buildings

MDPI

Article

Computer Vision-Based Hazard Identification of Construction Site Using Visual Relationship Detection and Ontology

Yange Li, Han Wei, Zheng Han *, Nan Jiang, Weidong Wang and Jianling Huang

School of Civil Engineering, Central South University, Changsha 410075, China; liyange@csu.edu.cn (Y.L.); weihan@csu.edu.cn (H.W.); jiangnan_0911@163.com (N.J.); weidong0530@126.com (W.W.); hjl1201@csu.edu.cn (J.H.)

* Correspondence: zheng_han@csu.edu.cn

Abstract: Onsite systematic monitoring benefits hazard prevention immensely. Hazard identification is usually limited due to the semantic gap. Previous studies that integrate computer vision and ontology can address the semantic gap and detect the onsite hazards. However, extracting and encoding regulatory documents in a computer-processable format often requires manual work which is costly and time-consuming. A novel and universally applicable framework is proposed that integrates computer vision, ontology, and natural language processing to improve systematic safety management, capable of hazard prevention and elimination. Visual relationship detection based on computer vision is used to detect and predict multiple interactions between objects in images, whose relationships are then coded in a three-tuple format because it has abundant expressiveness and is computer-accessible. Subsequently, the concepts of construction safety ontology are presented to address the semantic gap. The results are subsequently recorded into the SWI Prolog, a commonly used tool to run Prolog (programming of logic), as facts and compared with triplet rules extracted from using natural language processing to indicate the potential risks in the ongoing work. The high-performance results of Recall@100 demonstrated that the chosen method can precisely predict the interactions between objects and help to improve onsite hazard identification.

Keywords: hazard identification; ontology; safety management; visual relationship detection

Citation: Li, Y.; Wei, H.; Han, Z.; Jiang, N.; Wang, W.; Huang, J. Computer Vision-Based Hazard Identification of Construction Site Using Visual Relationship Detection and Ontology. *Buildings* **2022**, *12*, 857. https://doi.org/10.3390/buildings12060857

Academic Editor: Audrius Banaitis

Received: 11 May 2022
Accepted: 13 June 2022
Published: 19 June 2022

1. Introduction

Construction is a high-risk industry accompanied by frequent accidents. Onsite risk sources include hazardous chemicals, unsafe behaviors of workers, the unsafe state of materials, and a harmful environment. The hazards are often fatal, causing serious physical injuries, pecuniary losses, and schedule delays. Statistically, China had 734 accidents in housing and municipal projects and 840 workers died in 2018 [1].

Onsite systematic monitoring has a pivotal role in hazards prevention. Traditional supervision of onsite activities often requires manual work. Field observations are a commonly used approach to evaluate potential hazards [2–4]. However, manual supervision can be costly, time-consuming, and error-prone. Hence, it is difficult to satisfy the efficiency requirement of safety management.

Recent developments in the field of computer vision have led to an interest in applying computer vision to identify hazards. The application of computer vision in hazards identification can be majorly divided into two categories, i.e., hand-crafted features-based identification and deep learning-based identification.

Hand-crafted features-based identification using machine learning has been extensively used due to its impressive object detection and classification capacity [5–9]. Developments in the field of machine learning enable computers to better understand what they see [10]. However, the performance of the approaches is often limited due to the complex design process, poor generalization ability, and the fact that this kind of approach

can only process natural data in their raw form [11] and choose features artificially with a strong subjectivity.

With the rapid development of deep learning, the above limitations have been significantly addressed. Deep learning-based identification methods extract complicated features end to end, by learning from multiple data to simplify the detection process [10]. In this sense, instead of being designed artificially, complicated features can be extracted automatically from images obtained using computer vision [10]. Extensive studies [12–18] have shown remarkable results, impressive accuracy, and the expeditious speed of deep learning-based methods.

However, the hazards identification using object detection based on the method can only identify the category and the locations of the objects. Moreover, the detection method cannot represent the content of visual scenes sufficiently involving various objects that interact with each other [10]. Recently, the visual relationship detection methods based on deep learning can capture multiple interactions between objects in images, greatly enriching the semantic understanding of visual scenes [19–22]. Nevertheless, the method fails to assess the visual information and compare the information with safety rules or regulations to identify hazardous operations and conditions in the workplace. It is also limited to making managers aware, visually and intuitively, of the existence of incorrect and unsafe construction. Besides that, there is a semantic gap between visual information extracted from the images by computer vision methods and the textual information in safety rules [23]. Therefore, it is unable to satisfy the requirement of safety management only using computer vision-based methods.

Currently, some visual relationship detection methods based on deep learning have been developed in order to detect the relationships of objects. The method is much more consistent with the practice of safety management. However, there is a shortage of combinations of computer vision and ontology methods to address the semantic gap [24]. Ontology is the formal representation of specific domain knowledge that explicitly defines classes, relationships, functions, axioms, and instances [25]. It expresses knowledge with clear and abundant semantics and prepares for knowledge query and reasoning [26]. Hence, the establishment of ontology is a significant problem.

Recently, some previous studies have used ontology in the construction domain [27–29], but few have combined computer vision and ontology in the construction domain to detect dangerous objects, and unsafe operations, address the semantic gap, and prepare for logical reasoning in the case of data shortage. One remarkable study was done by Xiong et al. [24], who developed an automated hazards identification system to detect hazards from site videos against safety guidelines that combines computer vision and ontology. The system was tested successfully on two separate onsite video clips. Fang et al. [30] proposed a knowledge graph including an ontological model for knowledge extraction and inference for hazard identification. This method also integrates computer vision algorithms with ontology models and can effectively detect falls from height from images. Both methods use visual relationships but with different strategies. The first method considers limited types of onsite visual relationships and requires manual effort, which is labor-consuming, to extract semantic information. While the second one uses distance and coordinate information from 2D images to extract relationships between objects, more information, such as temporal and spatial information, as well as more types of images such as images from stereo cameras that include more data and depth information, were suggested to improve the research. Most of the construction ontologies were manually developed, which is time-consuming and error-prone. The stakeholders such as occupants, owners, and contractors have different views and understandings about the terms and relationships of onsite entities [30]. These challenges need to be addressed for the development of ontology.

