IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 14, 2021, accepted January 12, 2022, date of publication January 21, 2022, date of current version February 1, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3145465

Towards Open-Set Scene Graph Generation With Unknown Objects

MOTOHARU SONOGASHIRA[1], MASAAKI IIYAMA[2], (Member, IEEE), AND YASUTOMO KAWANISHI[1], (Member, IEEE)

[1]GRP, RIKEN, Seika, Kyoto 619-0288, Japan
[2]Faculty of Data Science, Shiga University, Hikone, Shiga 522-8522, Japan

Corresponding author: Motoharu Sonogashira (motoharu.sonogashira@riken.jp)

This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, through Grant-in-Aid for Scientific Research under Grant JP21H03519.

ABSTRACT Scene graph generation (SGG) aims to detect objects and their relationships in an image, thereby enabling a detailed understanding of a complex scene for various real-world applications. In SGG applications such as robot vision, it is important to correctly detect all objects without recognizing any object as another kind of object or ignoring it. However, previous studies on SGG do not consider unknown objects whose classes are unseen in training. Consequently, current SGG methods wrongly classify them as known object classes or overlook them. In this paper, we propose a new problem named ''open-set SGG'' with unknown objects, focusing on detecting even unknown objects and their relationships. Specifically, we formally define this new problem and propose an evaluation protocol, including an extended dataset with unknown objects and novel evaluation metrics designed for the open-set setting. We also build baseline methods by employing and extending existing SGG methods and compare them through experiments to establish the current baseline performance of open-set SGG. Finally, we discuss the limitations of the current SGG methodology in the open-set setting and point out future research directions.

INDEX TERMS Scene graph generation, open-set, object detection.

I. INTRODUCTION

Scene graphs are detailed descriptions of scenes by graphs consisting of objects as vertices and their relationships as edges [1]. Recently, the prediction of such a graph from an image [2] has been studied, which enables automated scene graph generation (SGG) from an image captured in the real world and thereby facilitates a detailed understanding of complex scenes. SGG has a wide range of applications such as image retrieval [1], visual question answering [3], human-robot interaction [4], and robot navigation [5].

For obtaining the complex mapping from an image to its scene graph, SGG methods rely on deep learning driven by big data, which uses a deep neural network as a prediction model and estimates its parameters from a large number of training images with ground-truth scene graphs. However, due to the difficulty of high-quality manual annotation [6], existing SGG datasets are limited in terms of the variety of the object classes with usable labels for training [2]. The difficulty results in the presence of unknown objects, which are absent in the training data. If such an object is present in the testing phase, the current SGG methods either classify it into one of the known object classes or completely fail to detect it by treating it as a part of the background, as shown in Fig. 1a. In practice, misclassified or overlooked objects lead to incorrect scene understanding and cause serious problems in applications; for example, if a robot recognizes an object as the different kinds of objects, it may take an inappropriate action on the object, or if it is unaware of the existence of the object, it may even exhibit dangerous behavior. Although such a problem setting of handling unseen and untargeted classes has been referred to as *open-set* [7] and addressed in tasks such as image recognition and object detection, it has never been tackled in the literature of SGG. It involves the detection of relationships as well as objects.

In this paper, we address the problem of *open-set SGG* with unknown objects. To the best of our knowledge, this is the first study on such a problem. This task predicts a scene graph where unknown objects are correctly localized and classified as ''unknown'', rather than classifying them as one of the known classes or ignoring them as background, as shown in Fig. 1b. In addition, this enables the detection

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan.

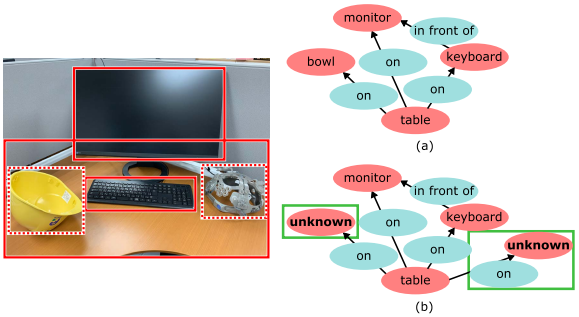

FIGURE 1. (Left) An example image. Solid and dotted red boxes denote known and unknown objects, respectively. (Right) Scene graphs for the image that should be predicted by (a) closed-set scene graph generation and (b) open-set scene graph generation. Green boxes denote the novel parts of the proposed open-set problem compared with the previous closed-set problem. Red and blue circles denote objects and relationships, respectively.

of relationships involving unknown objects, which previous studies have completely ignored. Specifically, we first provide a formal definition of the open-set SGG problem. Then, we propose an evaluation protocol, including a scene graph dataset with unknown objects. We construct the dataset from an existing large-scale dataset by defining unknown object classes and splitting a sufficient number of training images without unknown objects. Also, we propose novel evaluation metrics to quantitatively measure the open-set performance of SGG, focusing on the effect of unknown objects in both object and relationship detection. Furthermore, we develop baseline open-set SGG methods by modifying existing SGG methods for the open-set setting, introducing model-agnostic unknown detection by thresholding classification scores. Through extensive experiments based on the proposed protocol, we compare these methods and present initial evaluation results of the new problem, thereby establishing the baseline performance of open-set SGG and demonstrating the limitations of the current SGG methodology to be overcome in the future. The proposed open-set SGG extends the applicability of SGG in the real world, where the existence of unknown objects is inevitable.

Our contributions in this paper are summarized as follows:

- We propose the new problem of *open-set SGG* with unknown objects and provide its formal definition (Section III-A).
- We propose an evaluation protocol of open-set SGG, including an extended dataset (Section III-B), the *low-frequency-first unknown-class selection scheme* for splitting training and testing data, and novel evaluation metrics (Section III-C) named *open-set recalls* and *open-set object/relationship counts*.
- We develop multiple baseline methods of open-set SGG by employing representative SGG methods and extending them to unknown-aware versions with model-agnostic unknown detection (Section IV-A), which is simple yet effective as demonstrated experimentally.
- We demonstrate the current performance of open-set SGG by comparing the baseline methods through extensive experiments based on the proposed evaluation protocol (Section IV-C), discussing the limitations of the current SGG methodology, and pointing out future research directions (Section V).
- We will make our implementation (Section IV-B) for the dataset preparation, baseline methods, and experiments publicly available upon publication as a benchmark of open-set SGG to facilitate future research.

