Journal of Building Engineering 42 (2021) 103036

Contents lists available at ScienceDirect

Journal of Building Engineering

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

Combining computer vision with semantic reasoning for on-site safety management in construction

Haitao Wu[a], Botao Zhong[b,*], Heng Li[a], Peter Love[c], Xing Pan[b], Neng Zhao[b]

[a] Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong
[b] School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, 430074, Hubei, China
[c] School of Civil and Mechanical Engineering, Curtin University, GPO Box U1987, Perth, Western Australia, Australia

ARTICLE INFO

Keywords:
Computer vision
Ontology
Semantic reasoning
Hazard identification
Construction safety management

ABSTRACT

Computer vision has been utilized to extract safety-related information from images with the advancement of video monitoring systems and deep learning algorithms. However, construction safety management is a knowledge-intensive task; for instance, safety managers rely on safety regulations and their prior knowledge during a jobsite safety inspection. This paper presents a conceptual framework that combines computer vision and ontology techniques to facilitate the management of safety by semantically reasoning hazards and corresponding mitigations. Specifically, computer vision is used to detect visual information from on-site photos while the safety regulatory knowledge is formally represented by ontology and semantic web rule language (SWRL) rules. Hazards and corresponding mitigations can be inferred by comparing extracted visual information from construction images with pre-defined SWRL rules. Finally, the example of falls from height is selected to validate the theoretical and technical feasibility of the developed conceptual framework. Results show that the proposed framework operates similar to the thinking model of safety managers and can facilitate on-site hazard identification and prevention by semantically reasoning hazards from images and listing corresponding mitigations.

1. Introduction

People working in construction are globally recognized to be highly prone to experiencing injuries, accidents, and even fatalities [1–3]. The continued introduction and modification of statutory regulations and the implementation of management practices to redress safety, accidents, and fatalities remain a pervasive problem despite the considerable amount of research that has been undertaken. For instance, the Bureau of Labor Statistics (BLS) in the United States reported that more than 970 construction fatalities were reported in 2017 [4]. Approximately 35 serious claims are reported each day in the construction industry of Australia [5]. A total of 734 accidents were recorded in China in 2018, thereby leading to 840 death [2].

Computer vision has recently attracted increasing attention because of its potentials to overcome the drawbacks of manual observation of on-site hazards. Several computer vision approaches were proposed to detect unsafety behaviors or dangerous working conditions by extracting visual information automatically and continuously from on-site images or videos [6–11]. The development of deep learning algorithms further enhances the capability of computer vision on processing and analyzing visual imagery [12]; however, as a kind of end-to-end learning, computer vision approaches are limited in some areas requiring knowledge reasoning [13]. However, construction safety management is a type of knowledge-intensive job [14]. As shown in Fig. 1 (a), site managers and engineers must extract visual information using their perceptual capability when performing a jobsite safety inspection; this information is then reasoned to identify potential hazards and corresponding mitigation measurements based on existing safety rules in regulations and their experiences [8].

Computer vision can automate visual tasks as human visual systems (detecting the *two workers*, *concrete supports, ropes* form Fig. 1(b)). However, inferring that workers are prone to experience *falling from height* and *tripping* hazards based on safety regulations is difficult for computer vision. Additionally, compared with safety inspectors, computer vision cannot give suggestions to mitigate potential hazards. This condition can be summarized as a "semantic gap," which indicates the limitation of computer vision systems in extracting the visual data of what the cameras "see" and the substantial enhancement of the meaning of what they observe by considering the domain knowledge [15–17].

Ontology, a semantic technique, is used to provide safety domain

* Corresponding author.
E-mail address: dadizhong@hust.edu.cn (B. Zhong).

https://doi.org/10.1016/j.jobe.2021.103036
Received 29 October 2020; Received in revised form 10 March 2021; Accepted 26 July 2021
Available online 29 July 2021

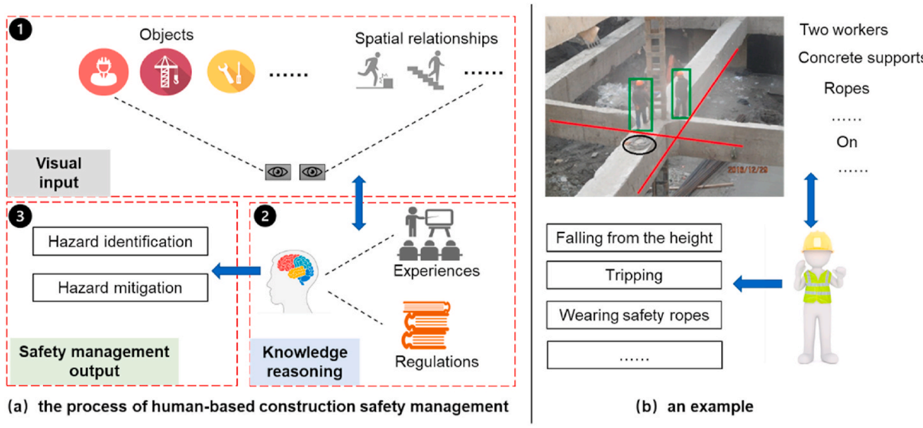

Fig. 1. Process of manual construction safety management and an example.

knowledge, which includes explicit and rich semantics, to support efficient knowledge management and reasoning on safety issues [14]. An ontology can provide a formal conceptualization of knowledge for a given domain [18]. The ontology technique provides a method for converting textual regulation knowledge of a specific domain into a specific and understandable format. Ontology has been widely used in construction safety management due to its benefits on knowledge representation and reasoning [19–22]. These studies show the potential of ontology technology on knowledge-based safety management, which may solve the "semantic gap" in current computer vision approaches [12,23]. [24] and Fang et al. [25] revealed evidence showing the necessity of knowledge and computer vision in construction safety management. Nevertheless, the combination of ontology and computer vision is still scarce in the construction industry. Additionally, previous studies lack detailed discussions on ontology development, which is important for construction safety management regarded as a knowledge-intensive task [14].

A framework that combines computer vision algorithms and a formal ontology model for construction safety management is developed in this paper. Computer vision algorithms are used in the proposed framework to capture visual information (e.g., objects and their spatial relationships) from onsite images, and an ontology model is established to represent the domain knowledge formally according to the five steps in Noy and McGuinness [26]. The visual information acquired by the computer vision would be transferred into instances in the developed ontology model for hazard inference. Specifically, potential hazards and their mitigation measurements would be inferred in a knowledge reasoner based on pre-defined Semantic Web Rule Language (SWRL) rules in the ontology model. Notably, this paper aims to illustrate the importance of ontology in helping computer vision function similarly to the human thinking model in construction safety management. Particularly, this paper focuses on hazards associated with falls from height (FFH) because they are the most common safety issues that arise during construction [2,3,27,28]. The developed framework is validated using data derived from construction projects, which form part of the Wuhan Rail Transit System (China).

The rest of this paper proceeds as follows. Section 2 reviews related works and summarizes the research gap. Section 3 comprehensively describes the entire framework integrating computer vision and ontology. Section 4 conducts a case study of FFH, while section 5 discusses the finding and limitations. Section 6 concludes with implications for construction safety management.

2. Related works

2.1. Computer vision for construction safety management

As previously mentioned, computer vision has been receiving widespread attention in construction and has been specifically used to examine hazards in construction, such as progress monitoring [29], productivity analysis [30], and health and safety monitoring [31–34]. Performing safety inspections is traditionally a time-consuming and labor-intensive process, requiring site managers and engineers to walk around the job site to determine hazards. Computer vision is regarded as a robust method to identify hazards from images automatically due to the inefficient and ineffective process of manual safety inspection [10,11,35–37]; d; [33].

Numerous algorithms were proposed for computer vision-based construction safety management, which can be classified as follows [37,38]: (1) shallow learning methods, such as support vector machine and histogram of oriented gradients; and (2) deep learning methods, such as the convolutional neural network (CNN) and recurrent neural network. Shallow learning methods are reliant on handcrafted features, which must be manually produced, thus adversely affecting the accuracy of their detection [39]. Compared with shallow learning approaches, deep learning can automatically extract sophisticated features from data with multiple levels of end-to-end representations [40,41]. In particular, CNN is the most commonly used to support computer vision due to its capability to detect objects accurately and reliably [42]. Various algorithms, such as the Single Shot Multibox Detector (SSD) [43], You Only Look Once [44], faster region-based CNN (Faster R–CNN) [44], and mask region-based CNN (Mask R–CNN) [45], have been developed to support the CNNs for object detection. Deep learning algorithms have considerably promoted computer vision-based applications, such as unsafe behavior detection [33,36,46], activity recognition [47,48], and object detection [11,49]. For example, Fang et al. [11] used faster R–CNN to detect non-hardhat-use behavior from surveillance videos. Ding et al. [36] demonstrated that a hybrid deep learning model can accurately identify unsafe actions during construction by integrating CNN and long short-term memory. Similarly, Fang et al. [33] created a Mask R–CNN algorithm to identify the unsafe behavior of people traversing structural supports. Luo et al. [48] proposed a discriminative model combining deep activity features and contextual information to recognize the activities of on-foot workers from on-site videos.

Semantic knowledge is absent in computer vision-based construction safety management despite the ability of site managers and engineers to detect objects automatically using deep learning due to previous research [50]. Neural network models enable the computer vision system to achieve advanced performance on AI tasks [41], such as detecting objects from images or videos, due to their good representation and

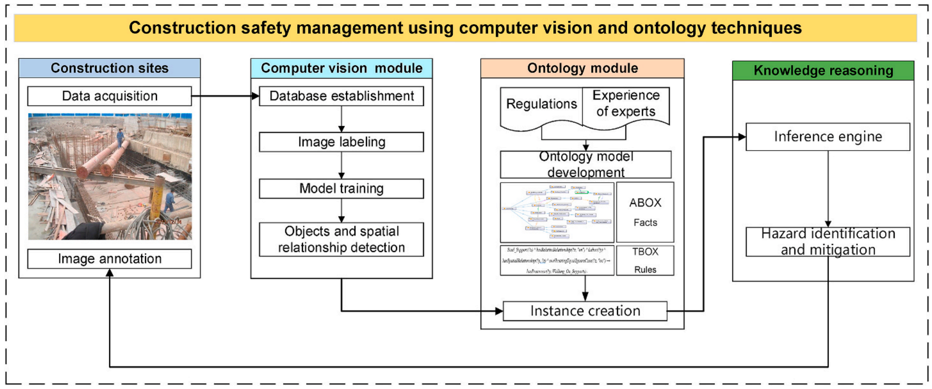

Fig. 2. Framework of the proposed method.

learning capabilities. However, construction safety management is a knowledge-intensive task [14]. The computer vision system cannot understand the hazard meaning behind the visual information without the semantic representation of safety knowledge. For instance, safety managers can use their prior knowledge to identify potential hazards and provide some mitigation measurements to prevent hazards in the workplace. Zhong et al. [51] also highlighted the semantic gap between high-level knowledge and low-level visual features obtained by computer vision systems. The upshot is an inability to judge hazards in complex scenes and accordingly set up mitigation strategies. Hence, safety domain knowledge is necessary in the design of computer vision-based construction safety management.

