMCV	0.7320	0.8088	0.3821	0.3229	0.5054	0.2345
	OHATP (Ours)	**0.8702**	**0.7606**	**0.4302**	**0.7929**	**0.8253**	**0.4163**

5 Conclusion and Future Work

In this paper, we investigated the shortcomings of existing contrastive learning methods in GAD, which usually cause anomaly distortion and disrupt critical dense substructures. To address these issues, we introduced a novel approach OHATP that combines feature-level orthogonal-hyperspherical augmentation with structure-level topology perception. We strengthened feature diversity and preserved feature discriminability via geometric constraints to protect anomaly signals during augmentation, while screening abnormal dense structures via the topology perception module. Extensive experiments on six datasets demonstrate that OHATP substantially outperforms state-of-the-art methods, revealing that geometric constraints are crucial for augmentation in GAD, and the synergy between feature and topology levels is of great significance to improve the final GAD performance.

Limitation and Future work. Despite its effectiveness, OHATP faces computational overhead from multi-view and multi-round screening, limiting scalability on massive graphs. Additionally, its k-core-based topology perception may underperform on extremely sparse graphs. Future work will optimize the architecture for efficiency and explore more universal topology perception methods.

6 Acknowledgments

This work was supported by the New Generation Artificial Intelligence-National Science and Technology Major Project (Nos. 2025ZD0123502), National Natural Science Fund of China (Nos. 62506155), Provincial Natural Science Fund of Jiangsu (Nos. BK20251985), Suzhou Municipal Leading Talents Fund (Nos. ZXL2025320), ZTE Industry-University Project (Nos. IA20260910004), Alipay Industry-University Project (Nos. 2026101543758), and Key Project of the State Key Laboratory for Novel Software Technology (Nos. ZZKT2026A14).

References

[1] S. Pei, L. Yu, G. Yu, and X. Zhang, "Graph alignment with noisy supervision," in *Proceedings of the ACM Web Conference 2022*, WWW '22, p. 1104–1114, 2022.

[2] H. Yang, H. Chen, S. Pan, L. Li, P. S. Yu, and G. Xu, "Dual space graph contrastive learning," in *Proceedings of the ACM Web Conference 2022*, WWW '22, p. 1238–1247, 2022.

[3] Y. Liu, Y. Zheng, D. Zhang, H. Chen, H. Peng, and S. Pan, "Towards unsupervised deep graph structure learning," in *Proceedings of the ACM Web Conference 2022*, WWW '22, p. 1392–1403, 2022.

[4] L. Shuaiyi, K. Wang, L. Zhang, and B. Wang, "Process-oriented heterogeneous graph learning in gnn-based ics anomalous pattern recognition," *Pattern Recognition*, vol. 141, p. 109661, 2023.

[5] C. Tian, F. Zhang, and R. Wang, "Adversarial regularized attributed network embedding for graph anomaly detection," *Pattern Recognition Letters*, vol. 183, pp. 111–116, 2024.

[6] M. Mesgaran and A. B. Hamza, "Graph fairing convolutional networks for anomaly detection," *Pattern Recognition*, vol. 145, p. 109960, 2024.

[7] X. Song, M. Wu, C. Jermaine, and S. Ranka, "Conditional anomaly detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 5, pp. 631–645, 2007.

[8] K. Ding, J. Li, and H. Liu, "Interactive anomaly detection on attributed networks," in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, WSDM '19, p. 357–365, 2019.

[9] H. Chen, H. Yin, W. Wang, H. Wang, Q. V. H. Nguyen, and X. Li, "Pme: Projected metric embedding on heterogeneous networks for link prediction," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, p. 1177–1186, 2018.

[10] H. Chen, H. YIN, X. Sun, T. Chen, B. Gabrys, and K. Musial, "Multi-level graph convolutional networks for cross-platform anchor link prediction," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, p. 1503–1511, 2020.

[11] Y. Zheng, M. Jin, Y. Liu, L. Chi, K. T. Phan, and Y.-P. P. Chen, "Generative and contrastive self-supervised learning for graph anomaly detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12220–12233, 2023.

[12] H. Chen, Y. Li, and H. Yang, "Graph data mining in recommender systems," in *Web Information Systems Engineering – WISE 2021*, (Cham), pp. 491–496, 2021.

[13] Y. Liu, Z. Li, S. Pan, C. Gong, C. Zhou, and G. Karypis, "Anomaly detection on attributed networks via contrastive self-supervised learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2378–2392, 2022.

[14] J. Liu, F. Xia, X. Feng, J. Ren, and H. Liu, "Deep graph learning for anomalous citation detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2543–2557, 2022.

[15] T. Zhao, T. Jiang, N. Shah, and M. Jiang, "A synergistic approach for graph anomaly detection with pattern mining and feature learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2393–2405, 2022.

[16] G. Pang, C. Aggarwal, C. Shen, and N. Sebe, "Editorial deep learning for anomaly detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2282–2286, 2022.

[17] L. Deng, D. Lian, Z. Huang, and E. Chen, "Graph convolutional adversarial networks for spatiotemporal anomaly detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2416–2428, 2022.

[18] L. He, G. Xu, S. Jameel, X. Wang, and H. Chen, "Graph-aware deep fusion networks for online spam review detection," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 5, pp. 2557–2565, 2023.

