Advanced Engineering Informatics 42 (2019) 100966

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Full length article

Onsite video mining for construction hazards identification with visual relationships

Ruoxin Xiong[a], Yuanbin Song[a,*], Heng Li[b], Yuxuan Wang[c]

[a] School of Naval Architecture, Ocean & Civil Engineering, Shanghai Jiao Tong Univ., 800 Dongchuan RD, Minhang District, Shanghai 200240, China
[b] Dept. of Building and Real Estate, Hong Kong Polytechnic Univ., Rm ZS734, Hung Hom, Kowloon 999077, Hong Kong Special Administrative Region
[c] Jiangsu Key Laboratory of Urban ITS, School of Transportation, Southeast Univ., Si Pai Lou #2, Nanjing 210096, China

ARTICLE INFO

Keywords:
Scene graph
Hazards identification
Safety regulations
Ontology
Video mining

ABSTRACT

Widely-used video monitoring systems provide a large corpus of unstructured image data on construction sites. Although previous developed vision-based approaches can be used for hazards recognition in terms of detecting dangerous objects or unsafe operations, such detection capacity is often limited due to lack of semantic representation of visual relationships between/among the components or crews in the workplace. Accordingly, the formal representation of textural criteria for checking improper relationships should also be improved. In this regard, an **Automated Hazards Identification System (AHIS)** is developed to evaluate the operation descriptions generated from site videos against the safety guidelines extracted from the textual documents with the assistance of the ontology of construction safety. In particular, visual relationships are modeled as a connector between site components/operators. Moreover, both visual descriptions of site operations and semantic representations of safety guidelines are coded in the three-tuple format and then automatically converted into Horn clauses for reasoning out the potential risks. A preliminary implementation of the system was tested on two separate onsite video clips. The results showed that two types of crucial hazards, i.e., failure to wear a helmet and walking beneath the cane, were successfully identified with three rules from Safety Handbook for Construction Site Workers. In addition, the high-performance results of Recall@50 and Recall@100 demonstrated that the proposed visual relationship detection method is promising in enriching the semantic representation of operation facts extracted from site videos, which may lead to better automation in the detection of construction hazards.

1. Introduction

Construction work is a hazardous field-based job. This involves occupational risk exposures (e.g., electricity), tools (e.g., nail guns), equipment (e.g., crane), and environment (e.g., dust and noise) in the daily workplace. Worldwide, the hazards have caused incredible physical injuries and pecuniary losses, as well as schedule delays in construction projects. In detail, the United States had 5,147 fatal workplace injuries and illnesses in 2017, while the construction occupational group accounted for 47 percent of worker deaths [1]. The National Academy of Social Insurance (NASI) estimated that compensation programs paid $2.23 billion for employees with work-related injuries and illnesses in the construction industry [2]. As suggested by Jehring and Heinrich [3], hazardous operations and conditions are the two direct causes of accidents on construction sites. Systematic monitoring of onsite activities thus plays a crucial role in hazards prevention and elimination by providing decision-support information. In practical applications, field observations and inspections are commonly conducted approaches to assessing the potential hazards occurred in operational tasks or site conditions [4–6]. During the investigation, qualified supervisors serve to visually examine the workplace and observe operators perform a specific work with the safety checklist, while it can be costly, time-consuming, and error-prone for manual observations and documentation. On the other hand, as they operate without requirements to attach any sensors and enable more comprehensive onsite information (e.g., locations and activities of construction entities), computer vision techniques bring operational and technical advantages over various wearable sensors, such as radio frequency identification (RFID) [7–10], global positioning system (GPS) [11–13], inertial measurement unit (IMU) [14–16], and ultra-wideband (UWB) [17,18].

Computer vision-based methods for hazards identification can be typically divided into three main groups: (1) entity-based; (2) location-

* Corresponding author.
E-mail addresses: xiongrxchn@sjtu.edu.cn (R. Xiong), ybsong@sjtu.edu.cn (Y. Song), heng.li@polyu.edu.hk (H. Li), wangyuxuan@seu.edu.cn (Y. Wang).

https://doi.org/10.1016/j.aei.2019.100966
Received 12 March 2019; Received in revised form 6 June 2019; Accepted 13 July 2019
Available online 02 August 2019
1474-0346/

based; and (3) operation-based. First, entity-based risk assessments require identifying specific entities in the static scene, such as improper storage of flammable and explosive materials or failure to wear the helmet and protective clothing. For these kinds of hazards, object detection algorithms have been useful for the identification of targeted entities in construction projects [19,20]. Secondly, location-based risks such as improper working position necessitate the dynamic trajectories of the entity in a sequence of images, which are usually implemented by object tracking methods in construction [21,22]. Lastly, operation-based hazards identification mainly emphases on the action recognition of equipment or workers in the workplace [23–25]. However, the hazards detection capacity is usually confined in the large-scale scene without extensive descriptions of visual descriptions between/among the site components or crews. Workers interact with tools, equipment, materials, or other workers frequently at construction sites, which offers supplementary evidence for operation inspections. In this view, the need for detailed relations detection, as well as semantic representation in a computer-accessible format is of particular importance.

In this paper, we aim at proposing a unified model for automated risk identification with onsite videos. This requires both perceptual and reasoning capabilities. For instance, computers have to semantically understand image scenes using their perceptual capabilities such as object detection/tracking, action recognition, or visual relations detection. Then, visual information should be assessed by comparing it with regulations, standards, or guidelines to recognize hazardous operations and conditions. To this end, we introduce the scene graph to "imitate" images understanding of humans and express detailed scenes in the structured language (data). The notion of scene graph visualizes and represents the richness of objects, and relationships between objects that can exist in an intuitive graph-structure, as illustrated in Fig. 1. Provided an onsite image, the visual relationship detection is to explicitly encode object instances and relationships between objects in the form of visual triplets (e.g., (object$_1$, relation, object$_2$)). Then, regulatory documents are semantically mapped into a set of three-tuples for information representations, which incorporates knowledge external to visual facts. The construction safety ontology model is established to fill the semantic gap between visual and textual information. Finally, we design a logic reasoner for automated hazards identification. The extracted information instances (in three-tuple formats) are further transformed into Horn clauses represented as facts and rules, and an automated reasoner is initiated by executing a query over these relations for rule checking. Our contributions are: (1) before this research, the valuable relations information between site objects in the image is not fully used. We use the scene graph to model rich vocabularies of operation elements, including objects and relationships that specify relations between objects. It provides deeper insights towards capturing the structural representations of the onsite scene; (2) we define three-tuple structures to represent concepts and relations involved in safety documents. This helps to enrich scene graphs derived purely from images with external knowledge sources and enable computer manipulation and reasoning; and (3) we propose an Automated Hazards Identification System (AHIS), which provides a unified model for multiple hazards recognition tasks, and this system could enhance the generality and automation in the process of construction safety management.

2. Related works

Over the past decade, computer vision technologies have been frequently applied to the automatic monitoring and analysis process in construction [26–28]. When it comes to image understanding, the Architecture, Engineering & Construction (AEC) industry is still struggling to figure out basic questions such as: how should we represent it? Usually, workers or equipment are detected and tracked (e.g., objects detection, objects tracking), or their activities are analyzed (e.g., action recognition). However, those detection approaches generally cannot provide sufficient information to represent site scenes involving multiple objects that interact with each other in the workplace [29]. An apparent alternative idea to represent the content of visual scenes is natural language descriptions. For instance, Socher et al. [30] induced detailed text sentences to represent the recorded images. The results showed that the explicit recognition of semantic elements benefits image understanding and improves image retrieval. However, those text-based representation methods are typically limited because of ambiguity and expressiveness [31].