There has been a limited number of studies in this aspect due to the scarcity of data. As such, most studies are focused on promoting the accuracy of the object detection methods using computer vision. At an early stage, due to the superiority of ontology in information and knowledge management, ontology was applied to project management. With the rapid

development of Building Information Model (BIM), ontology is combined with BIM to address the problem in the construction industry. Given the massive rise of computer vision, it is believed that the combination of computer vision and ontology used for systematic safety and health monitoring will be boosting topics.

Therefore, we proposed a framework that integrates computer vision, ontology, and natural language processing to improve systematic safety management in the construction domain. In the framework, visual relation detection algorithm can detect the interaction between entities and extract the visual information of construction site. Constructing the construction safety ontology model can reuse and share the safety management knowledge in construction domain and fill the semantic gap. Then, the entity relation extraction technology based on natural language processing is used to extract the entity relation from the construction safety regulation text by dependency analysis, and the construction safety rule information is output in triplet mode. Finally, combined with semantic reasoning based on SWI Prolog, the extracted visual information, safety management knowledge in the construction domain and rule information are evaluated to deduce the safety risk of the construction site, and realize the intelligent safety management of the construction site.

In the paper, we mainly realize the combination of visual relationship detection and ontology in the construction domain to improve onsite safety management and enhance automated systematic hazards identification. The visual relationship detection algorithm can detect the visual relationships between all the entities, and we use the safety ontology in the construction domain as an instance to illustrate our proposed method. We use a visual relationship detection model visual translation embedding network (VTransE) to detect all the categories and locations of objects in images and their interaction relationships with each other. The objects and the relationships in the newly detected images will be encoded in the form of triplets, which is object 1, relation predicate, object 2. Then, we establish the construction safety ontology model and take the detected object at the construction site as an example to address the semantic gap between visual information extracted from images and textual information in safety regulations and prepare for knowledge query and reasoning. The previous research is limited in that extracting regulatory documents and encoding them in a computer-processable format often requires manual work which can be costly and time-consuming. In this research, we will improve the framework by providing the natural language processing method and a more precise visual relationship detection model. The method can provide an opportunity to enhance safety management and improve automated hazards identification.

The paper begins by providing the method of the visual relationship detection model VTransE in Section 2. It will then go on to the description of the construction safety ontology in Section 3. Section 4 presents the experimental results of the study. Section 5 discusses the limitations of the study. Finally, we conclude our work in Section 6.

2. Visual Relationship Detection

Visual relationship detection methods can detect multiple interactions between objects such as "worker wear helmet" and "worker ride truck" and offer a comprehensive scene understanding of onsite images. The methods have demonstrated a marvelous ability to connect computer vision and natural language. A considerable amount of literature has been published on visual relationship detection [31–34]. In the paper, we use a state-of-art method VtransE [35] to detect visual relationships and lay a foundation for the construction safety ontology model and improve safety management.

The method dubbed Visual Translation Embedding network (VtransE) is chosen because it is the first end-to-end relation detection network that can detect objects and relations simultaneously, and it is competitive among the state-of-art visual relationship detection methods. The authors integrated translation embedding and knowledge transfer to propose an original visual relation learning model for VTransE. The method demonstrated great performance on two large-scale datasets: Visual Relationship [31] and Visual Genome [36].

The VTransE method involves two parts: an object detection module and a relation module. As illustrated in Figure 1, the VTransE network first builds an object detection module which is a convolutional localization network and then builds a relation module that integrates feature extraction and visual translation embedding. An image is input into the object detection module and a group of detected objects is the output. Then, objects are fed into the relation module. In the end, the detected images with objects and the relationships between objects in the form of the subject–predicate–object triplet will be output.