2.2. Ontology for representing construction safety knowledge

Compared with a database schema, an ontology can represent knowledge with explicit and rich semantics, which can enable knowledge reasoning [18]. The concept of ontology within construction has been widely applied, and the benefits of knowledge representation and reasoning have been adopted in areas, such as risk and knowledge management [52]. For instance, Tserng et al. [53] developed an ontology-based framework to enhance safety management performance throughout the lifecycle of a project. In stark contrast, Wang and Boukamp [54] used ontology to represent the knowledge ingrained in activities, job steps, and hazards to improve the ability to query and share job hazard analysis (JHA) knowledge. This approach enabled Wang and Boukamp [54] to develop an ontological mechanism for reasoning safety rules for activities.

With the ontology, domain knowledge can be converted into a machine-processable format, which can support the development of intelligent applications for hazard identification and safety management. For example, Lu et al. [20] established a construction safety checking ontology, in which safety checking constraints were encoded into semantic rules for safety checking based on semantic rule languages. Zhang et al. [21] developed a JHA ontology model to store and re-use construction safety-related knowledge. Xing et al. [14] proposed a domain ontology (SRI-Onto) to formalize safety knowledge considering that hazard identification of metro construction is a knowledge-intensive process. Wu et al. [22] used ontology to represent metro accidents to facilitate efficient knowledge retrieval and reasoning. Overall, these studies indicated that ontology provides a way to represent safety knowledge explicitly and formally to further develop information system processes, such as knowledge retrieval and inference.

2.3. Summary

Studies combining ontology and computer vision techniques to facilitate safety management in construction are limited. Tang and Golparvar-Fard [50] developed a design of a language–image framework to understand and detect semantic roles of activities mentioned in safety rules. Xiong et al. [24] initially attempted to apply safety knowledge to computer vision systems and created an automated hazard identification system for hazard identification based on the aforementioned work. Safety guidelines from textual regulations were represented in this system by scene graphs in a three-tuple format to evaluate the operation descriptions generated from site videos. However, this approach lacks discussions on ontology-based knowledge modeling, and the developed ontology model is simple. The literature review of computer vision by Zhong et al. [12] and Fang et al. [23] indicated the importance of ontology to computer vision. For instance, Fang et al. [23] suggested that the computer-oriented and logic-based features embedded within ontology can provide access to regulation knowledge for computer vision systems. Meanwhile, Zhong et al. [51] developed an ontology model to describe contents from images for following semantically retrieval; however, the hazard in images is manually identified, which is a time-consuming task. Similarly, Fang et al. [25] used the Neo4j database to visualize and store the visual information detected by the computer vision algorithm for hazard identification. The knowledge graph structure of Neo4j can support semantic query; however, this structure lacks a reasoning capability, which is important for knowledge-intensive safety management [14].

Overall, supplementing ontological safety knowledge can give meaning to these visual elements obtained by computer vision; for instance, reasoning hazards and listing mitigation suggestions, which are performed by safety managers. The ontology model should be properly developed to ensure its capability to represent knowledge in the domain of construction safety management formally. Additionally, the visual information acquired by computer vision algorithms should be linked with the instances in the ontology model. Developing an explicit, extendable, and accessible knowledge base for computer vision for construction safety management remains unclear despite insightful findings from previous works.

3. Research approach

A conceptual framework, which can combine ontology-based regulation knowledge with visual information extracted from images by computer vision algorithms, is developed in this paper based on our previous efforts [33]. The proposed framework comprises three modules (Fig. 2).

1. *Computer vision module*: Detects visual information, including objects and their spatial relationships from images collected from construction sites, which are then converted into instances in the ontology models for knowledge reasoning.
2. *Ontology module*: Formalizes the domain knowledge of construction safety management to provide extracted visual information with semantic concepts, which contain definitions of taxonomies and

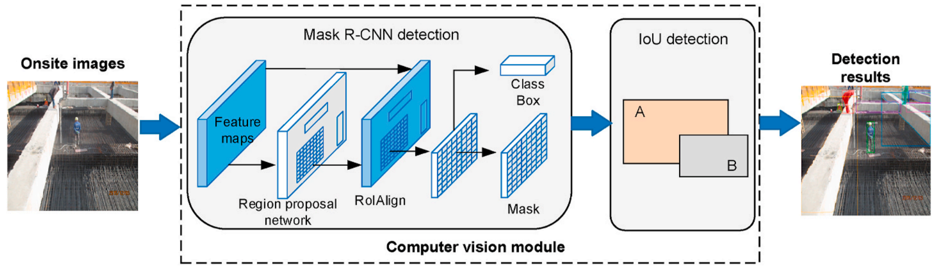

Fig. 3. Mask R–CNN based visual information detection.

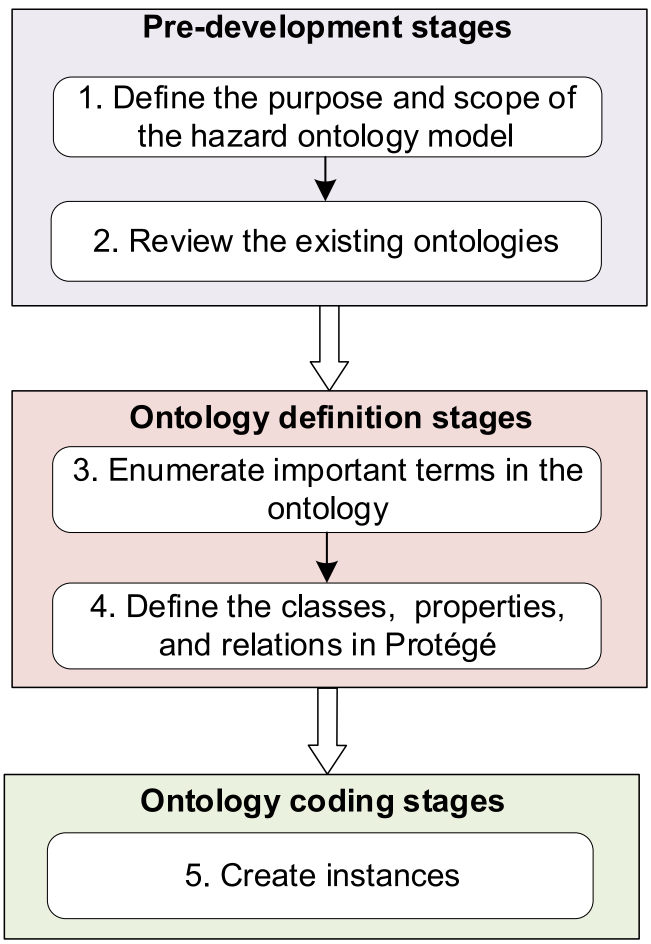

Fig. 4. Development method of the ontology model.

properties. Inference rules representing regulatory knowledge are needed for complex hazard identification because construction safety management is a knowledge-intensive process [14,52]. The SWRL rules for knowledge inference are encoded in the reasoning engine.

3. *Knowledge reasoning module*: Potential hazards from images and mitigation measures are inferred by performing rule checking on the visual inputs detected by the computer vision module. Additionally, the inferred information would be attached to the images for further efficient image retrieval or processing.

3.1. Computer vision-based visual information identification

The computer vision module, acting as "human eyes" in the integrated framework, aims to identify visual information from on-site images, which is regarded as inputs for ontology-based hazard inference. From the perspective of hazard identification, the following visual information should be extracted from on-site images: objects (e.g., people, equipment, materials, and environment elements) and their spatial relationships.

Various deep learning algorithms, such as SSD [43], faster R–CNN [44], and Mask R–CNN [45], have been developed for a variety of different purposes in construction. Among these algorithms, the Mask R–CNN can detect multiple objects with higher levels of accuracy than other deep-learning algorithms because it extends Faster R–CNN by adding a branch for predicting segmentation masks on each Region of Interest (RoI) [45]. The Mask R–CNN has good performance on instance segmentation and can recognize overlapped objects, which has been evaluated in our previous study [33]. Therefore, Mask R–CNN is selected to detect objects from images.

Fig. 3 shows the entire workflow of Mask R–CNN–based visual information detection. The network architecture of Mask R–CNN, which contains two steps [45]: (1) to extract feature maps of an entire image from the residual and the feature pyramid networks as inputs. A region proposal network is then used to define regions of interest (ROIs); (2) a RoIAlign layer is adopted to extract spatial locations from each candidate box and perform classification, bounding box regression, and mask generation. Parameters used in this research are similar to those of He et al. [45]. The RoIAlign, which uses bilinear interpolation, can generate exact spatial locations to compute the exact values of the input features and avoid any quantization of its boundaries or bins.

In addition to objects detection, spatial relationships between objects are needed for hazard reasoning. In this paper, three types of spatial relationships are considered by calculating Intersection over Union (IoU) with Eq. (1): (1) on; (2) overlap; and (3) away. Specifically, the overlapping areas between objects are used to determine the spatial relationships. For example, when the value of IoU (A, B) belongs to (0, 1), the spatial relationship of object A and object B is overlapped as shown in Fig. 3. Notably, the authors of this paper used the previous study [33] in which the Mask R–CNN method is used to detect workers traveling the structural supports. Therefore, only three spatial relationships are taken into considerations. We admit that there would be optimal algorithms for detecting visual information from images and more spatial relationships should be considered for reasoning different types of hazards on construction sites. However, rather than presenting a novel method to identify all objects and spatial relationships, this paper aims to propose a conceptual framework, which introduces a formal ontology for computer vision systems, to close the semantic gap between low-level visual features and high-level knowledge needed in construction safety management.

$$\mathrm{IoU}(\mathrm{A},\mathrm{B}) = \frac{area\ (A) \cap area\ (B)}{\min\{area(A),\ area(B)\}} = \begin{cases} 1, & on \\ (0,\ 1), & overlap \\ 0, & away \end{cases} \tag{1}$$

3.2. Ontology model for safety management

Safety knowledge and the experience of experts must be presented in an understandable format by a computer. Thus, the developed ontology

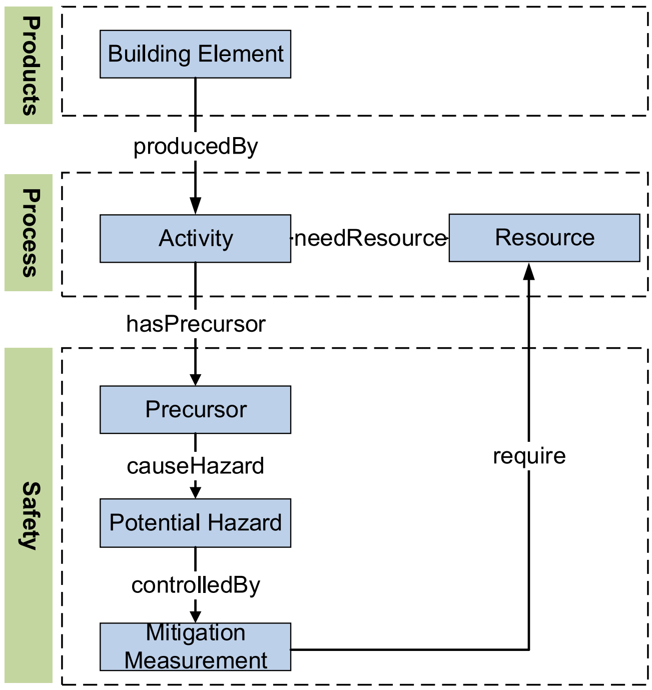

Fig. 5. Top-level classes in the meta-ontology model.

model enables the computable representation of safety regulation knowledge.

3.2.1. Ontology development for safety Hazard identification and mitigation

Ontology is used to standardize the description of safety knowledge and facilitate the reasoning of hazards. Safety knowledge is modeled and represented using the following five steps (Fig. 4) [26].