II. RELATED WORK

A. SCENE GRAPH GENERATION

Scene graphs provide a more detailed description of scenes than image recognition (image-wise object classification) and object detection (localization and region-wise classification), detecting not only individual objects but also their relationships [1]. First, we recall the definition of closed-set SGG, i.e., the previous problem setting that considers known objects only. Let $\mathcal{K}$ be a set of known object classes. Given an RGB image $I \in \mathbb{R}^{W\times H\times 3}$, where $W, H \in \mathbb{N}^+$ are its width and height, respectively, object detection, which is a subproblem of SGG, aims to localize and classify each i-th object by predicting bounding box $b_i = (x_i, y_i, w_i, h_i) \in \mathbb{R}^4$, which are the horizontal and vertical center locations, width, and height of the bounding box, respectively, and object class $o_i \in \mathcal{K}$. SGG further detects each l-th relationship for object pair (i_l, j_l) by predicting relationship class $r_l \in \mathcal{C}$, where $\mathcal{C}$ is the set of relationship classes. The goal of closed-set SGG is to build a model that can predict these bounding boxes and classes for all objects and predict relationships in the given image, i.e., a mapping from I to label $T = (\{b_i, o_i\}_{i=1}^n, \{i_l, j_l, r_l\}_{l=1}^m)$, where $n, m \in \mathbb{N}^+$ are the numbers of objects and relationships, respectively. This is typically achieved by data-driven learning using pairs of images and ground-truth labels as training data.

The most widely-used SGG dataset is Visual Genome (VG) [6], which is a large-scale dataset consisting of images from object detection datasets such as MS COCO [8] and labels made by crowdsourcing-based annotation. The majority of SGG studies [9]–[14] also employ the preprocessing proposed for the early SGG method named iterative message passing (IMP) [2], which removes noisy labels in VG and then randomly splits images into training and testing data.

As large-scale datasets such as VG have become available, many SGG methods have employed the modern deep-learning approach, and various models have been proposed [2], [9]–[12], [15]–[18]. These models make full use of the continuously evolving methodology of deep neural networks consisting of various components, e.g., convolution [17], graph convolution [10], long short-term memory [9], and transformers [18], resulting in quite different network architectures among models. Since the selection of the best model depends on the types of targeted scenes and individual applications, we do not aim to build a specific model for open-set SGG in this paper.

Apart from models, various SGG techniques have been proposed, e.g., losses [11], [14] and learning strategies [13], [19], aiming at improved performance regardless of model.

These topics are orthogonal to our open-set SGG, whose focus is on dealing with unknown objects rather than improving closed-set performance. Applying these techniques to the open-set setting is out of the scope of this paper.

While we consider the open-set setting of SGG for the first time, previous SGG studies addressed related topics called few-shot and zero-shot learning [20], [21]. These settings in the context of SGG are different from our open-set SGG since they aim to detect relationships that involve rarely-seen or unseen *combinations* of seen object classes that are present in training data. e.g., predicting the ''stand on'' relationship between the ''elephant'' and ''street'' objects when all these classes appear in the training data but their combination does not [22]. Instead, we deal with unseen object classes themselves in our open-set SGG setting. We naturally handle unseen class combinations in this setting since any class combinations involving unknown classes are necessarily unseen in training. Meanwhile, we do not separately consider the previously-addressed case, i.e., unseen combinations of seen classes only, since it requires a specialized train-test data split. Although there is a recent study [23] that claims to address ''open-set SGG'', its problem setting is closer to zero-shot learning in the non-SGG literature, e.g., image recognition and object detection [24], [25]. It attempts to classify individual unseen classes by associating them with seen classes using external knowledge such as language information. In contrast, our problem setting of open-set SGG is consistent with those of open-set recognition and detection described in Section II-B, i.e., we do not distinguish within unknown objects but aim to separate them from known classes (technically by assigning their instances to a special single ''unknown'' class) without the need of additional information.

SGG has been further extended using additional data, e.g., language information such as captions [22], [26], temporal information from videos [27]–[29], and 3D spatial information from depth images or point clouds [4], [5], [30]–[33]. Although the ability to handle unknown objects is also important in these augmented problem settings, we focus on the open-set generalization of the standard single-image SGG problem, leaving these advanced topics for future research.

B. OPEN-SET OBJECT DETECTION

Open-set image recognition [34] is a relatively new research topic that aims to deal with unknown classes in image-wise classification [7]. It typically consists of a conventional closed-set recognition part and an unknown detection part, and the unknown detection part is technically similar to anomaly detection [35] and novelty detection [36] in rejecting unknown classes, although open-set recognition also classifies known objects in the closed-set recognition part. While early studies employed traditional learning techniques such as the support vector machine [37], [38], motivated by recent advancements in closed-set recognition using deep learning, deep neural networks have become popular in open-set recognition [39], [40]. Recently, the open-set methodology of image-wise classification has been extended to region-wise classification after localization, thereby initiating the problem of open-set object detection [41]–[44].

The main difference of the proposed open-set SGG problem from the open-set object detection is that SGG classifies all objects and their relationships simultaneously, considering their contextual dependencies. Thus, we present novel experimental results for relationship-aware open-set object detection and relationship detection, both of which have not been evaluated by the previous studies. In addition, unlike a recent evaluation study [43] on open-set detection, we compare several baseline methods, including unknown-aware extensions of existing methods. Although more sophisticated unknown detection techniques have been proposed for open-set recognition [39], [40] and object detection [44], we leave integration of such advanced techniques with state-of-the-art SGG methods as a future research topic.

Another important difference from object detection is that SGG needs a specialized dataset with ground-truth relationship labels for training and testing. Thus, we cannot reuse the open-set detection datasets with unknown objects used in the previous studies, nor follow their dataset construction scheme [43], [44], which relied on the availability of multiple large-scale datasets with mutually exclusive class definitions. Instead, we propose a frequency-based class selection scheme for defining unknown classes, which enables us to split training images without unknown classes while maintaining sufficient training data as part of our novel evaluation protocol for open-set SGG.

Open-set problems have been further extended to open-world problems [44]–[46], where unknown classes incrementally turn into new known classes. Extending open-set SGG to open-world is an interesting but advanced topic, thus being out of the scope of this work.

The differences between the proposed open-set SGG compared with closed-set SGG and open-set object detection are summarized in Table 1. This table highlights the novelty of this work.

