

Multi-dimensional Edge-based Graph Representation Learning for Obstructed Prohibited Items Detection in X-ray Images

Haolin Tang¹, Hongxia Gao² and Runze Lin³

¹ South China University of Technology, Guangdong, China

Abstract. X-ray security inspection has been widely used to maintain safety in public places and transportation systems. Due to the imaging characteristics of X-ray images, the stacking of items can cause translucency interference in the images, making it challenging to detect contraband items in backpacks or suitcases during security checks. Most existing methods have improved detection by adjusting the combination of features without considering the relationships between targets. In this paper, we propose a novel prohibited item graph representation learning algorithm to explicitly model inter-item relationships, aiming at improving their detection performance. Our approach starts with GTG module which generates a graph topology structure connecting the proposals output by the detection backbone network, where each proposal is treated as a node describing a candidate object. Then, the MDE module creates a set of multi-dimensional edge features to comprehensively and explicitly describe the relationships between each pair of connected nodes, allowing context information to be used for their detection. Extensive experiments validate the effectiveness of our method which not only enhances the detection accuracy, but also better identifies hard-to-distinguish objects in complex scenarios. This exploration opens up an uncharted graph-based direction previously unexplored in prior research, providing a new path for future studies in graph-based X-ray security inspection detection. Our code is provided in the Supplementary Material.

Keywords: Prohibited items detection, X-ray image, Graph Representation Learning and Multi-dimensional Edge Feature.

1 Introduction

In the past years, the increasing density of crowds in public transportation hubs has made the security checks in public spaces increasingly important to effectively suppress terrorism and criminal incidents[1]. X-ray security imaging can describe the internal information of objects in a non-contact manner. Since X-ray images provide excellent recognition ability, clarity, and visualization capabilities[2], this technology has been commonly used to check luggage for prohibited items in the past decades. To alleviate the pressure of human staff at checkpoints and reduce public safety hazards, intelligent prohibited item detection based on X-ray images holds great research value.

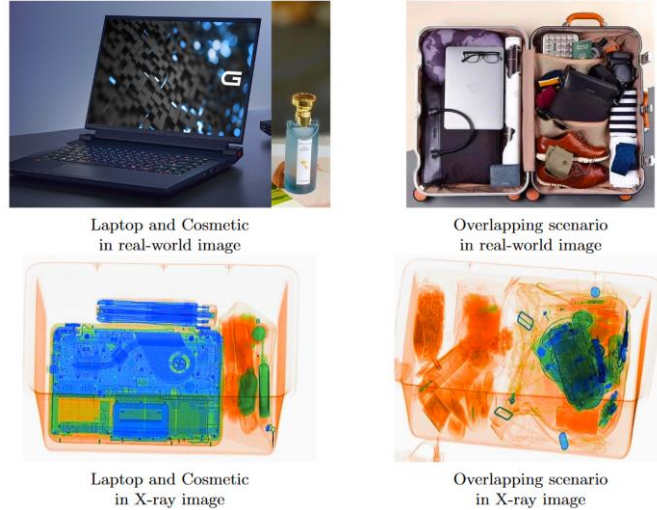


Fig. 1. Comparison of Obstruction in Real-World and X-ray Images: This figure illustrates the difference in obstruction challenges faced in X-ray image prohibited item detection versus general object detection. In X-ray images, objects often overlap, creating a unique challenge not typically observed in standard object detection scenarios.

To address the issue of translucency interference, Rao et al. [3] attempted to incorporate edge and colors, which is prominent features of X-ray images, into their design. They utilized the traditional edge detection operator, Sobel, to obtain edge images, generate attention maps, and weight the backbone features, thereby solving some overlapping problems. Zhao et al. [4] introduced a new label-aware mechanism to separate overlapping objects in high-level feature maps. By assigning labels to different anchor boxes and adaptively adjusting the corresponding features, this approach can handle overlaps between objects and similar backgrounds, as well as among multiple objects. Although these two methods have improved detection accuracy to some extent by introducing prior information and adjusting label allocation mechanisms, targets are often interfered with by other stacked items due to the translucent nature of X-ray images. In other words, these methods only consider feature reconstruction on a single target level, which means that the extracted features will always contain elements of the background or other items, hindering further performance enhancement. In general, prior approaches primarily concentrated on modifying the front network structure to mitigate the effects of translucency interference, while neglecting the relationships between objects.