Recently, graph-structured methods have developed to address structured representations of visual scenes [32,33]. Scene graph, as defined by Johnson et al. [36], is a detailed and formal representation for image compositions, where entities are coded with nodes that connect with edges denoting pairwise relationships between entities. These efficient scene representations have shown wide-spread use to improve various computer vision tasks, such as image retrieval [37], image captioning [38], object detection [39,40], image generation [41], and visual question answering [42]. In this paper, we apply the graph-structured method to represent the onsite construction scene because: (1) as compared to other vision-based methods (e.g., object detection, object tracking, and action recognition), the scene graph can generate comprehensive descriptions for multiple interactions of entities in the workplace; and (2) The efficient and structured representation format (in three-tuple) enables automated manipulation and evaluation of computers. In addition, existing datasets for visual relationship detection at a large scale are Visual Relationships dataset [43] and Visual Genome dataset [44], which are typically collected from daily life scenarios. The cluttered construction scenarios, characterized by diverse categories of specialized equipment and materials, and continuously changing working conditions, may result in several technical issues for those vision-based methods, such as illumination and occlusion. In this regard, whether the graph-structured representations model would perform well on the real construction dataset remains to be evaluated in this paper.

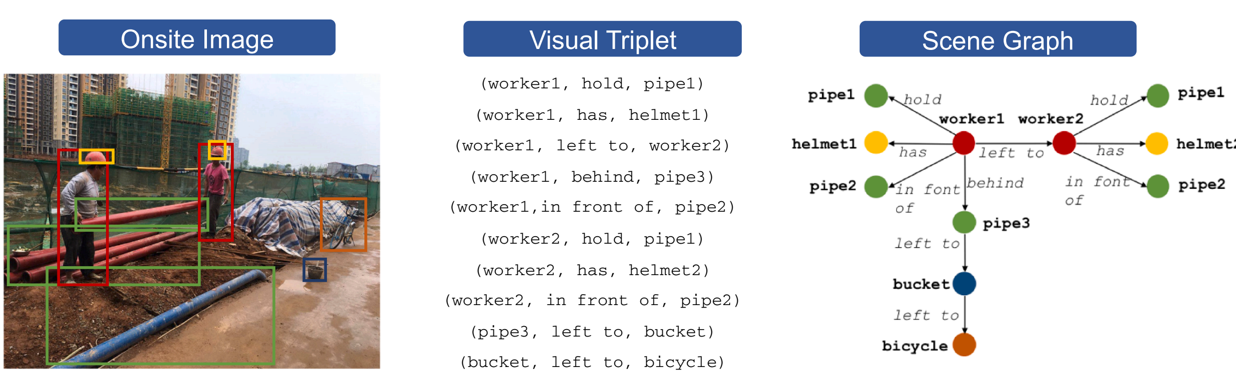

Fig. 1. An example of on-site construction image and its scene graph.

While the richness of scene graph descriptions has helped the complete extraction of onsite images, a missing link between visual facts and external knowledge (e.g., regulatory information) become apparent. Onsite operations must comply with various regulatory documents from current safety regulations and industry standards for hazards prevention. For example, Chi and Caldas [21] identified spatial safety risks on earthmoving operations based on safety assessment regulations: (1) speed limitations, (2) in hazardous areas; and (3) proximity between entities. Zhang et al. [45] performed ergonomic posture recognitions from site cameras based on the Ovako Working Posture Analyzing System (OWAS). As such, a considerable number of safety rules exist, with each rule having its own semantic contents and formatting structures. Therefore, computer vision methods should aim to extract scene compositions, but also transform heterogeneous rules into computer-accessible formats to evaluate potential hazards. However, safety regulations are documented in arbitrary text fields rather than constrained values. The difficulty of formally extracting and transforming such text information has led to less consideration in previous studies. Inspired by semantic representations of knowledge base (KB) [46,47], a three-tuple format is selected to regulations representation because: (1) those formally-defined tuples have sufficient expressiveness to represent concepts and relations involved in documents; and (2) once the regulatory information is correctly transformed in a logic format, the hazards reasoning process can be conducted with derived visual triplets of onsite images in a fully-automated way. In this way, the new transformation of regulatory rules and standards (in three-tuple) allow with the generality of hazards for diverse tasks and facilitates the automated reasoning process of rule checking.

3. Methodology

To make full use of semantic information in images, we propose a general framework for hazards identification on construction sites, as shown in Fig. 2. The visual relationship detection module uses a deep learning method to capture the rich of entities and relationships between entities by a visual triplet in the form of (a, b, c), where a, b, and c represent the object$_1$ category, the relation predicate, and the object$_2$ category respectively. Construction regulatory statements are coded with semantic phrases (in three-tuple) and logical connectives to form commonly accessible representations available for the computer. Construction ontology model is also involved in sharing basic concepts and relationships for deep reasoning. Finally, the joint knowledge inference engine automatically identifies hazards by simultaneously examining visual facts against semantic rules.

3.1. Scene graph generation

Relationship facts, standing for relationships between two entities, play a major role in representation and reasoning in scene understanding. In this paper, we focus on such relations as geometry (e.g., beneath), possession (e.g., has), and actions (e.g., hold) to describe where, what, and how entities are related to each other in the workplace. Our visual relationship detection method is developed based on the work of Dai et al. [48] because it has shown impressive results on Visual Genome dataset [44]. However, our model adapts the previous one in three aspects: (1) we obtain object detections using SSD of Liu et al. [49] instead of Faster R-CNN [50] for its faster detection and comparable accuracy; (2) statistical relationships in the construction domain are integrated to enhance the entity pair detections; and (3) we visualize and represent the captured visual triplets in a scene graph for its intuitive format.

The adapted visual relationship detection method involves four sequential modules: **entity detection**, **pair filtering**, **joint recognition**, and **graph representation**, as illustrated in Fig. 3. Given an image, individual construction entities are first located by an object detector. The pair filter module is designed to filter out meaningless pairs, thus substantially reducing the computational cost of relationship recognition. For each pair of entities, both entity categories and spatial configurations are extracted and further compressed through fully-connected (FC) layers. Then, compressed features of pairs, as well as appearance features, will be jointly inputted to the joint recognition module, thus producing multiple triplets that can exist as the output. Finally, we visualize semantic representations of an image in the form of a scene graph.

Entity Detection SSD network is applied to locate construction entities with a bounding box and extract appearance features. It first provides a set of fixed-size bounding boxes in the image. Let $x_{ij} = \{1,0\}$ be an indicator for assigning the i-th bounding box to the j-th ground-truth box. Both center offsets and confidences f are also predicted for each default box. Then, we generate scores for predicted instances in those boxes. The overall objective loss function is given as follows:

$$\mathrm{L}(x, f, p, t) = (\mathrm{L}_\alpha(x, f) + k\mathrm{L}_\beta(x, p, t))/N \quad (1)$$

where L_α is a Smooth L_1 loss between the ground-truth box t and the predicted box p, L_β is a softmax loss on label confidences, N is the number of matched boxes, and k is the weight coefficient.

Pair Filtering The next step is to match meaningful pairs with considerable entities detected in the image. However, it is almost impossible to explore all those potential pairs. To filter out useless pairs, we mainly consider two aspects:

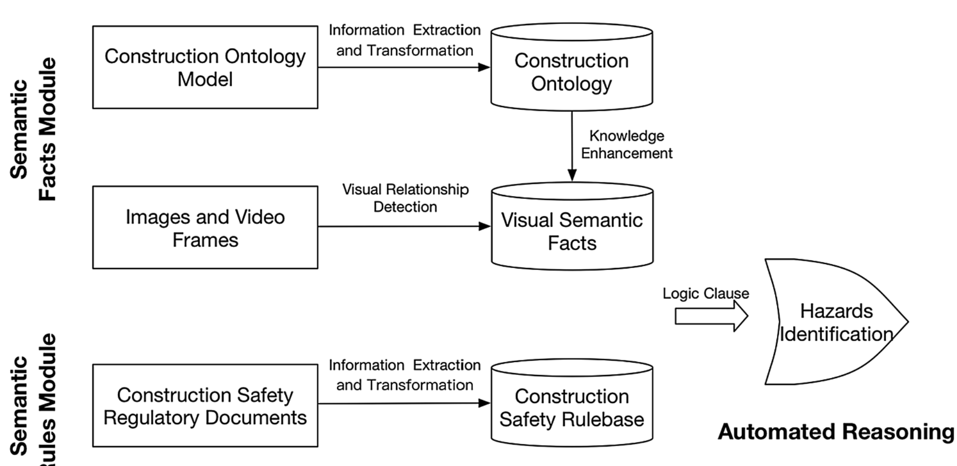

Fig. 2. The overall pipeline of the proposed method.

