Remote Sensing Image Scene Graph Generation Method Based on Knowledge Graph Enhancement and Relationship Filtering

Yu Geng [1], Jingguo Lv [1]*

[1] School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing,100044, China - lvjingguo@bucea.edu.cn

Keywords: Remote sensing; Knowledge graph, Relationship recognition; Scene graph generation.

Abstract

The generation of scene maps of remote sensing images is very important for the understanding of image depth, but its development is limited by the characteristics of wide image size, significant changes in target scale and dense distribution. In this paper, a scene graph generation method based on knowledge graph enhancement and relationship filtering is proposed. Firstly, the method takes the Reltr model as the backbone framework, constructs and integrates the domain knowledge graph, and uses the typical spatial relationships (such as orientation and topology) and semantic associations (such as functional constraints) of the encoded remote sensing targets as structured prior knowledge to guide the model to understand the more likely reasonable relationship patterns between targets. Secondly, the semantic information entropy mechanism was introduced to calculate the information entropy value of the probability distribution of the relationship class predicted by the model to quantify the uncertainty of the prediction. Based on this, an adaptive threshold is set to effectively filter and suppress low-confidence fuzzy or erroneous relationship predictions, focusing on mining deep complex relationships with high certainty between targets. Experiments on the STAR dataset show that the accuracy of the proposed method in the task of remote sensing scene map generation is significantly improved: the accuracy of the R@100, R@200, and R@500 reaches 23.1%, 25.6%, and 26.1%, respectively, and the mR@100, mR@200, and mR@500 reach 13.1%, 15.6%, and 17.1%, respectively, and the accuracy is better than that of the existing algorithms. This verifies the effectiveness of the method and provides a new and effective solution for the scene understanding of remote sensing images.

1. Introduction

Scene graph generation refers to the automatic generation of three tuple detection results that describe the detection results of image scene target box and a series of target relationships, namely scene graph, according to the input remote sensing image. Scene graph generation mainly includes two core stages: target detection and relationship prediction. Firstly, in the target detection phase, the salient instance objects and their spatial positions in the image are identified, which provides the basis for subsequent analysis. Then, in the relationship prediction stage, based on the target instance detected above, the semantic or spatial relationship (such as "building side road") between any two related targets is inferred and predicted, and a specific (subject, relationship, object) triple is generated. Finally, the series of triples from the relationship prediction output together constitute the graph structure representation of the image scene, that is, the scene graph. Scene maps can be used in downstream tasks such as image retrieval (Johnson, 2015) , natural language description (Chang, 2021) and visual question answering (Li, 2024a) , and have immeasurable application prospects.

Visual genome (Krishna, 2017) in the field of natural images is a well-known dataset that annotates natural images with scene icons. Lu (Lu, 2016) introduced the language priori and decoupling training paradigm for the first time in their pioneering work. By independently modeling the visual features of subject/object and predicate, and using the semantic constraints of word vector space to solve the problem of long tail relationship distribution, they laid a theoretical foundation for the generation of modern scene maps. 2017 ushered in Architecture Innovation: Dai (Dai, 2017) proposed the deep relational network to build an end-to-end differentiable relational reasoning module and realize high-order interaction modeling between objects; The iterative message passing framework (IMP) developed by Xu (Xu, 2017) gradually optimizes the joint representation of the target node and the relationship edge through multi round graph neural network message propagation, significantly improving the structured reasoning ability. Tang (Tang, 2020) proposed a bias correction mechanism based on causal intervention to effectively balance the prediction accuracy of the relationship between high frequency and rarity in view of the semantic bias in training data; Zarian (Zareian, 2020) innovatively introduced external knowledge mapping, used graph neural network to bridge visual scene and common sense knowledge, realized multi-hop semantic reasoning across modes, and promoted the breakthrough of scene graph generation task. After 2021, the research will deepen in the direction of high efficiency and robustness: Liu(Liu, 2021)'s full convolution scene graph generation abandoned the traditional region proposal mechanism and directly generated triples through pixel level prediction, which increased the reasoning speed by three times; Suhail (Suhail,, 2021) pioneered the energy learning paradigm to transform relationship discrimination into an energy function optimization problem, significantly enhancing the robustness of the model to fuzzy relationships. Recently, the relational transformer (Reltr) designed by Cong (Cong, 2023) replaces graph convolution with multiple heads' self attention, and uses Recall@50 Up to 38.6% refresh SOTA; Yang (Yang, 2023) further proposed panoramic video scene graph generation, modeling cross frame target trajectories through spatio-temporal graph convolution network, and solving the problem of temporal consistency of dynamic scenes; Im (Im, 2024) designed a lightweight Graph Extraction header to share parameters at the transformer coding layer to achieve real-time reasoning, paving the way for industrial applications.