To address the aforementioned issues, this paper proposes a novel graph-based X-ray image prohibited items detection approach. It starts with an object detection network [5] which detects potential contraband items (i.e., candidate objects) within the X-ray images by individually capturing each single target object whose features may be intertwined with background elements or obscured by other objects. Since objects with mutual translucent obstruction in a X-ray image may share common regions of interest, while X-ray security inspection images usually involve various types of objects such as luggage, electronic devices, liquids, and metal products, we assume the

interactions and relationships between these objects differ from those in regular color images recorded in natural conditions. To model such interactions/relationships, we treat each the regions of interest (ROIs) output by the detection network as a node, and explicitly explore the relationship between each pair of candidate objects via a multi-dimensional edge feature learned by our novel graph representation learning framework inspired by GRATIS [6]. Specifically, our strategy determines the connectivity between nodes based on the size, shape, and position of the ROIs. Building upon this, we learn task-specific multi-dimensional edge features to better represent the relationships between nodes. In this manner, we can effectively simulate the real interactions and connections between objects in X-ray images, making it more likely for the graph neural network to capture potential prohibited items. As a result, our proposed approach can accurately interpret the complex structures of these objects, thereby further enhancing the detection accuracy. The main contributions of this paper can be summarized as follows:

- By exploring a previously uncharted strategy, i.e., reducing the impact of translucency occlusion by investigating the relationships between target candidate objects, this paper proposes the first graph-based approach that learns a multi-dimensional edge feature for the detection of prohibited items in a X-ray imagery.
- The proposed GTG (Graph Topology Generation) and MDE (Multi-dimensional Edge) modules dynamically generate graph topology structures and multidimensional edge features to explicitly describe the relationship and interaction between each pair of detected objects.
- Extensive experiments demonstrate that the superiority of our proposed method compared to state-of-the-art (SOTA) approaches. This research offers a novel direction for future X-ray security inspection detection studies.

2 Methodology

In this section, we propose a novel method that integrates the object detection framework with the graph representation learning. This method generates a graph representation with a dynamic topology and multidimensional edge features to describe the regions of interest and their relationships, the overall framework of our method as shown in Figure 2.

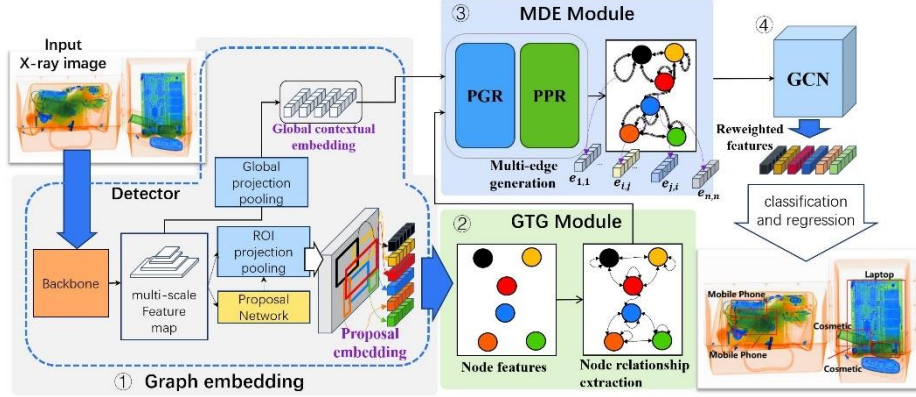


Fig. 2. The overall pipeline of the proposed approach. ① **Graph Embedding module** first takes the X-ray image as the input and outputs features describing N object proposals. Within this module, multi-scale spatial features of the input X-ray image are extracted by the backbone network, where the output of each layer is subjected to a pooling operation to obtain a global contextual representation. ② **The Graph Topology Generation (GTG) module** then generates a basic graph topology structure based on these input proposals and their coordinates used for calculating IOU. ③ **The Multidimensional Edge (MDE) module** creates multidimensional edges, utilizing the global contextual representation and the basic graph topology structure provided by GTG. ④ The nodes and multidimensional edge features are processed through a **Graph Convolutional Network (GCN)** to obtain the learned object-specific features, based on which the final predictions are made.