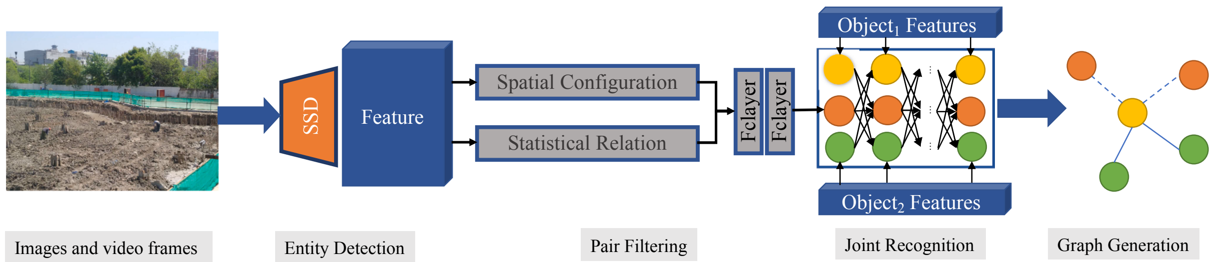

Fig. 3. The proposed framework for scene graph generation.

- Spatial configurations: the spatial features (e.g., locations and sizes) can reflect the correlation between part members. Entities too far away are almost impossible to be related as pairs. Such cues are invariable to dynamic conditions, e.g., the changes in illumination.
- Entity categories: strong statistical dependencies exist between entity categories. For example, some certain entities are unlikely to construct a meaningful pair, such as (crane, helmet) or (gloves, truck).

Thus, we introduce a pair filtering network (see Fig. 4) to solve this problem. Dual spatial masks derived from bounding boxes are used as entity representations to leverage spatial configurations. For each possible pair, there exists two binary masks, one for the object$_1$ and the other for the object$_2$. To train this network, we randomly sample pairs of bounding boxes and labels from the targeted image and divide the dataset into two samples: the positive sample and negative sample, based on Intersection over Union (IoU) with ground-truth pairs and labels.

Joint Recognition Visual relationships are then detected by jointly learning from multiple factors above. Typically, the joint inference of visual relationship can be formulated based on Conditional Random Field (CRF) [51],

$$p(b, a, c|\mathbf{x}_b, \mathbf{x}_a, \mathbf{x}_c) = \exp(\mathrm{J}(b, a, c|\mathbf{x}_b, \mathbf{x}_a, \mathbf{x}_c; \mathbf{U}))/M \tag{2}$$

where $\mathbf{x}_b$ is the compressed feature of pairs; $\mathbf{x}_a$ and $\mathbf{x}_c$ are appearance features for a and c, respectively; and M is the normalizing matrix, depending on the parameters $\mathbf{U}$. The overall potential J can be obtained from a sum of single potentials,

$$\begin{aligned} \mathrm{J} = {} & \psi_e(a|\mathbf{x}_a; \mathbf{U}_e) + \psi_e(c|\mathbf{x}_c; \mathbf{U}_e) + \psi_b(b|\mathbf{x}_b; \mathbf{U}_b) + \varphi_{ba}(b, a|\mathbf{U}_{ba}) \\ & + \varphi_{bc}(b, c|\mathbf{U}_{bc}) + \varphi_{ac}(a, c|\mathbf{U}_{ac}) \end{aligned} \tag{3}$$

where ψ_e is the potential of individual entities with appearance features; ψ_b is the potential of the relationship predicate with compressed pair features; and φ_{ba}, φ_{bc}, and φ_{ac} respectively represents statistical relations among b, a, and c.

Then the posterior distribution of r is given as follows

$$p(\mathrm{b}|a, c, \mathbf{x}_b; \mathbf{U}) \propto \exp(\psi_b(\mathrm{b}|\mathbf{x}_b; \mathbf{U}_b) + \varphi_{ba}(b, a|\mathbf{U}_{ba}) + \varphi_{bc}(b, c|\mathbf{U}_{bc})) \tag{4}$$

Let $\mathbf{v}_b$ be the posterior probability vector for b, the Eq. (4) thus can be represented as

$$\mathbf{v}_b = \boldsymbol{\sigma}(\mathbf{U}_b\mathbf{x}_b + \mathbf{U}_{ba}\mathbf{v}_a + \mathbf{U}_{bc}\mathbf{v}_c) \tag{5}$$

where $\boldsymbol{\sigma}$ is the softmax function; and $\mathbf{v}_a$, $\mathbf{v}_b$ are probabilistic vectors of a and b.

Similar derivation can infer a and c. As a consequence, we can get a series of updating formulas, which take the current probability vectors $\mathbf{v}_a$, $\mathbf{v}_b$, and $\mathbf{v}_c$ as inputs, and output the updated vectors $\mathbf{v}'_a$, $\mathbf{v}'_b$ and $\mathbf{v}'_c$,

$$\mathbf{v}'_a = \boldsymbol{\sigma}(\mathbf{U}_e\mathbf{x}_a + \mathbf{U}_{ab}\mathbf{v}_b + \mathbf{U}_{ac}\mathbf{v}_c) \tag{6}$$

$$\mathbf{v}'_b = \boldsymbol{\sigma}(\mathbf{U}_b\mathbf{x}_b + \mathbf{U}_{ba}\mathbf{v}_a + \mathbf{U}_{bc}\mathbf{v}_c) \tag{7}$$

$$\mathbf{v}'_c = \boldsymbol{\sigma}(\mathbf{U}_e\mathbf{x}_c + \mathbf{U}_{ea}\mathbf{v}_a + \mathbf{U}_{cb}\mathbf{v}_b) \tag{8}$$

Scene Graph Generation Since the entity categories and corresponding pairwise relationships have been predicted, we generate a directed graph for each image so as to visualize the captured relationships in an intuitive structure. Formally, provided a set of entities O and relationship types S, we represent the scene graph with a matrix G suggested by Li et al. [32]. In the matrix G, the element G_{ii} is the i-th object O_i at diagonal position (i, i) and the element G_{ij} $(i \neq j)$ is the phrase representing the relationship S_{ij} between object O_i and O_j. If the object O_i and O_j are relevant, then two objects are connected through the predicate G_{ij}. In this way, we will obtain a scene graph based on the matrix G.

3.2. Regulatory information representation and ontology development

Existing regulatory rules in construction domains are mostly documented with natural language sentences. Accordingly, this research provides a textual information representation and transformation method for computer manipulation and evaluation. The representation format consists of two basic parts: semantic phrases and logical connectives. For each safety rule, semantic phrase elements are defined as a three-tuple (e.g., (object$_1$, relation, object$_2$)). "object$_1$" or "object$_2$" is subjected to a particular ontology in regulations, which can be a "thing" (e.g., equipment, building structure) or "personnel" (e.g., worker). "Relation" semantically connects objects with limitations, such as actions (e.g., hold, carry), geometry (e.g., beneath, in front of, on), and possession (e.g., has). For those complex rules with multiple restrictions, regulatory statements are decomposed and transformed to serval semantic phrases with logical connectives. Three forms of logical connectives are used:

- Logical conjunction (∧): $S_1 \wedge S_2$ is true only if S_1 is true and S_2 is true;
- Logical disjunction (∨): $S_1 \vee S_2$ is true if S_1 is true, or S_2 is true;

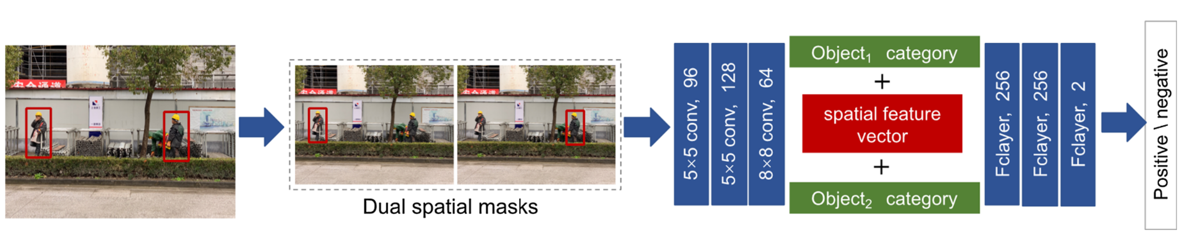

Fig. 4. The network structure of pair filter module.