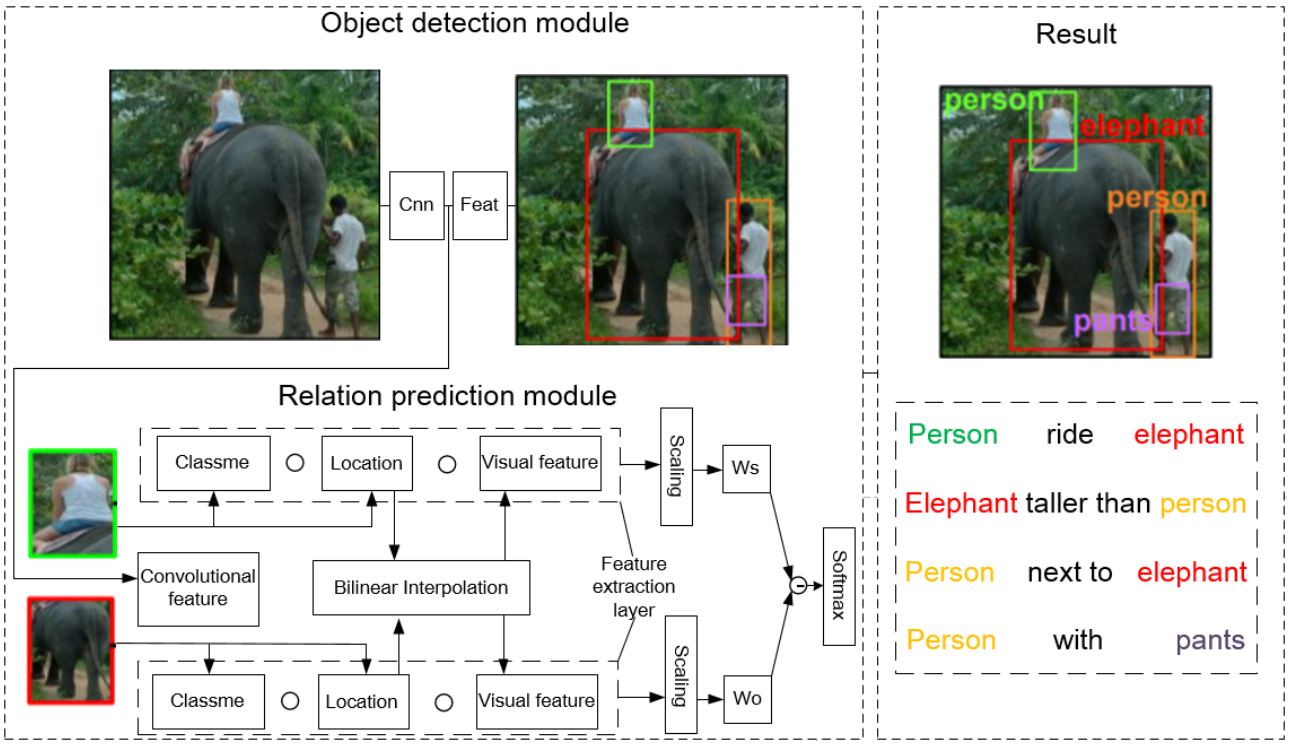

Figure 1. The VTransE Network (modified from [35]).

The VTransE method refers to the visual relationship as the subject–predicate–object triplet, where the predicate can be a verb, spatial (under), preposition (with), and comparative (higher). Inspired by Translation Embedding (TransE) [37], the authors map the features of objects and predicates in a low-dimensional space, in which the relation triplet is explained as a vector translation, e.g., person + wear ≈ helmet. For knowledge transfer in relation, the authors present a unique feature extraction layer that extracts three kinds of object features: classeme (i.e., class probabilities), locations (i.e., bounding boxes coordinates and scales), and RoI visual features. Especially, the bilinear feature interpolation is utilized rather than RoI pooling for differentiable coordinates.

Visual Translation Embedding. TransE depicts subject–predicate–object in low-dimensional vectors s, p, and o accordingly. The relation is represented as $s + p \approx o$ when the relation is established. Suppose $x_s, x_o \in \mathbb{R}^M$ are the M-dimensional features of subject and object. VTransE learns the relation translation vector $t_p \in \mathbb{R}^r$ $(r \ll M)$ as in TransE and two projection matrices W_S, $W_O \in \mathbb{R}^{r \times M}$ from the feature space to the relation space. Therefore, the visual relation is expressed as:

$$W_s x_s + t_p \approx W_o x_o \tag{1}$$

A simple softmax is used for prediction loss:

$$\mathcal{L} = \mathcal{L}_{obj} + 0.4\mathcal{L}_{rel} \tag{2}$$

$$\mathcal{L}_{rel} = \sum_{(s,p,o)\in R} -log\ softmax\left(t_p^T(W_o x_o - W_s x_s)\right) \quad (3)$$

The ultimate result for relation detection is the summation of object detection score and predicate prediction score: $S_{s,p,o} = S_s + S_p + S_o$.

Feature Extraction. The VTransE method proposed a feature extraction layer to extract x_s and x_o. Three types of features are used to represent the multiple aspects of objects in relations:

Classeme. The classme is a vector which means object classification probabilities. The dimension of the vector is $(N+1)$, (i.e., N classes and 1 background) from the object detection network. In relation detection, it is used to reject impossible relations such as horse–drive–person.

Location. The dimension of the vector location is 4, (t_x, t_y, t_w, t_h). (t_x, t_y) represents the scale-invariant translation and (t_w, t_h) represents the log-space height/width transformation relative to the subject or object. Take subject as an instance:

$$t_x = \frac{x - x'}{w'}, t_y = \frac{y - y'}{h'}, t_w = log\frac{w}{w'}, t_h = log\frac{h}{h'} \quad (4)$$

where (x, y, w, h) and (x', y', w', h') are the box coordinates of subject and object.

Visual Feature. It is a $D - d$ vector converted from a convolutional feature of the shape $X \times Y \times C$. The features are bilinearly interpolated from the last convolution feature map. Thus, end-to-end training with knowledge transfer function can be realized.

The overall feature x_s or x_o is a weighted join of the above three features $(M = N + D + 5)$, where the weights are scaling layers that can be learned since the feature contribution changes from one relation to another. As shown in Figure 1, the presented feature extraction layer couples the two modules.

Architecture details. The object detection network of VTransE uses the VGG-16 architecture [38] from the Faster-RCNN. Then, the final pooling layer of VGG-16 is removed and the last convolutional feature map F of the shape $W' \times H' \times C$ is used. C = 512, the number of channels. $W' = [W/16]$, and $H' = [H/16]$, where W and H represent the width and height of the input image. F represents the visual appearance of the image. The relation error is back-propagated to the object detection network to polish the objects for object-relation knowledge transfer. Thus, the ROI pooling layer is replaced by bilinear interpolation. As for optimization, the VTransE network is trained end-to-end by stochastic gradient descent.

3. Construction Safety Ontology

Although there is no uniform definition of ontology at present, a commonly used definition of ontology is "an explicit and formal specification of a conceptualization" [39], "a particular system of categories accounting for a certain vision of the world" [40]. It can describe varieties of concepts and the relationships among the concepts which belong to a certain domain. It is an integration of certain domain knowledge in the form of formal conceptualization with an immense expressive capacity [41]. It contributes to representing, sharing, and reusing domain knowledge semantically and provides an opportunity to transform human knowledge into a format that computers can understand and process explicitly and easily. By this means, the interactions between humans and computers can be expedited. The ontology construction can play a vital role in knowledge query and reasoning and can address the problem of the semantic gap between visual information obtained from using computer vision and textual information in safety regulations.

Ontology usually describes five concepts: classes, relationships, functions, axioms, and instances in a specific domain. Concepts that represent things in the special domain will be defined as classes. The interactions or connections between existing classes and instances will be defined as relationships. The function is a special kind of relationship. Axioms are constraints on the attribute's value of a concept and the attribute's value of a relationship, or on the relationship between concept objects that represent a statement that

is always true. The instances are particular individuals in the classes. However, in practical applications, it is not necessary to strictly follow the above five modeling primitives to create an ontology. It can be constructed according to the actual situation.