2.1 X-ray Image Graph Representation Learning

Node Feature and Adjacency Matrix Learning. Graph representation learning includes the graph embedding, graph topology structure generation and multi-dimensional Edge learning. We use the features of the regions of interest output by the object detection network as node features, forming the node set $p \subseteq \{p_i \in R^{1 \times K}\}, i = 1, 2, \dots, N$, and then map the full-scale features through RoI pooling to obtain the global context $GC \in R^{1 \times K}$. Unlike the processing of graph data, there is no predefined graph topology in detection tasks. Therefore, we propose a graph topology structure generation module to generate the graph topology. This module takes the basic node features $\setminus(P)$ and the coordinates of the proposal boxes as input and outputs the adjacency matrix $A \in R^{1 \times K}$ that delineates the topology of graph G , where the elements of A are binary values: 0 indicates that there is no connection between the corresponding nodes, and 1 indicates a connection. In this case, A is a symmetric matrix, as the connections between nodes are mutual. It can be expressed as:

$$A = GTG(P) \quad (2)$$

$$E \subseteq \{e_{i,j} = 1 | p_i, p_j \in \text{and } A_{i,j} = 1\}$$

The node features derived from the detection network depend on the learning of the detection network itself. It encodes the region proposal boxes into node features. Due

to the translucent imaging characteristics of X-ray images, the learned node features include not only the characteristics of the target but also certain background features. Inputting these raw features into subsequent classification and regression networks can introduce translucent interference. Therefore, we introduce a graph neural network to further update node features, filtering out background features while retaining prohibited item features. After integrating the graph topology structure, the node features learned by the network ultimately contain not only the characteristics of the regions of interest themselves but also those of related nodes (other regions of interest). The existence of each edge can be determined by a specific rule applied to the corresponding pair of vertices (for example, the distance between vertices, their similarity, or the relative position of the corresponding frames in the original image), where each existing edge feature is 1, and non-existing ones are 0. Grounded in a key observation that in X-ray images, items with mutual translucency occlusions tend to share common regions in the original imagery, our graph topology construction strategy connects nodes that share a significant proportion of the same area. In other words, when the Intersection over Union (IoU) of two proposed boxes exceeds a predetermined threshold, they are interconnected.

Multidimensional Edge Feature Learning. We hypothesize that the complex web of relationships between objects, as well as between objects and their background, can not be sufficiently captured by basic feature representations, be they binary or numerical. Therefore, upon acquiring the complete set of node features P and creating a tailored graph topology A , our methodology introduces an advanced module for the generation of multidimensional edge features. This innovative module substantially enriches the understanding of inter-node dynamics within a $1 \times K$ dimensional space, facilitating detailed characterization of each edge through the assignment of multidimensional features $\hat{e}_{(i,j)} \in R^{1 \times K}$. These assignments play a crucial role in enabling a comprehensive message-passing network across the graph, thereby integrating relational insights both within the individual node features and throughout the wider global context GC .

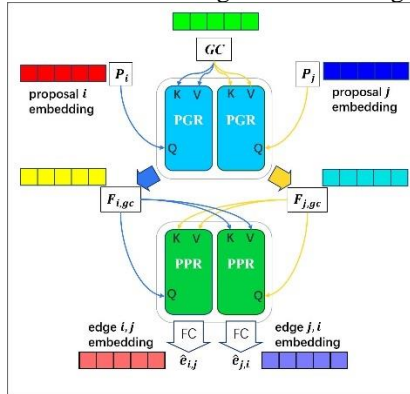


Fig. 3. Description of the Proposed MDE Module.