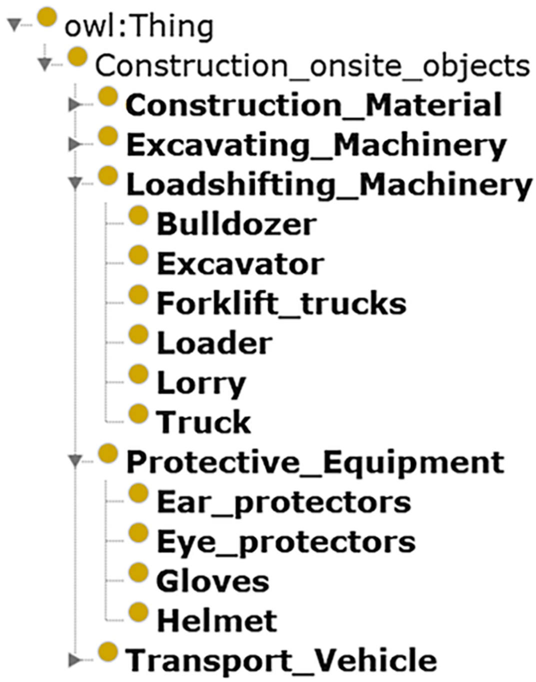

Fig. 5. Partial view of construction safety ontology.

- Negation (¬): an operation that takes a statement S to another statement "not S".

On the other hand, there may exist a semantic gap between visual representations and textual definitions. For example, we usually use a specific label to represent the entities in images (e.g., helmets, gloves), while those entities may be documented as "Personal Protective Equipment" in regulations. Ontology is a knowledge representation structure that can provide an unambiguous terminology that can be shared by all involved in the reasoning process [52]. In this regard, this paper develops an ontology model that extracts knowledge from the safety domain. The ontology model is developed as follows: (1) Identify terms that most commonly used in construction safety domains, and (2) Develop a taxonomy of those terms with their functions and relations. The knowledge sources used to identify relevant concepts and construct the safety ontology model include Safety Handbook for Construction Site Workers [53], The Construction (Design and Management) Regulations 2015 [54], and Recommended Practices for Safety & Health Programs in Construction [55]. To facilitate the development of an ontology model, we use Protégé, a free and open-source ontology editor and framework, to define the taxonomy and relations [56]. Fig. 5 shows a partial view of the developed ontology model.

Generally, as described in previous, the rules in construction safety domain can be classified into three groups: (1) entity-based; (2) location-based; and (3) operation-based, as shown in Table 1. Entity-based rules mainly concern possession relationships (e.g., has, part of). Location-based requirements describe improper locations of project entities in the workplace, which are engaged with geometric relationships (e.g., beneath, inside). Action-based rules require action relationships (e.g., contact, hold) to describe the hazardous operations of workers or equipment. Besides, a semantic mapping step is required to match the extracted information elements to their logic representations: (1) the subject of rules should be completed in constructing a three-tuple; and (2) the element instances in logic representation should be used as their simplest forms.

3.3. Automated reasoning for rule checking

The hazards in the workplace can be identified by comparing extracted visual triplets from images with the transformed rules from documents. In addition, the resultant visual triplets, regulatory phrases, and ontology model are automatically converted into the Horn clauses with a developed batch file. In this section, Prolog, a widely-used logic programming language, is utilized to automated reasoning for hazards identification. In Prolog, the program logic is represented in terms of relations and expressed as facts and rules [61], which is well-suited for this logical hazards-identification task that benefits from rule-based queries.

Therefore, visual information obtained from the job sites can be treated as logic facts, and a reasoner is initiated by executing a query over these relationships to automatically recognize rule violations. As indicated in Table 2, relations and queries are using Prolog's Horn clauses. For example, the three-tuple representation of visual facts and regulations are transformed as relation (object$_1$, object$_2$). Logical connectives also have their corresponding syntax in the Prolog. As for ontology concepts in rule representations, their element instances are assigned to relevant concepts and relations using unary predicates. For example, the "helmet" is represented by the "protective_equipment (helmet)" with "helmet" being the instance name and the element "protective_equipment" being the argument for the instance. Then, we use queries to initiate execution with SWI-Prolog, which offers a comprehensive Prolog environment [62]. Given a query, the inference engine attempts to search a refutation of the negated query. If the negated query is refuted, it will output an appropriate variable in place, which is a logic result of the program, and continue to search the query. The query is set to have succeeded until all generated variables are reported to users.

4. Experiments

4.1. Onsite visual relation dataset

To train and test the proposed method in this paper, we constructed a visual relationship dataset collected from real construction sites considering points of views, illumination, weather, and occlusion condition to avoid potential bias. The dataset covers several common construction activities, such as "Transporting", "TieRebar", "Drilling", "Bolting", and "Plastering". Construction site entities were categorized into six groups: (1) personnel (e.g., worker), (2) material (e.g., pipe, rebar, brick), (3) tool (e.g., hammer, shovel, barrow), (4) equipment (e.g., bulldozer, crawler crane, tower crane), (5) facility or structure (e.g., wall, window, stair), and (6) general (e.g., car, bicycle, motorcycle). For each image, we manually annotated all existed entities with bounding boxes and described relationships with (object$_1$, relation, object$_2$) tuples using fixed labels. In total, we have annotated 18

Table 1
Examples of regulatory requirements and logic representations.

Type	Regulatory requirements	Logic representations
Object-based	"Wear a safety helmet on a construction site." (Personal Protective Equipment Regulation)	has (worker, helmet)
Space-based	"Do not work beneath any suspended load." (Lifting Appliance and Gear Regulation)	¬beneath (worker, suspended load)
Operation-based	"Remember to wear gloves when contacting chemicals." (Personal Protective Equipment Regulation)	has (worker, gloves) ∧ contact (worker, chemicals)

*Regulatory information source: Safety Handbook for Construction Site Workers, 2004 [53].

Table 2
Basic syntax in Prolog.

Element	Logic representation	Syntax in SWI-Prolog
Fact	(object$_1$, relation, object$_2$)	Relation (object$_1$, object$_2$)
Query	(object$_1$, relation, object$_2$)	-Relation (object$_1$, object$_2$)
Ontology	(Cluster, objects)	Cluster (objects)
Logical conjunction	$\wedge$	,
Logical disjunction	$\vee$	;
Negation	$\neg$	not

relationships and 40 entities in the dataset.

Then, we randomly split our dataset into two parts: 70% images for training and 30% images for testing. The data augmentation strategy was used to ensure the model more robust to diverse input scenes, so as to improve the detection performance [63]. Each image from training dataset was randomly processed as follows:

– color jittering;
– randomly add Gaussian noise;
– use the original input image.

In this paper, the code was written in python, and all networks were implemented using Caffe. We performed our algorithms on a PC equipped with a 4.00 GHz Intel(R) i7-6700 K CPU, two NVIDIA GeForce GTX 1080 GPUs and 64G RAM. The time for computing visual relationships was around 2 frames/s.

To integrate all relevant functions in an automated way, we developed a prototype system called “AHIS” (automated hazards identification system), as shown in Fig. 6. The scene graph was generated on the screen for its intuitive expressiveness to represent entities and relations involved in images. Construction regulatory statements were coded with semantic phrases (in three-tuple) and logical connectives for rule checking. Note that more rules can be emerged to the automated reasoner and images with the same scene graph are skipped in hazards checking. Finally, we could execute the system and output an appropriate variable in place for the violated image and rule, as well as export the results in text format.

4.2. Performance evaluation

To demonstrate and assess the validity of the proposed AHIS, we randomly chose three rules from Safety Handbook for Construction Site Workers [53].

Rule 1: Wear a safety helmet on a construction site.
Rule 2: Do not work beneath any suspended load.
Rule 3: Wear suitable eye protectors while operating an electric drill.