Although there is no clearly established, united, and completed method to build the construction safety ontology, some current methods that are researched and concluded from specific ontology construction programs by scholars are as follows: IDEF-5 methodology [42], Skeletal Methodology [43], TOVE methodology [44], Methontology methodology [45], and seven-steps methodology [46]. However, there still exist some problems in ontology construction of specific domains. Ontology construction is too subjective, relatively arbitrary, and lacks a scientific management and evaluation mechanism when manually constructing ontology by using the methods above. When reusing existing ontologies, there are problems such as: (1) there are few existing ontologies that can be reused without being modified; (2) there are many domains where ontology resources are not available; (3) it takes a lot of investment to transform some ontology resources and it needs to be studied whether the transformation is worthwhile.

In the paper, we choose to manually build the construction safety ontology because there is no available existing ontology to reuse. Through the horizontal comparison of common ontology construction methods, the seven-steps methodology is more complete and mature among the methods above. Hence, the construction safety ontology is built according to the characteristics of the program and mainly based on the seven-steps methodology proposed by Stanford University School of Medicine. The knowledge sources used to construct the safety ontology include Safety Handbook for Construction Site Workers [47] and the previous experience of the scholars.

The first step is to determine the domain and scope of the ontology. Considering that no available and suitable existing ontology has been found, step two, "Consider reusing existing ontologies", is omitted. Then, transform step three, "Enumerate important terms in the ontology", into "Enumerate terms that are commonly used in construction safety domains" and transform step four, "Define the classes and the class hierarchy", into "Categorize these terms and their class hierarchy relationships". Then, step five, "Define the properties of classes—slots", and step six, "Define the facets of the slots", are retained, because in the research the slots are important and will be discussed. The class hierarchy relationships are what we need to address the semantic gap. The last step, "Create instances", is retained because in the research it requires individual instances of classes to conduct semantic reasoning. Therefore, in this paper, we build the construction safety ontology using the following steps:

- Determine the domain and scope.
- Enumerate terms that are most commonly used in construction safety domains.
- Categorize these terms and their class hierarchy relationships.
- Define the properties of classes and the facets of the slots.
- Create instances.

The ontology is in the construction domain and is used for construction safety management. The ontology can address the semantic gap between visual information extracted from the images by computer vision methods and the textual information in safety rules. For example, the entity at the construction site detected by deep learning object model is "helmet", while what is identified by humans from the construction site and found in safety rules is "personal protective equipment". Thus, the representation of entities at the construction site is the domain of the ontology. The concepts describing different construction entities and their hierarchical relationships will figure into the ontology.

The terms or the entities are the particular onsite entities in the regulations which can be a "thing" (e.g., equipment, building structure) or "personnel" (e.g., worker). The terms emphasize the construction onsite objects including construction material, excavating machinery, load shifting machinery, protective personal equipment, suspended load, transport vehicle, and construction workers. Then, a top-down approach is adopted to start from the top concepts and gradually refine them. The subclass is the lower hierarchy

entity of the top concept. For instance, the subclasses of Personal Protective Equipment include ear protectors, eye protectors, gloves, helmets, etc., and the top terms are disjointed from each other. According to the characteristics of the program, the properties of the classes and the facets of the slots will be added accordingly. The entities detected at the construction site using the visual relationship detection model will be created as instances in the construction safety ontology model.

The construction safety ontology is developed by using Protégé, which is commonly used as a free and open-source ontology editor to build a domain ontology. The ontology will be represented in Web Ontology Language (OWL), which can explicitly define the terms and the relationships between and express the hierarchical ontology structure. Figure 2 shows the mind map of construction safety ontology model built in Protégé.

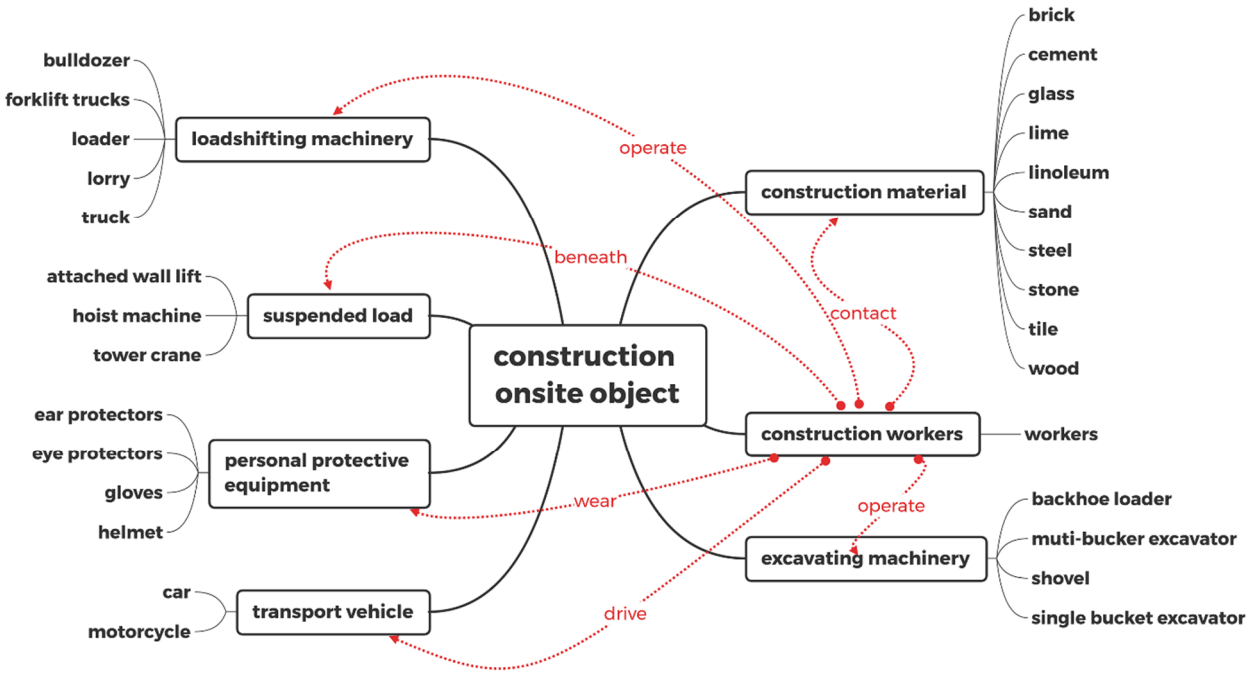

Figure 2. Mind map of construction safety ontology model.