As a result, for each linked edge where $A_{(i,j)} = 1$, we craft a specific multidimensional edge feature $\hat{e}_{(i,j)}$. This feature incorporates insights from the individual node features p_i and p_j , alongside the overarching global context GC , delineated as:

$$\hat{e}_{(i,j)} = MDE(GC, p_i, p_j) \quad (2)$$

$$\hat{E} \subseteq \{\hat{e}_{(i,j)} = \hat{e}(p_i, p_j) | p_i, p_j \in P \text{ and } A_{i,j} = 1\}$$

The MDE module is divided into two primary components: the Proposal-Global context Relation (PGR) block, which initially pinpoints cues relevant to each vertex within the global context GC ; and the Proposal-Proposal Relation (PPR) block, dedicated to extracting contextual relational features between proposals from the output of PGR. This process is targeted at developing the final multidimensional edge features. The entire sequence of operations is depicted in Figure 3.

PGR: In the PGR module, relational cues between various proposal boxes, denoted as P_i and P_j , are discerned and then transformed into multidimensional edge attributes $e_{(i,j)}$ or $e_{(j,i)}$. This component leverages the node properties of P_i and P_j , in conjunction with the global context GC , as its primary inputs. Within this framework, P_i and P_j function autonomously as queries to pinpoint specific node-context relational features $F_{(i,gc)}$ and $F_{(j,gc)}$, within GC , utilizing GC itself as both key and value in an attention mechanism. The procedural mathematics of this operation are depicted as:

$$F_{(i,gc)} = PGR(P_i, GC) \quad F_{(j,gc)} = PGR(P_j, GC) \quad (3)$$

where the cross-attention mechanism within PGR is formulated as:

$$PGR(A, B) = softmax\left(\frac{(AW_q(BW_k)^T)}{\sqrt{d_k}}\right)(BW_v) \quad (4)$$

Here, W_q , W_k , and W_v are the learnable weights for encoding the query, key, and value respectively. The selection of these weights is contingent upon the input data's structure, and d_k acts as a scaling factor, aligning with the channel count in B . The ensuing outputs $F_{(i,gc)}$ and $F_{(j,gc)}$ encapsulate task-specific indicators pertinent to vertices P_i and P_j , as derived from the global context GC .

PPR: Utilizing $F_{(i,gc)}$ and $F_{(j,gc)}$, the PPR module delves deeper into extracting context-specific cues pertinent to each pair of vertices. Like PGR, PPR employs a cross-attention mechanism. This approach uniquely leverages $F_{(i,gc)}$ and $F_{(j,gc)}$ as queries, keys, and values, generating two reciprocal node-to-node relational features $F_{(i,gc,j)}$ and $F_{(j,gc,i)}$. Specifically, $F_{(i,gc,j)}$ encapsulates the relational insights from $F_{(j,gc)}$ relevant to $F_{(i,gc)}$, and vice versa for $F_{(j,gc,i)}$. These features, $F_{(i,gc,j)}$ and $F_{(j,gc,i)}$, amalgamate global context and specific details pertaining to nodes P_i and P_j , capturing their intricate relationships. The formulation is as given:

$$F_{(i,gc,j)} = PPR(F_{(i,gc)}, F_{(j,gc)}) \quad (5)$$

$$F_{(j,gc,i)} = PPR(F_{(j,gc)}, F_{(i,gc)})$$

In the concluding step, a fully connected layer (denoted by the operation L) is utilized to morph the features into multidimensional edge vectors $\hat{e}_{(i,j)}$ and $\hat{e}_{(j,i)}$, effectively finalizing the transformation as delineated:

$$\hat{e}_{(i,j)} = L(F_{(i,gc,j)}) \quad \hat{e}_{(j,i)} = L(F_{(j,gc,i)}) \quad (6)$$

Consequently, every multidimensional edge feature $\hat{e}_{(i,j)}$ encompasses tailored cues from the full context GC of the input data I , pertinent to vertices p_i and p_j . This approach allows edge features to be expressed as $\hat{e}_{(i,j)} = [\hat{e}_{(i,j)}(1), \hat{e}_{(i,j)}(2), \dots, \hat{e}_{(i,j)}(K)]$, effectively capturing the intricate inter-node relationships that single-valued edge representations may overlook.