In the field of remote sensing, due to the lack of relevant data sets, the task of generating remote sensing image scene map

was still blank in the early years. In recent years, some scholars gradually began to pay attention to the task of remote sensing image scene map generation. Chen (Chen, 2021) proposed a three tuple representation method based on message passing, which is used for reasoning the relationship between geographical objects in high-resolution remote sensing images. This research laid the foundation for the generation of remote sensing scene map. By constructing an effective information transmission mechanism, the relationship between geographical objects can be more accurately captured and represented. On this basis, Lin (Lin, 2022a) further proposed SRSG and s2sg models, which made a further breakthrough in the task of scene map generation by integrating segmentation results and graph generation technology, and expanded the method framework of remote sensing image scene map generation. In the same year, Lin (Lin, 2022b) further optimized the effect of remote sensing image scene map generation by fusing context information and statistical knowledge. Li (Li, 2024b) released the star dataset, providing the first large-scale benchmark dataset for the generation of scene maps of large-scale satellite images, which greatly promoted the research process in this field. In addition, the semantic relationship model and data set for remote sensing scene understanding are constructed, which provides important resources for in-depth study of the semantic relationship of remote sensing images; Tang (Tang, 2025) focused on improving the performance of remote sensing scene map generation and retrieval by using spatial relationships. These studies complement each other, and jointly promote the continuous progress and development of remote sensing image scene map generation technology, making the understanding and application ability of remote sensing image continuously improved.

However, in the process of generating remote sensing scene map, there are mainly two core stages: target detection and relationship prediction. As one of the core links, relationship prediction still has many limitations. First of all, the existing models are not deep enough in semantic understanding, and it is difficult to accurately capture and express the complex semantic associations between geographical objects in remote sensing images. The context information is not fully utilized, and the models fail to fully mine and utilize the spatial layout, interaction and other key context features of geographical objects in remote sensing images, which affects the accuracy of relationship prediction. Secondly, there is the problem of relationship redundancy, that is, the model often produces a large number of repeated or unnecessary relationships when extracting the relationships between geographical objects, which not only increases the computational burden, but also reduces the accuracy and interpretability of the scene map. However, the existing research lacks an effective mechanism in relation selection, and it is difficult to effectively identify and remove redundant relations, which leads to the limited quality of the generated scene map, and it is difficult to meet the needs of precise semantic relation representation in complex remote sensing images.

In view of this, this paper proposes a remote sensing image scene map generation method based on knowledge map enhancement and relationship screening. This method is based on the Reltr model (Cong, 2023)and introduces the knowledge map and semantic information entropy, which significantly enhances the recognition ability of the model for the complex relationship between targets, thus effectively improving the accuracy and quality of remote sensing image scene map generation, and providing strong support for the in-depth understanding and analysis of remote sensing images. The innovations of this paper are as follows:

(1) This paper proposes a semantic enhanced relationship prediction module based on knowledge map, which introduces knowledge map to provide rich semantic information and prior knowledge for Reltr model, effectively enhances the model's ability to understand the complex relationships between geographical objects, fully excavates and uses the key context features of geographical objects' spatial layout and interaction in remote sensing images to improve the accuracy of relationship prediction, and widens the research idea of remote sensing scene map generation.

(2) This paper introduces semantic information entropy to quantify the uncertainty of relationship, innovatively constructs the relationship screening mechanism, solves the problem of relationship redundancy in remote sensing scene map generation, optimizes the structure of scene map, and improves the efficiency and performance of the model, which is of great significance to improve the quality of remote sensing image scene map generation.

The schedule of this paper is as follows: Section 2 introduces the related work, section 3 describes the principle of the method, section 4 analyzes the experimental results, and section 5 gives the conclusion.

2. Related work

2.1 Reltr Model

Reltr model, short for Relational Transformer Model, is a deep learning model for relational prediction. It is widely used in the computer vision field for scene graph generation task (Cong, 2023), which effectively captures the complex relationships between target objects in a visual scene by introducing an attention mechanism. In the scene graph generation task, the Reltr model is able to generate more accurate and richer semantic relationship graphs by combining target detection results and contextual information.

Shao (Shao, 2022) achieved the perception of image region-object relationships by introducing the Transformer architecture, which improved the accuracy and richness of the dense description of the image. Gao (Gao, 2023) further extended the application of the Transformer architecture by combining relational modelling with target tracking, which showed that the Transformer architecture's ability to deal with various relationships in images has been further explored and optimised, providing a useful reference for the integration of scene graph generation techniques with dynamic vision tasks. In the same year, Cong (Cong. 2023) proposed the Reltr model (i.e., Relation Transformer), which innovatively uses the Transformer architecture to model the relationships in scene graphs, capturing the complex interactions and semantic associations between the objects in the images more efficiently, and significantly improving the accuracy and completeness of scene graph generation, which becomes an It significantly improves the accuracy and completeness of scene graph generation, and becomes an important milestone in this field, which promotes the development of natural image scene graph generation technology in the direction of more refinement and intelligence.

Although the Reltr model has excellent performance in natural image scene graph generation, it still faces significant

challenges in remote sensing image processing. The inherent characteristics of remote sensing imagery, such as the vast size of the area, the drastic change of target scale, and the dense distribution of targets, make it difficult to directly migrate existing methods based on the design of natural imagery to the application. The aim of this study is to explore the improvement potential of the Reltr model and provide new solutions for tasks such as remote sensing scene map generation.

