Knowledge Graph for Identifying Hazards on Construction Sites: Integrating Computer Vision with Ontology

Weili Fang[1,2,3,4], Ling Ma[5*],Peter E.D. Love[4], Hanbin, Luo[1,2], Lieyun Ding[1,2], Ao Zhou[1,2]

[1]School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, 430074, Hubei, China

[2]Hubei Engineering Research Center for Virtual, Safe and Automated Construction, 430074, Hubei, China

[3]School of Civil Engineering and Engineering Mechanics, Columbia University in the City of New York, 10027, New York, USA

[4] Dept. of Civil and Mechanical Engineering, Curtin University, Perth, WA 6845, Australia

[5] The Bartlett School of Construction and Project Management, University College London, London WC1E 6BT, United Kingdom

Corresponding Author: Ling Ma, Email: l.ma@ucl.ac.uk

Knowledge Graph for Identifying Hazards on Construction Sites: Integrating Computer Vision with Ontology

Abstract

Hazards potentially affect the safety of people on construction sites include falls from heights (FFH), trench and scaffold collapse, electric shock and arc flash/arc blast, and failure to use proper personal protective equipment. Such hazards are significant contributors to accidents and fatalities. Computer vision has been used to automatically detect safety hazards to assist with the mitigation of accidents and fatalities. However, as safety regulations are subject to change and become more stringent prevailing computer vision approaches will become obsolete as they are unable to accommodate the adjustments that are made to practice. This paper integrates computer vision algorithms with ontology models to develop a knowledge graph that can automatically and accurately recognise hazards while adhering to safety regulations, even when they are subjected to change. Our developed knowledge graph consists of: (1) an ontological model for hazards: (2) knowledge extraction; and (3) knowledge inference for hazard identification. We focus on the detection of hazards associated with FFH as an example to illustrate our proposed approach. We also demonstrate that our approach can successfully detect FFH hazards in varying contexts from images.

Keywords: Hazards; ontology; computer vision; safety; knowledge graph database

1.0 Introduction

Over 60,000 fatal injuries are reported to occur every year from construction projects worldwide [44]. According to the Occupation Safety and Health Administration (OSHA), for example, the construction industry is responsible for more than 20% of fatalities in the United States [53]. In the United Kingdom, for example, a similar scenario occurs where construction accounts for the highest number of fatalities across all sectors [16].

Typically hazard analysis is undertaken before construction and may be performed using manual methods and/or three-dimensional (3D) models [27, 50]. Hazards can change once construction commences, and their identification then needs to be undertaken manually, which can be a labour-intensive and time-consuming process. Several automatic computer vision-based approaches have been developed to overcome the drawbacks of manually identifying hazards [62, 20-24]).

Despite the success of being able to deploy computer vision to identify hazards, it is unable to recognise those that are newly defined as a result of changes to safety regulations and procedures as (1) typically one computer vision algorithm is used to identify a single hazard in a scene. For example, identifying a person who is not wearing their safety helmet, and (2) current computer vision approaches are unable to extract semantic relationships between detected objects. As a result, a 'semantic gap' is formed between the low-level features derived from images and the high-level semantic information that people obtain.

This paper combines computer vision algorithms with ontology to construct a knowledge graph that can automatically detect hazards to address the 'semantic gap' that prevails. We aim to determine whether our as-built semantic vision-based knowledge graph can identify risks with complex rules. In doing so, we develop a knowledge graph that integrates computer-vision with ontology. An ontology is used to help experts annotate knowledge and is used to describe the relationships between the entities. Describing these relationships enables computer applications to represent and reason about safety knowledge efficiently. When an ontology is used in conjunction with computer vision, knowledge can be extracted (i.e., entity recognition and relationship extraction) from images automatically.

We commence our paper by providing a review of computer vision-based object detection approaches and applications of ontology-based risk management that have been developed in construction (Section 2). Then, we introduce and describe our proposed knowledge graph framework for identifying hazards (Section 3). Following a description of the developed framework, we then demonstrate and test the validity of our developed framework using hazards identified during the construction of the Wuhan Rail Transit System in China (Section 4). Next, we discuss our research findings, specifically highlighting the benefits and limitations of our framework. We conclude our paper by identifying the paper's contributions to the field of computer vision in construction.

2.0 Research Methodology

2.1 Computer Vision-based Object Detection

Computer vision has been utilised to perform a variety of tasks in construction such as productivity analysis [26], progress monitoring [29], as well as the recognition of unsafe behaviour [10,20,22]. Vision-based object detection within the domain of construction has focused on utilising the following approaches: (1) hand-crafted features; and (2) deep learning. In Table 1, we present a summary of critical vision-based object detection studies that have been undertaken.

Hand-crafted feature-based methods employ a three-stage procedure, which consists of: (1) feature extraction; (2) feature representation; and (3) classification. Descriptors typically used to extract features from images and videos include Histogram of Oriented Gradients (HOG) [8], Histogram of Optical Flow (HOF) [57], and Scale Invariant Feature Transform (SIFT) [45]. Once features are extracted, they are then inserted into a classifier such as Support Vector Machine (SVM) and k-Nearest Neighbour. There exists a considerable body of work that has used hand-crafted feature approaches to detect objects in construction.