2.2 Object Feature Reconstruction

After determining the node connections and creating multidimensional edges, we can employ a Graph Convolutional Network (GCN) to refine node features. This architecture is structured into several layers, with each layer tasked with distinct transformations. **Edge and Node Feature Transformation:** For each GNN layer, there's a series of linear transformations applied to both node and edge features. This is achieved through linear layers. These transformations are crucial for learning complex relationships in the graph. **Residual Connections:** The architecture incorporates residual connections to combat the issue of vanishing gradients, facilitating the effective training of more profound network layers. **Node and Edge Feature Update:** The updates of node and edge features are intertwined within the graph structure. Edge features are refined using the transformed node features and existing edge information, involving operations like einsum and batch normalization for efficiency and stability. Simultaneously, node features are updated through aggregating neighbor information, guided by edge attention weights. **Output:** The final output is a set of transformed node features, which can be used for downstream tasks like classification or regression.

3 Experiments

3.1 Experimental settings

Table 1. An Overview of the Category Distribution Statistics in the OPIXray and HiXray Dataset.

Dataset	Categories	train- ing	test- ing	total
OPIXray	Folding Knife, Straight Knife, Scissor, Utility Knife, Multi-tool Knife	7109	1776	8885
HiXray	Portable Charge 1, Portable Charge 2, Water, Laptop, Mobile Phone, Tablet, Cosmetic, Nonmetallic Lighter	82452	20476	102928

Datasets: We evaluate our approach on the OPIXray [3] and HiXray [7] datasets as our primary data sources. The OPIXray dataset stands as a pioneering high-quality dataset tailored for security object detection, incorporating a diverse range of bladed instruments across 8885 X-ray images. In parallel, the HiXray dataset encompasses

44,364 images derived from routine security screenings at global airports, cataloging eight types of prohibited items including, but not limited to, lithium batteries, electronic gadgets, liquids, and lighters, commonly found in everyday scenarios. These datasets are universally acknowledged as among the foremost dependable public resources in the domain of X-ray object detection. Their detailed information is presented in Table 1.

Implementation details: The experiments are conducted utilizing the AdamW optimizer, setting the initial learning rate at 0.0001, with betas configured to (0.9, 0.999), and implementing a weight decay of 0.05. We resize the input images to a resolution of 1000x600 pixels, and initialize our model's backbone with ResNet50, which is pre-trained on the ImageNet 1K dataset. The experimental setup is built on the mmdetection [8] framework for object detection. The programming for these experiments is done in Python, and we leverage five NVIDIA RTX3090 GPUs, each equipped with 24GB of VRAM. Our computations leverage the parallel computing capabilities of CUDA 11.1, utilizing the Pytorch 1.7.1 framework for deep learning.

Evaluation metrics: We utilize the Mean Average Precision (mAP) metric to evaluate the efficacy of our models.

3.2 Comparison with existing methods

Table 2. Comparisons on OPIXray, where FO, ST, SC, UT and MU denote “Folding Knife”, “Straight Knife”, “Scissor”, “Utility Knife” and “Multi-tool Knife”.

models	OPIXray					mAP
	FO	ST	SC	UT	MU	%
DOAM [3]	86.7	68.5	90.2	78.8	87.6	82.4
	1	8	3	4	7	1
CHR [9]	87.9	84.5	95.2	50.9	74.4	78.6
	4	3	3	9	7	3
FBS [10]	86.3	88.2	95.4	57.9	80.6	81.7
	8	9	5	9	2	5
ATSS [4]	87.7	74.9	97.6	85.7	90.2	88.2
	2	9	0	0	6	6
Faster RCNN [5]	88.7	77.5	90.1	86.2	89.7	86.4
	2	9	0	5	8	9
FAPID [11]	89.8	84.2	90.2	88.0	89.6	88.4

	3	6	8	0	0	0
Ours	91.1	85.3	93.1	90.7	90.2	90.1
	6	3	7	3	7	3

Comparisons on the OPIXray dataset: In Table 2, our approach achieved superior performance. With the highest overall mAP⁵⁰ of 90.13, it outperforms the other models. Notably, it achieves top precision in FO (91.16), UT (90.73), and MU (90.27), showcasing its effectiveness across various categories. Comparatively, while models like FBS show strong results in certain categories, such as ST and SC, they do not consistently maintain this high level of precision across all categories. Furthermore, compared to the standard Faster RCNN, our method demonstrates improvements across all categories to varying degrees, particularly in the weaker categories of the original model (ST), where it shows a notable increase of 7.74%. This performance is even 1.07% higher than methods that reconstruct features, such as FAPID.