2.2 Knowledge Graph

In the cutting-edge research on knowledge graph-assisted remote sensing scene map generation, the introduction of knowledge graph brings a revolutionary meaning of semantic understanding and structured representation to the field. From the early introduction of basic concepts to the expansion of today's complex applications, knowledge graphs provide a powerful structured representation of complex semantic relationships in remote sensing images, enabling models to more accurately identify and understand geographic entities and their interrelationships in remote sensing images.

Chen (Chen, 2019) proposed knowledge embedded routing network for the first time, which significantly improved the accuracy of scene graph generation by introducing external semantic information through knowledge mapping.Zareian (Zareian. 2020) further explored the fusion method of knowledge mapping and scene graph generation, which strengthened the semantic associations through knowledge mapping.Yu (Yu, 2022) proposed zero sample scene graph generation method, which extends the semantic boundary by using knowledge graph complementation technology and opens up a new path for less sample scene graph generation. Wang (Wang, 2024) proposed a knowledge-enhanced context representation method, which optimises the effect of the application of knowledge graph in the generation process.

3. Methodology

To address the issues of insufficient semantic understanding and relationship redundancy in the generation of scene graphs from remote sensing images, this paper proposes a network framework based on the Reltr model, embedding a knowledge graph from the remote sensing field.

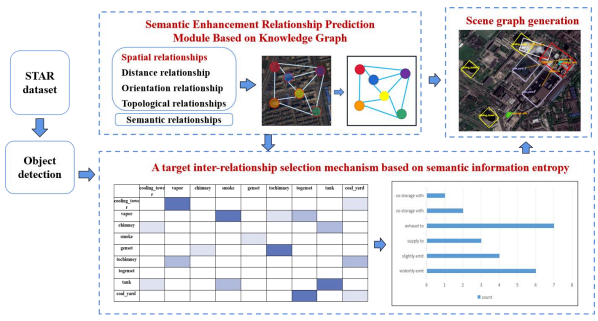


Figure 1. Technical flowchart

This approach utilizes the spatial and semantic prior information of remote sensing targets from the knowledge graph to guide relationship prediction, enhancing semantic understanding and making full use of contextual features. Additionally, a semantic information entropy mechanism is introduced to quantify the uncertainty in relationship predictions and establish adaptive thresholds for filtering high-confidence relationships, thereby suppressing redundant relations. The main process is illustrated in Figure 1.

3.1 Semantic Enhancement Relationship Prediction Module Based on Knowledge Graph

In order to enhance the semantic understanding capability of relation prediction in remote sensing scene generation, this paper proposes a knowledge graph-based semantic enhancement relation prediction module. This module systematically organizes the semantic and spatial prior information among targets in the field by constructing a remote sensing knowledge graph that includes categories of remote sensing objects, attribute features, spatial relationships, and other relevant content. Simultaneously, the knowledge graph is embedded into the Reltr model, guiding the model to better capture the semantic associations and spatial interaction patterns between geographic objects in relation prediction through the fusion of graph entities and relation vectors. By integrating knowledge graph priors, this module effectively enhances the model's ability to understand the complex relationships in remote sensing images, thereby improving the accuracy and robustness of relation prediction.

3.1.1 The construction of a knowledge graph for remote sensing

Unlike natural graphs, which mainly rely on the extraction of entities and relationships from general corpora and image scenes, the construction of remote sensing knowledge graphs faces problems such as multi-source heterogeneous data fusion, large changes in target scale, ambiguous category semantics, and complex spatial layout. The construction of remote sensing knowledge graph usually includes three steps: entity extraction, relationship definition, and graph triplet generation. Firstly, based on the artificial annotation or detection results in the remote sensing image, the target entity set $O=\{o_1,o_2,...o_N\}$ was extracted, in which each entity o_i represented a remote sensing target object, including its category label $c_i \in C$, spatial location $l_i=(x_i,y_i,w_i,h_i)$ (representing the center point coordinates and width and height, respectively) and attribute feature vector $a_i \in \Re^d$. Then, the possible set of semantic and spatial relation $R=\{\mathrm{r}_1,\mathrm{r}_2,...,\mathrm{r}_\mathrm{K}\}$ between the targets is defined, and the correlation degree of any entity to the relation (o_i,o_j) between r_k is calculated by the following relation-scoring function, as shown in equation (1):

$$s_{ijk}=f_r(o_i,o_j,r_k)=\alpha\cdot sim(a_i,a_j)+\beta\cdot IoU(l_i,l_j)+\gamma\cdot P(r_k|c_i,c_j) \quad (1)$$

When, S_{iik} exceeds the set threshold τ , the structured triad (o_i,r_k,o_j) can be generated as the edge constituents in the knowledge graph. The remote sensing knowledge graph constructed through this process not only encodes the target categories and attribute semantics, but also explicitly fuses the spatial layout and relational a priori, forming a structured knowledge support oriented to remote sensing scene graph construction.