III. OPEN-SET SCENE GRAPH GENERATION

A. PROBLEM FORMULATION

In closed-set SGG, if the assumption $o_i \in \mathcal{K}$ is violated, i.e., if an object does not belong to any known class in $\mathcal{K}$ is present in an image I, either (1) the model will classify it to one of the known classes, or (2) the model will treat it as background and not detect it as an object. This has not been regarded as a failure in previous studies. On the other hand, in this study, we consider that the prediction for an unknown object has failed if (1) all predicted objects overlapping with it are classified into known classes. We also consider so if (2) no predicted objects overlap with it. Here, we assume that the ground-truth label of the unknown object is available. Such a failure in object detection also has a negative impact on relationship detection since the prediction of relationship classes is typically conditioned on predicted object classes [9].

TABLE 1. Difference of problems addressed in this study and previous studies in terms of detectable entities.

Problem	Objects		Relationships	
	Known	Unknown	With known objects only	With unknown objects
Closed-set SGG	✓	✗	✓	✗
Open-set object detection	✓	✓	✗	✗
Open-set SGG (proposed)	✓	✓	✓	✓

Now, we provide the formal definition of open-set SGG by extending that of the closed-set SGG. In open-set SGG, we also have a set of unknown object classes $\mathcal{U}$, which is excluded from the known classes, i.e., $\mathcal{K} \cap \mathcal{U} = \varnothing$, and each object class may be either known or unknown, i.e., $o_i \in \mathcal{K} \cup \mathcal{U}$. This is the essential difference from the closed-set setting. By the definition of the open-set recognition setting [7], any object of the unknown classes cannot appear in training images, i.e., the model cannot see objects of the unknown classes in training and thus cannot learn how to classify objects into these classes. Hence, we do not include the individual unknown classes in the target classes of object classification and aim to assign objects of any unknown classes to the special single class "unknown" in testing. This class assignment is typically achieved by introducing some training-free mechanisms of unknown detection to the model, which is only enabled in testing. Moreover, in open-set SGG, the object pair of each relationship consists of two known objects (as in closed-set SGG), one known object and one unknown object, or two unknown objects. This new problem formulation allows us to tackle the issues of the closed-set SGG in the presence of unknown objects, i.e., wrong classification and failure in detection.

B. DATASET

In order to bypass the difficulty of large-scale annotation from scratch and to naturally extend the previous methodology of closed-set SGG such as noise-label removal and known-class selection to the open-set setting, we make full use of the existing data by employing the combination of the VG dataset and the IMP preprocessing, introducing unknown objects to it. Specifically, we alter the IMP preprocessing to select unknown object classes and then extract images without unknown objects for training, which is a requirement of the open-set setting. To avoid using noisy class labels, we first discard low-frequency classes by selecting the most frequent 1,500 object classes in terms of the number of objects in VG (after removing small or overlapping objects in the original IMP preprocessing). This results in at least 100 objects per class. Among them, we use the same 150 object classes as in IMP as known object classes to facilitate comparison with the previous closed-set setting, and we select unknown classes from the other 1,350 classes.

The issue here is that an image must belong to testing data if it contains *any* objects belonging to unknown classes; otherwise, a model would learn to treat unknown objects as background since they have no bounding box labels in training data. Such violation of the open-set assumption leads to low performance for unknown objects in testing. However, a random selection of unknown classes may assign most images in the dataset to testing, leading to unsuccessful training due to insufficient data. Indeed, as shown in Fig. 2, the number of testing images rapidly grows as we randomly add classes to the unknown set, leaving almost no training images. Note that previous studies on open-set object detection did not face this issue, since they could combine multiple datasets with mutually exclusive class definitions to ensure that images from one dataset do not contain the classes from the others [43], [44], while we do not have other large-scale datasets like VG to be combined.

To overcome this issue, we propose to select unknown classes from low-frequency classes, which results in almost linear correspondence between the number of unknown object classes and that of testing images, as shown in Fig. 3. We name this the *low-frequency-first unknown-class selection scheme*. This enables us to easily control the ratio between the numbers of training and testing images by that of known and unknown classes. Here, we approximate the ratio of the original IMP split (the image splitting scheme of the VG dataset employed by the IMP preprocessing), i.e., 7:3 between training and testing, by selecting 30% of the lowest frequency classes for 406 unknown classes from the 1,350 classes.

After defining known and unknown classes, we remove the object whose classes are neither known nor unknown by dropping their bounding boxes and also remove their relationships from each image. Then, we also remove the images that consequently have no objects or relationships, which would not contribute much to SGG training, following previous studies [10], [13]. Other parts of the IMP preprocessing, e.g., removing invalid images in VG and selecting 50 relationship classes, are unchanged.

C. METRICS

1) CLOSED-/OPEN-SET RECALLS

For quantitative SGG evaluation, recall-based metrics are often used. They count correctly-detected ground-truth relationships in each image. Specifically, we use the following types of commonly-used recall-based metrics that perform prediction differently [22]:

- SGCls (scene graph classification) is a recall of the prediction of object classes and relationship classes given ground-truth bounding boxes.
- SGDet (scene graph detection) is a recall of the prediction of bounding boxes along with the classes without using any ground-truth labels of bounding boxes nor classes.

We simply refer to the collection of metrics of these two types as *recalls* in this paper. Note that we do not use PredCls (predicate classification), another common SGG metric,

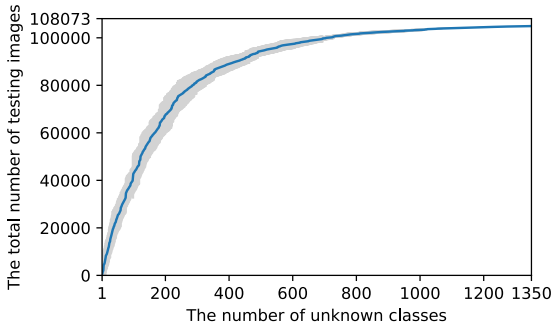

FIGURE 2. The number of unknown object classes vs. the total number of testing images (i.e., images with any objects of unknown classes) when unknown classes are selected randomly. The mean number of images over 10 trials is plotted with its standard deviation represented by the gray area.