[19] S. Qiu, J. Ye, J. Zhao, L. He, L. Liu, X. Huang, *et al.*, "Video anomaly detection guided by clustering learning," *Pattern Recognition*, vol. 153, p. 110550, 2024.

[20] Z. Liu, Y. Zhou, Y. Xu, and Z. Wang, "Simplenet: A simple network for image anomaly detection and localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20402–20411, June 2023.

[21] X. Sun, C. Zhang, and L. Li, "Dynamic emotion modelling and anomaly detection in conversation based on emotional transition tensor," *Information Fusion*, vol. 46, pp. 11–22, 2019.

[22] K. Liu, Y. Dou, Y. Zhao, X. Ding, X. Hu, R. Zhang, *et al.*, "Bond: Benchmarking unsupervised outlier node detection on static attributed graphs," in *Advances in Neural Information Processing Systems*, vol. 35, pp. 27021–27035, 2022.

[23] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, "High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning," *Pattern Recognition*, vol. 58, pp. 121–134, 2016.

[24] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," *SIGMOD Rec.*, vol. 29, p. 93–104, May 2000.

[25] S. Zhou, Q. Tan, Z. Xu, X. Huang, and F.-l. Chung, "Subtractive aggregation for attributed network anomaly detection," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, CIKM '21, p. 3672–3676, 2021.

[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[27] B. Chen, J. Zhang, X. Zhang, Y. Dong, J. Song, P. Zhang, K. Xu, E. Kharlamov, and J. Tang, "Gccad: Graph contrastive coding for anomaly detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 8, pp. 8037–8051, 2023.

[28] H. Qiao and G. Pang, "Truncated affinity maximization: One-class homophily modeling for graph anomaly detection," in *Advances in Neural Information Processing Systems*, vol. 36, pp. 49490–49512, 2023.

[29] Y. ZHENG, S. Pan, V. C. Lee, Y. Zheng, and P. S. Yu, "Rethinking and scaling up graph contrastive learning: An extremely efficient approach with group discrimination," in *Advances in Neural Information Processing Systems*, vol. 35, pp. 10809–10820, 2022.

[30] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 1735–1742, 2006.

[31] Y. Tian, D. Krishnan, and P. Isola, "Contrastive multiview coding," in *Computer Vision – ECCV 2020*, pp. 776–794, 2020.

[32] C. Yang, Z. An, L. Cai, and Y. Xu, "Mutual contrastive learning for visual representation learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 3045–3053, 2022.

[33] C. Yang, Z. An, H. Zhou, F. Zhuang, Y. Xu, and Q. Zhang, "Online knowledge distillation via mutual contrastive learning for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 10212–10227, 2023.

[34] C. Yang, Z. An, and Y. Xu, "Multi-view contrastive learning for online knowledge distillation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3750–3754, 2021.

[35] M. Jin, Y. Liu, Y. Zheng, L. Chi, Y.-F. Li, and S. Pan, "Anemone: Graph anomaly detection with multi-scale contrastive learning," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, CIKM '21, p. 3122–3126, 2021.

[36] J. Zhang, S. Wang, and S. Chen, "Reconstruction enhanced multi-view contrastive learning for anomaly detection on attributed networks," in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, IJCAI-2022, p. 2376–2382, 2022.

[37] J. Duan, S. Wang, P. Zhang, E. Zhu, J. Hu, H. Jin, Y. Liu, and Z. Dong, "Graph anomaly detection via multi-scale contrastive learning networks with augmented view," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, pp. 7459–7467, 2023.

[38] X. Wang, S. Lai, S. Zhu, Y. Chen, L. Lv, and Y. Qi, "Graph anomaly detection via multiscale contrastive self-supervised learning from local to global," *IEEE Transactions on Computational Social Systems*, vol. 12, no. 2, pp. 485–497, 2025.

[39] X. Ma, J. Wu, S. Xue, J. Yang, C. Zhou, Q. Z. Sheng, H. Xiong, and L. Akoglu, "A comprehensive survey on graph anomaly detection with deep learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12012–12038, 2023.

[40] J. Hu, S. Wang, J. Duan, H. Jin, X. Liu, and E. Zhu, "Higher-order enhanced contrastive-based graph anomaly detection without graph augmentation," *Pattern Recognition*, vol. 167, p. 111666, 2025.

[41] P. Velickovic, W. Fedus, W. L. Hamilton, P. Liò, Y. Bengio, and R. D. Hjelm, "Deep graph infomax," *stat*, vol. 1050, p. 21, 2018.

[42] F.-Y. Sun, J. Hoffmann, V. Verma, and J. Tang, "Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization," *arXiv preprint arXiv:1908.01000*, 2020.

[43] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph contrastive learning with augmentations," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 5812–5823, 2020.

[44] Y. Jiao, Y. Xiong, J. Zhang, Y. Zhang, T. Zhang, and Y. Zhu, "Sub-graph contrast for scalable self-supervised graph representation learning," in *2020 IEEE International Conference on Data Mining (ICDM)*, pp. 222–231, 2020.

[45] H. Hafidi, M. Ghogho, P. Ciblat, and A. Swami, "Graphcl: Contrastive self-supervised learning of graph representations," *arXiv preprint arXiv:2007.08025*, 2020.