Where s_{ijk} = the score of entity pair under (o_i,o_j) relationship type.

$sim(a_i,a_j)$ = the semantic similarity between entity attribute features.

$IoU(l_i,l_j)$ = the degree of spatial overlap of entities to quantify positional proximity.

$P(r_k|c_i,c_j)$ = the a priori probability of occurrence of relationship under the combination of target categories (c_i,c_j).

α,β,γ = the weight coefficient, which is used for balancing the information of different dimensions.

3.1.2 Embedding of remote sensing knowledge graph

After completing the construction of the remote sensing knowledge graph, it needs to be further embedded into the relationship prediction model to fully utilize the spatial a priori and semantic associations embedded in the graph, and to assist in the accurate modeling and reasoning of complex relationships in the remote sensing scene graph. The embedding of knowledge graph is essentially to map the discrete graph entities and relationship triples into a low-dimensional continuous vector space representation, aiming to maintain the consistency of the graph structure and semantic information. Let the constructed remote sensing knowledge graph be represented as set $T=\{(o_i,r_k,o_j)\}$ of triples, where $o_i,o_j\in O$ denotes the remote sensing target entities and $r_k\in R$ is the semantic or spatial relationship between them. In order to realize the embeddability of the atlas, the *TransE* model is introduced for vector modeling of the ternary set, whose basic assumption is that the relationship is equivalent to the translation between entities, as shown in Equation (2):

$$e_{o_i}+e_{r_k}\approx e_{o_j} \tag{2}$$

Where e_{o_i},e_{o_j} = the embedding vector of entities o_i and o_j.

$e_{r_k}\in R^d$ = the embedding vector of relation r_k.

d = the dimension of the embedding space.

In the training phase, embedding optimization is achieved by minimizing the following scoring function as shown in Equation (3):

$$L_{KG}=\sum_{(o_i,r_k,o_j)\in T}\left[\begin{array}{l}\left\|e_{o_i}+e_{r_k}-e_{o_j}\right\|_2^2 \\ +\sum_{(o_i,r_k,o_j')\in T}\max(0,\gamma+\left\|e_{o_i}+e_{r_k}-e_{o_j}\right\|_2^2 \\ \qquad\qquad -\left\|e_{o_i}+e_{r_k}-e_{o_j'}\right\|_2^2)\end{array}\right] \tag{3}$$

where T^- = the set of constructed negative sample triples.

γ = the set interval hyperparameter.

After completing the graph embedding, the obtained entity embedding and relationship embedding can be used as external a priori knowledge, which can be fused with the visual features extracted from remote sensing images through the feature splicing or attention mechanism to assist the context modeling process of the relationship decoder in the Reltr model. Eventually, the structural and semantic information of the knowledge graph can be deeply integrated into the visual relational reasoning, providing strong knowledge support and semantic guidance for the remote sensing scene graph generation task.

3.2 A target inter-relationship selection mechanism based on semantic information entropy

On the basis of integrating remote sensing knowledge map to enhance semantic understanding and improve the accuracy of relationship prediction, it is still necessary to deal with the problem of relationship redundancy that exists in the generation process of remote sensing scene graph. In order to further optimize the structure of the scene graph and improve the efficiency of model inference and the quality of graph expression, this paper introduces a semantic information entropy-based inter-target relationship screening mechanism to quantify the uncertainty of the predicted relationships, sift out the low-confidence redundant relationships, and retain the semantically clear and representative relationship structure. Specifically, let the relationship distribution predicted by the model for any target entity pair (o_i,o_i) be the probability vector $p_{ii}=\left[p_{ii}^{(1)},p_{ii}^{(2)},...,p_{ii}^{(K)}\right]\in[0,1]^K$, According to the principle of information theory, the semantic information entropy of this predicted distribution is defined as shown in Eq. (4):

$$H_{ij}=-\sum_{k=1}^{K}p_{ij}^{(k)}\log p_{ij}^{(k)} \tag{4}$$

Where K = the total number of relationship categories.

$p_{ij}^{(k)}$ = the probability that the entity pair is predicted to be a relationship of the k^{th} category, which satisfies $\sum_{k=1}^{K}p_{ij}^{(k)}=1$.

H_{ij} = the predicted uncertainty of the relationship of the target to (o_i,o_j).

The higher value of H_{ii} indicates that the relationship prediction is more decentralized and less discriminative, and vice versa represents more centralized and credible prediction. To realize adaptive relationship screening, the threshold function is further defined as shown in Equation (5):

$$\tau_{ij}=\mu\cdot H_{\max}+(1-\mu)\cdot\bar{H} \tag{5}$$

Where $H_{\max}=\log K$ = the theoretical maximum entropy.

$\bar{H}$ = the average entropy of the current batch prediction relationship.

$\mu\in[0,1]$ = the weight factor that controls the threshold bias.

Eventually, all the target-pair relations that satisfy $H_{ii}<\tau_{ii}$ are retained as a set of high-confidence relations for scene graph construction. This mechanism can not only explicitly suppress invalid or semantically ambiguous relationship connections and reduce the redundancy of the graph structure, but also improve the efficiency and robustness of the overall relational reasoning, providing a semantically clear and structurally simple high-quality basis for the generation of remote sensing scene graphs.