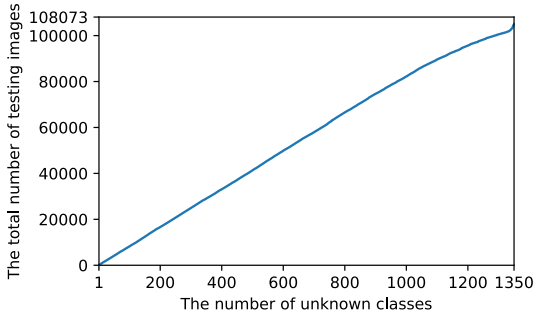

FIGURE 3. The number of unknown object classes vs. the total number of testing images when unknown classes are selected from low frequency classes.

since it needs the support of ground-truth object class labels in relationship detection, which is generally nontrivial for unknown classes and requires model-dependent modifications (e.g., when class-wise embeddings are needed [9], [12]). Also, note that, by following the previous SGG studies [2], we do not use precision metrics for SGG evaluation. This is because they may penalize the detection of unlabeled objects and relationships in VG, whose annotation is incomplete due to the limitation of crowdsourcing and yield uninterpretable metric values.

To compute a specific recall metric, we count each ground-truth relationship where it is correctly localized and classified. That is, it has at least one predicted relationship whose two corresponding bounding boxes overlap with the ground-truth boxes respectively with intersection over union (IoU) over 0.5 and whose two object and one relationship classes match the ground-truth classes. Note that in the case of SGCls, all predicted relationships are estimated using the ground-truth bounding boxes, and thus all ground-truth relationships are always correctly localized. In addition, we only consider top-K predicted relationships sorted by the product of the classification scores corresponding to the three classes in each relationship and denote each recall-based metric by its type suffixed with the K value, e.g., SGDet@100 when $K = 100$.

In our evaluation protocol, we adapt each recall metric to the open-set setting and consider the following two versions:

The closed-set version ignores ground-truth unknown objects and does not count the relationships involving them. This is equivalent to the recall-based metric used in previous closed-set SGG studies.

The open-set version regards all unknown classes as the single "unknown" class and treats it in the same manner as individual known classes when matching ground-truth and predicted object classes.

By comparing these two versions, we can see the effect of unknown objects in SGG and highlight problems in previous evaluation protocols. Note that, while the *closed-set recalls* have been extensively used in previous studies on SGG, the *open-set recalls* are a novel collection of evaluation metrics proposed in this paper, which is designed specifically for the new problem of open-set SGG.

2) OPEN-SET OBJECT/RELATIONSHIP COUNTS

In addition to these metrics as a natural extension of previous closed-set SGG, we also propose recall-like metrics designed for detailed analysis of object and relationships detection in open-set SGG with unknown objects, inspired by previous studies in open-set object detection [41], [43], [44]. As in the case of the recall-based metrics, we count ground-truth objects or relationships in each image. For the first of the two metric collections that we propose, we count ground-truth objects while distinguishing whether each of them belongs to (0) known or (1) unknown classes (where we enumerate these cases using the number of unknown objects to be consistent with the relationship-counting metrics described below) and whether its prediction is

(a) correct: the ground-truth object is correctly localized and classified (possibly into the single "unknown" class) by a predicted object. Here, we define the correct localization and classification for objects in the same manner as object detection. That is, for correct localization, the predicted object must overlap with the ground-truth object with IoU over 0.5. For correct classification, the overlapping object, or if multiple overlapping predicted objects exist, at least one of them must have the same class as the ground-truth object.
(b) wrong: it is correctly localized by one or more predicted objects but not correctly classified by any of them.
(c) background: it is not correctly localized by any predicted objects.

We also consider only top-K predicted objects sorted by object classification scores. By considering all possible combinations of the two ground-truth categories (0/1) and the three prediction categories (correct/wrong/background), we obtain six scores in total. We denote each count-based metric by the combination of two categories, e.g., "0-correct" for known and correctly-classified objects and "1-background" for unknown and undetected objects. Also, we call the proposed collection of these six count-based metrics *open-set object counts* (OSOC), which are suffixed by an actual K value, e.g., OSOC@100. We note that the number of unknown objects classified as known ("1-wrong") coincides with absolute open-set error proposed for open-set

object detection [41], which was also used in the recent study proposing a state-of-the-art open-set object detection method [44]. Meanwhile, we do not use precision-like metrics such as another open-set detection metric called wilderness ratio [43] since they are not suitable to the sparsely annotated VG dataset as described above.

Similarly, we count the number of ground-truth relationships by distinguishing the number of unknown objects that are involved in it, i.e., (0) zero, (1) one, or (2) two, and whether its prediction is (a) correct (correctly localized and classified), (b) wrong (correctly localized but not correctly classified), or (c) background (not detected). Here, the definition of the correct localization and classification of relationships, as well as the top-K selection, are the same as SGDet. We call the collection of the resulting nine metrics *open-set relationship counts* (OSRC).

IV. EXPERIMENTS

A. BASELINE METHODS

To establish the baseline for the new problem of open-set SGG, we evaluated several different SGG methods in our experiments. More specifically, we compared the following representative models originally proposed for closed-set SGG:

- Freq [9]is a simple model that predicts relationship classes by using their frequencies given object classes. It takes the object classes of each pair predicted by object detection and returns the most probable relationship class given them by referring to the object-conditioned relationship-class distribution learned from training data. It is called a strong baseline [10] because it often achieves surprisingly high performance without using other information to classify relationships.
- IMP [2]is the model of one of the earliest SGG methods. It uses iterative message passing on image-wise graphs to predict both object and relationship classes in consideration of their context. Though it is relatively simple, its performance is reportedly comparable to more recent models [29].
- VCTree [12] is a recent model that uses dynamic tree structures to perform context-dependent message passing while considering hierarchical relationships of objects. This model has been employed in more recent studies that focus on SGG techniques other than models, including losses and learning strategies, to achieve state-of-the-art performance [13], [14].

For a fair comparison, we fixed the network architecture other than the relationship detection part of these models. In particular, for the object detection part, which precedes the relationship detection part and predicts bounding boxes and object classes, we employed Faster R-CNN [47] as in the majority of SGG studies [9], [13]. It has been known that the two-stage design of Faster R-CNN is advantageous for open-set object detection [43] since its first region-proposal stage relies only on class-agnostic objectness, thereby being able to localize unknown objects similar to the objects in the training data. The relationship detection part may further update the outputs of the object detection part, depending on models, and yields final outputs, i.e., bounding boxes, object classes, and relationship classes. The objects and relationships are ordered by classification scores for these classes.