1. Define the purpose and scope of the hazard ontology model.
2. Review existing ontologies.
3. Enumerate important terms.
4. Define classes, properties, and relations.
5. Create instances based on the output of the computer vision module.

Guided by the aforementioned steps, a meta-ontology model can be developed to standardize the description of construction safety knowledge.

After reviewing existing ontologies [14,21,28], the JHA ontology introduced by Zhang et al. [21] is used to develop the meta-ontology model, which comprises the following: (1) products, (2) process, and (2) safety. Six top-level classes are used to represent safety knowledge on construction sites. These classes include *Building Elements*, *Activity*, *Resource*, *Precursor*, *Potential Hazard, and Mitigation Measurement.* The meta-ontology and relations between these classes are presented in Fig. 5.

3.2.2. Ontology definition of safety Hazard knowledge

Engineering regulations and manuals in construction form a source of domain knowledge, which can be utilized for hazard identification and safety management (Fang et al., 2020). Therefore, the class taxonomies were identified based on the selected knowledge sources, including "Unified Regulation for Construction Quality Acceptance of Construction Engineering (gb50300-2017)," "Building Engineering Measurement Regulations," "Industry Foundation Classes (IFC) standards," "Occupational Injury and Illness Classification Manual," and "Building Construction Safety Inspection Regulation." A part of the taxonomy of the developed ontology is presented as shown in Fig. 6.

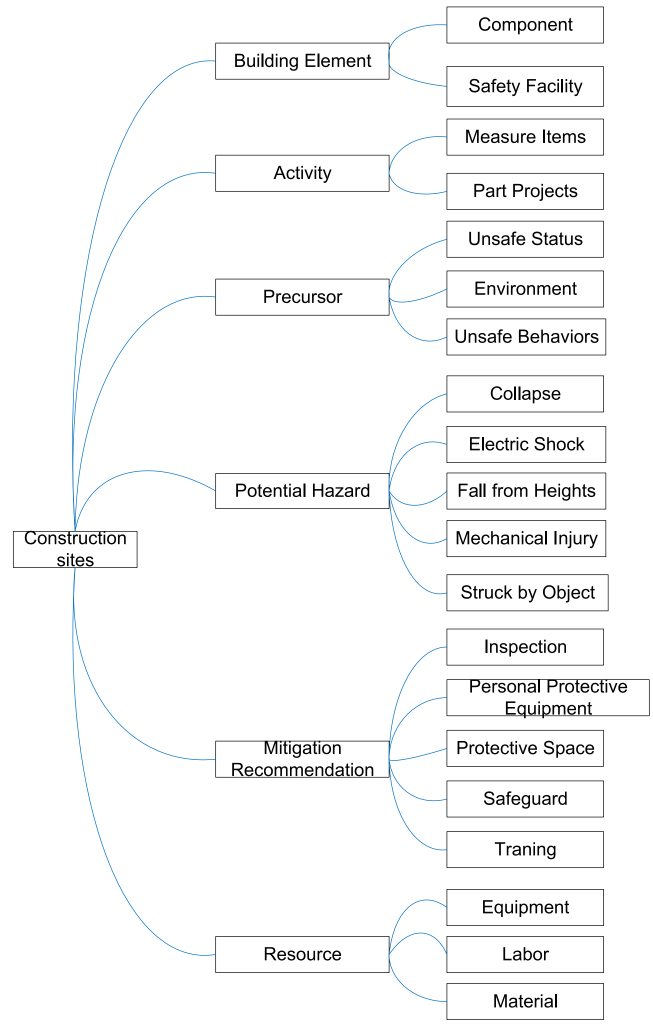

Fig. 6. Part of the taxonomy of construction safety knowledge.

Notably, no perfect ontology and optimum concept hierarchy are available [55]. Thus, the proposed ontology does not completely cover a domain of interest. Consequently, the taxonomy of ontologies can be developed for different applications.

The Products Model, which accommodates building elements, relates to its components (e.g., beam, column, door, and wall that are aligned with the IFC schema) and safety (e.g., temporary facilities and support structures). The Process Model contains activities and the necessary building resources. The taxonomy of activity is based on the regulation of the "Unified Regulation for Construction Quality Acceptance of Construction Engineering (gb50300-2017)" and the "Building Engineering Measurement Regulations." The resource is classified into labor, equipment, and material, which provide the connection between semantic concepts and objects extracted from images.

The Safety Model includes the potential hazards on construction sites and the corresponding precursors and mitigation measurements. The taxonomy of this model is based on safety regulations, such as the "Occupational Injury and Illness Classification Manual" and the "Building Construction Safety Inspection Regulation." For example, the regulation of the "Quality and Safety Inspection Guide of Urban Rail Transit Engineering" can be used as a reference to examine hazardous events of the construction procedures in China. The potential hazards relating to tasks performed during construction and techniques employed can be classified into 21 inspection categories, such as safety management, fastener-type scaffolding, full-style scaffolding, foundation pit support engineering, template engineering, and hoisting.

Table 1
Part of the object properties.

Object Property	Domain	Range	Characteristic
producedBy	Building element	Activity	Inverse functional
Produce	Activity	Building element	Functional
needResource	Activity	Resource	Inverse functional
needLabor	Activity	Labor	Inverse functional
beUsedIn	Resource	Activity	Functional
hasPrecursor	Activity	Precursor	Inverse functional
causeHazard	Precursor	Potential hazard	Inverse functional
controlledBy	Potential hazard	Mitigation Recommendation	Inverse functional
hasUnsafeBehavior	Labor	Unsafe behavior	Functional
hasUnsafeStatus	Element	Unsafe status	Functional
isPartOf	–	–	Functional

Table 2
Part of the data properties.

Object Property	Domain	Value	Characteristic
hasE_ID	Building element	int	Functional
hasE_Location	Building element	string	Functional
hasDimension	Building element	decimal	Functional
hasRelativeRelationship	Building element	string	Functional
hasFrequency	Activity	int	Functional
hasTime	Activity	dateTime	Functional
isOccupationOf	Labor	string	Functional
hasSpatialRelationship	Labor	string	Functional
hasBodilyReaction	Labor	string	Functional
hasL_ID	Labor	int	Functional
hasSeverity	Potential hazard	string	Functional
hasLikelihood	Potential hazard	string	Functional
hasMaxArrestForce	Safeguard	decimal	Functional
hasS_ID	Safeguard	int	Functional

Table 3
Mapping relations between XML and OWL.

XML Schema	OWL Schema
Complex Type	Class
Simple Datatype	Datatype Declaration
Element	Datatype Property; Object Property
Attribute	Datatype Property
Sequence	Unnamed Class-Intersection
Choice	Unnamed Class-Union
Annotation	Comment

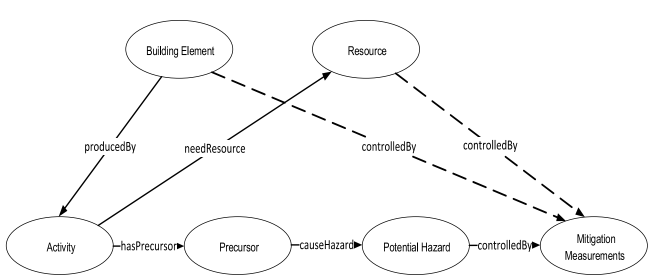

Fig. 7. Property rules between top-level classes.

Each precursor is associated with multiple potential hazards through the "cause hazards" property. Each potential hazard is controlled by some mitigation recommendation to eliminate or reduce the hazard.

The class taxonomies cannot provide sufficient information to describe the overall safety hazard. The properties refer to the characteristics that describe each of the classes [55]. Two types of properties are identified in this paper: (1) object property reflecting relations between two classes and (2) data property reflecting relations between classes and values. For example, some essential object and data properties between the concepts defined in the meta-ontology respectively presented in Tables 1 and 2 are listed.

3.3. Knowledge reasoning for Hazard identification and mitigation

Hazards and the corresponding mitigation measures are inferred by comparing the visual information detected by the computer vision module with the safety rules defined in the ontology model. Herein, the visual information must be mapped with the defined instance in the developed ontology model. Additionally, safety knowledge should be encoded into rules based on the relevant regulations.

3.3.1. Information mapping from visual information to ontology instances

Visual information obtained from images, including objects and their spatial relationships, is identified and extracted as the input for creating instances in the ontology model for knowledge reasoning. The identified objects in the computer vision-based module and their relationships are outputted in the eXtensible Markup Language (XML), while the safety knowledge in the ontology model is described in the Ontology Web Language (OWL). OWL is a W3C recommended language for ontology representation on the semantic web, which enables the users to describe the information in separate scopes and different schemas [56]. The XS2OWL is used as the conversion method between XML and OWL. The mapping relations of XS2OWL are listed in Table 3. The classification label in XML is converted into "owl:class" in OWL using the XS2OWL. The spatial relationships are transformed into "owl: dataProperty" in OWL. Therefore, the visual information detected by a specific computer vision algorithm can be stored in the ontology model for safety hazard knowledge reasoning.

3.3.2. Encoding safety knowledge into rules

Safety hazard identification is a knowledge-intensive process that adheres to rules [52], which can be formatted into a specific representation language, such as N3Logic, Rule Interchange Format, and SWRL. The SWRL is used to represent rules for safety knowledge reasoning because this rule language is tightly integrated with ontology [57]. Moreover, the SWRL offers powerful deductive reasoning abilities by allowing users to write Horn-like rules based on OWL concepts. The context-driven hazard identification can be implemented because the SWRL rule is expressed in terms of ontology concepts (classes, properties, and instances). Each SWRL rule contains an antecedent and consequent part. Both parts comprise positive conjunctions of atoms; for example, the typical SWRL rule is shown as follows:

$$C(x) \wedge P(x,y) \wedge SameAs(y,z) \rightarrow M(x,z),$$

where C(x) is an atom, "*C*" and "*M*" are classes in OWL ontology, "*x*" is the instance of class "*C*," "*P*" represents data or object properties, "*y*" can be a value or instance, and "*SameAs*" is used to describe equal relationships. The following SWRL rule can be written based on these atoms to define the constraint for representing construction safety knowledge. The rule is triggered to execute knowledge inference once the antecedents of a rule are satisfied. The new facts are deduced and stored after executing the rule.

Specific safety regulations and the experience of experts are coded using the SWRL rule formats, which are compatible with ontology classes and relationships. For example, Fig. 7 shows some hidden property rules between the top-level classes in the meta-ontology model (i.e., "Building Elements," "Activity," "Resource," "Precursor," "Potential Hazard," and "Mitigation Measurement"), which can be represented in the SWRL rules as follows:

SWRL Rule1:

Fig. 8. Structural support for deep foundation pits.

Fig. 9. Examples of visual information detection in computer vision module.

Building_Element(?x) ˆ Activity(?y) ˆ Resource(?z) ˆ Precursor(?a) ˆ Potential_Hazard(?b) ˆ Mitigation_Recommendation(?c) ˆ producedBy(?x, ?y) ˆ needResource(?y, ?z) ˆ hasPrecursor(?y, ?a) ˆ causeHazard(?a, ?b) ˆ controlledBy(?b, ?c) → controlledBy(?x, ?c).

If *X* belongs to the "Building Elements," *X* is generated by "Activity" *Y*, *Y* requires "Resource" *Z*, *Y* has "Resource" *A*, *A* may generate "Potential Hazard" *B*, and *B* is controlled by "Mitigation Measurement" *C*, then *X* should also be controlled by preventive measures *C*.