4. Experimental Results and Analysis

4.1 Dataset and Experimental Configuration

This article utilizes the publicly available STAR dataset, which has a spatial resolution ranging from 0.15 meters to 1 meter, as shown in Figure 2. This dataset is specifically designed for studying multi-target semantic relationship modeling and scene graph generation tasks in remote sensing images. It consists of high-resolution remote sensing images that cover various typical geographical scenes such as airports, ports, wind farms, nuclear power plants, thermal power plants, construction sites, stadiums, service areas, toll stations, traffic bridges, and dams. It contains over 210,000 target objects and 400,000 relational triples, as illustrated in Figure 3, with image sizes ranging from 512×768 to 27860×31096 pixels, featuring a rich variety of land cover target types and complex spatial interaction relationships.

Figure 2. Sample of the STAR dataset

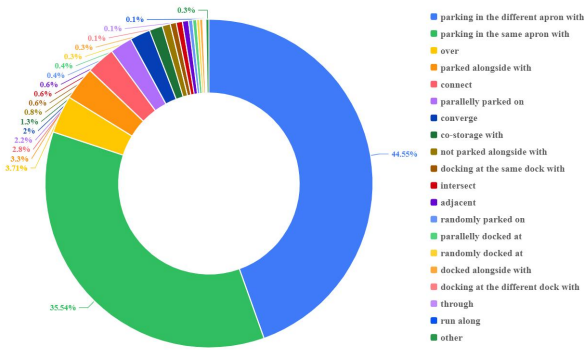


Figure 3. Classification of relational labels

4.2 Experimental results

4.2.1 Evaluation indicators

This experiment continues to verify the effectiveness of large-scale remote sensing image scene graph generation methods through three sub-tasks used in natural image scene graph generation. The three sub-tasks are: Predicate Classification (PredCls), which requires predicting the relationships between objects based on their positions and categories in the given image; Scene Graph Classification (SGCls), which involves predicting object categories and relationships based on object positions in the given image; and Scene Graph Generation (SGGen), which requires generating a scene graph solely from the given image, predicting object positions, categories, and relationships.

In the task of generating remote sensing scene maps, in order to evaluate the performance of the model in relation prediction, $Recall@K$ ($R@K$) and $Mean\ Recall@K$ ($mR@K$) are often used as the core evaluation indicators. $Recall@K$ is used to measure whether the model successfully overwrites the correct items in the labeled relationship in the top K most likely relationships predicted, reflecting the model's ability to identify significant relationships. Suppose there are M true relationship triples in an image, each triplet is in the form (o_i, r_k, o_j), and the model outputs the prediction with the highest confidence in the first K in the candidate relationship set for each pair of entities (o_i, o_j) as $\hat{R}_{ij}^{(K)}$, then the $Recall@K$ definition, as shown in equation (6):

$$R@K = \frac{1}{M} \sum_{(o_i, r_k, o_j) \in \tau} \mathrm{II}\left[r_k \in \hat{R}_{ij}^{(K)} \right] \tag{6}$$

Where τ =the set of true triples labeled

$\mathrm{II} \cdot [\ ]$ =the indicator function

The value is 1 when the true relation r_k exists in the first K predictions, otherwise it is 0.

In contrast, the $Mean\ Recall@K$ ($mR@K$) pays more attention to the equilibrium prediction ability of each relationship category, which is especially suitable for alleviating the bias problem caused by the uneven distribution of remote sensing relationship categories. Specifically, for each relation category $r_k \in R$, the $Recall@K$ of that category is calculated separately and denoted as $R \in K_{r_k}$, and then the arithmetic mean is taken for all relation categories, as shown in Eq. (7):

$$mR@K = \frac{1}{|R|} \sum_{r_k \in R} R@K_{r_k} \tag{7}$$

Where $|R|$ = the total number of relationship categories

This metric effectively measures the model's generalization ability in long-tail relationships and the overall predictive balance, thus holding significant evaluative value in remote sensing scene generation tasks.

For all tasks, only those triple predictions that are fully consistent with the labels <subject, relation, object> and have an Intersection over Union (IoU) greater than 0.5 between the subject and object targets and their bounding box annotations will be considered correct predictions. Both of the above metrics are used to evaluate the proportion of completely correct predicted triples within the image.

4.2.2 Comparison results of different methods

In order to explore the effectiveness of the proposed method, the mainstream remote sensing scene graph algorithms in this paper include the message-passing-driven triplet representation (MP-Triplet) (Chen, 2021)and the segmentation-based model to generate remote sensing image scene graphs (SRSG) (Lin, 2022a), Remote Sensing Scene Graph Generation by Fusing Contextual Information and Statistical Knowledge (RSSGG_CS) (Lin, 2022b) , The Remote Sensing Scene Graph Generation for Improved Retrieval Based on Spatial Relationships (IRSR)

(Tang, 2025) and Relation Transformer (Reltr) (Cong, 2023) were compared experimentally, as shown in Tables 1 and 2.