Furthermore, we built an unknown-aware version of each previous method by introducing a simple technique for unknown object detection, thereby enabling the measurement of the baseline performance of unknown-aware SGG. This also enables detection of the relationships of unknown objects without further modification to models, exploiting the similarity of known and unknown objects. By noticing that classification scores represent the confidence of being known classes, we applied thresholding to the class-wise scores of each predicted object, and if the scores for all classes were below a threshold, we updated the object class to the single ''unknown'' class. Note that this thresholding was not applied in training since the open-set setting assumes that unknown objects are only present in testing. Also, note that similar thresholding techniques have been widely used in open-set image recognition and object detection [34], [37], [38], [40], [44], although they relied on more complicated strategies for score calculation, etc., which are out of scope of this paper. We denote the new unknown-detecting version of each previous method with suffix ''+'', e.g., Freq+.

Given the VG dataset with unknown objects described in Section III-B, we trained each model using the training data and evaluated it quantitatively by computing the metrics in Section III-C over the testing data. Here, we performed prediction by the two versions for each of the three models, thereby comparing six baseline methods. In addition, we performed qualitative evaluation by visualizing predicted scene graphs on several images.

B. IMPLEMENTATION

We employed the publicly-available implementation[1] of a previous study on closed-set SGG [13], which supports multiple SGG models, including the above-mentioned ones and several metrics such as SGCls and SGDet. Note that the unbiasing technique, which was the main focus of the previous study [13], was not used since it is orthogonal to our study. We modified the code of this implementation to support the thresholding-based unknown detection described in Section IV-A and the proposed open-set metrics in Section III-C. This implementation also depends on the VG dataset, which is publicly-available.[2] Additionally, to introduce unknown objects to the dataset as described in Section III-B, we modified the code of the implementation[3]

[1]https://github.com/KaihuaTang/Scene-Graph-Benchmark.pytorch/tree/master/maskrcnn_benchmark

[2]https://visualgenome.org

[3]https://github.com/danfeiX/scene-graph-TF-release

of IMP [2], whose preprocessing is also assumed by the main implementation [13]. We also modified the IMP code for the visualization mentioned in Section IV-A.

For the hyperparameters of the previous methods, we reused the default values in the implementation [13]. Meanwhile, we tuned the only hyperparameter introduced in this study, which is the threshold value for each unknown-detecting method. Specifically, we optimized open-set SGDet@100 over validation data, which were split from the training images by selecting additional unknown classes in the same manner as the testing data in Section III-B. Note that the closed-set version of this metric is also used for validation for early stopping, etc. in the implementation [13]. Here, we set the ratio between the numbers of training and validation images to 6:1 and performed a grid search over the threshold values from 0.1 to 0.9 with stride 0.1, based on the fact that scores are bounded in [0, 1]. After validation, we retrained the method using the original training data, including the validation data, to maximize the amount of training data. Then, we tested it with the best threshold value.

C. RESULTS

1) CLOSED-/OPEN-SET RECALLS

First, we show the quantitative results of the previous metrics, i.e., closed-set recalls, computed over the testing data of our new training-testing data split of the VG dataset for each previous closed-set method, in Table 2. Here, we computed each metric for each image and averaged over all images in the testing data while using the same K values 20, 50, and 100 as previous studies [13], [14]. We can see that the metric values are close to previously-reported results [13] for the original split defined by the IMP preprocessing, where unknown objects were not considered, indicating that our new split of the VG dataset itself does not affect the closed-set performance so much.

TABLE 2. Closed-set recalls.

	SGCls			SGDet		
Model	@20	@50	@100	@20	@50	@100
Freq [9]	27.52	32.63	34.91	18.53	24.76	29.38
IMP [2]	**31.51**	34.66	35.61	17.10	24.40	29.68
VCTree [12]	30.90	**34.84**	**36.30**	**24.21**	**30.98**	**35.14**

Next, we show the results measured by the proposed new metrics, i.e., open-set recalls, in Table 3. Here, we compare both the original versions of the previous methods and their unknown-detecting versions (denoted by the suffix "+"). We first observe that the scores of the original methods were significantly lower than those in the closed-set setting in Table 2. This result reveals the limitation of the closed-set SGG evaluation protocols used in previous studies for open-set SGG, i.e., they can yield unrealistically high-performance scores in the presence of unknown objects, which is often the case in practice and thus problematic in applications. We believe that our new open-set evaluation protocol better reflects the real-world performance of SGG. Another observation is that each method's performance could be improved consistently for all metrics when the thresholding-based unknown detection was enabled. Thus, despite being simple, the thresholding technique can be effectively used to build baseline methods of open-set SGG by turning any closed-set method into an unknown-aware version.

TABLE 3. Open-set recalls. The "+" suffix indicates the methods with thresholding-based unknown detection.

	SGCls			SGDet		
Model	@20	@50	@100	@20	@50	@100
Freq [9]	8.39	10.84	12.27	6.84	9.17	10.91
Freq+	8.91	12.11	14.44	6.98	9.63	11.74
IMP [2]	10.12	12.17	13.06	6.60	9.32	11.23
IMP+	10.54	13.21	14.63	6.59	9.43	11.66
VCTree [12]	10.24	12.29	13.24	8.90	11.50	13.09
VCTree+	**11.43**	**15.17**	**17.93**	**9.27**	**12.32**	**14.57**

Overall, the methods based on the VCTree model achieved the highest performance in both the closed- and open-set settings, although our open-set methodology proposed in this paper is orthogonal to models and can be applied to any newer methods.

2) OPEN-SET OBJECT/RELATIONSHIP COUNTS

To perform a quantitative analysis of open-set SGG in more detail, we invoke another collection of new metrics, i.e., OSOC. Similar to recalls, each count-based metric was computed for each image and averaged over all testing images. We show a plot of the results only where K is equal to 100 in Fig. 4, as we observed that other K values yielded similar results. From this plot, we can clearly see that each unknown-aware version successfully recovered a significant proportion of the unknown objects (areas of "1-correct" of the "+"-suffixed methods) that were wrongly classified to any known classes by their original versions ("1-wrong" areas of the non-suffixed methods). This demonstrates the effectiveness of the simple thresholding-based unknown detection in dealing with unknown objects in SGG. Meanwhile, the unknown-aware versions slightly reduced the number of correctly-classified known objects ("0-correct"). This result can be considered as a side effect of the unknown detection, suggesting room for improvement. We also observe that the simple thresholding could not recover undetected objects, which were treated as background ("1-background"). Overcoming this limitation requires the redesign of the object detection part of each model, thereby being another future research direction.