4. Experiment

The visual relationship detection dataset we used is the benchmark dataset: VRD [31]. VRD is the Visual Relationships Dataset which includes 5000 images with 100 object categories and 70 predicates. In the aggregate, it includes 37,993 relationships with 6672 relationship types and 24.25 predicates per object category. A total of 4000 images are divided into the training set and the remaining 1000 images are divided into the test set [31]. The evaluation metric is the benchmark metric: Recall@100. Recall@x computes the fraction of times that true positive relationship is predicted in the top x confident relationship predictions in images.

The algorithms we used in the paper are all written in Python [35], and we implemented the algorithms using Tensorflow framework. We experimented on the computer with a high-performance computing server equipped with Nvidia Quadro series professional GPU card and 640 tensor kernels, which can meet the deep learning training requirement.

The VTransE algorithm is trained with the 4000 images in the training set and the remaining 1000 images are used to evaluate the ability of the model. The training set and the test set are trained and tested respectively. The accuracy of the trained results using the implemented method is 49.91%, which calculates whether the prediction result of the predicate and the ground truth are equal and whether the maximum number in top-1 in the prediction results of each sample contains the real label in prediction. The predicate recall@100 to evaluate the implemented method is 49.13%. The recall@100 computes the

proportion that the correct predicate is predicted in the top-100 confident predictions. The high-performance results of the accuracy and the recall@100 demonstrate that the VTransE algorithm is competitive among the various visual relationship detection methods and can predict precisely the interactions between objects, thus contributing to automatic onsite hazard identification.

The detected images demonstrate the detected objects, the predicted labels, and the triplets (object 1, relation predicate, object 2). There are five typical detection results shown in Figure 3.

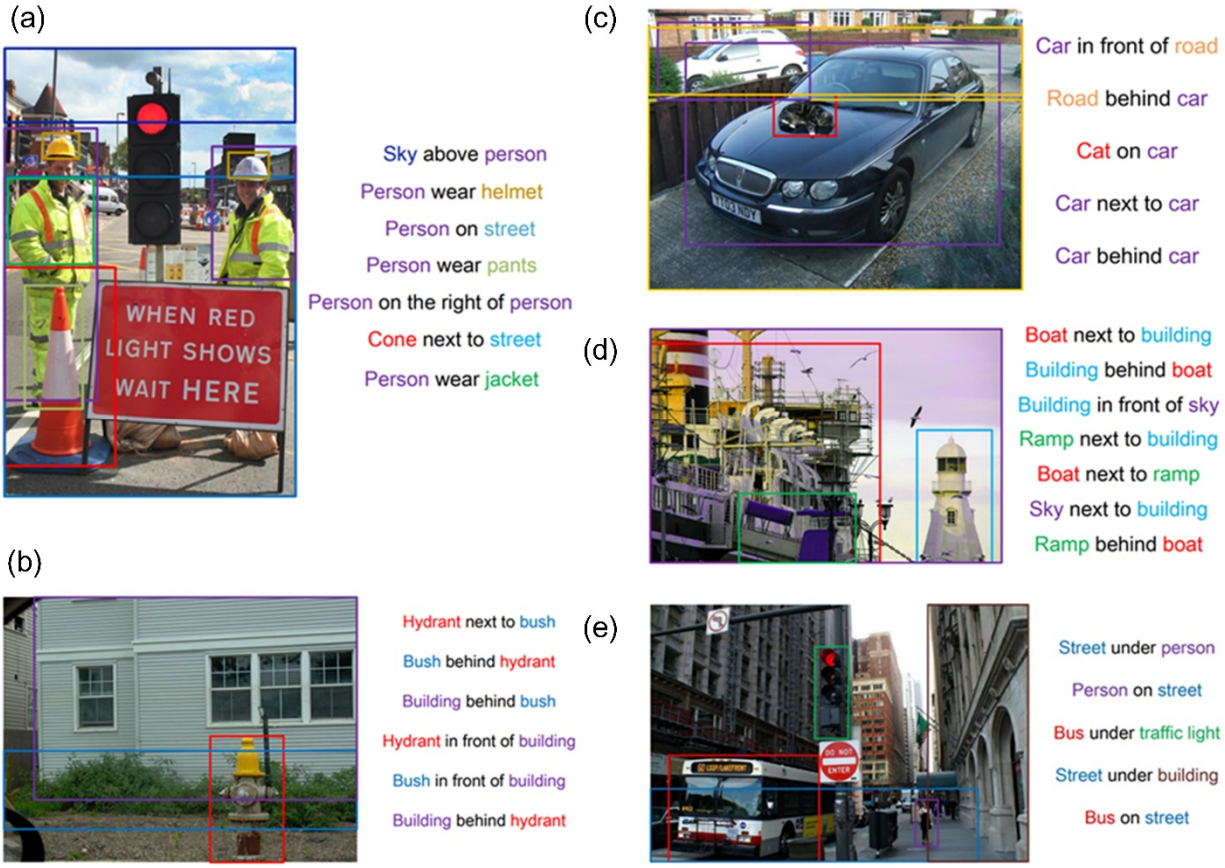

Figure 3. Detection results (images are available in VRD dataset [31]).