[46] K. Hassani and A. H. Khasahmadi, "Contrastive multi-view representation learning on graphs," in *International Conference on Machine Learning*, vol. 119, pp. 4116–4126, PMLR, 2020.

[47] Y. Liu, X. Yang, S. Zhou, X. Liu, Z. Wang, K. Liang, W. Tu, L. Li, J. Duan, and C. Chen, "Hard sample aware network for contrastive deep graph clustering," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, pp. 8914–8922, 2023.

[48] J. Zhang and S. Chen, "Topology reorganized graph contrastive learning with mitigating semantic drift," *Pattern Recognition*, vol. 159, p. 111160, 2025.

[49] S. Thakoor, C. Tallec, M. G. Azar, R. Munos, P. Veličković, and M. Valko, "Bootstrapped representation learning on graphs," in *ICLR 2021 workshop on geometrical and topological representation learning*, pp. 1–14, 2021.

[50] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, *et al.*, "Bootstrap your own latent - a new approach to self-supervised learning," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 21271–21284, 2020.

[51] Y. Liu, W. Tu, S. Zhou, X. Liu, L. Song, X. Yang, and E. Zhu, "Deep graph clustering via dual correlation reduction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, pp. 7603–7611, 2022.

[52] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, "Barlow twins: Self-supervised learning via redundancy reduction," in *Proceedings of the 38th International Conference on Machine Learning*, vol. 139, pp. 12310–12320, PMLR, 2021.

[53] B. Perozzi and L. Akoglu, "Scalable anomaly ranking of attributed neighborhoods," in *Proceedings of the 2016 SIAM international conference on data mining*, pp. 207–215, 2016.

[54] J. Li, H. Dani, X. Hu, and H. Liu, "Radar: Residual analysis for anomaly detection in attributed networks.," in *Ijcai*, vol. 17, pp. 2152–2158, 2017.

[55] Z. Peng, M. Luo, J. Li, H. Liu, Q. Zheng, *et al.*, "Anomalous: A joint modeling approach for anomaly detection on attributed networks.," in *Ijcai*, vol. 18, pp. 3513–3519, 2018.

[56] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021.

[57] Y. Pei, T. Huang, W. Van Ipenburg, and M. Pechenizkiy, "Resgcn: attention-based deep residual modeling for anomaly detection on attributed networks," *Machine Learning*, vol. 111, no. 2, pp. 519–541, 2022.

[58] K. Ding, J. Li, R. Bhanushali, and H. Liu, "Deep anomaly detection on attributed networks," in *Proceedings of the 2019 SIAM international conference on data mining*, pp. 594–602, 2019.

[59] H. Fan, F. Zhang, and Z. Li, "Anomalydae: Dual autoencoder for anomaly detection on attributed networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5685–5689, 2020.

[60] X. Luo, J. Wu, A. Beheshti, J. Yang, X. Zhang, Y. Wang, and S. Xue, "Comga: Community-aware attributed graph anomaly detection," in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, WSDM '22, p. 657–665, 2022.

[61] T. Huang, Y. Pei, V. Menkovski, and M. Pechenizkiy, "Hop-count based self-supervised anomaly detection on attributed networks," in *Machine Learning and Knowledge Discovery in Databases*, pp. 225–241, 2023.

[62] Q. Wang, G. Pang, M. Salehi, W. Buntine, and C. Leckie, "Cross-domain graph anomaly detection via anomaly-aware contrastive alignment," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 4676–4684, 2023.

[63] J. Chen, G. Zhu, C. Yuan, and Y. Huang, "Boosting graph anomaly detection with adaptive message passing," in *The Twelfth International Conference on Learning Representations*, 2024.

[64] X. Wei, L. Qi, H. Li, J. Zhou, and Y. Cao, "Graph anomaly detection based on multi-aspect contrastive learning," in *2024 4th International Symposium on Computer Technology and Information Science (ISCTIS)*, pp. 51–55, 2024.

[65] H. Qiao, Q. Wen, X. Li, E.-P. Lim, and G. Pang, "Generative semi-supervised graph anomaly detection," in *Advances in Neural Information Processing Systems*, vol. 37, pp. 4660–4688, 2024.

[66] T. Guo, Y. Fan, C. Cui, J. Liang, J. Zhao, and D. Wang, "Topological anomaly quantification for semi-supervised graph anomaly detection," in *The Fourteenth International Conference on Learning Representations*, 2026.

[67] N. Chen, Z. Liu, B. Hooi, B. He, R. Fathony, J. Hu, and J. Chen, "Consistency training with learnable data augmentation for graph anomaly detection with limited supervision," in *The Twelfth International Conference on Learning Representations*, 2024.

[68] J. Hu, B. Xiao, H. Jin, J. Duan, S. Wang, Z. Lv, S. Wang, X. Liu, and E. Zhu, "Samcl: Subgraph-aligned multiview contrastive learning for graph anomaly detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 1, pp. 1664–1676, 2025.

[69] A. Cayley, "Sur quelques propriétés des déterminants gauches.," 1846.

[70] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2019.

[71] X. Cui, Y. Qin, Y. Gao, E. Zhang, Z. Xu, T. Wu, K. Li, X. Sun, W. Zhou, and H. Li, "Sinkd: Sinkhorn distance minimization for knowledge distillation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 7, pp. 11887–11901, 2025.