We also plot the results of OSRC@100 in Fig. 5. We first observe that all the original methods without unknown detection could not detect most ground-truth relationships and treated them as background ("0/1/2-background"), confirming the well-known difficulty of SGG compared with object detection. Consequently, their unknown-detecting versions, which can only change the object classes in detected relationships, could not improve so much. Still, these methods, especially VCTree+, managed to fix some wrongly-classified unknown objects ("1-correct" and "2-correct"). Expanding these areas of successful predictions is a main future issue of open-set SGG to go beyond the baseline established in this paper.

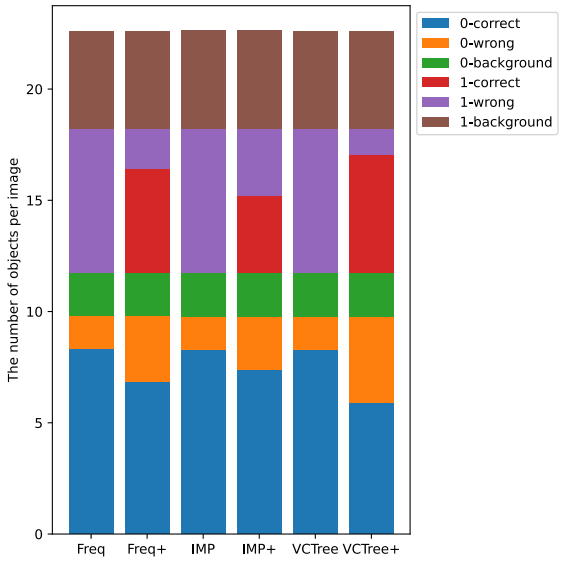

FIGURE 4. Open-set object counts@100. The "+" suffix indicates the methods with thresholding-based unknown detection. The "0-" and "1-" prefixes indicates known and unknown ground-truth objects, respectively.

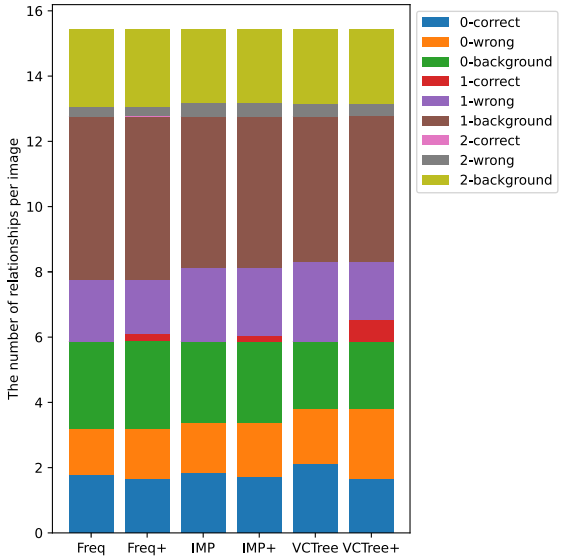

FIGURE 5. Open-set relationship counts@100. The "+" suffix indicates the methods with thresholding-based unknown detection. The "0-", "1-", and "2-" prefixes indicates the number of unknown objects in each ground-truth relationship.

3) VISUALIZATION

For qualitative evaluation, we visualize the ground-truth and predicted scene graphs on examples of testing images in Figs. 6 to 8. Here, we show the predictions by the best-performing model in Section IV-C1, i.e., VCTree+. Following the IMP visualization [2], we show each predicted relationship only if both of its objects overlap with any ground-truth relationships, similarly to recall metrics such as SGDet. Here, to avoid cluttered visualization while focusing on open-set-specific factors, we consider only ground-truth relationships with any unknown object. Note that the object indices (e.g., "1" of "unknown1") in these figures are added just for the purpose of explanation and did not exist neither in ground-truth nor predicted graphs.

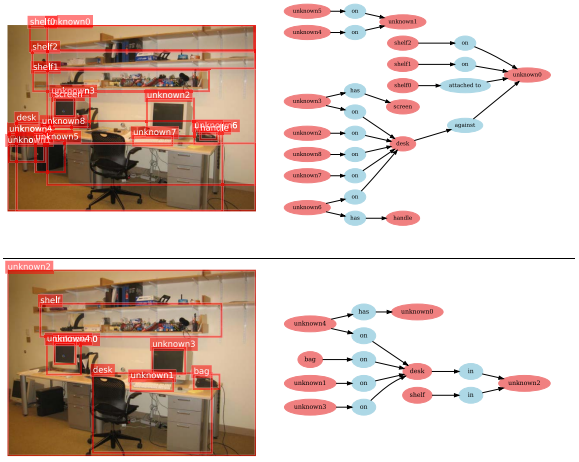

FIGURE 6. Visualization of scene graphs for a testing image. (top) Ground-truth scene graph. (bottom) Predicted scene graph by VCTree+.

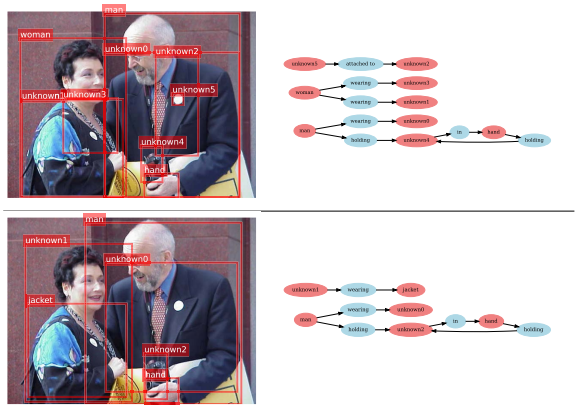

FIGURE 7. Visualization of scene graphs for a testing image. (top) Ground-truth scene graph. (bottom) Predicted scene graph by VCTree+.