Overall, the visual information extracted from the construction site using the computer vision module can be converted into the instances in the ontology model. The SWRL rules can then be executed in an inference engine to recognize hazards and justify the mitigation measures.

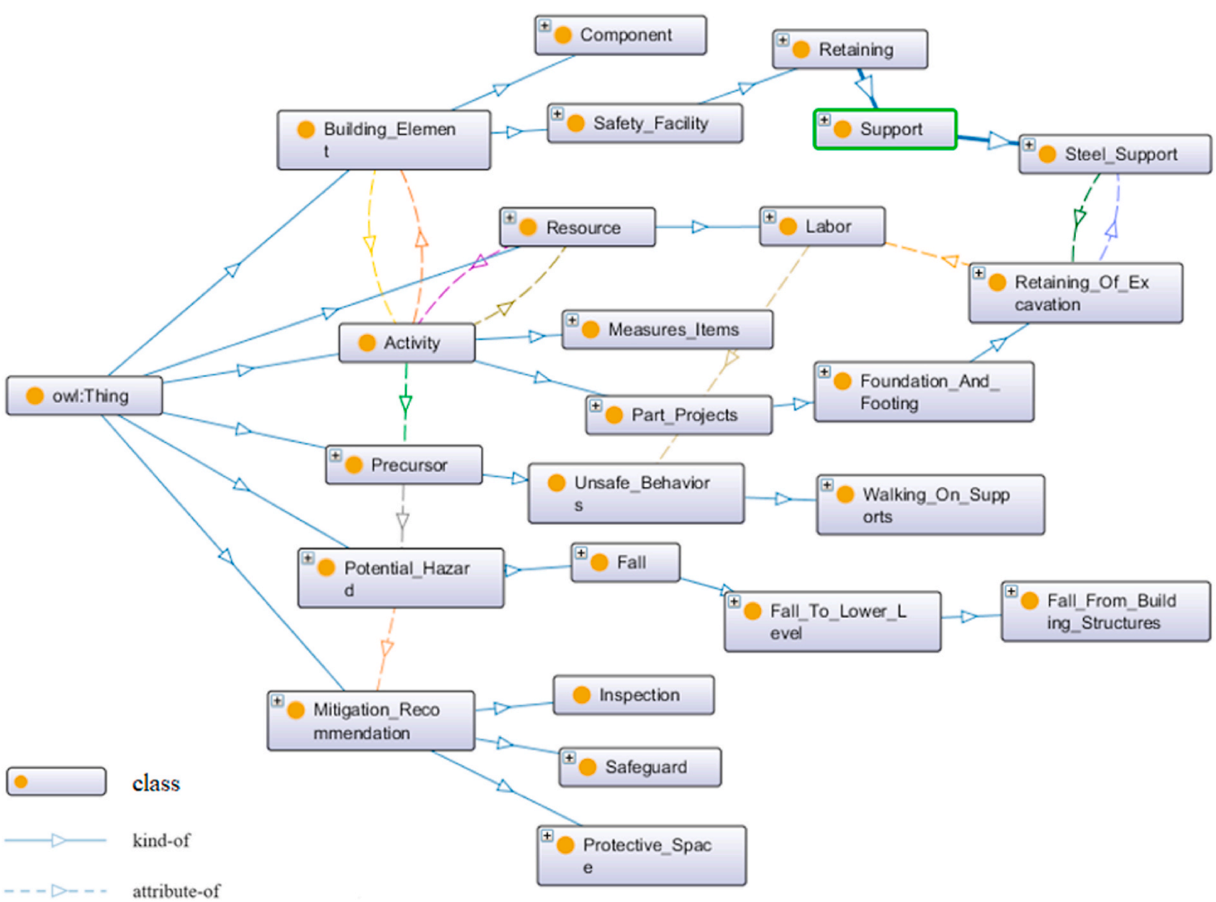

Fig. 10. Developed prototype ontology in Protégé.

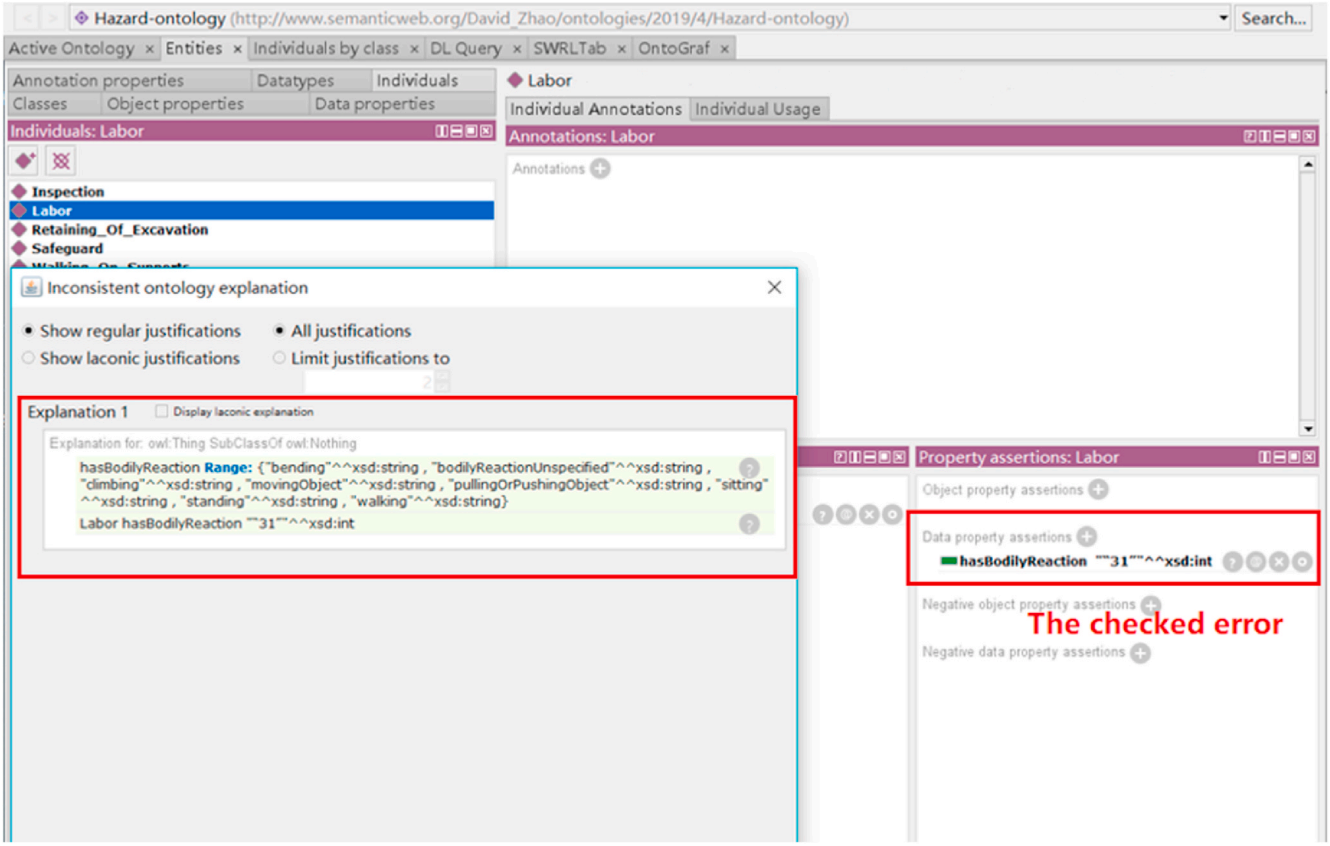

Fig. 11. Ontology checking consistent error in Pellet.

4. Case study

Deep foundation pits are the major activity that must be performed for the underground construction of a metro-system. The construction of the pits invariably requires the deployment of concrete and steel supports to stabilize soil and transfer loads. People in China tend to traverse the structural supports over deep foundation pits (Fig. 8). Therefore, site management and engineers must regularly police this unsafe behavior, particularly because this behavior is prone to not wearing a safety harness even when required to work at heights [58]. Therefore, FFH hazards are selected to test the validity of the proposed framework.

4.1. Site visual information detection

In line with previous research [33], Mask R–CNN is used to detect various entities and their spatial relationships from images between people traversing structural supports over deep foundation pits.

4.1.1. Dataset establishment and model training

Working in close collaboration with the main contractor of the Wuhan Metro Project, monocular cameras were deployed across several construction sites. A database containing 2018 images of individuals walking and not walking on structural supports over deep foundation-pits was also constructed. The database is randomly divided into two parts: (1) training and (2) testing. The images for training and testing respectively contain objects of individuals and structural supports. A total of 1461 and 450 images are respectively used for training and validating. The other 107 images containing individuals walking on concrete/steel supports are used to test the trained Mask R–CNN model.

Training images would be initially annotated manually based on the Labeling Tool example Flask App developed by Python. Furthermore, Microsoft's Common Objects in Context (MS COCO) database containing more than 330 k images is adopted to train the Mask R–CNN to avoid bias. The pre-trained model would be fine-tuned by extracting features from annotated 1461 training images, and then a set of 450 testing images is used to test the Mask R–CNN model.

4.1.2. Results performance of visual information detection

The Mask R–CNN modular is conducted on a server equipped with a 2.5 GHz Intel® Xeon® E5-2680 CPU, an NVIDIA(R) Tesla (TM) K80GPU, and 64 RAM. The learning rate is 0.001, which is decreased by 10 after all layers are fined tuned. The weight decay regularization is 0.0001 and the momentum is 0.9. The RPN network adopts five scales (32×32, 64×64, 128×128, 256×256, and 512×512 pixels) and three aspect ratios (1:1, 1:2, and 2:1) while the stride of anchors is 1. The precision rates on detecting workers, steel supports, and cement supports are 100%, 100%, and 99%, while the recall rates are 84%, 74%, and 81%, respectively. The precision and recall rates on spatial relationship detection are 75% and 90%, respectively. For example, Fig. 8 shows that objects (i.e., person and the bounding box) and their spatial relationships are detected using the Mask R–CNN algorithm. The left side of Fig. 9 indicates the mask for semantic segmentation. The different colored bounding boxes in Fig. 9 indicate the location of targets. The upper left corner of the bounding box indicates the target category label (such as the concrete support) and confidence (0.999). The three types of information form the input into the overlap detection for attribute extraction, and the spatial relationship between the "*labor*" and the "*concrete support*" is "*on.*" The extracted visual information is converted into the OWL format to create instances in the ontology model for safety knowledge reasoning. Numerous advanced algorithms extracting visual information are available. However, the experiment aims to validate the proposed conceptual framework rather than examine the performance of Mask R–CNN with other deep learning algorithms. Built on our previous study of Fang et al. [33] who used the Mask R–CNN approach to