As shown in Figure 3a, seven visual relationships are detected, such as "sky above person", "person wear helmet", "person on street", "person on the right of person", "cone next to street". The relationships present the major relationships between the objects onsite, such as actions (e.g., hold, wear, carry), geometry (e.g., beneath, in front of, on, in). The detected objects are mainly person, street, cone, helmet, jacket, and pants. In Figure 3a,e, all the objects, predicates and relationships are detected and predicted precisely, which suggests the feasibility of the method. In Figure 3b, there are three detected objects: hydrant, building, and bush. The interactions between the three objects in the image are predicted accurately such as "bush behind hydrant", "building behind bush". Despite the success of the detection and prediction results, the predicted relationships between the same two objects are multiple because the relationships are interactional. "Building behind bush" and "bush in front of building" are both right. Nevertheless, the interaction and relationship between such objects in safety rules are explicitly stipulated in advance. Hence, although the detection results are correct, the difference in triplet expression can cause problems in the comparison between visual information extracted from the images and the textual information in safety rules. In Figure 3c, the relationships between the detected objects cat, car, and road are predicted precisely. However, the predicates between the

cars are "next to" and "behind". The two interactions can both describe the relationships between the two cars but can also be problematic in our framework. Similar to the problem in Figure 3c, the relationships between the objects boat and building in Figure 3d are different but will not influence the understanding of humans. However, it is perplexing for computer-understanding tasks.

There also exist some errors in visual relationship detection. As shown in Figure 4, the detected triplets are "people wear phone" in (a), "sky in sky" in (b), "coat wear hat" in (c), "jacket wear hat" in (d) and "dog wear glasses" in (e). In the five examples, all the objects are detected correctly, but it went wrong in the relationship matching. The relationships are impossible in real life. The locations of the objects in the images are very close and it may cause false relationship detections. The below unlikely relations should be rejected through classeme features in relation detection module. It indicates that the knowledge transfer part needs to be improved.

Figure 4. Detection errors (images are available in VRD dataset [31]).

In thousands of detection results, although some relationships and predicates between subject and object are predicted improperly, all the objects in the images are detected correctly. It indicates that the object detection module of the method is effective and potent, and the relation prediction module is feasible but not perfect. In general, the VTransE method is practicable and feasible for visual relationship detection tasks and can contribute to improving construction site safety management and enhancing onsite automated identification detection.

To prepare for logical reasoning, the relationships extracted from using computer vision and the three tuple formats generated from visual relationship detection (object 1, relation, object 2) will be represented as "Relation (object 1, object 2)" which meets the syntax requirements in SWI Prolog. For example, the (person, wear, helmet) will be represented as "wear (person, helmet)". The logic representation Relation (object 1, object 2) will be presented as the facts in SWI Prolog to conduct logic queries and reasoning. To address the semantic gap, the unary predicates are used to assign the element instances in the ontology.

For example, the "helmet" will be represented as the "personal_protective_equipment (helmet)". The results of the visual relationship detection and the ontology model will be both used as the facts and input into the SWI Prolog which is a logic programming language. When facts and rules are given, it analyzes the logic relation automatically and allows users to perform complex logic operations by querying.

5. Discussion

5.1. Limitations of the Proposed Method

The VTransE algorithm is trained with the 4000 images in the training set and the remaining 1000 images are used to evaluate the ability of the model. The accuracy of the trained results using the implemented method is 49.91%. The predicate recall@100 to evaluate the implemented method is 49.13%. The high-performance results of the accuracy and the recall@100 demonstrate that the VTransE algorithm is competitive among the various visual relationship detection methods and can predict precisely the interactions between objects, thus contributing to automatic onsite hazard identification.

The visual relationship detection method can detect multiple interactions between objects such as "worker wear helmet" and "worker ride truck" and connect computer vision and natural language. The images are detected and the predicted triplet relationships (object 1, relation predicate, object 2) will be presented as "Relation (object 1, object 2)". The high predicate detection results and the visual results of the methods suggest the feasibility and effectiveness of the methods.

However, there exist some limitations and challenges. First, since it is expensive and time-consuming to construct a dataset in the construction domain that includes objects at the construction site and onsite interactions between objects, the dataset we used in visual relationship detection is the public benchmark dataset VRD. The domain-specific dataset can concentrate on construction safety management and helps to improve the accuracy of onsite detection tasks.

Secondly, in the research, we consider the visual relationship detection which detects multiple interactions between objects. However, some relationships and interactions such as "in the front of, on the right of" are hard to describe and interpret in the ontology model. Thus, the relationships cannot be checked in later rule-checking tasks. Therefore, a better and more exact way needs to be found to define the relationship between entities. An intersection over union can be used to identify the spatial relationships between objects (i.e., within, or overlap) by using geometric and spatial features. Additionally, in further studies, ergonomic analysis can also be considered and added to onsite systematic monitoring. The visual triplet representations can include posture detections such as (worker, body_part, location) to detect and predict the workers' motion.

5.2. Future Research

In this paper, we proposed a framework that integrates computer vision, ontology, and natural language processing to improve systematic safety management. We realized the combination of the visual relation detection algorithm and the construction safety ontology. The visual relation detection algorithm can detect the interaction between entities and extract the visual information of construction site. The construction safety ontology model can reuse the safety management knowledge in the construction domain and fill the semantic gap.

In our ongoing studies, we are working on the triplet extraction from regulatory information using natural language process methods and logic reasoning using SWI Prolog. The entity relation extraction technology based on natural language processing is used to extract the entity relation from the construction safety regulation text by dependency analysis, and the construction safety rule information is output in triplet form. The triplet form (object 1, relation predicate, object 2) will be extracted from regulations such as Safety Handbook for Construction Site Workers [47] and will then be presented as Relation (object 1, object 2). The results will be used as the rules and be input into the SWI Prolog. The element instances

in the construction safety ontology will be presented as Class (subclass) through semantic mapping to address the semantic gap. The results of the visual relationship detection and the semantic representation of ontology will be input into the SWI Prolog as facts. Finally, the extracted visual information and rule information are evaluated to deduce the safety risk of the construction site by semantic reasoning based on SWI Prolog. Thus, whether the onsite activities violate the regulations will be reasoned out.