In Fig. 6, the unknown-aware method successfully detected the keyboard object, along with its "on" relationship with the "desk" object. Meanwhile, the PC object ("unknown4" in the ground truth) under the desk could not be detected, which is the limitation of the current thresholding-based unknown detection that cannot recover the objects treated as background, as discussed in Section IV-C2. In Fig. 7, the method succeeded in detecting relationships involving an unknown object ("unknown0" and "unknown1" in the ground truth and prediction, respectively), where the "man" and his "hand" are "holding" it. Here, the method could also find the reversed relationship that the object is "in" the "hand". Meanwhile, the "woman" was wrongly classified as "unknown", which explains the side effect of the unknown detection, i.e., the decreased number of correctly-classified known objects (the "0-correct" metric) observed in Section IV-C2. In Fig. 8, the unknown object ("unknown5" and "unknown0", respectively) on the "building" wall was detected, but the relationship between the unknown object and the "building" object was predicted as "on" instead of the true class "in front of", which is semantically not critically wrong but still affects the performance measured by metrics such as SGDet. This kind of class

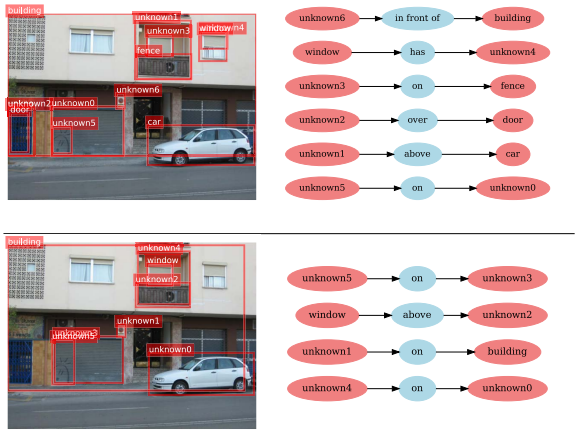

FIGURE 8. Visualization of scene graphs for a testing image. (top) Ground-truth scene graph. (bottom) Predicted scene graph.

ambiguity is also an issue in closed-set SGG and has been addressed in recent studies [13], [14], [48], and may also be of interest in the future research of open-set SGG.

V. CONCLUSION

Previous SGG studies have ignored the existence of unknown objects, and thus the real-world performance of SGG has been limited. In this paper, we addressed the new problem of open-set SGG, which allows us to detect unknown objects and also relationships involving them. Specifically, we formalized the problem and proposed an evaluation protocol including a dataset and metrics. We also presented the first experimental results on open-set SGG by comparing original and modified versions of previous methods to establish the baseline of open-set SGG. We believe that these contributions facilitate future researches in this unexplored yet important problem and also extend the applicability of SGG to various real-world scenarios.

Finally, we point out several future research directions of open-set SGG. While we employed the simple thresholding technique to build the baseline of unknown-aware versions, various unknown detection techniques have been proposed for open-set image recognition and object detection [38], [40], [44]. By appropriately combining these techniques with open-set SGG, we will be able to enhance unknown object detection and thereby relationship detection, hopefully dealing with the issues observed in Section IV, i.e., known objects classified into the "unknown" class and completely undetected objects. Meanwhile, importing the actively-developed techniques of conventional closed-set SGG into open-set SGG, e.g., network architectures, losses, and learning techniques, which are orthogonal to this study, is also important to enhance the performance against both known and unknown objects. Among them, we believe that techniques to deal with the ambiguity of relationship classes [13], [14], [48], which we observed in Section IV-C3, are particularly beneficial for open-set SGG. Although relationships have relatively less variety compared with objects, allowing unknown relationship classes as a further generalization of open-set SGG may also help mitigate difficulties due to ambiguous annotations inherent in large-scale SGG datasets. Inspired by attribute-based zero-shot learning [24], the use of object attributes, which are already available in datasets like VG [6] but currently unexploited for SGG, may be useful in the classification of unknown objects e.g. by distinguishing unknown classes using their common visual attributes. Recent advancements in the use of 3D-spatial and temporal information [31]–[33] may further benefit open-set SGG targeted at real-world applications such as robot navigation.

REFERENCES

[1] J. Johnson, R. Krishna, M. Stark, L.-J. Li, D. A. Shamma, M. S. Bernstein, and L. Fei-Fei, "Image retrieval using scene graphs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3668–3678.

[2] D. Xu, Y. Zhu, C. B. Choy, and L. Fei-Fei, "Scene graph generation by iterative message passing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5410–5419.

[3] Y. Zhu, O. Groth, M. Bernstein, and L. Fei-Fei, "Visual7W: Grounded question answering in images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4995–5004.

[4] B. Yu, C. Chen, F. Zhou, F. Wan, W. Zhuang, and Y. Zhao, "A bottom-up framework for construction of structured semantic 3D scene graph," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 8224–8230.

[5] I. Armeni, Z.-Y. He, A. Zamir, J. Gwak, J. Malik, M. Fischer, and S. Savarese, "3D scene graph: A structure for unified semantics, 3D space, and camera," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5664–5673.

[6] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma, M. S. Bernstein, and L. Fei-Fei, "Visual genome: Connecting language and vision using crowdsourced dense image annotations," *Int. J. Comput. Vis.*, vol. 123, no. 1, pp. 32–73, 2017.

[7] C. Geng, S.-J. Huang, and S. Chen, "Recent advances in open set recognition: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3614–3631, Oct. 2021.

[8] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.

[9] R. Zellers, M. Yatskar, S. Thomson, and Y. Choi, "Neural motifs: Scene graph parsing with global context," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5831–5840.

[10] J. Yang, J. Lu, S. Lee, D. Batra, and D. Parikh, "Graph R-CNN for scene graph generation," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 670–685.

[11] J. Zhang, K. J. Shih, A. Elgammal, A. Tao, and B. Catanzaro, "Graphical contrastive losses for scene graph parsing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11535–11543.

[12] K. Tang, H. Zhang, B. Wu, W. Luo, and W. Liu, "Learning to compose dynamic tree structures for visual contexts," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6619–6628.

[13] K. Tang, Y. Niu, J. Huang, J. Shi, and H. Zhang, "Unbiased scene graph generation from biased training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3716–3725.

[14] M. Suhail, A. Mittal, B. Siddiquie, C. Broaddus, J. Eledath, G. Medioni, and L. Sigal, "Energy-based learning for scene graph generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13936–13945.

[15] Y. Li, W. Ouyang, B. Zhou, K. Wang, and X. Wang, "Scene graph generation from objects, phrases and region captions," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1261–1270.

[16] H. Zhang, Z. Kyaw, S.-F. Chang, and T.-S. Chua, "Visual translation embedding network for visual relation detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5532–5540.

[17] H. Liu, N. Yan, M. Mortazavi, and B. Bhanu, "Fully convolutional scene graph generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11546–11556.