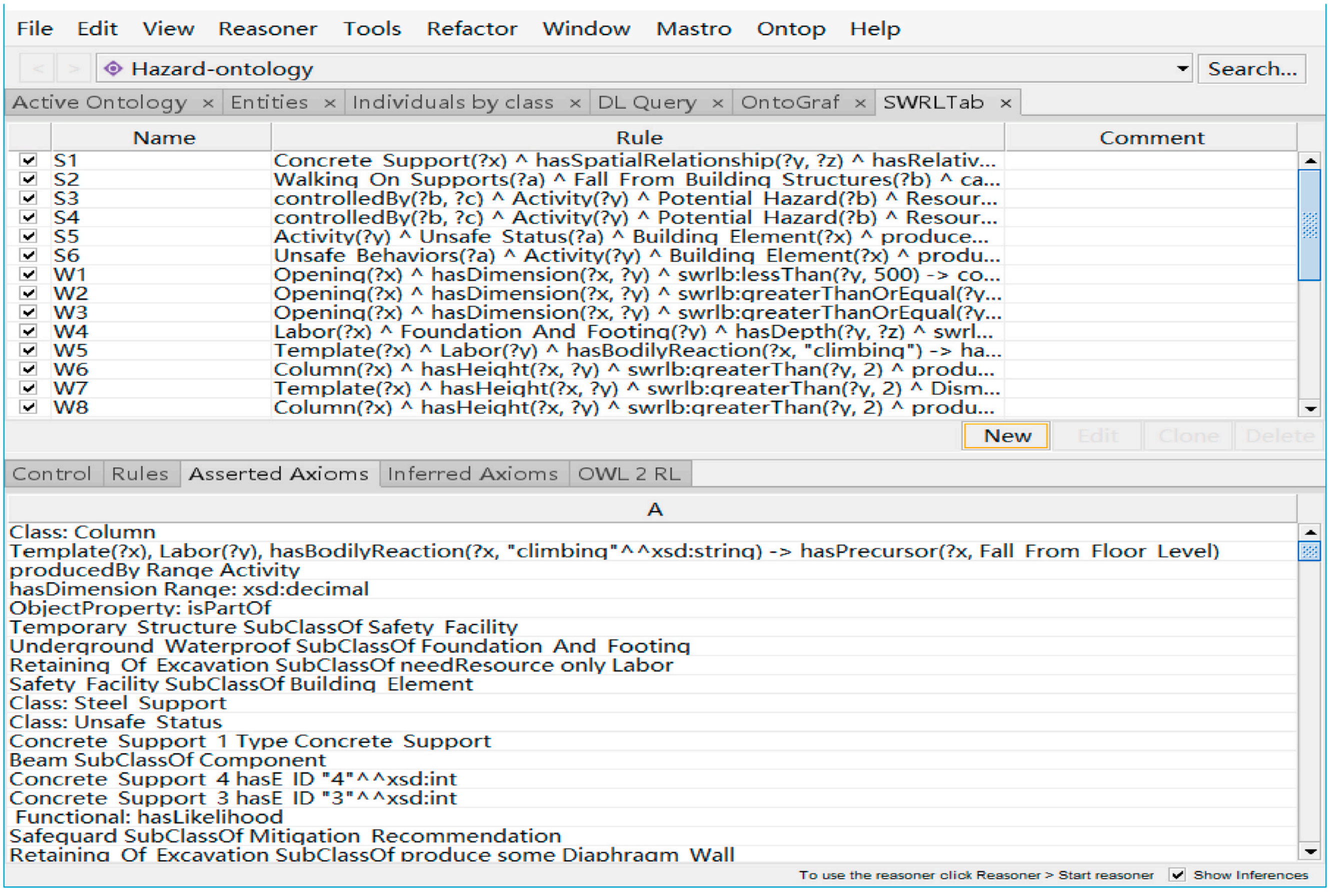

Fig. 12. Developed SWRL rules in the Protégé.

detect workers traveling structural supports, this paper proposed a conceptual framework to help computer vision function similarly to the human thinking model in construction safety management by integrating a formal ontology model. Specific details of the Mask R–CNN algorithm are presented in our previous study [33].

4.2. Ontology-based knowledge modeling for FFH safety

4.2.1. Definition of FFH ontology

A prototype FFH ontology model was developed in Protégé 5.0 to represent the domain knowledge for FFH safety. Fig. 10 presents the class hierarchy of the ontology model.

The consistency of the developed ontology model in Protégé is checked using the Pellet reasoner [59]. The consistency criterion determines whether the developed ontology includes any contradictory facts. This criterion also ensures that the conclusions are inferential and semantically consistent. Pellet, as a logic reasoner, not only detects inconsistent ontology concepts but also supports the diagnosis and resolution of errors. Fig. 11 shows the result of consistent checking using Protégé. The value of "*hasBodyReaction*" data property should be "string" rather than the "int."

4.2.1. Development of FFH rules

The SWRL rules were used to reason hazards and their corresponding mitigations, which have been encoded in the SWRLTab plug-in in Protégé 5.0. The SWRLTab plug-in allows users to add or modify SWRL rules in Protégé. Fig. 12 shows the SWRL rules. The FFH safety knowledge sources of these rules include "Technical code for the safety of working at the height of building construction (JGJ 80–2016)" and "Technical code for the safety of deep building foundation excavations (JGJ 311–2013)." For example, the regulation "Technical code for the safety of working at the height of building construction (JGJ 80–2016)" explicitly stipulates that people are not allowed to traverse the structural supports over a deep foundation on supports because this action easily leads to FFH hazards. Additionally, some mitigation measurements of FFH, such as wearing appropriate personal protective equipment and setting reminders or the protective railing shall at the edge, are available in the "Technical code for the safety of working at the height of building construction (JGJ 80–2016)." The regulatory knowledge can be formally represented by the SWRL rules as follows.

Rule 1:

Concrete_Support(?x) ˆ hasRelativeRelationship(?x, "on") ˆ Labor(?y) ˆ hasSpatialRelationship(?y, ?z) ˆ swrlb:stringEqualIgnoreCase(?z, "on") → hasPrecursor(?y, Walking_On_Supports) ˆ causeHazard (Fall_From_Building_Structures).

Rule 2:

Walking_On_Supports(?a) ˆ Fall_From_Building_Structures(?b) ˆ causeHazard(?a, ?b) → controlledBy(?b, Protective_Space) ˆ controlledBy(?b, Safeguard).

With the ontology model and SWRL rules, FFH safety knowledge from textual regulations can be formally represented to allow computer access. The developed ontology model provides a knowledge base for the visual information extracted by computer vision algorithms and helps the computer vision system understand the meaning behind the objects. Considering construction safety management, the ontological knowledge base can infer potential FFH hazards and list corresponding mitigation measures based on the visual inputs obtained by the computer vision module.

4.3. Mapping of visual information and ontology instances

The visual information extracted by the computer vision module was

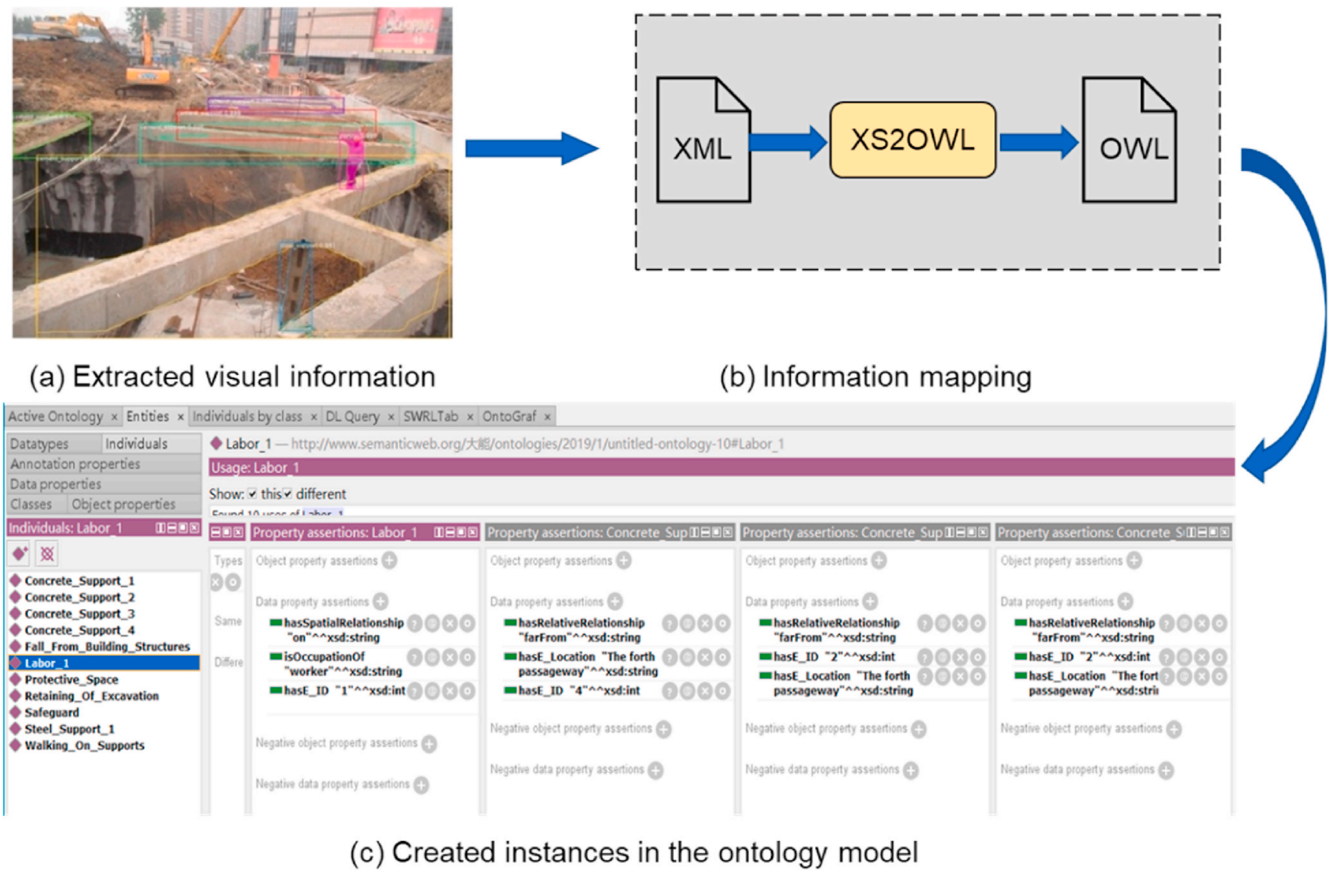

Fig. 13. Mapping visual information with instances in the ontology.

connected with the developed ontology model by creating an automated instance. Fig. 13 shows the information mapping process. Fig. 13(a) shows that the objects (i.e., concrete supports, steel supports, and labors) and their spatial relationship (i.e., the labor is on the concrete support) are detected in the computer vision module using the Mask R–CNN. The visual information is outputted into an XML format and then converted into the OWL format using the XS2OWL as shown in Fig. 13(b). The classification label of detected objects in XML is converted into "owl:class" in OWL, while spatial relationships are transformed into "owl: dataProperty" in OWL. As shown in Fig. 13(c), the extracted visual information is finally mapped into instances in Protégé for following knowledge reasoning.

4.4. Safety knowledge reasoning from images

Knowledge reasoning can be implemented in Protégé by using the rule engine named Drools after mapping the visual information into instances in the ontology model (Fig. 14).

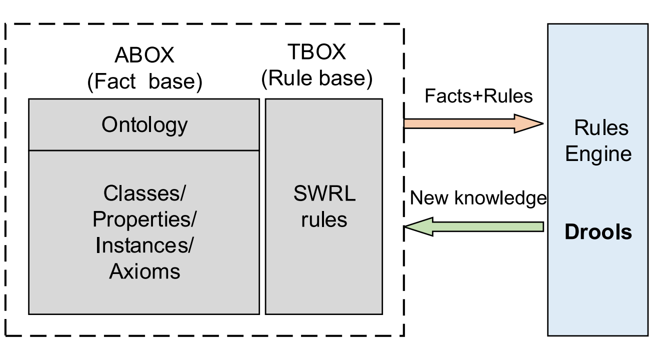

Fig. 14. Implementing soft environment for knowledge in Protégé.