Those rules were semantically extracted and represented in three-tuple formats: (worker, ¬has, helmet), (worker, beneath, suspended load), and (worker, ¬has, eye protectors) ∧ (worker, use, electric drill), respectively. Meanwhile, the visual facts of two separate onsite video clips were extracted with the visual relation detection network. After conducting the semantic mapping process of visual facts and rules, these semantic phrases were further converted into Horn-Clause-type logic representations. Finally, we implemented the rule checking process with the developed AHIS. It turned out that two types of hazards were identified, including failing to wear a helmet and walking beneath the cane, as illustrated in Fig. 7. After that, those two video clips were manually reviewed and inspected by two researchers to verify its correctness. The result showed that all involving hazards were successfully identified.

We then evaluated the accuracy of visual relationship detection on our dataset and compared with the random results in the experiments.

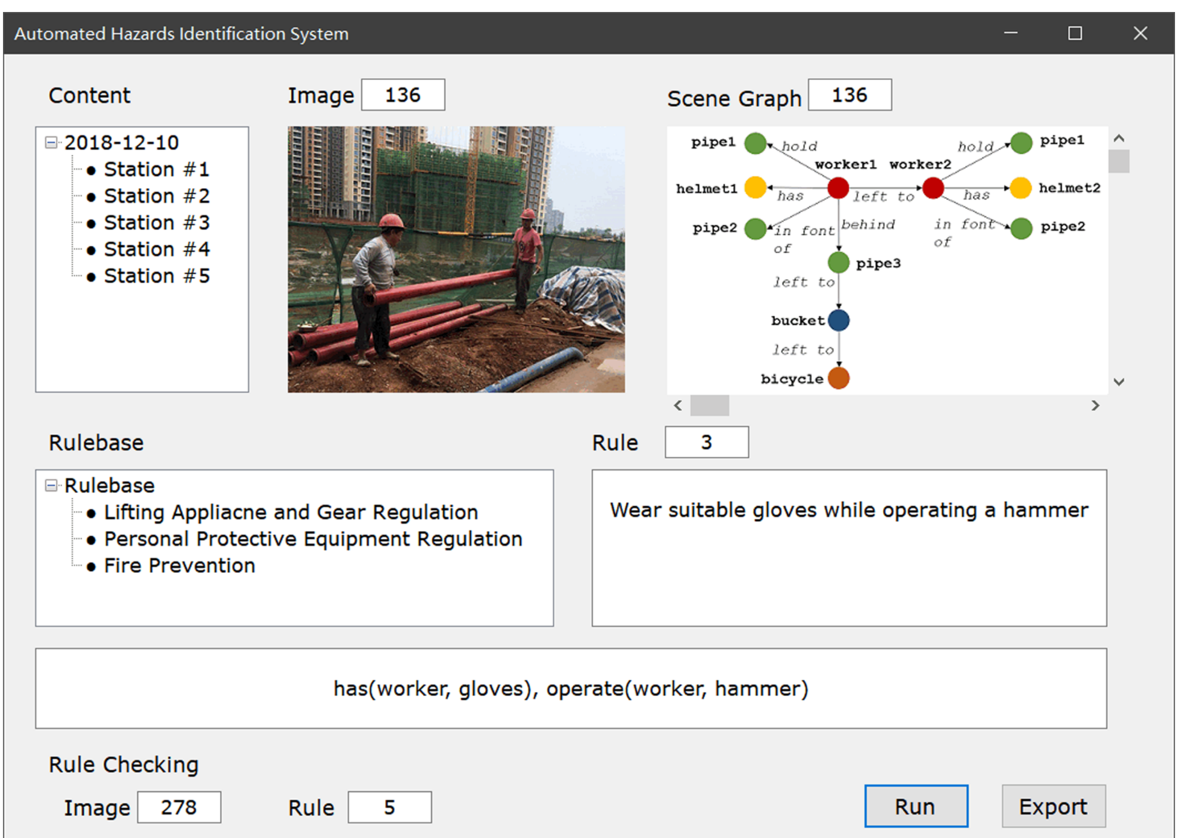

Fig. 6. The prototype system of AHIS.

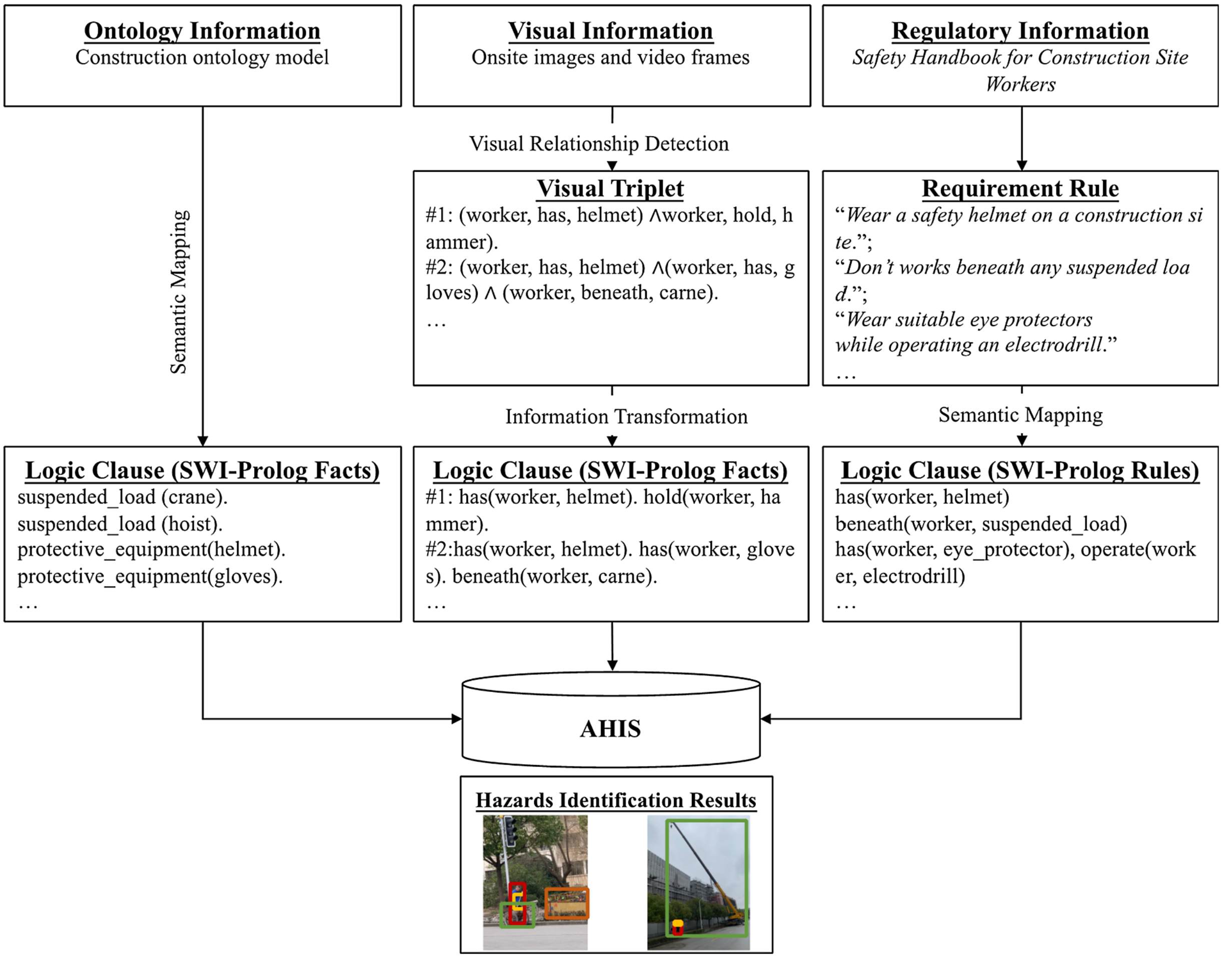

Fig. 7. The hazards identification process of the automated reasoner.

The reference results used random permutation in ranking the answers. Detecting visual relationships involves the localization of both entities, recognition of entity categories, and prediction of the relation between entities. To evaluate how the proposed model performs on the dataset, we focused on the prediction results under two conditions:

- Entity Detection: the detection involves a set of entity pairs and bounding boxes for each entity;
- Relation Detection: the detection requires a set of relationship triplets and bounding boxes of for each entity.