Method	PredCls			SGCls			SGGen		
	R@100	R@200	R@500	R@100	R@200	R@500	R@100	R@200	R@500
MP-Triplet	28.1	36.5	38.9	24.7	31.8	35.8	14.3	17.6	18.9
SRSG	31.8	42.3	44.7	27.5	35.6	41.2	17.2	20.1	21.4
RSSGG_CS	33.5	47.2	49.6	29.8	39.1	45.3	19.3	22.4	23.7
IRSR	35.2	50.8	53.1	31.6	42.7	49.5	21.3	24.3	25.5
Reltr	30.3	46.3	49.4	28.9	36.4	40.6	16.8	21.1	22.3
Ours	38.6	55.9	57.3	34.6	46.2	52.9	23.1	25.6	26.1

Table 1. Comparison results $R@K$ different methods (unit: %)

Method	PredCls			SGCls			SGGen		
	mR@100	mR@200	mR@500	mR@100	mR@200	mR@500	mR@100	mR@200	mR@500
MP-Triplet	15.2	20.8	24.1	10.7	16.3	19.5	8.2	10.3	11.8
SRSG	16.9	22.1	25.8	11.8	17.9	21.2	9.4	11.8	13.1
RSSGG_CS	19.4	26.7	30.1	15.6	22.3	26.8	11.7	14.5	16.4
IRSR	20.3	28.1	32.5	16.9	24.6	28.7	12.6	15.1	16.8
Reltr	18.6	25.2	28.3	13.9	20.5	24.3	10.8	13.4	15.0
Ours	21.5	30.6	34.8	18.3	26.7	31.2	13.1	15.6	17.1

Table 2. Comparison results $mR@K$ different methods (unit: %)

The experimental results presented in Tables 1 and 2 demonstrate that the method proposed in this paper has achieved significantly superior performance in all three sub-tasks of remote sensing scene generation (PredCls, SGCls, SGGen), validating the effectiveness and advancement of the proposed model in semantic understanding and relational modeling. In the PredCls task, without predicting the location of targets, the proposed method achieved 38.6%, 55.9%, and 57.3% at R@100, R@200, and R@500, respectively, exceeding the existing best comparative method IRSR by approximately 3 percentage points; furthermore, in terms of the more generalized mR@100 and mR@500, the results reached 21.5% and 34.8%, surpassing all comparative models. This indicates that the model has a stronger discriminative capability in capturing complex semantic relationships.

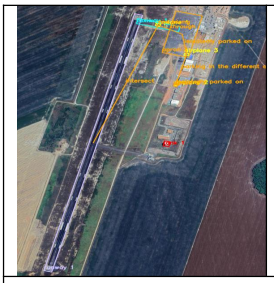
In the SGCls task, the model is required to recognize both the target categories and their relationships simultaneously. The method proposed in this paper still achieves 52.9% and 31.2% in R@500 and mR@500 respectively, both of which are the highest values, indicating stable performance in multi-task joint modeling.

Especially in the most challenging SGGen task, the method in this paper achieves 17.1% in mR@500, representing an increase of approximately 1.6% and 0.3% compared to KGC-SGG and KECR, respectively, demonstrating the robustness of this algorithm.

Overall, this paper significantly improves the overall performance of scene graph generation by introducing knowledge graphs to enhance semantic representation and combining it with an information entropy-based relationship selection mechanism, showcasing superior modeling capability and generalization potential in the identification and mapping tasks of complex object relationships in remote sensing images.

4.2.3 Results visualization

In order to visualize and qualitatively analyze the effect of the remote sensing scene map generation model in practical applications, this paper conducts visualization experiments in a number of typical remote sensing scenarios, covering representative geographical areas such as airports, parking lots, ports and power plants, aiming to demonstrate the target recognition and relationship prediction capabilities of the method in different complex environments. This is shown in Figure 4.

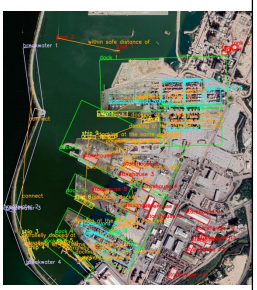


1) airplane run along taxiway
2)airplane isolatedly parked on apron
3)airplane isolatedly parked on apron
4) runway intersect taxiway
5) taxiway through apron
6) airplane parking in the different apron with airplane

Airport

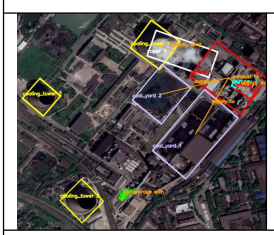
1) truck isolatedly parked on truck_parking
2) car parallelly parked on car_parking
3) car isolatedly parked on car_parking
4) car in the same parking with car
5) car not parked alongside with car
6) car in the different parking with car