[18] N. Dhingra, F. Ritter, and A. Kunz, "BGT-Net: Bidirectional GRU transformer network for scene graph generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 2150–2159.
[19] R. Li, S. Zhang, B. Wan, and X. He, "Bipartite graph network with adaptive message passing for unbiased scene graph generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11109–11119.
[20] R. Krishna, V. Chen, P. Varma, M. Bernstein, C. Re, and L. Fei-Fei, "Scene graph prediction with limited labels," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2580–2590.
[21] A. Dornadula, A. Narcomey, R. Krishna, M. Bernstein, and F.-F. Li, "Visual relationships as functions: Enabling few-shot scene graph prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1730–1739.
[22] C. Lu, R. Krishna, M. Bernstein, and L. Fei-Fei, "Visual relationship detection with language priors," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 852–869.
[23] Y. Zhong, J. Shi, J. Yang, C. Xu, and Y. Li, "Learning to generate scene graph from natural language supervision," 2021, *arXiv:2109.02227*.
[24] Y. Xian, B. Schiele, and Z. Akata, "Zero-shot learning—The good, the bad and the ugly," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4582–4591.
[25] A. Bansal, K. Sikka, G. Sharma, R. Chellappa, and A. Divakaran, "Zero-shot object detection," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 384–400.
[26] K. Ye and A. Kovashka, "Linguistic structures as weak supervision for visual scene graph generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8289–8299.
[27] R. Herzig, E. Levi, H. Xu, H. Gao, E. Brosh, X. Wang, A. Globerson, and T. Darrell, "Spatio-temporal action graph networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 2347–2356.
[28] Y.-H.-H. Tsai, S. Divvala, L.-P. Morency, R. Salakhutdinov, and A. Farhadi, "Video relationship reasoning using gated spatio-temporal energy graph," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10424–10433.
[29] J. Ji, R. Krishna, L. Fei-Fei, and J. C. Niebles, "Action genome: Actions as compositions of spatio-temporal scene graphs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10236–10247.
[30] A. Rosinol, A. Gupta, M. Abate, J. Shi, and L. Carlone, "3D dynamic scene graphs: Actionable spatial perception with places, objects, and humans," 2020, *arXiv:2002.06289*.
[31] C. Zhang, J. Yu, Y. Song, and W. Cai, "Exploiting edge-oriented reasoning for 3D point-based scene graph analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9705–9715.
[32] S.-C. Wu, J. Wald, K. Tateno, N. Navab, and F. Tombari, "SceneGraphFusion: Incremental 3D scene graph prediction from RGB-D sequences," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7515–7525.
[33] J. Ost, F. Mannan, N. Thuerey, J. Knodt, and F. Heide, "Neural scene graphs for dynamic scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2856–2865.
[34] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1757–1772, Jul. 2013.
[35] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, Jul. 2009.
[36] K. Lee, K. Lee, K. Min, Y. Zhang, J. Shin, and H. Lee, "Hierarchical novelty detection for visual object recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1034–1042.
[37] W. J. Scheirer, L. P. Jain, and T. E. Boult, "Probability models for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2317–2324, Nov. 2014.
[38] E. M. Rudd, L. P. Jain, W. J. Scheirer, and T. E. Boult, "The extreme value machine," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 762–768, Mar. 2018.
[39] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1563–1572.
[40] L. Shu, H. Xu, and B. Liu, "DOC: Deep open classification of text documents," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 2911–2916.
[41] D. Miller, L. Nicholson, F. Dayoub, and N. Sunderhauf, "Dropout sampling for robust object detection in open-set conditions," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 3243–3249.
[42] Y. H. Kim, D. K. Shin, M. U. Ahmed, and P. K. Rhee, "Hierarchical open-set object detection in unseen data," *Symmetry*, vol. 11, no. 10, p. 1271, Oct. 2019.
[43] A. R. Dhamija, M. Gunther, J. Ventura, and T. E. Boult, "The overlooked elephant of object detection: Open set," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 1021–1030.
[44] K. J. Joseph, S. Khan, F. S. Khan, and V. N. Balasubramanian, "Towards open world object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5830–5840.
[45] A. Bendale and T. Boult, "Towards open world recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1893–1902.
[46] L. Shu, H. Xu, and B. Liu, "Unseen class discovery in open-world classification," 2018, *arXiv:1801.05609*.
[47] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," 2015, *arXiv:1506.01497*.
[48] G. Yang, J. Zhang, Y. Zhang, B. Wu, and Y. Yang, "Probabilistic modeling of semantic ambiguity for scene graph generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12527–12536.

MOTOHARU SONOGASHIRA received the B.S. degree in engineering and the M.S. and Ph.D. degrees in informatics from Kyoto University, Kyoto, Japan, in 2013, 2015, and 2018, respectively. He is currently a Researcher at the Multimodal Data Recognition Research Team, RIKEN Guardian Robot Project. His research interests include image restoration, blind deconvolution, super resolution, optical flow, and variational Bayes.

MASAAKI IIYAMA (Member, IEEE) received the B.S. degree in engineering informatics and the M.S. and Ph.D. degrees in informatics from Kyoto University, in 1998, 2000, and 2006, respectively. From 2003 to 2006, he was a Research Associate at ACCMS and an Assistant Professor, from 2009 to 2015. From 2006 to 2009, he was an Assistant Professor at the Graduate School of Economics, Kyoto University, and an Associate Professor, from 2009 to 2015. He is currently a Professor at the Faculty of Data Science, Shiga University. His research interests include computer vision, 3D modeling, and pattern recognition.

YASUTOMO KAWANISHI (Member, IEEE) received the B.Eng. degree in engineering and the M.Inf. and Ph.D. degrees in informatics from Kyoto University, Japan, in 2006, 2008, and 2012, respectively. In 2012, he became a Postdoctoral Fellow with Kyoto University. In 2014, he moved to Nagoya University, Japan, as a Designated Assistant Professor, where he became an Assistant Professor, in 2015, and a Lecturer, in 2020. Since 2021, he has been the Team Leader of the Multimodal Data Recognition Research Team, RIKEN Guardian Robot Project. His main research interests include robot vision for environmental understanding and pattern recognition for human understanding, especially pedestrian detection, tracking, retrieval, and recognition. He is a member of IIEEJ and IEICE. He received the Best Paper Award from SPC2009 and the Young Researcher Award from the IEEE ITS Society Nagoya Chapter.