Following Xu et al. [33], the evaluation metric used in this paper was Ranked Recall, which assesses the performance of the classifier to detect the highest number of relevant samples. Recall@x denotes the proportion of ground-truth (GT) relationships that appear on the top-x confident results. Here, we measure Recall@100 and Recall@50 for evaluation. Since we have 18 predicates and 40 entities in the dataset, the possible relationship predictions are $40 \times 18 \times 40$, which suggests that the random combination will result in a Recall@100 of 0.0035. As shown in Table 3, our model is able to perform good results of effective learning between objects and relations, e.g., on relationship detection and entity detection measured by Recall@50, we were 13.6 and 36.3 respectively.

Fig. 8 shows the qualitative results of the scene graph model trained with our dataset and the reference model. As compared to other vision-based methods like object detection in Fig. 8(a), the scene graph can generate structural relationship descriptions and has sufficient expressiveness for multiple interactions of entities in the workplace. Moreover, we compared the scene graph with the text description method, which uses description sentences to represent the content of visual scenes. Since the same onsite scene can be described in myriad levels, the text-based representation method is typically limited because of ambiguity and expressiveness. Our scene graph model is able to resolve some of the ambiguity with semantic triplets, which embody rich vocabularies of visual elements. For example, it predicts semantic triplets, e.g., (worker, has, helmet), (worker, use, barrow), (pipe, on, barrow) and visualizes an intuitive scene graph in Fig. 8(c). So those formally-defined tuples enable automated manipulation and evaluation of computers.

Table 3
Performances of entity detection and relationship detection.

Task	Entity Det.		Relationship Det.	
Metric	Recall@50	Recall@100	Recall@50	Recall@100
Results	**36.3**	**50.2**	**13.6**	**14.3**
Rand	0.005	0.01	0.0017	0.0035

5. Limitations and discussion

The developed AHIS is successful in automatically performing rule checking for hazards identification. However, there are some current limitations and challenges. First, the necessary, but expensive dataset to recognize rich visual relationships could limit the application of the proposed method in the construction industry. There are various publicly datasets for multiple tasks. For example, PASCAL VOC [64] and

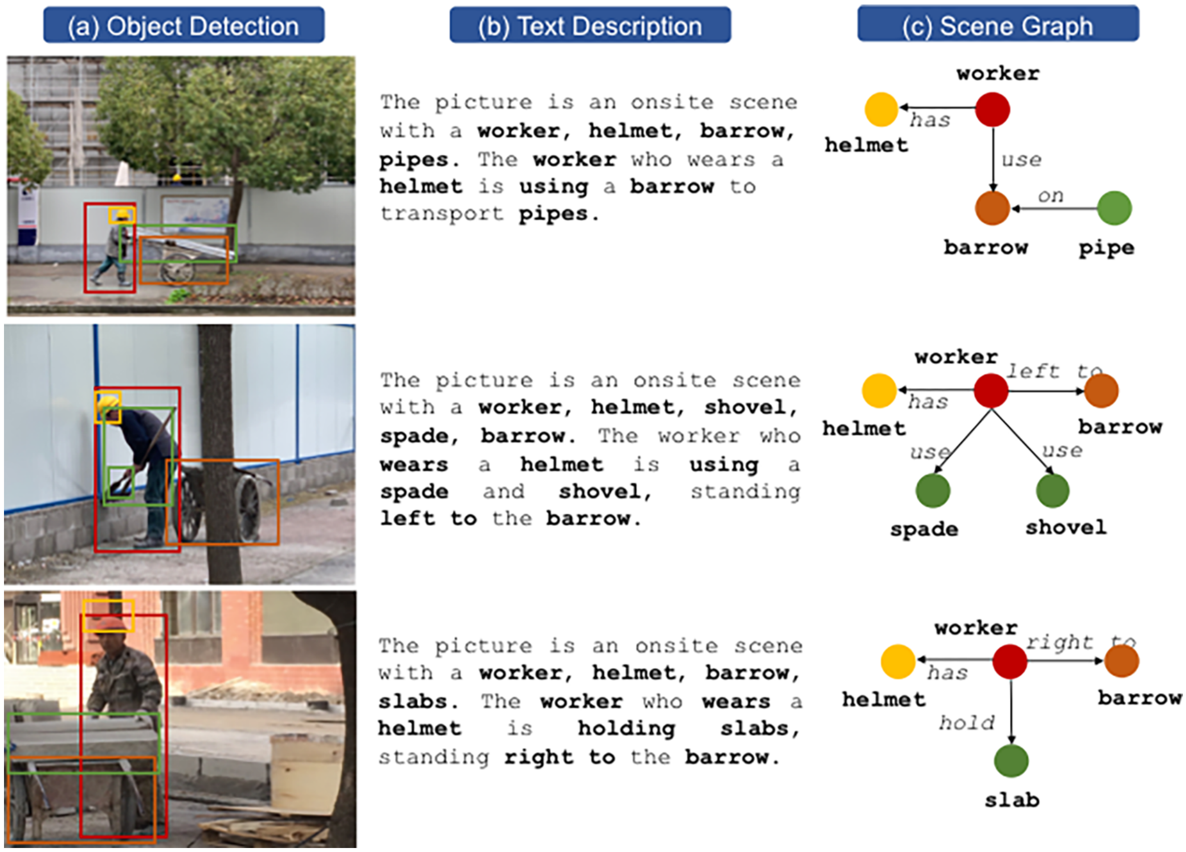

Fig. 8. Sample representations from the reference model and our scene-graph model.

MS COCO [65] are built for object detection and classification, and UCF-101 [66], and SPORTS-1M [67] are developed for activity recognition. However, there is a lack of such a publicly available dataset for visual relations detection in construction domains. Such a dataset with sufficient samples is critical in training and testing, while it is expensive and time-consuming to develop a domain-specific dataset for the heavy labeling tasks. This study only investigated a small experimental dataset for visual relationship detection collected from construction sites. A larger dataset offers the potential to improve the accuracy of detection tasks.

Secondly, in this paper, we consider limited types of visual relationships, which are the most commonly used interactions in rule checking tasks. Note that the detail of relationships can be altered according to the monitoring applications. For example, if an ergonomic analysis is required for comprehensive construction management applications, posture detections can be added to visual triplet representations (e.g., (worker, body_part, location)) to describe their motion (see Fig. 9).

Lastly, it requires manual effort to extract regulatory documents and encode them in a computer-processable format, which can be time-consuming, costly, and error-prone. Recently, many attempts have already been explored for automated information extraction in many domains [69–71]. For example, Zhang et al. [72] presented a rule-based natural language processing (NLP) method to extract semantic information from construction regulatory documents automatically. In the future, we will utilize advanced semantic NLP methods to facilitate the automated process for extracting safety rules from textual documents and formalizing these regulations in three-tuples.

6. Conclusion

This paper offers a novel AHIS framework to automatically identify hazards in the workplace and facilitate the generation of insights into the process of construction safety management. In this framework, the source video frames are encoded into graph-structured representations by a visual relation detection network, while construction regulatory documents are extracted and transformed into semantic phrases and logical connectives with the assistance of construction safety ontology. Subsequently, the resultant graph-structured and semantic representations are further converted into Horn clauses for automated rule checking. In particular, we employ a deep learning approach to generate scene graphs, which can organize video contents according to semantic relationships. Visual relationship detection network embodies rich vocabularies of visual elements, including object categories and relationships between objects. We then transform the rule codes with semantic phrases and logical connectives to evaluate the onsite facts with safety guidelines. This helps to enrich scene graphs with external safety guidelines and enable computer manipulation and reasoning. Besides, both visual contents and textual documents are extracted and

Fig. 9. Posture detection based on OpenPose [68].

transformed into the three-tuple representation for its straightforward and sufficient expressiveness to represent concepts and relations.

The proposed AHIS was tested on two separate video clips against three rules from Safety Handbook for Construction Site Workers. Two types of crucial hazards, i.e., failing to wear a helmet and walking beneath the cane, were successfully identified. Using Recall@x evaluation matrices, our model achieved 13.6 and 36.3, respectively on relationship detection and entity detection measured by Recall@50. These high-performance results indicated that our model is promising in effective learning between objects and relations. As compared to the text description method, our scene graph model was able to resolve some of the ambiguity. As part of future/ongoing research, we will continue to construct a larger dataset to improve the accuracy of detection tasks and utilize NLP method to extract semantic phrases from construction regulatory documents automatically.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to acknowledge the financial support of the National Natural Science Foundation of China (Grant No.71271137). Any opinions, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily represent those of the National Natural Science Foundation of China.