[72] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[73] Y. Zhu, Y. Xu, F. Yu, Q. Liu, S. Wu, and L. Wang, "Graph contrastive learning with adaptive augmentation," in *Proceedings of the Web Conference 2021*, WWW '21, p. 2069–2080, 2021.

[74] S. B. Seidman, "Network structure and minimum degree," *Social networks*, vol. 5, no. 3, pp. 269–287, 1983.

[75] V. Batagelj and M. Zaversnik, "An o (m) algorithm for cores decomposition of networks," *arXiv preprint cs/0310049*, 2003.

[76] C. C. Noble and D. J. Cook, "Graph-based anomaly detection," in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, p. 631–636, 2003.

[77] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Galligher, and T. Eliassi-Rad, "Collective classification in network data," *AI magazine*, vol. 29, no. 3, pp. 93–93, 2008.

[78] X. Yuan, N. Zhou, S. Yu, H. Huang, Z. Chen, and F. Xia, "Higher-order structure based anomaly detection on attributed networks," in *2021 IEEE International Conference on Big Data (Big Data)*, pp. 2691–2700, 2021.

[79] J. Li, X. Hu, J. Tang, and H. Liu, "Unsupervised streaming feature selection in social media," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, p. 1041–1050, 2015.

[80] Y. Liu, S. Li, Y. Zheng, Q. Chen, C. Zhang, and S. Pan, "Arc: A generalist graph anomaly detector with in-context learning," in *Advances in Neural Information Processing Systems*, vol. 37, pp. 50772–50804, 2024.

[81] Y. Lin, J. Tang, C. Zi, H. Zhao, Y. Yao, and J. Li, "Unigad: Unifying multi-level graph anomaly detection," in *Advances in Neural Information Processing Systems*, vol. 37, pp. 136120–136148, 2024.

A Pseudo Code of OHATP

The overall training and evaluation process of our proposed OHATP model is summarized in Algorithm 1. It mainly consists of representation learning with orthogonal-hyperspherical augmentation and topology perception. By jointly optimizing multi-view contrastive losses and attribute reconstruction loss, the model learns robust node representations, and screens dense topological substructures to compute the final anomaly scores for all nodes.

Algorithm 1 Proposed Model OHATP

Input: Graph: $\mathcal{G} = (\mathbf{A}, \mathbf{X})$, Balance Coefficients: α, β, γ, Noise Magnitude ϵ, Batch Size: B, Epochs: E, Learning Rate: lr, Subgraph Size: K_1 and K_2.

Output: Anomaly score function $score(\cdot)$.

Representation learning with orthogonal-hyperspherical augmentation
Randomly initialize the model weights W_0.
for e from 1 to E **do**
    Form two views from the original graph and the orthogonal-hyperspherical augmented graph via Eq. (1), (2).
    For each node v_i, generate each views' subgraphs with size K_1 and K_2 respectively.
    Obtain $\mathcal{B}$ by randomly split $\mathcal{V}$ into batches with B.
    for batch $b \in \mathcal{B}$ **do**
        Calculate the node-subgraph contrastive loss via Eq. (5), (6).
        Calculate the inter-view contrastive loss via Eq. (7).
        Calculate the g-EMD between non-aligned subgraphs via Eq. (9) and the intra-view contrastive loss via Eq. (10).
        Calculate attribute reconstruction loss via Eq. (11).
        Combine these modules, calculate the total loss $\mathcal{L}_{\text{total}}$ via Eq. (12).
        Update model parameter.
    end for
end for
# Topology Perception
Calculate the average degree k_{init} of nodes in graph.
for k from k_{init} to k_{max} **do**
    Detect dense topological substructures $\{C_0, \ldots, C_b\}$.
    for C_j **do**
        Calculate the average node-pair similarity of C_j via Eq. (13).
    end for
end for
By multiple round detection, calculate the final anomaly score for each node via Eq. (14), (15).

B Detailed Dataset Description

Table 5 summarizes the primary characteristics of the utilized datasets. A comprehensive introduction of each dataset is provided below.

- Cora [77] is a widely used citation network dataset consisting of scientific publications in machine learning. Each node represents a paper, and edges denote citation relationships between them. Node features are derived from paper text and used to characterize the research content of each paper.
- Citeseer [77] is also a widely used citation network dataset, similar to Cora, where each node represents a paper and edges indicate citation relationships. It contains many computer science papers categorized into various topics, with feature vectors extracted from text content. Compared to Cora, Citeseer is sparser and more challenging.
- BlogCatalog [79] is a social network dataset collected from an online blogging platform. Nodes represent users, and edges denote social relationships such as friendships or interactions. Each user is associated with interest tags forming multi-label attributes.

- ACM [78] is a publication network dataset from a computer science academic paper database, constructed using papers selected from the ACM Digital Library to form a graph structure. Nodes represent research papers, while edges capture citation or co-citation relationships. Node attributes are generated from keywords or textual information.
- Pubmed [77] is a large citation network dataset derived from biomedical research papers. Each node corresponds to a scientific publication, and edges represent citation relationships. Node features are extracted from paper abstracts using text processing techniques. Due to its large scale and complex structure, the Pubmed dataset is usually used to evaluate the performance of graph neural networks on large-scale graphs.
- Citation [78] is a scientific publication network dataset where nodes represent academic papers and edges denote citation relationships. Node attributes are typically constructed from textual content such as titles, abstracts, or keywords. This dataset contains thousands of nodes and numerous citation links, featuring clear community structures and sparsity. It reflects citation patterns and knowledge dissemination paths across different research fields.