Table 3. Comparisons on HiXray, where PO1, PO2, WA, LA, MP, TA, CO and NL denote “Portable Charger 1”, “Portable Charger 2”, “Water”, “Laptop”, “Mobile Phone”, “Tablet”, “Cosmetic”, and “Nonmetallic Lighter”. **Bold** indicates the best performance overall, while underline denote the best performance within the same baseline.

Ba se- lin e	Models	HiXray								mA P%
		PO 1	PO 2	WA	LA	MP	TA	CO	NL	
-	SCM[13]	96. 0	95. 0	93. 9	98. 3	98. 5	95. 8	65. 6	20. 0	83. 2
YO LO v5s [12]	YOLOv5s	95.	94.	92.	97.	98.	94.	63.	16.	81.
		5	5	8	9	0	9	7	3	7
	DOAM	95.	94.	93.	98.	98.	95.	65.	16.	82.
		9	7	7	1	1	8	0	1	2
	LIM	96.	95.	<u>93.</u>	98.	<u>98.</u>	96.	65.	21.	83.
		1	1	<u>8</u>	2	<u>3</u>	<u>4</u>	8	3	2
ZPGNet	95.	95.	92.	96.	97.	94.	66.	33.	83.	
		7	2	5	5	7	4	4	0	9

	Ours	<u>96.</u>	<u>95.</u>	93.	<u>98.</u>	98.	95.	<u>68.</u>	<u>34.</u>	<u>85.</u>
		<u>6</u>	<u>4</u>	1	<u>3</u>	3	7	<u>4</u>	<u>0</u>	<u>0</u>
Faster RCNN	Faster RCNN	89.	88.	<u>88.</u>	90.	90.	<u>89.</u>	65.	27.	78.
		8	4	<u>9</u>	2	1	<u>9</u>	7	9	9
Faster RCNN	FAPID	89.	89.	88.	89.	90.	88.	63.	36.	79.
		7	1	2	5	0	8	3	5	4
Faster RCNN	Ours	<u>90.</u>	<u>88.</u>	88.	<u>90.</u>	<u>90.</u>	89.	<u>70.</u>	<u>66.</u>	<u>84.</u>
		<u>5</u>	<u>7</u>	4	<u>7</u>	<u>4</u>	3	<u>3</u>	<u>2</u>	<u>3</u>

Results on the HiXray dataset: Table 3 presents a comprehensive comparison of our method (Ours) against various improved strategies based on YOLOv5s and Faster RCNN models, evaluated on the HiXray dataset. By examining the $mAP@_{50}$ and the precision across different categories, our approach demonstrates superiority on multiple fronts. Specifically, our method achieves the highest $mAP@_{50}$ scores of 85.0 (post-YOLOv5s improvements) and 84.3 (post-Faster RCNN improvements), outperforming the best YOLOv5s enhancement, ZPGNet, by 1.1 percentage points, and surpassing the top Faster RCNN enhancement, FAPID, by 4.9 percentage points. This underscores our method's exceptional comprehensive performance. In terms of individual categories, our approach shows remarkable competitiveness in PO1 (96.6%), PO2 (95.4%), LA (98.3%), MP (98.3%), CO (68.4%), and NL (34.0%), notably in the typically challenging CO and NL categories. Here, our method exceeds the next best competitor, ZPGNet, by 1.8 and 0.6 percentage points, respectively. These results not only highlight our method's capability in identifying features in clearer categories but also, more importantly, its significant advantages in detecting categories with less obvious features. This advantage is more pronounced in our improved methods based on Faster RCNN. Compared to the baseline Faster RCNN model, our method demonstrates improvements across all categories, especially in the challenging CO and NL categories, with a huge performance boosts of 4.6 and 28.3 percentage points, respectively. This significantly enhances the model's ability to recognize occluded and overlapped items.