6. Conclusions

In this paper, we proposed a framework to improve construction safety management. In the framework, the visual relationship detection methods based on deep learning are used to detect multiple interactions between objects in images. The accuracy of the trained results using the implemented method is 49.91%. The predicate recall@100 to evaluate the implemented method is 49.13%. The high-performance results of the accuracy and the recall@100 demonstrate that the VTransE algorithm is competitive among the various visual relationship detection methods and can predict precisely the interactions between objects. The results suggest the effectiveness and feasibility of the method. The presented method offers an effective solution to detect the onsite activities and identify the hazard at the construction site automatically.

Author Contributions: Conceptualization, Y.L. and W.W.; writing—original draft preparation, H.W.; methodology, Z.H., Y.L. and J.H.; review and editing, N.J.; resources, W.W. and J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the National Key R&D Program of China (grant number 2018YFD1100401); the National Natural Science Foundation of China (grant number 52078493); the Natural Science Foundation for Outstanding Youth of Hunan Province (grant number 2021JJ20057); the Innovation Provincial Program of Hunan Province (grant number 2020RC3002); the Scientific and Technological Project of Changsha (grant number kq2106018).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. General Office of the Ministry of Housing and Urban-Rural Development of RPC. Reports on the Special Action to Address Production Safety Accidents and Construction Safety in Housing and Municipal Projects in 2018. Available online: http://zbxsgaq.com/xinwen/gonggao/190.html (accessed on 15 April 2019).
2. Buchholz, B.; Paquet, V.; Punnett, L.; Lee, D.; Moir, S. PATH: A work sampling-based approach to ergonomic job analysis for construction and other non-repetitive work. *Appl. Ergon.* **1996**, *27*, 177–187. [CrossRef]
3. Rozenfeld, O.; Sacks, R.; Rosenfeld, Y.; Baum, H. Construction job safety analysis. *Saf. Sci.* **2010**, *48*, 491–498. [CrossRef]
4. Gouett, M.C.; Haas, C.T.; Goodrum, P.M.; Caldas, C.H. Activity analysis for direct-work rate improvement in construction. *J. Constr. Eng. Manag.* **2011**, *137*, 1117–1124. [CrossRef]
5. Du, S.; Shehata, M.; Badawy, W. Hard hat detection in video sequences based on face features, motion and color information. In Proceedings of the 2011 3rd International Conference on Computer Research and Development, Shanghai, China, 11–13 March 2011. [CrossRef]
6. Azar, E.R.; McCabe, B. Part based model and spatial–temporal reasoning to recognize hydraulic excavators in construction images and videos. *Autom. Constr.* **2012**, *24*, 194–202. [CrossRef]
7. Yang, J.; Arif, O.; Vela, P.A.; Teizer, J.; Shi, Z. Tracking multiple workers on construction sites using video cameras. *Adv. Eng. Inform.* **2010**, *24*, 428–434. [CrossRef]
8. Wang, X.; Han, T.X.; Yan, S. An HOG-LBP human detector with partial occlusion handling. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009. [CrossRef]
9. Lin, Y.; Lv, F.; Zhu, S.; Yang, M.; Cour, T.; Yu, K.; Cao, L.; Huang, T. Large-scale image classification: Fast feature extraction and svm training. In Proceedings of the Conference on Computer Vision and Pattern Recognition 2011, Colorado Springs, CO, USA, 20–25 June 2011. [CrossRef]