Table 5: Statistics of the six datasets used in our experiments.

Datasets	#Nodes	#Edges	#Attributes	#Anomalies(Rate)
Cora	2,708	5,429	1,433	150(5.54%)
Citeseer	3,327	4,732	3,703	150(4.51%)
BlogCatalog	5,196	171,743	8,189	300(5.77%)
ACM	16,484	71,980	8,337	600(3.63%)
Pubmed	19,717	44,338	500	600(3.04%)
Citation	8,935	15,098	6,775	450(5.03%)

C Supplementary Information on Comparative Methods

Below, we elaborate on the twelve GAD models employed in our comparative experiments.

- LOF [24] shifts outlier detection from a binary property to a relative degree of isolation. By evaluating a target instance against its local neighborhood, LOF ranks isolation degrees and identifies local anomalies that traditional global approaches fail to detect.
- ANOMALOUS [55] proposes a joint framework integrating CUR decomposition with residual analysis to spot anomalies on attributed networks. By collectively exploiting attribute and network structure information, it concurrently performs robust attribute selection and effectively identifies anomalies across synthetic and real-world datasets.
- CoLA [13] pioneers the application of contrastive self-supervised learning for anomaly detection. It features a novel framework comprising instance pair sampling, a GNN-based contrastive model, and multi-round scoring. By capturing node-neighborhood relationships through an anomaly-related objective, it effectively identifies graph anomalies.
- ANEMONE [35] presents a comprehensive framework leveraging multi-scale contrastive learning to capture anomalies across various structural levels on graphs. Additionally, it integrates a novel statistical estimation strategy to predict abnormality scores, demonstrating effectiveness across multiple benchmark datasets.
- Sub-CR [36] proposes a multi-view self-supervised framework integrating contrastive and attribute reconstruction modules. It captures local and global structural anomalies by contrasting dual multi-scale views, while rebuilding raw target attributes from localized subgraphs, facilitating effective anomaly detection.
- GRADATE [37] introduces a graph anomaly detection framework via multi-scale contrastive learning with an augmented view. By incorporating subgraph-subgraph contrast and investigating diverse augmentation strategies, it effectively identifies anomalous nodes.
- GADMCLG [38] proposes a multiscale contrastive self-supervised anomaly detection model that integrates local-level and global-level contrasts. To capture discrepancies from

underlying node contexts, it incorporates a subgraph centralization strategy to neutralize spatial influences during global structural evaluations.

- SAMCL [68] introduces a novel anomaly detection approach rooted in multi-view subgraph-aligned contrastive learning. By aligning diverse subgraph views, it conducts node-subgraph contrast and attribute reconstruction within these contexts. This synergy detects topological and attribute anomalies across different representations.
- HECGAD [40] proposes an augmentation-free, multi-view framework that avoids harmful structural or attribute perturbations. To circumvent traditional graph augmentations, it introduces a novel subgraph sampling strategy, jointly integrating intra-view contrast, inter-view contrast, and attribute reconstruction to precisely capture anomalous deviations.
- UniGAD [81] is a unified graph anomaly detection framework jointly addressing node, edge, and graph-level anomalies. By integrating MRQSampler for critical information capture and the GraphStitch network for multi-level information sharing, it achieves superior performance and strong zero-shot transferability.
- ARC [80] is a pioneering "one-for-all" generalist graph anomaly detection model. Utilizing in-context learning, it detects anomalies on-the-fly across diverse datasets without dataset-specific fine-tuning, relying only on few-shot normal nodes to achieve strong generalizability and high efficiency.
- GADAM [63] is a novel unsupervised graph anomaly detection framework that decouples local inconsistency mining from message passing. By incorporating adaptive message passing and global consistency discernment, it offers a clear, highly efficient, and scalable solution that achieves superior performance on large-scale datasets.

D Time Complexity Analysis

This subsection analyzes the time complexity of our proposed OHATP in detail. Specifically, let N and E denote the numbers of nodes and edges in graph G, respectively. Let T be the number of training epochs, R be the number of evaluation rounds, c be the local subgraph size, and ω be the average node degree of G. The number of layers of the encoder and the dimension of node embeddings are considered constant for simplicity.

According to the Algorithm 1, the orthogonal-hyperspherical augmentation is applied to the local subgraph representations of each node, resulting in $O((T+R)Nc)$ time complexity. The subgraph sampling step is implemented by neighborhood sampling, leading to $O((T+R)Nc\omega)$. The contrastive learning module includes node-subgraph contrast, inter-view contrast, and intra-view contrast. Its main cost comes from computing the $c \times c$ cost matrix between local subgraphs, giving $O((T+R)Nc^2)$. The attribute reconstruction module costs $O((T+R)Nc)$. In addition, the topology perception module iteratively performs k-core decomposition and computes pairwise similarities within detected dense substructures, with complexity $O(I(N+E)+\sum|C|^2)$, where I is the number of k-core iterations and C denotes a detected dense substructure. Therefore, the overall time complexity of OHATP is:$O((T+R)(Nc\omega+Nc^2)+I(N+E)+\sum|C|^2)$. We compare the time complexity of our OHATP against representative baselines in Table 6.