3.3 Ablation studies

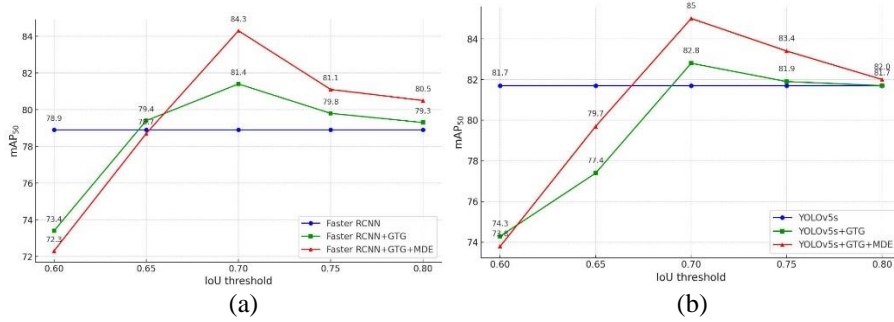


Fig. 4. Results (mAP%) achieved by different module combinations and IOU threshold settings. Figure (a) shows the ablation experiments on method based on Faster RCNN, while (b) presents the experiments on method based on YOLOv5s. The addition of the MDE module is predicated on the inclusion of the GTG module; the MDE module cannot be used in isolation.

Table 4. The average number of connected node under different baseline model.

IoU threshold	0.6	0.65	0.7	0.75	0.8
Faster RCNN	3.11	1.82	1.17	1.06	1
YOLOv5s	3.42	1.75	1.21	1.05	1

Contributions of different modules under different IOU threshold settings. As illustrated in Figure 4 and Table 4, the best performance is achieved when the IoU threshold for connecting boxes is set to 0.7, resulting in a 2.5% improvement with the addition of just the graph topology structure, and further increasing to 84.3% with the inclusion of multidimensional edges on Faster RCNN based model. Moreover, in the experiments on method based on YOLOv5s, incorporating the GTG module leads to an improvement of 1.1%, and further addition of the MDE module results in an enhancement of 3.3%. At this point, the average number of connected boxes is 1.17 and 1.21 respectively, suggesting that each box was average connected to 0.17 and 0.21 other boxes in addition to itself. When the IoU threshold is increased to 0.75, both the addition of the GTG module and the further inclusion of the MDE module exhibit a decrease in performance compared to that at an IoU threshold of 0.7. When the IoU threshold is set to 0.8, the average number of connected boxes was reduced to 1, indicating that each box was only connected to itself. This result suggests that considering only the individual box, without taking into account information from other boxes, does not yield the best results in X-ray image detection. As the IoU threshold is lower to 0.65 and 0.6, the average number of connected boxes continue to increased, and performance began to noticeably decline with the addition of more connected boxes. This indicates that as the number of connections increases, more interference is introduced, diminishing the model's performance. It's evident that the model with multidimensional edges experienced

a more significant performance decrease with an increase in connected boxes. We speculate this is because multi-dimensional edges amplify interference from irrelevant nodes, leading to a further decline in model performance.

Visualization.

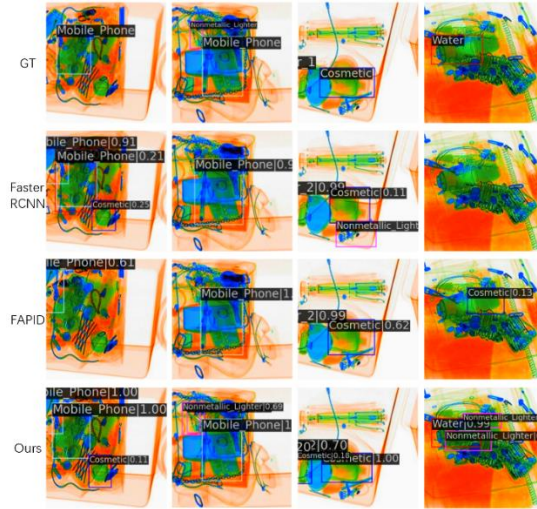


Fig. 5. A figure caption is always placed below the illustration. Short captions are centered, while long ones are justified. The macro button chooses the correct format automatically.