10. Fang, W.; Love, P.E.D.; Luo, H.; Ding, L. Computer vision for behaviour-based safety in construction: A review and future directions. *Adv. Eng. Inform.* **2020**, *43*, 100980. [CrossRef]
11. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
12. Fang, W.; Zhong, B.; Zhao, N.; Love, P.E.; Luo, H.; Xue, J.; Xu, S. A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network. *Adv. Eng. Inform.* **2019**, *39*, 170–177. [CrossRef]
13. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Luo, H.; Rose, T.M.; An, W. Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Autom. Constr.* **2018**, *85*, 1–9. [CrossRef]
14. Ding, L.; Fang, W.; Luo, H.; Love, P.E.D.; Zhong, B.; Ouyang, X. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Autom. Constr.* **2018**, *86*, 118–124. [CrossRef]
15. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Rose, T.M.; An, W.; Yu, Y. A deep learning-based method for detecting non-certified work on construction sites. *Adv. Eng. Inform.* **2018**, *35*, 56–68. [CrossRef]
16. Anjum, S.; Khan, N.; Khalid, R.; Khan, M.; Lee, D.; Park, C. Fall Prevention From Ladders Utilizing a Deep Learning-Based Height Assessment Method. *IEEE Access* **2022**, *10*, 36725–36742. [CrossRef]
17. Khan, N.; Khan, M.; Cho, S.; Park, C. Towards the Adoption of Vision Intelligence for Construction Safety: Grounded Theory Methodology based Safety Regulations Analysis. In Proceedings of the EG-ICE 2021 Workshop on Intelligent Computing in Engineering, Berlin, Germany, 30 June–2 July 2021; Universitätsverlag der TU Berlin: Berlin, Germany, 2021; p. 250. [CrossRef]
18. Khan, N.; Saleem, M.R.; Lee, D.; Park, M.-W.; Park, C. Utilizing safety rule correlation for mobile scaffolds monitoring leveraging deep convolution neural networks. *Comput. Ind.* **2021**, *129*, 103448. [CrossRef]
19. Zhang, J.; Kalantidis, Y.; Rohrbach, M.; Paluri, M.; Elgammal, A.; Elhoseiny, M. Large-scale visual relationship understanding. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 9185–9194. [CrossRef]
20. Wan, H.; Liang, J.; Du, J.; Liu, Y.; Ou, J.; Wang, B.; Pan, J.Z.; Zeng, J. Iterative Visual Relationship Detection via Commonsense Knowledge Graph. *Big Data Res.* **2021**, *23*, 100175. [CrossRef]
21. Gan, M.-G.; He, Y. Adaptive depth-aware visual relationship detection. *Knowl.-Based Syst.* **2022**, *247*, 108786. [CrossRef]
22. Cui, Z.; Xu, C.; Zheng, W.; Yang, J. Context-dependent diffusion network for visual relationship detection. In Proceedings of the of the 26th ACM International Conference on Multimedia, Seoul, Korea, 22–26 October 2018; pp. 1475–1482. [CrossRef]
23. Gouthaman, K.V.; Nambiar, A.; Srinivas, K.S.; Mittal, A. Linguistically-aware attention for reducing the semantic gap in vision-language tasks. *Pattern Recognit.* **2021**, *112*, 107812. [CrossRef]
24. Xiong, R.; Song, Y.; Li, H.; Wang, Y. Onsite video mining for construction hazards identification with visual relationships. *Adv. Eng. Inform.* **2019**, *42*, 100966. [CrossRef]
25. Zhong, B.; Wu, H.; Li, H.; Sepasgozar, S.; Luo, H.; He, L. A scientometric analysis and critical review of construction related ontology research. *Autom. Constr.* **2019**, *101*, 17–31. [CrossRef]
26. Anumba, C.J.; Issa, R.; Pan, J.; Mutis, I. Ontology-based information and knowledge management in construction. *Constr. Innov.* **2008**, *8*, 218–239. [CrossRef]
27. Lu, Y.; Li, Q.; Zhou, Z.; Deng, Y. Ontology-based knowledge modeling for automated construction safety checking. *Saf. Sci.* **2015**, *79*, 11–18. [CrossRef]
28. El-Diraby, T.E.; Osman, H. A domain ontology for construction concepts in urban infrastructure products. *Autom. Constr.* **2011**, *20*, 1120–1132. [CrossRef]
29. Zhong, B.; Gan, C.; Luo, H.; Xing, X. Ontology-based framework for building environmental monitoring and compliance checking under BIM environment. *Build. Environ.* **2018**, *141*, 127–142. [CrossRef]
30. Fang, W.; Ma, L.; Love, P.E.D.; Luo, H.; Ding, L.; Zhou, A.O. Knowledge graph for identifying hazards on construction sites: Integrating computer vision with ontology. *Autom. Constr.* **2020**, *119*, 103310. [CrossRef]
31. Lu, C.; Krishna, R.; Bernstein, M.; Li, F.F. Visual relationship detection with language priors. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 852–869. [CrossRef]
32. Yin, G.; Sheng, L.; Liu, B.; Yu, N.; Wang, X.; Shao, J.; Loy, C.C. Zoom-net: Mining deep feature interactions for visual relationship recognition. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 322–338. [CrossRef]
33. Qi, S.; Wang, W.; Jia, B.; Shen, J.; Zhu, S.C. Learning human-object interactions by graph parsing neural networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 401–417. [CrossRef]
34. Yang, X.; Zhang, H.; Cai, J. Shuffle-then-assemble: Learning object-agnostic visual relationship features. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 36–52. [CrossRef]
35. Zhang, H.; Kyaw, Z.; Chang, S.F.; Chua, T.S. Visual translation embedding network for visual relation detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5532–5540. [CrossRef]
36. Krishna, R.; Zhu, Y.; Groth, O.; Johnson, J.; Hata, K.; Kravitz, J.; Chen, S.; Kalantidis, Y.; Li, L.J.; Shamma, D.A. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *Int. J. Comput. Vis.* **2017**, *123*, 32–73. [CrossRef]
37. Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; Yakhnenko, O. Translating embeddings for modeling multi-relational data. *Adv. Neural Inf. Process. Syst.* **2013**, *2*, 2787–2795. [CrossRef]

38. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556. [CrossRef]
39. Gruber, T.R. Toward principles for the design of ontologies used for knowledge sharing? *Int. J. Hum. Comput. Stud.* **1995**, *43*, 907–928. [CrossRef]
40. Maedche, A. Ontology—Definition & Overview. In *Ontology Learning for the Semantic Web*; The Kluwer International Series in Engineering and Computer Science; Springer: Boston, MA, USA, 2002; Volume 665, pp. 11–27. [CrossRef]
41. Zhang, J.; El-Diraby, T.E. Social semantic approach to support communication in AEC. *J. Comput. Civil. Eng.* **2012**, *26*, 90–104. [CrossRef]
42. Ye, Y.; Yang, D.; Jiang, Z.; Tong, L. Ontology-based semantic models for supply chain management. *Int. J. Adv. Manuf. Technol.* **2008**, *37*, 1250–1260. [CrossRef]
43. Uschold, M.; Gruninger, M. Ontologies: Principles, methods and applications. *Knowl. Eng. Rev.* **1996**, *11*, 93–136. [CrossRef]
44. Fernández-López, M. Overview of methodologies for building ontologies. *Knowl. Eng. Rev.* **1999**, *17*, 129–156. [CrossRef]
45. Fernández-López, M.; Gómez-Pérez, A.; Juristo, N. Methontology: From Ontological Art Towards Ontological Engineering. In Proceedings of the AAAI-97 Spring Symposium Series, Menlo Park, CA, USA, 24–26 March 1997; Available online: http://oa.upm.es/5484/ (accessed on 30 May 2022).
46. Noy, N.F.; McGuinness, D.L. *Ontology Development 101: A Guide to Creating Your First Ontology. Knowledge Systems Laboratory*; Stanford University: Stanford, CA, USA, 2001; p. 32.
47. Safety Handbook for Construction Site Workers. Occupational Safety and Health Branch Labour Department, Hong Kong, China. 2019. Available online: https://www.labour.gov.hk/eng/public/os/D/ConstrutionSite.pdf (accessed on 15 June 2022).