Parking



1) ship isolatedly docked at dock
2) ship parallelly docked at dock
3)ship parallelly docked at breakwater
4) goods_yard adjacent goods_yard
5) storehouse over dock
6) breakwater connect breakwater
7) dock connect dock
8) boat within safe distance of boat
9) ship docking at the same dock with ship
10) ship docking at the different dock with ship
11) ship docking at the same breakwater with ship

Port



1) cooling_tower violently emit vapor
2) chimney slightly emit smoke
3) coal_yard supply to genset
4) genset exhaust to chimney
5) tank co-storage with tank

Power Plant

Figure 4. Visualization result

As shown in Fig. 4, the method in this paper achieves high-quality target detection and relationship prediction in multiple complex geographic scenarios (e.g., airports, parking lots, ports, and power plants), and demonstrates good semantic mapping ability and spatial structure expression. In the airport scenario, the model not only accurately identifies key targets such as runways, airplanes and terminals, but also effectively constructs typical spatial relationships such as “aircraft-parked-on-apron”, “runway-connects-to-taxiway”, and so on. It also effectively

constructs typical spatial relationships such as “aircraft-parked on-apron” and “runway-connects to-taxiway”, which demonstrates the model's ability to understand the functional and spatial neighbor relationships; in the map of the logistics park, the targets such as vehicles, containers, and parking zones are accurately detected, and the relationships such as “vehicle-parked-in-slot” and The relationships such as “vehicle-parked-in-slot” and “entrance-connected-to-road” form a clear traffic semantic structure; in the port scenario, the model successfully extracts the complex “dock-berths-ship”, In the port scenario, the model successfully extracts complex relationships such as “dock-berths-ship” and “stack area-near-warehouse”, and constructs a scenario diagram with hierarchical organization, which significantly reflects the organization and operation logic between spatial entities; the power plant diagram further demonstrates the advantage of the model in semantic scene recognition, and accurately predicts “cooling_tower-emit-vapor”, “coal_yard-supply-genset”, “tank-storage with -tank”, and “tank-storage with”, indicating that the model is not only capable of static structure perception, but also of portraying the semantic behavioral chains implicit in the scene. Taken together, the proposed method can generate semantically clear, well-structured and richly relational scene graphs in multiple types of remote sensing scenes, which provides effective support for the in-depth understanding and automated interpretation of remote sensing images.

4.3 Ablation experiments

An ablation experiment was also designed for this experiment. As shown in Tables 3 and 4, the specific contribution of each module is clarified through the semantic enhancement relationship prediction module with sequential elimination of knowledge graph (M_1) and the inter-target relationship screening mechanism based on semantic information entropy (M_2). The first line is the benchmark model: This is the traditional Reltr model.

M1	M2	PredCls			SGCls			SGGen		
		R@100	R@200	R@500	R@100	R@200	R@500	R@100	R@200	R@500
×	×	30.3	46.3	49.4	28.9	36.4	40.6	16.8	21.1	22.3
√	×	32.6	47.7	50.2	29.7	38.9	45.1	20.1	22.1	23.7
×	√	34.2	48.8	52.1	31.8	42.6	47.4	21.3	23.3	24.6
√	√	38.6	55.9	57.3	34.6	46.2	52.9	23.1	25.6	26.1

Table 3. R@K results of ablation experiments (unit: %)

M1	M2	PredCls			SGCls			SGGen		
		mR@100	mR@200	mR@500	mR@100	mR@200	mR@500	mR@100	mR@200	mR@500
×	×	18.6	25.2	28.3	13.9	20.5	24.3	10.8	13.4	15.0
√	×	18.8	26.7	29.1	15.7	22.4	26.9	11.9	13.8	15.6
×	√	19.3	27.1	30.5	17.9	24.1	28.6	12.8	14.7	16.8
√	√	21.5	30.6	34.8	18.3	26.7	31.2	13.1	15.6	17.1

Table 4. mR@K results of ablation experiments (unit: %)

The experimental results in Table 3 and Table 4 show that the knowledge graph and relationship screening modules are removed in this experiment, and the changes of R@K and mR@K indicators under the three types of subtasks (PredCls, SGCls, and SGGen) are observed. The results show that when neither module is enabled (i.e., the traditional Reltr model), the indicators are at the lowest level. The introduction of any module can bring stable improvement in each task, especially the relationship screening mechanism has a particularly obvious inhibitory effect on the low-confidence relationship in SGGen task, mR@500 from 15.0% to 16.8%. Furthermore, when the two are used together, the model performance is optimal: in the PredCls task, the mR@100, 200, and 500 reach 21.5%, 30.6%, and 34.8%, respectively. 18.3%, 26.7% and 31.2% in SGCls tasks; In the most challenging SGGen mission, the mR@K targets of 13.1%, 15.6% and 17.1% were also achieved. On the whole, the knowledge graph embedding module effectively enhances the model's ability to understand the target semantic relationship, while the information entropy-driven relationship screening mechanism improves the accuracy and robustness of relationship prediction.