Knowledge reasoning is conducted in the Drools rule engine, in which the general structure is "When A, Then B" [60]. The basic framework of the Drools rules engine comprises a fact base, a rule base, and an execution engine. Hazards and corresponding mitigations can be inferred using the Drools reasoning engine. As shown in Fig. 15, the hazard scene in the image of Fig. 9 was inferred on the basis of developed SWRL rules and created instances. The reasoning results are regarded as new knowledge, which can be stored in the fact base.

Fig. 16 presents the reasoning results in the scene of Fig. 9. Person_1 has the precursor named "*walking on supports*," which would cause the hazard "*Falling_From_Building_Structures*." Visual information captured by the computer vision system can be endowed with safety meanings by knowledge reasoning using the developed ontology model (known as the fact base) and the SWRL rules (known as the rule base). Specifically, potential hazards in the image can be identified, and the corresponding mitigations preventing these hazards can also be reasoned similarly to humans. Different from traditional computer vision approaches, the FFH hazard in this scene was detected on the basis of rules. Moreover, the mitigations (e.g., wearing safeguards and setting protective space) can be inferred for safety managers, which can prevent potential hazards.

The reasoning results demonstrate the theoretical and technical feasibility of the proposed conceptual framework. The safety knowledge can be explicitly represented and combined with the visual information extracted by computer vision algorithms through the integration of ontology techniques. The FFH hazard and its mitigation measures can then be reasoned by performing rule checking on the site visual information extracted by Mask R–CNN. The ontology model with SWRL rules provides a formal and explicit safety knowledge base to computer vision systems, which can help computer vision operate similarly to a human thinking model in construction safety management. Therefore, the

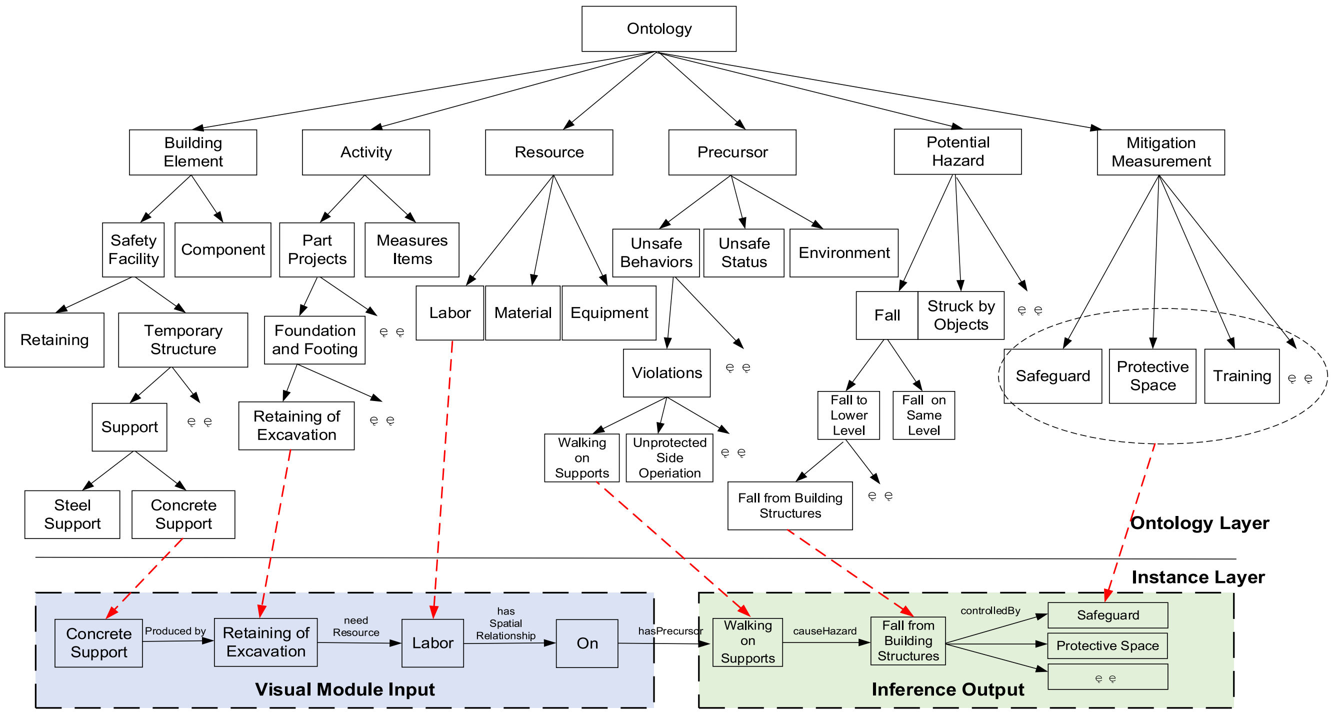

Fig. 15. Inference process of the scene presented in Fig. 11 (a person is traversing the concrete support).

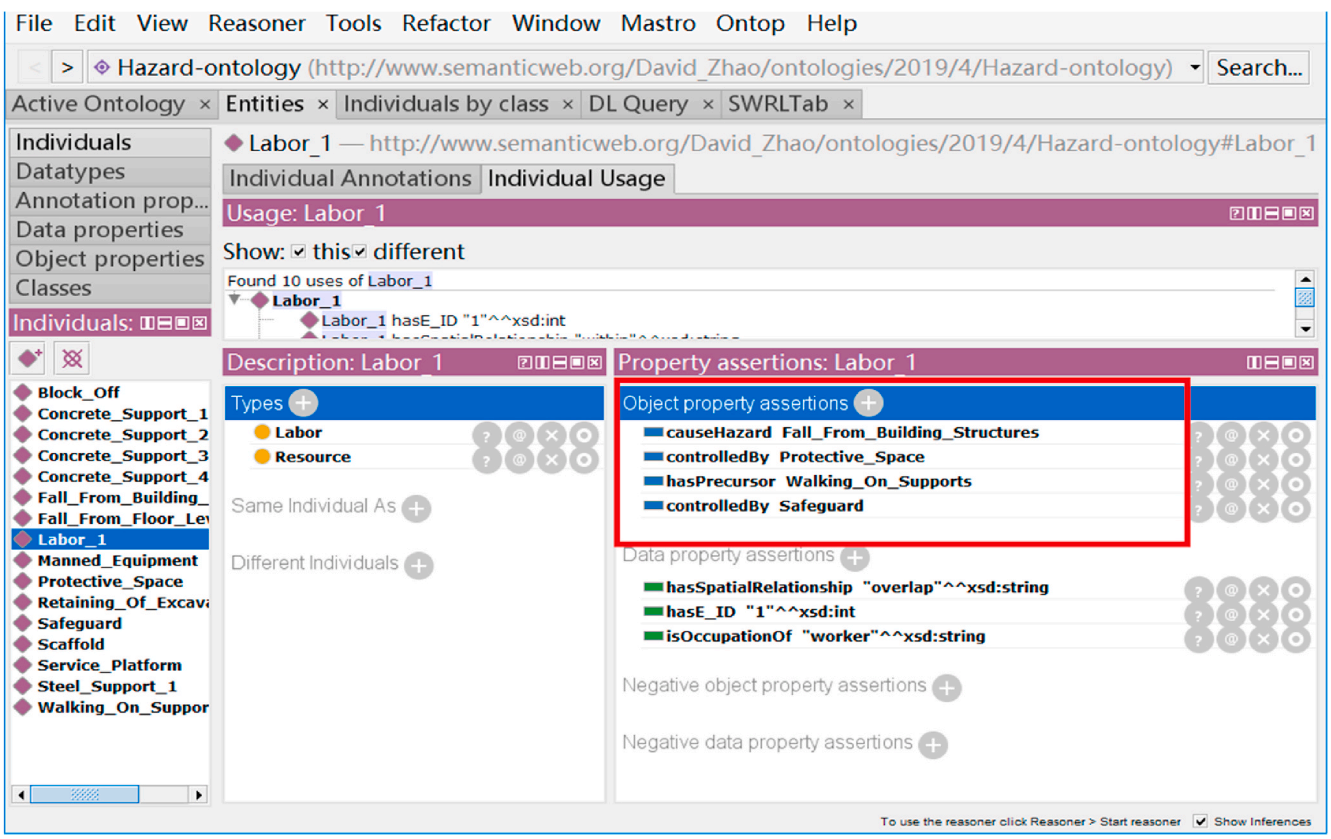

Fig. 16. Reasoning results of hazards and mitigations in the scene of Fig. 9.

proposed framework can facilitate safety management by automatically reasoning the hazards and corresponding mitigation measures for safety managers.

Furthermore, the proposed framework can attach the extracted visual information (low-level features) and semantic hazard information (high-level knowledge) to initial images, which can facilitate the image annotation and semantic retrieval [61]. The extracted objects can be linked with semantic concepts through the developed ontology model,

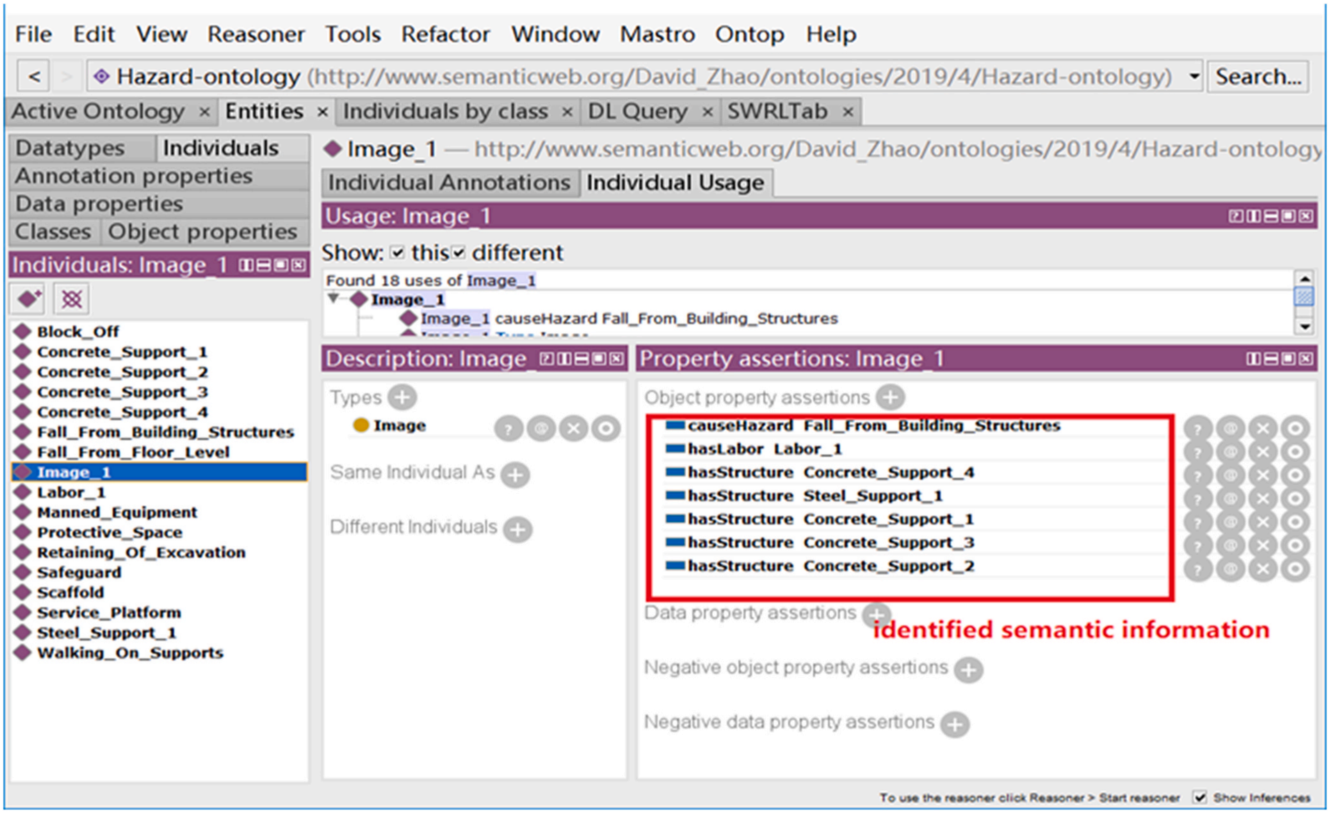

Fig. 17. Information annotation in the ontology model.

and the hazard information can enrich the description of images. For example, Fig. 17 shows the extracted information (e.g., labor_1, concrete supports, and steel support) and reasoned hazard information (e.g., *Fall_From_Building_Structures*) of the image (Fig. 9). The obtained images from the site can be re-used considering this semantic annotation. The explicit semantic may facilitate further image index and retrieval in a vast collection of images.