References

[1] U.S. Bureau of Labor Statistics (BLS), Employer-Reported Workplace Injury and Illnesses, November 2018. < https://www.bls.gov/news.release/pdf/cfoi.pdf > (last access March 5, 2019).
[2] CPWR – The Center for Construction Research and Training, The Construction Chart Book, sixth ed., February 2018. < https://www.cpwr.com/publications/research-findings-articles/construction-chart-book > (last access March 7, 2019).
[3] J. Jehring, H.W. Heinrich, Industrial accident prevention: a scientific approach, Ind. Labor Relat. Rev. 4 (1951) 609, https://doi.org/10.2307/2518508.
[4] B. Buchholz, V. Paquet, L. Punnett, D. Lee, S. Moir, PATH: a work sampling-based approach to ergonomic job analysis for construction and other non-repetitive work, Appl. Ergon. 27 (1996) 177–187, https://doi.org/10.1016/0003-6870(95)00078-X.
[5] O. Rozenfeld, R. Sacks, Y. Rosenfeld, H. Baum, Construction job safety analysis, Saf. Sci. 48 (2010) 491–498, https://doi.org/10.1016/j.ssci.2009.12.017.
[6] Michael C. Gouett, Carl T. Haas, Paul M. Goodrum, Carlos H. Caldas, Activity analysis for direct-work rate improvement in construction, J. Constr. Eng. Manage. 137 (2011) 1117–1124, https://doi.org/10.1061/(ASCE)CO.1943-7862.0000375.
[7] W. Qin, R.Y. Zhong, H.Y. Dai, Z.L. Zhuang, An assessment model for RFID impacts on prevention and visibility of inventory inaccuracy presence, Adv. Eng. Inf. 34 (2017) 70–79, https://doi.org/10.1016/j.aei.2017.09.006.
[8] A. Costin, N. Pradhananga, J. Teizer, Leveraging passive RFID technology for construction resource field mobility and status monitoring in a high-rise renovation project, Autom. Constr. 24 (2012) 1–15, https://doi.org/10.1016/j.autcon.2012.02.015.
[9] Z. Ren, C.J. Anumba, J. Tah, RFID-facilitated construction materials management (RFID-CMM) – a case study of water-supply project, Adv. Eng. Inf. 25 (2011) 198–207, https://doi.org/10.1016/j.aei.2010.02.002.
[10] L.-C. Wang, Y.-C. Lin, P.H. Lin, Dynamic mobile RFID-based supply chain control and management system in construction, Adv. Eng. Inf. 21 (2007) 377–390, https://doi.org/10.1016/j.aei.2006.09.003.
[11] N. Pradhananga, J. Teizer, Automatic spatio-temporal analysis of construction site equipment operations using GPS data, Autom. Constr. 29 (2013) 107–122, https://doi.org/10.1016/j.autcon.2012.09.004.
[12] H. Li, Z. Chen, L. Yong, S.C.W. Kong, Application of integrated GPS and GIS technology for reducing construction waste and improving construction efficiency, Autom. Constr. 14 (2005) 323–331, https://doi.org/10.1016/j.autcon.2004.08.007.
[13] A.A. Oloufa, M. Ikeda, H. Oda, Situational awareness of construction equipment using GPS, wireless and web technologies, Autom. Constr. 12 (2003) 737–748, https://doi.org/10.1016/S0926-5805(03)00057-8.
[14] X. Yan, H. Li, A.R. Li, H. Zhang, Wearable IMU-based real-time motion warning system for construction workers' musculoskeletal disorders prevention, Autom. Constr. 74 (2017) 2–11, https://doi.org/10.1016/j.autcon.2016.11.007.
[15] M. Ibrahim, O. Moselhi, Inertial measurement unit based indoor localization for construction applications, Autom. Constr. 71 (2016) 13–20, https://doi.org/10.1016/j.autcon.2016.05.006.
[16] K. Yang, C.R. Ahn, M.C. Vuran, S.S. Aria, Semi-supervised near-miss fall detection for ironworkers with a wearable inertial measurement unit, Autom. Constr. 68 (2016) 194–202, https://doi.org/10.1016/j.autcon.2016.04.007.
[17] A. Shahi, A. Aryan, J.S. West, C.T. Haas, R.C.G. Haas, Deterioration of UWB positioning during construction, Autom. Constr. 24 (2012) 72–80, https://doi.org/10.1016/j.autcon.2012.02.009.
[18] T. Cheng, M. Venugopal, J. Teizer, P.A. Vela, Performance evaluation of ultra wideband technology for construction resource location tracking in harsh environments, Autom. Constr. 20 (2011) 1173–1184, https://doi.org/10.1016/j.autcon.2011.05.001.
[19] Q. Fang, H. Li, X. Luo, L. Ding, T.M. Rose, W. An, Y. Yu, A deep learning-based method for detecting non-certified work on construction sites, Adv. Eng. Inf. 35 (2018) 56–68, https://doi.org/10.1016/j.aei.2018.01.001.
[20] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T.M. Rose, W. An, Detecting non-hardhat-use by a deep learning method from far-field surveillance videos, Autom. Constr. 85 (2018) 1–9, https://doi.org/10.1016/j.autcon.2017.09.018.
[21] S. Chi, C.H. Caldas, Image-based safety assessment: automated spatial safety risk identification of earthmoving and surface mining activities, J. Constr. Eng. Manage. 138 (2012) 341–351, https://doi.org/10.1061/(ASCE)CO.1943-7862.0000438.
[22] W. Fang, L. Ding, H. Luo, P.E.D. Love, Falls from heights: a computer vision-based approach for safety harness detection, Autom. Constr. 91 (2018) 53–61, https://doi.org/10.1016/j.autcon.2018.02.018.
[23] J. Yang, Z. Shi, Z. Wu, Vision-based action recognition of construction workers using dense trajectories, Adv. Eng. Inf. 30 (2016) 327–336, https://doi.org/10.1016/j.aei.2016.04.009.
[24] S. Han, S. Lee, A vision-based motion capture and recognition framework for behavior-based safety management, Autom. Constr. 35 (2013) 131–141, https://doi.org/10.1016/j.autcon.2013.05.001.
[25] L. Ding, W. Fang, H. Luo, P.E.D. Love, B. Zhong, X. Ouyang, A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory, Autom. Constr. 86 (2018) 118–124, https://doi.org/10.1016/j.autcon.2017.11.002.
[26] M. Memarzadeh, M. Golparvar-Fard, J.C. Niebles, Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors, Autom. Constr. 32 (2013) 24–37, https://doi.org/10.1016/j.autcon.2012.12.002.
[27] M. Memarzadeh, A. Heydarian, M. Golparvar-Fard, J.C. Niebles, Real-time and automated recognition and 2D tracking of construction workers and equipment from site video streams, Computing in Civil Engineering (2012), American Society of Civil Engineers, Clearwater Beach, Florida, United States, 2012, pp. 429–436, , https://doi.org/10.1061/9780784412343.0054.
[28] Y.M. Ibrahim, T.C. Lukins, X. Zhang, E. Trucco, A.P. Kaka, Towards automated progress assessment of workpackage components in construction projects using computer vision, Adv. Eng. Inf. 23 (2009) 93–103, https://doi.org/10.1016/j.aei.2008.07.002.
[29] J. Seo, S. Han, S. Lee, H. Kim, Computer vision techniques for construction safety and health monitoring, Adv. Eng. Inf. 29 (2015) 239–251, https://doi.org/10.1016/j.aei.2015.02.001.
[30] R. Socher, A. Karpathy, Q.V. Le, C.D. Manning, A.Y. Ng, Grounded compositional semantics for finding and describing images with sentences, Trans. Assoc. Comput. Ling. 2 (2014) 207–218, https://doi.org/10.1162/tacl_a_00177.
[31] R. Socher, L. Fei-Fei, Connecting modalities: Semi-supervised segmentation and annotation of images using unaligned text corpora, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 966–973, , https://doi.org/10.1109/CVPR.2010.5540112.
[32] Y. Li, W. Ouyang, B. Zhou, K. Wang, X. Wang, Scene graph generation from objects, Phrases and Region Captions, 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, 2017, pp. 1270–1279, , https://doi.org/10.1109/ICCV.2017.142.
[33] D. Xu, Y. Zhu, C.B. Choy, L. Fei-Fei, Scene graph generation by iterative message passing, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, 2017, pp. 3097–3106, , https://doi.org/10.1109/CVPR.2017.330.
[36] J. Johnson, R. Krishna, M. Stark, L.-J. Li, D.A. Shamma, M.S. Bernstein, L. Fei-Fei, Image retrieval using scene graphs, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, 2015, pp. 3668–3678, , https://doi.org/10.1109/CVPR.2015.7298990.
[37] S. Schuster, R. Krishna, A. Chang, L. Fei-Fei, C.D. Manning, Generating semantically precise scene graphs from textual descriptions for improved image retrieval, Proceedings of the Fourth Workshop on Vision and Language, Association for Computational Linguistics, Lisbon, Portugal, 2015, pp. 70–80, , https://doi.org/10.18653/v1/W15-2812.
[38] P. Anderson, B. Fernando, M. Johnson, S. Gould, SPICE: Semantic propositional image caption evaluation, 2016 European Conference on Computer Vision (ECCV), Springer International Publishing, Cham, 2016, pp. 382–398, , https://doi.org/10.1007/978-3-319-46454-1_24.
[39] Y. Liu, R. Wang, S. Shan, X. Chen, Structure inference net: Object detection using scene-level context and instance-level relationships, ArXiv:1807.00119 [Cs], 2018. http://arxiv.org/abs/1807.00119 (accessed March 2, 2019).
[40] M.A. Sadeghi, A. Farhadi, Recognition using visual phrases, 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Colorado Springs, CO, USA, 2011, pp. 1745–1752, , https://doi.org/10.1109/CVPR.2011.5995711.
[41] J. Johnson, A. Gupta, L. Fei-Fei, Image generation from scene graphs, ArXiv:1804.01622 [Cs], 2018. http://arxiv.org/abs/1804.01622 (accessed March 2, 2019).
[42] P. Lu, L. Ji, W. Zhang, N. Duan, M. Zhou, J. Wang, R-VQA: Learning visual relation