5. Conclusion

Focusing on the problems of insufficient semantic understanding and relationship redundancy in remote sensing scene map generation, this paper proposes a remote sensing image scene map generation method based on knowledge graph enhancement and relationship screening. In this method, the semantic enhancement relationship prediction module of the knowledge graph and the relationship screening mechanism based on semantic information entropy are introduced, which effectively improves the expression ability and structure optimization level of the model in the relationship modeling of complex geographical objects. By constructing a remote sensing knowledge graph that integrates spatial prior and semantic knowledge, and embedding it into the relational decoder, the model's ability to construct semantic and spatial interaction between targets is significantly enhanced. At the same time, the information entropy mechanism is introduced to adaptively screen the high-confidence relationship and suppress the low-confidence redundancy relationship, which further improves the accuracy and structural rationality of the scene graph. Experimental results on the STAR remote sensing public dataset show that the proposed method is superior to the existing representative methods in multiple indicators of PredCls, SGCls and SGGen subtasks, which verifies the effectiveness of the method.

However, there are still two limitations in this paper: (1) the current work mainly focuses on relational modeling, and the design and optimization of the object detection module are not discussed in depth, and the object detection performance has a significant impact on the quality of scene graph generation; (2) The time efficiency and inference speed of the proposed method in large-scale remote sensing data processing have not been systematically analyzed, and the follow-up work will further explore the lightweight and real-time optimization strategies of the model.

References

Chang, X., Ren, P., Xu, P., 2021. A comprehensive survey of scene graphs: Generation and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 1 – 26.

Chen, J., Zhou, X., Zhang, Y., 2021. Message-passing-driven triplet representation for geo-object relational inference in HRSI. *IEEE Geoscience and Remote Sensing Letters*, 19, 1 – 5.

Chen, T., Yu, W., Chen, R., 2019. Knowledge-embedded routing network for scene graph generation. *In. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6163 – 6171.

Cong, Y., Yang, M.Y., Rosenhahn, B., 2023. Reltr: Relation transformer for scene graph generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9), 11169 – 11183.

Dai, B., Zhang, Y., Lin, D., 2017. Detecting visual relationships with deep relational networks. *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3076 – 3086.

Gao, S., Zhou, C., Zhang, J., 2023. Generalized relation modeling for transformer tracking. *In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18686 – 18695.

Im, J., Nam, J.Y., Park, N., 2024. Egtr. Extracting graph from transformer for scene graph generation. *In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 24229 – 24238.

Johnson, J., Krishna, R., Stark, M., 2015. Image retrieval using scene graphs. *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3668 – 3678.

Krishna, R., Zhu, Y., Groth, O.,, 2017. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 123, 32 – 73.

Li, H., Zhu, G., Zhang, L., 2024a. Scene graph generation: A comprehensive survey. *Neurocomputing*, 566, 127052.

Li, Y., Wang, L., Wang, T., 2024b. Star: A first-ever dataset and a large-scale benchmark for scene graph generation in large-size satellite imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(5), 6.

Lin, Z., Zhu, F., Kong, Y., 2022a. SRSG and S2SG: A model and a dataset for scene graph generation of remote sensing images from segmentation results. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1 – 11.

Lin, Z., Zhu, F., Wang, Q., 2022b. RSSGG_CS: Remote sensing image scene graph generation by fusing contextual information and statistical knowledge. *Remote Sensing*, 14(13), 3118.

Liu, H., Yan, N., Mortazavi, M., 2021. Fully convolutional scene graph generation. *In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11546 – 11556.

Lu, C., Krishna, R., Bernstein, M., 2016. Visual relationship detection with language priors. *In: Computer Vision – ECCV 2016: Proceedings, Part I* 14, 852 – 869.

Shao, Z., Han, J., Marnerides, D., 2022. Region-object relation-aware dense captioning via transformer. *IEEE Transactions on Neural Networks and Learning Systems*.

Suhail, M., Mittal, A., Siddiquie, B., 2021. Energy-based learning for scene graph generation. *In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13936 – 13945.

Tang, J., Tong, X., Qiu, C., 2025. Remote sensing scene graph generation for improved retrieval based on spatial relationships. *ISPRS Journal of Photogrammetry and Remote Sensing*, 220, 741 – 752.

Tang, K., Niu, Y., Huang, J., 2020. Unbiased scene graph generation from biased training. *In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3716 – 3725.

Wang, Y., Liu, Z., Zhang, H., 2024. Knowledge-enhanced context representation for unbiased scene graph generation. *In: Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*, 248 – 263.

Xu, D., Zhu, Y., Choy, C.B., 2017. Scene graph generation by iterative message passing. *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5410 – 5419.

Yang, J., Peng, W., Li, X., 2023. Panoptic video scene graph generation. *In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18675 – 18685.

Yu, X., Chen, R., Li, J., 2022. Zero-shot scene graph generation with knowledge graph completion. *In: 2022 IEEE International Conference on Multimedia and Expo (ICME)*, 1 – 6.

Zareian, A., Karaman, S., Chang, S.F., 2020. Bridging knowledge graphs to generate scene graphs. *In: Computer Vision – ECCV 2020: Proceedings, Part XXIII 16*, 606 – 623.