Although the iterative detection within the topology perception module of OHATP introduces extra computational overhead, the dominant computational cost stems from operations performed on local subgraphs and detected dense structures. Since these local structures are usually much smaller than the whole graph, i.e., $c \ll N$ and $|C| \ll N$, the dominant cost of OHATP increases approximately linearly with N on sparse graphs. The quadratic term in multi-view contrastive learning is also restricted to the local subgraph size c, rather than the total number of nodes N. Therefore, compared with global contrastive methods, OHATP has better scalability on large-scale graphs.

E Additional Experimental Descriptions

E.1 More Hyperparameter Settings

We implement OHATP using PyTorch 1.12.0 with Python 3.7 and conduct all experiments on an NVIDIA GeForce RTX 4090 GPU with 48GB memory. To ensure reproducibility, this section

provides the extended hyperparameter configurations omitted from the main text. Table 7 summarizes the optimal settings across all evaluated datasets, including learning rate, subgraph sampling scales K_1 and K_2, noise magnitude ϵ, and balance parameters α, β, γ.

To mitigate the hyperparameter tuning burden, we also provide detailed tuning guidelines in this subsection. Specifically, ε controls the intensity of injected noise and can be fixed as a small constant between 0 and 1 across all datasets. δ and τ yield stable performance within a reasonable range and generally work well with default values. K_1 and K_2 can be selected according to data scale or neighborhood size requirements, and a simple empirical rule is provided as follows: set $K_1 = 2, K_2 = 4$ for sparse graphs; set $K_1 = 4$ or 6, and $K_2 = 6$ or 8 for relatively dense graphs. The fusion weight η for representation learning and topology perception modules is adaptively learned by the model during training. Consequently, only a small number of loss weighting coefficients (e.g., α, β, γ) require careful tuning. We list all hyperparameters primarily for experimental reproducibility rather than to mandate full-range parameter searching.

Table 6: Time complexity comparison of different methods.

Method	Overall Complexity
CoLA	$\mathcal{O}(cNR(c+\omega))$
ANEMONE	$\mathcal{O}(cNR(c+\omega))$
Sub-CR	$\mathcal{O}(N+cN(c+\omega))$
GRADATE	$O((T+R)Nc\omega)$
GADMCLG	$O((T+R)Nc\omega)$
SAMCL	$O\big(TN(c\omega+c^2)+RNc\omega\big)$
HECGAD	$O((T+R)Nc\omega)$
UniGAD	$O((T+R)Nc\omega)$
ARC	$O\big(T(N\omega+N^2)\big)$
GADAM	$\mathcal{O}(\mathcal{E}+N)$
OHATP (ours)	$O\big((T+R)(Nc\omega+Nc^2)+I(N+E)+\sum\|C\|^2\big)$

Table 7: Hyperparameters for reproducibility.

Dataset	Learning rate	K_1	K_2	α	β	γ	ϵ
Cora	2×10^{-3}	2	4	0.8	0.2	0.9	0.1
Citeseer	2×10^{-3}	2	8	0.6	0.4	0.9	0.1
BlogCatalog	1×10^{-2}	6	8	0.6	0.6	0.7	0.1
ACM	5×10^{-3}	6	8	0.8	0.2	0.9	0.1
Pubmed	2×10^{-3}	4	8	0.8	0.2	0.7	0.1
Citation	2×10^{-3}	4	6	0.8	0.8	0.9	0.1

Table 8: Performance comparison on OGB-Arxiv dataset

Method	AUROC	AUPRC
CoLA	0.8041	0.3348
HECGAD	0.8248	0.3983
GADAM	0.8122	0.3537
OHATP (ours)	**0.8325**	**0.4158**

E.2 Scalability Analysis

To evaluate the scalability of our OHATP on large-scale graphs, we further conducted an additional experiment on the OGB-Arxiv dataset, which contains 169K nodes. The experimental results in Table 8 show that OHATP still achieves stable performance, with an AUROC of 0.8325 and an

AUPRC of 0.4158. This indicates that the proposed method remains feasible on relatively large real-world graph data.

We also acknowledge that the practical efficiency on million-node graphs may be affected by implementation details, GPU memory, the efficiency of sparse matrix operations, batch size, the local subgraph size c, and the size of dense substructures. For million-scale graphs, OHATP can be further scaled by fixing the local subgraph size, adopting mini-batch training, using sparse adjacency storage, and limiting pairwise computations within dense substructures.

E.3 Training Loss and Performance Curves

To evaluate the learning dynamics of our OHATP model, we first visualize the training loss in Fig. 7 to verify its smooth and stable convergence. In Fig. 8 and Fig. 9, we plot the AUROC and AUPRC curves to comprehensively evaluate the model's detection performance. Together, these results clearly demonstrate that our proposed OHATP not only achieves efficient optimization during training but also maintains a highly effective discriminative capability for identifying complex graph anomalies.