To validate the efficacy of our method in mitigating the challenges posed by translucency occlusions, we have presented visualizations of detection results across various categories on the HiXray dataset. As seen in Figure 5 our method outperforms the baseline model and other approaches, especially in scenarios with severe item overlap. In all four columns of the figure, which depict items heavily occluded, other methods either fail to detect them or yield detections with very low confidence. In contrast, our model accurately identifies these items with exceptionally high confidence. Notably, as shown in the second column, our model can detect the Nonmetallic Lighter with a confidence score of 0.69, even though it is almost invisible to the human eye in the original image. Additionally, as evident from the fourth column, while our method does introduce some false positives, it is aligned with the stringent operational standards of security inspection, which prioritize minimizing missed detections while tolerating a certain level of false positives.

4 Conclusion

This paper explores a novel direction in addressing the unique translucency interference problem in X-ray images. To enhance the capability of learning more discriminative object representation, our approach focuses on the relationships between

targets, where an effective multi-dimensional edge-based graph representation learning approach is proposed. Our proposed GTG module generates a graph topology to connect the proposed boxes outputted by the detection network, while the MDE module generates multidimensional edges to further delineate the nodes' relationships. Extensive experiments demonstrate that our approach has significant advantages in X-ray security inspection detection despite the employed backbone is just a standard Faster RCNN and YOLOv5s. It not only enhances detection accuracy but also more effectively identifies hard-to-distinguish objects in complex scenarios. The main limitation of our experiments is that it only considered the IoU threshold between boxes as the condition for connection, without exploring other potential selection methods. Additionally, the introduction of a highly sparse adjacency matrix resulted in excessive VRAM usage by the model during computation.

References

- [1] Mery, D., Svec, E., Arias, M., Riffo, V., Saavedra, J. M., & Banerjee, S. (2016). Modern computer vision techniques for x-ray testing in baggage inspection. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(4), 682-692.
- [2] Mery, Domingo. "Computer vision technology for X-ray testing." *Insight-non-destructive testing and condition monitoring* 56.3 (2014): 147-155.
- [3] Tao, Renshuai, et al. "Over-sampling de-occlusion attention network for prohibited items detection in noisy x-ray images." *arXiv preprint arXiv:2103.00809* (2021).
- [4] Zhao, Cairong, et al. "Detecting overlapped objects in X-ray security imagery by a label-aware mechanism." *IEEE transactions on information forensics and security* 17 (2022): 998-1009.
- [5] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems* 28 (2015).
- [6] Song, Siyang, et al. "Gratis: Deep learning graph representation with task-specific topology and multi-dimensional edge features." *arXiv preprint arXiv:2211.12482* (2022).
- [7] Tao, Renshuai, et al. "Towards real-world X-ray security inspection: A high-quality benchmark and lateral inhibition module for prohibited items detection." *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.
- [8] Chen, Kai, et al. "MMDetection: Open mmlab detection toolbox and benchmark." *arXiv preprint arXiv:1906.07155* (2019).
- [9] Miao, Caijing, et al. "Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.
- [10] Shao, Fangtao, et al. "Exploiting foreground and background separation for prohibited item detection in overlapping X-Ray images." *Pattern Recognition* 122 (2022): 108261.
- [11] Liao, Hongyu, Bin Huang, and Hongxia Gao. "Feature-Aware Prohibited Items Detection for X-Ray Images." *2023 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2023.
- [12] Jocher, Glenn, et al. "ultralytics/yolov5: v5.0-YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations." *Zenodo* (2021).
- [13] Liu, Dongsheng, et al. "Handling occlusion in prohibited item detection from X-ray images." *Neural Computing and Applications* 34.22 (2022): 20285-20298.