5. Discussion and limitations

The computer vision system is currently limited in linking the low-level visual information with high-level semantic meaning considering construction safety management, which is a knowledge-intensive task. A conceptual framework is proposed in this paper to imitate the human-based safety inspections, which depend on their prior knowledge to detect hazards in the workplace and organize corresponding measurements to prevent these hazards. In the developed framework, the computer vision acts as "human eyes," which is responsible for extracting objects and their spatial relationships. The obtained visual information is regarded as inputs and is mapped to instances in the ontology model acting as the "human brain" to facilitate hazard inference. The result shows the developed framework is successful in knowledge-intensive construction safety management by semantically reasoning hazards and listing corresponding mitigations from on-site images.

However, some limitations still exist in current research. First, a simple but common FFH hazard scene in China is used to test the theoretical feasibility of combining computer vision and ontology for knowledge-intensive hazard identification and mitigation, in which only three types of spatial relationship (on, overlap, and away) are examined. However, other spatial relationships, such as "left," "right," "above," "blow," and "aligned" must be detected for different types of hazard identification. Additionally, the developed conceptual framework has the potential to reason different types of hazards for safety managers if the computer vision algorithm can accurately detect visual information from images and the ontology model is enriched with various SWRL rules. Examples are shown in Fig. 18. Computer vision algorithms can be used to extract visual information from images; for example, detecting workers who travel the structural supports [33] and don't wear the hardhat [11]. The extracted visual information is used as the input for rule-based hazard reasoning. With the extracted visual information and pre-defined rules, different types of hazards can be inferred. For instance, there are two types of hazards in Fig. 18 (a) and (b), namely, falling from height and head injuries. In this paper, we built on our previous study [33], and only a simple but common FFH hazard scene is used to show the theoretical feasibility of the proposed conceptual framework. Substantially complex scenes involving different types of hazards should be used for further validation of the framework in future studies. Moreover, prototype systems should be developed for further validation on the effectiveness of the proposed framework in construction practices.

Second, the capability of the computer vision module on capturing visual information is important for knowledge reasoning in the proposed framework. Specifically, acting as "human eyes," accurate visual information detection from images is important because it is the input in following hazard inference. The ontological knowledge base is ineffective without accurate inputs. Therefore, the accuracy and reliability of the computer vision module are essential. However, the computer vision is still slightly weak for practical implementation in the construction domain. On the one hand, the lack of sized construction image databases could affect its performance of deep learning algorithms [8]. On the other hand, occlusions also hinder this capability because construction sites are usually congested with materials, equipment, and structures of

Fig. 18. Process of detecting various types of hazards.

different heights and shapes, thus leading to massive blind spots for the cameras. Additionally, the camera cannot be installed if construction workers do not agree due to privacy concerns [12]. The Mask R–CNN is used in this paper due to its high performance in object detection and instance segmentation [45]. An array of robust vision-based algorithms are available; however, improving the accuracy of computer vision algorithms or finding the optimal algorithm is not the aim of this paper.

Third, a finite set of rules are encoded in this paper to validate the theoretical feasibility of the proposed framework. More SWRL rules for different types of hazards should be encoded in the future, aiming to satisfy the requirement of construction safety management in different hazardous scenes. Essentially, the ontology model and SWRL rules act as 'human brain' in the proposed framework, which aim to provide safety regulatory knowledge for hazard identification and prevention. Hence, the enrichment of SWRL rules is important to its effectiveness. However, encoding knowledge of numerous safety regulations into SWRL rules is time-consuming, which may hinder the practical application of the developed framework.

6. Conclusion

This paper introduces a conceptual framework integrating computer vision and ontology for construction safety management. The ontology-based representation of safety knowledge gives meaning to the acquired inputs of computer vision algorithms. Similar to humans, these algorithms can support safety inspection by automatically reasoning hazards from on-site photos and listing corresponding mitigation suggestions. In particular, a Mask R–CNN is used to recognize objects and spatial relationships because it can accurately detect multiple objects from images. The extracted visual information is connected with the ontology concepts by creating instances using XS2OWL. Hazards from images and their corresponding mitigations are then inferred using a reasoning engine based on the encoded SWRL rules. Finally, an FFH example is used to validate the feasibility of the proposed framework and demonstrate its capability to infer hazards by combining regulation knowledge automatically.

Overall, the developed theoretical framework can help computer vision function similarly to the human thinking model in construction safety management. The framework also provides site managers and engineers with a creative method to complement the cadre of hazard management approaches used to manage safety in construction. The accuracy of detecting objects and spatial relationships from images or videos can be markedly enhanced with the advancement of deep learning algorithms. For example, more than 1000 projects have used the commercial product named SmartVid's AI engine for detecting hard hats, gloves, and safety vests of workers [62]. Today, hundreds to thousands of images were acquired in construction sites on a daily basis [63], and daily construction activities can be recorded by the deployed cameras. The visual data can be used as the foundation for safety management. Hazards or near-misses in daily construction tasks and corresponding suggestions can be reasoned automatically during the application of the proposed framework to construction projects. The inference information can be sent to safety managers for review and further processing. Therefore, this information provides a means of frequent safety inspection and can reduce the workloads of safety managers. Moreover, the framework offers a probability for the construction company to conduct large-scale and frequent safety inspections. Relying on the advanced computer vision algorithms and the complete ontological knowledge base, different projects can statistically near-miss with high frequency and low severity and build their hazardous patterns for further proactive prevention. Hazardous patterns may differ from projects and construction companies. For instance, Raviv et al. [64] quantitatively analyzed tower crane-related accidents from two construction companies in Israel. They found that one suffered from technical failure, while the other suffered from rigging and signaling failures. Additionally, the raw data would be annotated with inferred information, which helps improve the completeness and accuracy of data for future machine learning training and development purposes. Real-time identification of unsafe acts during construction may lead to perfunctory intervention by management, which can result in immediate behavior modification. In addition, the acquired video can be used to provide people with direct visual feedback and be used as a tool for safety education.

The scientific contributions of the paper lie in the following: (1) the integrated framework combining advantages of computer vision (acting as the human eyes) and ontology (acting as the human brain), producing intelligent construction safety management; (2) the developed ontological model, a formal way to represent safety domain knowledge with SWRL rules, can be easily extended to the future hazardous scene. Relying on advanced computer vision algorithms and enriched ontology models, different hazards and these mitigations from images or videos can be automatically identified for safety inspectors on basis of the proposed framework.

Author statement

Haitao Wu: Writing-original draft, Review & editing, Formal analysis, Software.
Botao Zhong: Funding acquisition, Review& editing.
Heng Li: Writing, Review & editing.
Peter Love: Writing, Review & editing.
Xing Pan: Formal analysis, Software.
Neng Zhao:Formal analysis, Software.

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.

Acknowledgments

This research is partly supported by the "National Natural Science Foundation of China" (No.71732001, No. 51878311).

References

[1] OSHA, Commonly used Statistics. https://www.osha.gov/oshstats/commonstats.html2014. (Accessed 17 January 2015).
[2] MOHURD (Ministry of Housing and Urban-Rural Development of the People's Republic of China), Bulletin on the Safety Accidents of Housing Municipal Engineering, Retrieved from, http://www.mohurd.gov.cn/wjfb/201903/t20190326_239913.html, 2018.
[3] P.E.D. Love, P. Teo, J. Smith, F. Ackermann, Y. Zhou, The nature and severity of workplace injuries in construction: engendering operational benchmarking, Ergonomics 62 (10) (2019) 1273–1288, https://doi.org/10.1080/00140139.2019.1644379.
[4] Bureau of Labor Statistics (BLS), 2017 Census of Fatal Occupational Injuries (Final Data)– Industry by Event or Exposure. https://www.bls.gov/iif/oshwc/cfoi/cfch0016.pdf, 2018.
[5] Safe Work Australia (SWA), Construction Industry Profile. https://www.safeworkaustralia.gov.au/doc/construction-industry-profile, 2015.
[6] M.-W. Park, I. Brilakis, Construction worker detection in video frames for initializing vision trackers, Autom. ConStruct. 28 (2012) 15–25, https://doi.org/10.1016/j.autcon.2012.06.001.
[7] J. Teizer, Status quo and open challenges in vision-based sensing and tracking of temporary resources on infrastructure construction sites, Adv. Eng. Inf. 29 (2) (2015) 225–238, https://doi.org/10.1016/j.aei.2015.03.006.
[8] 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.
[9] C. Lu, R. Krishna, M. Bernstein, F. Li, Visual relationship detection with language priors, European Conference on Computer Vision (2016) 852–869, https://doi.org/10.1007/978-3-319-46448-0_51.
[10] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, C. Li, Computer vision aided inspection on falling prevention measures for steeplejacks in an aerial environment, Autom. ConStruct. 93 (2018) 148–163, https://doi.org/10.1016/j.autcon.2018.05.022.
[11] 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. ConStruct. 85 (2018) 1–9, https://doi.org/10.1016/j.autcon.2017.09.018.
[12] B. Zhong, H. Wu, L. Ding, L. Peter, H. Li, H. Luo, L. Jiao, Mapping computer vision research in construction: developments, knowledge gaps and implications for research, Autom. ConStruct. 107 (2019) 102919, https://doi.org/10.1016/j.autcon.2019.1029.
[13] X. Sun, Y. Zhang, J. Chen, RTPO: a domain knowledge base for robot task planning, Electronics 8 (10) (2019) 1105, https://doi.org/10.3390/electronics8101105.
[14] X. Xing, B. Zhong, H. Luo, H. Li, H. Wu, Ontology for safety risk identification in metro construction, Comput. Ind. 109 (2019) 14–30, https://doi.org/10.1016/j.compind.2019.04.001.
[15] R. Marroquin, J. Dubois, C. Nicolle, Know beyond seeing: combining computer vision with semantic reasoning, in: 2018 IEEE 12th International Conference on Semantic Computing (ICSC), IEEE, 2018, pp. 310–311, https://doi.org/10.1109/ICSC.2018.00061.
[16] D. Conigliaro, R. Ferrario, C. Hudelot, D. Porello, Integrating computer vision algorithms and ontologies for spectator crowd behavior analysis, Group and Crowd Behavior for Computer Vision (2017) 297–319, https://doi.org/10.1016/B978-0-12-809276-7.00016-3.
[17] H. Bannour, C. Hudelot, Building and using fuzzy multimedia ontologies for semantic image annotation, Multimed. Tool. Appl. 72 (3) (2014) 2107–2141, https://doi.org/10.1007/s11042-013-1491-z.
[18] C.J. Anumba, R.R.A. Issa, J. Pan, I. Mutis, Ontology-based information and knowledge management in construction, Construct. Innovat. 8 (2008) 218–239, https://doi.org/10.5220/0003053903310334.
[19] B. Zhong, H. Wu, H. Li, S. Semad, H. Luo, L. He, A scientometric analysis and critical review of construction related ontology, Autom. ConStruct. 101 (2019) 17–31, https://doi.org/10.1016/j.autcon.2018.12.013.
[20] Y. Lu, Q. Li, Z. Zhou, Y. Deng, Ontology-based knowledge modeling for automated construction safety checking, Saf. Sci. 79 (2015) 11–18, https://doi.org/10.1016/j.ssci.2015.05.008.
[21] S.J. Zhang, F. Boukamp, J. Teize, Ontology-based semantic modeling of construction safety knowledge: towards automated safety planning for job hazard analysis (JHA), Autom. ConStruct. 52 (2015) 29–41, https://doi.org/10.1016/j.autcon.2015.02.005.
[22] H. Wu, B. Zhong, B. Medjdoub, X. Xing, L. Jiao, An ontological metro accident case retrieval using CBR and NLP, Appl. Sci. 10 (15) (2020) 5298, https://doi.org/10.3390/app10155298.