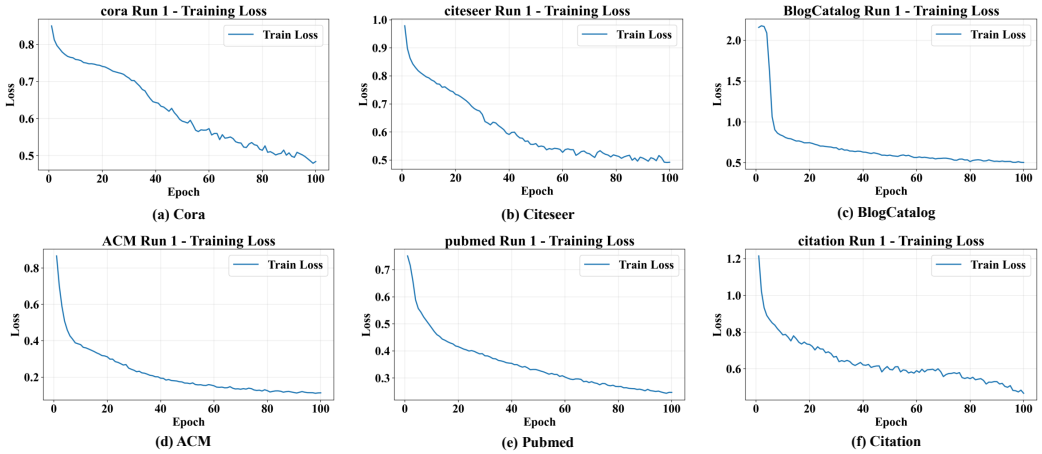

Figure 7: Loss curve of different datasets in OHATP.

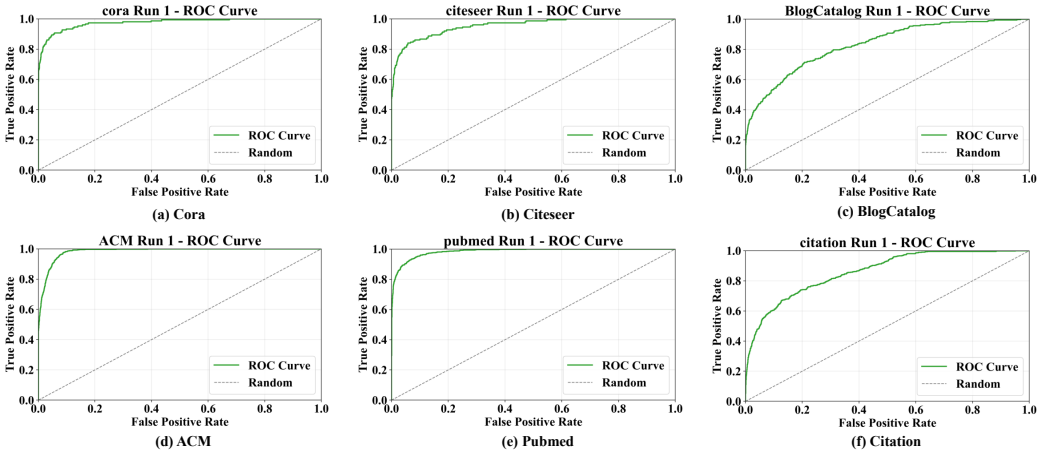

Figure 8: AUROC curve of different datasets in OHATP.

E.4 Parameters Sensitivity Analysis

Noise Magnitude ϵ. Appropriately increasing the noise magnitude ϵ allows the model to explore a broader semantic neighborhood and generate more diverse augmented views, thereby effectively facilitating the learning of robust node representations. However, an excessively large ϵ can disrupt the underlying graph information and sign structure, hindering the model's ability to accurately detect anomalies. As shown in Fig. 10a, we adjust ϵ to analyze its impact, and the observed performance trends align with our theoretical expectations. OHATP achieves optimal performance across all

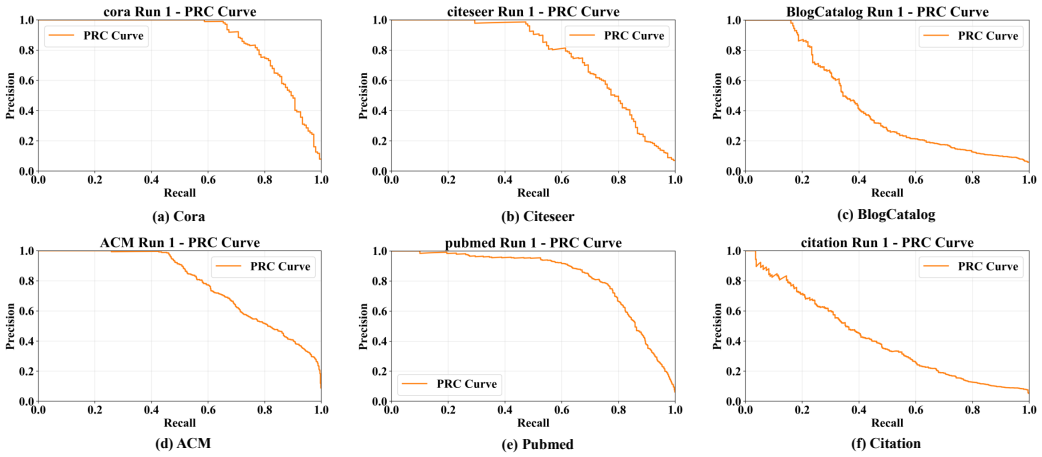

Figure 9: AUPRC curve of different datasets in OHATP.

benchmark datasets when ϵ is set near 0.1, striking a fine balance between enhancing view diversity and preserving core information.