facts with semantic attention for visual question answering, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 1880–1889, , https://doi.org/10.1145/3219819.3220036.
[43] C. Lu, R. Krishna, M. Bernstein, L. Fei-Fei, Visual relationship detection with language priors, ArXiv:1608.00187 [Cs], 2016. http://arxiv.org/abs/1608.00187 (accessed March 5, 2019).
[44] 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, F.-F. Li, Visual genome: connecting language and vision using crowdsourced dense image annotations, ArXiv:1602.07332 [Cs], 2016. http://arxiv.org/abs/1602.07332 (accessed March 3, 2019).
[45] H. Zhang, X. Yan, H. Li, Ergonomic posture recognition using 3D view-invariant features from single ordinary camera, Autom. Constr. 94 (2018) 1–10, https://doi.org/10.1016/j.autcon.2018.05.033.
[46] J. Hoffart, F.M. Suchanek, K. Berberich, G. Weikum, YAGO2: a spatially and temporally enhanced knowledge base from Wikipedia, Artif. Intell. 194 (2013) 28–61, https://doi.org/10.1016/j.artint.2012.06.001.
[47] S.B. Hedges, J. Dudley, S. Kumar, TimeTree: a public knowledge-base of divergence times among organisms, Bioinformatics 22 (2006) 2971–2972, https://doi.org/10.1093/bioinformatics/btl505.
[48] B. Dai, Y. Zhang, D. Lin, Detecting visual relationships with deep relational networks, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, 2017, pp. 3298–3308, , https://doi.org/10.1109/CVPR.2017.352.
[49] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, SSD: Single shot multiBox detector, ArXiv:1512.02325 [Cs] 9905 (2016) 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.
[50] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2015. https://arxiv.org/abs/1506.01497v3 (accessed February 28, 2019).
[51] J.D. Lafferty, A. McCallum, F.C.N. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, Proceedings of the Eighteenth International Conference on Machine Learning, Learning, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2001, pp. 282–289 http://dl.acm.org/citation.cfm?id=645530.655813.
[52] S. Zhang, F. Boukamp, J. Teizer, Ontology-based semantic modeling of construction safety knowledge: towards automated safety planning for job hazard analysis (JHA), Autom. Constr. 52 (2015) 29–41, https://doi.org/10.1016/j.autcon.2015.02.005.
[53] Occupational Safety and Health Branch Labour Department, Safety Handbook for Construction Site Workers, March 2004. < https://www.labour.gov.hk/eng/public/os/D/ConstrutionSite.pdf > (last access March 6, 2019).
[54] Health and Safety Executive, The Construction (Design and Management) Regulations 2015, January 2015. < http://www.legislation.gov.uk/uksi/2015/51/pdfs/uksi_20150051_en.pdf > (last access March 6, 2019).
[55] Occupational Safety and Health Administration (OSHA), Recommended Practices for Safety & Health Programs in Construction, October 2016. < https://www.osha.gov/shpguidelines/docs/8524_OSHA_Construction_Guidelines_R4.pdf > (last access March 6, 2019).
[56] Protégé, Protégé wiki. < https://protegewiki.stanford.edu/wiki/Main_Page > (last access March 4, 2019).
[61] W.F. Clocksin, C.S. Mellish, Programming in Prolog, Springer Berlin Heidelberg, Berlin, Heidelberg, 1984 https://doi.org/10.1007/978-3-642-96661-3.
[62] J. Wielemaker, T. Schrijvers, M. Triska, T. Lager, SWI-Prolog, Theory Pract. Logic Program. 12 (2012) 67–96, https://doi.org/10.1017/S1471068411000494.
[63] S. Hauberg, O. Freifeld, A.B.L. Larsen, J.W. Fisher III, L.K. Hansen, Dreaming more data: Class-dependent distributions over diffeomorphisms for learned data augmentation, ArXiv:1510.02795 [Cs], 2015. http://arxiv.org/abs/1510.02795 (accessed March 10, 2019).
[64] M. Everingham, S.M.A. Eslami, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The pascal visual object classes challenge: a retrospective, Int. J. Comput. Vision 111 (2015) 98–136, https://doi.org/10.1007/s11263-014-0733-5.
[65] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft COCO: common objects in context, 2004 European Conference on Computer Vision (ECCV), Springer International Publishing, Cham, 2014, pp. 740–755, , https://doi.org/10.1007/978-3-319-10602-1_48.
[66] K. Soomro, A.R. Zamir, M. Shah, UCF101: A dataset of 101 human actions classes from videos in the wild, ArXiv:1212.0402 [Cs], 2012. http://arxiv.org/abs/1212.0402 (accessed March 5, 2019).
[67] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, L. Fei-Fei, Large-scale video classification with convolutional neural networks, 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Columbus, OH, USA, 2014, pp. 1725–1732, , https://doi.org/10.1109/CVPR.2014.223.
[68] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, Y. Sheikh, OpenPose: realtime multi-person 2D pose estimation using part affinity fields, ArXiv:1812.08008 [Cs], 2018. http://arxiv.org/abs/1812.08008 (accessed February 27, 2019).
[69] J. Zhang, N.M. El-Gohary, Automated information transformation for automated regulatory compliance checking in construction, J. Comput. Civil Eng. 29 (2015) B4015001, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000427.
[70] A. Hogenboom, F. Hogenboom, F. Frasincar, K. Schouten, O. van der Meer, Semantics-based information extraction for detecting economic events, Multim. Tools Applic. 64 (2013) 27–52, https://doi.org/10.1007/s11042-012-1122-0.
[71] E. Soysal, I. Cicekli, N. Baykal, Design and evaluation of an ontology based information extraction system for radiological reports, Comput. Biol. Med. 40 (2010) 900–911, https://doi.org/10.1016/j.compbiomed.2010.10.002.
[72] J. Zhang, N.M. El-Gohary, Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking, J. Comput. Civ. Eng. 30 (2016) 04015014, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000346.