[23] W. Fang, L. Ding, E. Peter, H. Luo, H. Li, F. Pena-Mora, B. Zhong, C. Zhou, Computer vision applications in construction safety assurance, Autom. ConStruct. 110 (2020) 103013, https://doi.org/10.1016/j.autcon.2019.103013.

[24] R. Xiong, Y. Song, H. Li, Y. Wang, Onsite video mining for construction hazards identification with visual relationships, Adv. Eng. Inf. 42 (2019) 100966, https://doi.org/10.1016/j.aei.2019.100966.

[25] W. Fang, L. Ma, E. Peter, H. Luo, L. Ding, A. Zhou, Knowledge graph for identifying hazards on construction sites: integrating computer vision with ontology, Autom. ConStruct. 119 (2020) 103310, https://doi.org/10.1016/j.autcon.2020.103310.

[26] N.F. Noy, D.L. McGuinness, Ontology development 101: a guide to creating your first ontology. http://www.corais.org/sites/default/files/ontology_development_101_aguide_to_creating_your_first_ontology.pdf, 2001.

[27] L. Orlando, Detecting steel rods and micro-piles: a case history in a civil engineering application, J. Appl. Geophys. 81 (2012) 130–138, https://doi.org/10.1016/j.jappgeo.2012.01.009.

[28] B.H.W. Guo, Y.M. Goh, Ontology for the design of active fall protection systems, Autom. ConStruct. 82 (2017) 138–153, https://doi.org/10.1016/j.autcon.2017.02.009.

[29] K.K. Han, M. Golparvar-Fard, Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs, Autom. ConStruct. 53 (2015) 44–57, https://doi.org/10.1016/j.autcon.2015.02.007.

[30] J. Gong, C.H. Caldas, Computer vision-based video interpretation model for automated productivity analysis of construction operations, J. Comput. Civ. Eng. 24 (2010) 252–263, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000027.

[31] M. Golparvar-Fard, A. Heydarian, J.C. Niebles, Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers, Adv. Eng. Inf. 27 (4) (2013) 652–663, https://doi.org/10.1016/j.aei.2013.09.001.

[32] M.-W. Park, I. Brilakis, Continuous localization of construction workers via integration of detection and tracking, Autom. ConStruct. 72 (2016) 129–142, https://doi.org/10.1016/j.autcon.2016.08.039.

[33] W. Fang, B. Zhong, N. Zhao, P.E. Love, H. Luo, J. Xue, S.J. Xu, A deep learning-based approach for mitigating falls from height with computer vision: convolutional neural network, Adv. Eng. Inf. 39 (2019) 170–177, https://doi.org/10.1016/j.aei.2018.12.005.

[34] X. Yang, H. Li, Y. Yu, X. Luo, T. Huang, X. Yang, "Automatic pixel-level crack detection and measurement using fully convolutional network, Comput. Aided Civ. Infrastruct. Eng. 33 (12) (2018) 1090–1109, https://doi.org/10.1111/mice.12412, 10.1016/j.autcon.2017.05.002.

[35] I. Brilakis, M.W. Park, G. Jog, Automated vision tracking of project related entities, Adv. Eng. Inf. 25 (4) (2011) 713–724, https://doi.org/10.1016/j.aei.2011.01.003.

[36] 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. ConStruct. 86 (2018) 118–124, https://doi.org/10.1016/j.autcon.2017.11.002.

[37] Q. Fang, H. Li, X.C. Luo, L.Y. Ding, T.M. Rose, W.P. An, Y.T. 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.

[38] Z. Zhu, X. Ren, Z. Chen, Integrated detection and tracking of workforce and equipment from construction job site videos, Autom. ConStruct. 81 (2017) 161–171, https://doi.org/10.1016/j.autcon.2017.05.005.

[39] L. Nanni, S. Ghidoni, S. Brahnam, Handcrafted vs. non-handcrafted features for computer vision classification, Pattern Recogn. 71 (2017) 158–172, https://doi.org/10.1016/j.patcog.2017.05.025.

[40] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444, https://doi.org/10.1038/nature14539.

[41] N. Duan, D. Tang, M. Zhou, November). "Machine reasoning: technology, dilemma and future, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts, 2020, pp. 1–6, https://doi.org/10.18653/v1/2020.emnlp-tutorials.1.

[42] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Proc., Communications of the ACM 60 (6) (2012) 84–90, https://doi.org/10.1145/3065386.

[43] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, European Conference on Computer Vision (2016) 21–37, https://doi.org/10.1007/978-3-319-46448-0_2.

[44] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. (2016) 91–99, https://doi.org/10.1109/tpami.2016.2577031.

[45] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN." the IEEE International Conference On Computer Vision (ICCV). http://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html, 2017, 2961-2969.

[46] Y. Yu, H. Guo, Q. Ding, H. Li, M. Skitmore, "An experimental study of real-time identification of construction workers' unsafe behaviors, Autom. ConStruct. 82 (2017) 193–206.

[47] X. Luo, H. Li, D. Cao, Y. Yu, X. Yang, T. Huang, Towards efficient and objective work sampling: recognizing workers' activities in site surveillance videos with two-stream convolutional networks, Autom. ConStruct. 94 (2018) 360–370, https://doi.org/10.1016/j.autcon.2018.07.011.

[48] X. Luo, H. Li, Y. Yu, C. Zhou, D. Cao, Combining deep features and activity context to improve recognition of activities of workers in groups, Comput. Aided Civ. Infrastruct. Eng. (2020), https://doi.org/10.1111/mice.12538.

[49] H. Kim, K. Kim, H. Kim, Vision-based object-centric safety assessment using fuzzy inference: monitoring struck-by accidents with moving objects, J. Comput. Civ. Eng. 30 (2016), https://doi.org/10.1061/(ASCE)CP.1943-5487.0000562, 04015075.

[50] S. Tang, M. Golparvar-Fard, Joint reasoning of visual and text data for safety hazard recognition, J. Comput. Civ. Eng. (2017) 450–457, https://doi.org/10.1061/9780784480847.056.

[51] B. Zhong, H. Li, H. Luo, J. Zhou, W. Fang, X. Xing, Ontology-based semantic modeling of knowledge in construction: classification and identification of hazards implied in images, J. Construct. Eng. Manag. 146 (4) (2020), 04020013, https://doi.org/10.1061/(ASCE)CO.1943-7862, 0001767.

[52] L.Y. Ding, B.T. Zhong, S. Wu, H.B. Luo, Construction risk knowledge management in BIM using ontology and semantic web technology, Saf. Sci. 87 (2016) 202–213, https://doi.org/10.1016/j.ssci.2016.04.008.

[53] H.P. Tserng, S.Y. Yin, R. Dzeng, B. Wou, M. Tsai, W. Chen, A study of ontology-based risk management framework of construction projects through project life cycle, Autom. ConStruct. 18 (7) (2009) 994–1008, https://doi.org/10.1016/j.autcon.2009.05.005.

[54] H.-H. Wang, F. Boukamp, Ontology-based representation and reasoning framework for supporting job hazard analysis, J. Comput. Civ. Eng. 25 (6) (2011) 442–456, https://doi.org/10.1061/(asce)cp.1943-5487, 0000125.

[55] N.M. El-Gohary, T.E. El-Diraby, Domain ontology for processes in infrastructure and construction, J. Construct. Eng. Manag. 136 (7) (2010) 730–744, https://doi.org/10.1061/(asce)co.1943-7862.0000178.

[56] P. Pauwels, D. Van Deursen, R. Verstraeten, J. De Roo, R. De Meyer, "A Semantic Rule Checking Environment for Building Performance Checking" *Automation in Construction*, 2009, pp. 1–13.

[57] I. Horrocks, P.F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, M. Dean, SWRL: a semantic web rule language combining OWL and RuleML, 21, 2004, pp. 1–31, 79.

[58] M. Zhang, D. Fang, A cognitive analysis of why Chinese scaffolders do not use safety harnesses in construction, Construct. Manag. Econ. 31 (3) (2013) 207–222, https://doi.org/10.1080/01446193.2013.764000.

[59] E. Sirin, B. Parsia, B.C. Grau, A. Kalyanpur, Y. Kata, Pellet: a practical OWL-DL reasoner, Web Semant. Sci. Serv. Agents World Wide Web 5 (2) (2007) 51–53, https://doi.org/10.2139/ssrn.3199351.

[60] K. Schwabe, J. Teizer, M. König, Applying rule-based model-checking to construction site layout planning tasks, Autom. ConStruct. 97 (2019) 205–219, https://doi.org/10.1016/j.autcon.2018.10.012.

[61] N. Magesh, P. Thangaraj, Semantic image retrieval based on ontology and SPARQL query, in: International Conference on Advanced Computer Technology, 2011. https://pdfs.semanticscholar.org/8d08/e0f95f7c8dda24ebefd88da583e7d26425e6.pdf.

[62] S. Tang, D. Roberts, M. Golparvar-Fard, Human-object interaction recognition for automatic construction site safety inspection, Autom. ConStruct. 120 (2020) 103356, https://doi.org/10.1016/j.autcon.2020.103356.

[63] K.K. Han, M. Golparvar-Fard, Potential of big visual data and building information modeling for construction performance analytics: an exploratory study, Autom. ConStruct. 73 (2017) 184–198, https://doi.org/10.1016/j.autcon.2016.11.004.

[64] G. Raviv, B. Fishbain, A. Shapira, Analyzing risk factors in crane-related near-miss and accident reports, Saf. Sci. 91 (2017) 192–205, https://doi.org/10.1016/j.ssci.2016.08.022.