Balance Parameter α, β and γ. We further investigate the sensitivity of the model to the balance factors α, β, and γ. Specifically, α is designed to regulate the contributions of the original and augmented views, while β and γ are employed to balance the contrastive learning and attribute reconstruction modules within the OHATP framework. During experimentation, we tune α, β in a range of {0.2, 0.4, 0.6, 0.8}, and configure γ within {0.3, 0.5, 0.7, 0.9}. As shown in Fig. 11 and Fig. 12, the optimal performance varies across datasets. For instance, some datasets favor a stronger emphasis on reconstruction, while others benefit more from contrastive signals. Similarly, Fig. 10b shows diverse view weight preferences on different graph structures. It is crucial to find an appropriate trade-off between these modules. The optimal α, β, and γ for each dataset are summarized in Table 7. Notably, OHATP consistently yields satisfactory detection results under various parameter settings.

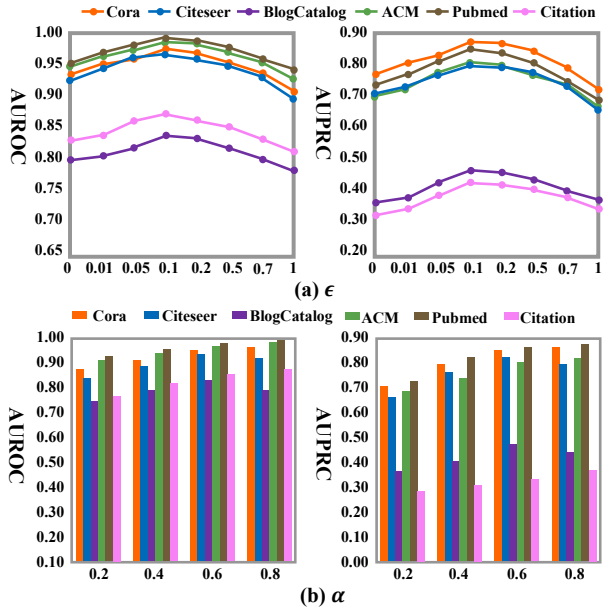

Figure 10: AUPRC and AUROC of OHATP under different ϵ and α.

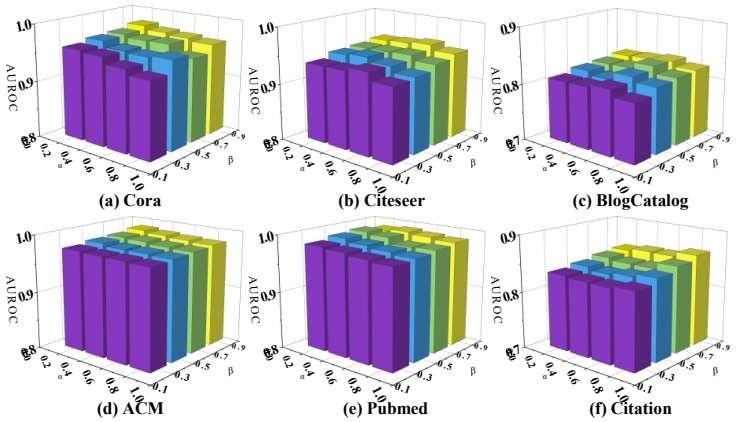

Figure 11: AUROC on parameters of contrastive module β and reconstruction module γ.

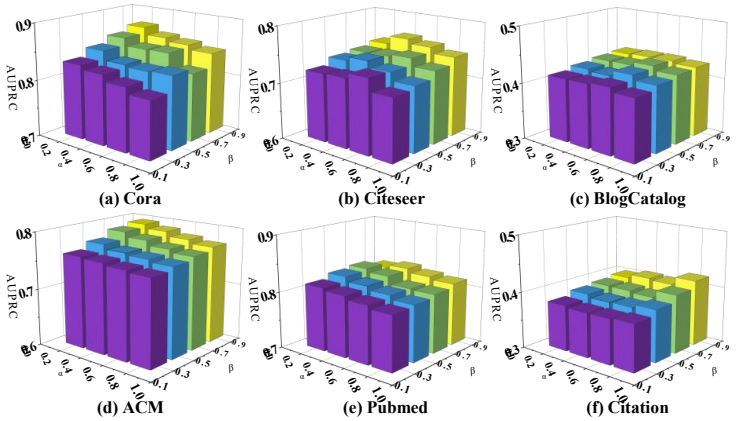

Figure 12: AUPRC on parameters of contrastive module β and reconstruction module γ.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state our contributions and scope.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitation of our proposed OHATP in Conclusion.

   Guidelines:

- The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

   Justification: The paper does not include theoretical results.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provide a detailed implementation detail in the paper for reproducing our experimental results.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.

- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We provide the code of our proposed method with detailed instructions for the reproduction. The datasets are also publicly available.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: We specify all the hyperparameters and optimizers in the experimental settings section and appendix.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: All the experimental results are acquired by multiple rounds of experiments.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Our model is implemented in PyTorch 1.12.0 with Python 3.7 and executed on GeForce RTX 4090 GPU.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.

- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in the paper conforms to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: There is no societal impact of the work performed.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We properly credit all referenced baseline works and datasets.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.