Chi and Caldas [6], for example, applied a background subtraction algorithm to extract features from images. Then, using a naïve Bayes classifier and neural network, people, loaders, and backhoes were identified [6]. Contrastingly, Park and Brilakis [55] and Azar and McCabe [2] have utilised HOG and Haar-like feature descriptors to detect individuals and equipment (e.g., machinery). Similarly, Memarzadeh [3] combined a HOG and colour features with new multiple binary SVM classifiers to automatically detect and distinguish between a person and equipment using videos. Despite the success of hand-crafted feature-based approaches, they are manually created. Therefore, there is a trade-off between detection accuracy and computational efficiency (i.e., speed) arises [52]. The uncertainties and changing conditions that prevail on a construction site can also impact the extraction of features from images. For example, view-point scale, intraclass and variance as well background clutter can lead to lower levels of accuracy for object detection [33,56].

With the advent of large-scale data sets such as ImageNet [9], improved designs for modelling and training deep networks, and the development of computer architectures, deep learning has provided the ability to automatically extract and learn features in an end to end manner from images with higher levels of accuracy [39]. A Convolutional Neural Network (CNN) can be used for object detection or action recognition and can automatically extract features due to their ability to stack multiple convolutional (i.e., detects local conjunctions of features from the previous layer) and pooling layers [39].

Several studies have demonstrated the potential of CNN's for object detection and action recognition on construction sites [61,21,23-24]. For example, Fang *et al*. [21] developed an improved Faster R-CNN to identify objects from images and have achieved accuracy with 91% and 95% when detecting individuals and excavators, respectively [21]. Likewise, Fang *et al*. [22] applied a computer vision approach with Mask Region-Based CNN (Mask R-CNN) to identify the unsafe behaviour of individuals that traversed structural supports. In this research, a Mask R-CNN was used to accurately identify people and structural supports, which achieved satisfactory levels of performance [22].

A review of computer vision-based studies in construction reveals that acceptable levels of accuracy (i.e., precision, recall) on object detection and attributes (e.g., distance measure) exist. For example, Kim *et al*. [36] applied a transformation matrix to determine the distance between objects from a single image. Here Kim *et al*. [36] applied a transformation matrix to represent the geometric relationship between objects. The distance between objects was estimated by measuring the pixel distance between them, where an object's reference geometric was known and used [37]. Drawing on the research of Fang *et al*. [22], we can observe that a Mask R-CNN is a suitable approach to detect objects from two-dimensional (2D) images, and the production of a transformation matrix [36-38] is appropriate for computing an object's distance from a single image.

Table 1. Key object detection studies

Authors (Year)	**Target of interest**	**Visual object detection methods**	**Type of detection approach**
Kim *et al*. [35]	Concrete mixer truck	Three-dimensional (3D) Reconstruction and HOG	Hand-crafted feature
Fang *et al*. [20]	People, Safety harness	Faster R-CNN	Deep learning
Fang *et al*. [21]	People, Excavator	Improved Faster R-CNN	Deep learning
Azar and Mccabe [2]	Hydraulic excavator	HOG	Hand-crafted feature
Park and Brilakis [55]	People	Background subtraction, HOG, HSV colour histogram	Hand-crafted feature

2.2 Ontology-based Risk Knowledge Management

Ontology is a formal conceptualisation of knowledge. It is a simplified view of a domain that describes objects, concepts, and relationships between them [15]. Traditional ontology relies on the experiences of the individual, knowledge of domain experts, and relevant managerial personnel to support the decision-making process. Semantic Web technology, for example, can allow various sources of information to be made available in a format that can be searched and retrieved from the Internet [18]. Thus, the combination of semantic web technology with ontology can enable the following advantages to be realised [11,18]:

- improved model flexibility, enabling the extension of knowledge, which can be readily changed and adapted by application requirements;
- robust semantic representation, and promotion of the semantical interaction between different computers; and
- support semantic inference and retrieval through improving requests from a concept level.

Ontology-based approaches have been extensively applied to numerous aspects of construction, such as energy management [7,31], building cost estimation [40] and risk management [63]. For example, Jia and Issa [32] proposed a synthesised methodology for taxonomy development in the domain of contractual semantics to support the development of an ontology model. Similarly, Wang *et al*. [59] used ontology technology to structure knowledge, such as activities, job steps, and hazards, to form a Job Hazard Analysis (JHA) database, and then developed the ontological reasoning mechanism to determine safety rules. The studies, as mentioned earlier, demonstrate the potential of ontology technology in supporting risk management, primarily as it can be used to raise the level of safety awareness. By organising knowledge as a logical semantic expression, it can be shared using ontology technologies and therefore enable semantic interoperability. As a result, the structured and unified knowledge in the ontology can be understood and readily operated by different parties and computer applications and thus ensure the re-use and promotion of knowledge. To the best of our knowledge, however, there has been no research that has integrated computer vision with ontology to identify hazards on construction sites.

3.0 Knowledge Graph Framework for Hazard Identification

In Figure 1, we present the workflow for implementing our proposed knowledge graph framework, which comprises three steps:

1. *Ontology modelling:* Engineering documents, historical accident reports, experts' experience, and safety codes are used to create a hazard taxonomy is constructed, which contains both the specialisation and relations between entities.
2. *Knowledge extraction:* Computer vision approaches are used to automatically detect a set of entities and attributes, using the data derived from step one. In doing so, object types and their attributes (i.e., geometric, coordinates in images) are identified so that they can be stored in Neo4j for reasoning and querying. After identifying objects and their attributes, an intersection over union (IoU) is used to extract the spatial relationships between objects (i.e., within, away, or overlap) by using geometric and spatial features. Here, the relationships between objects for hazards are defined in step one using the hazard taxonomy that is established.

.

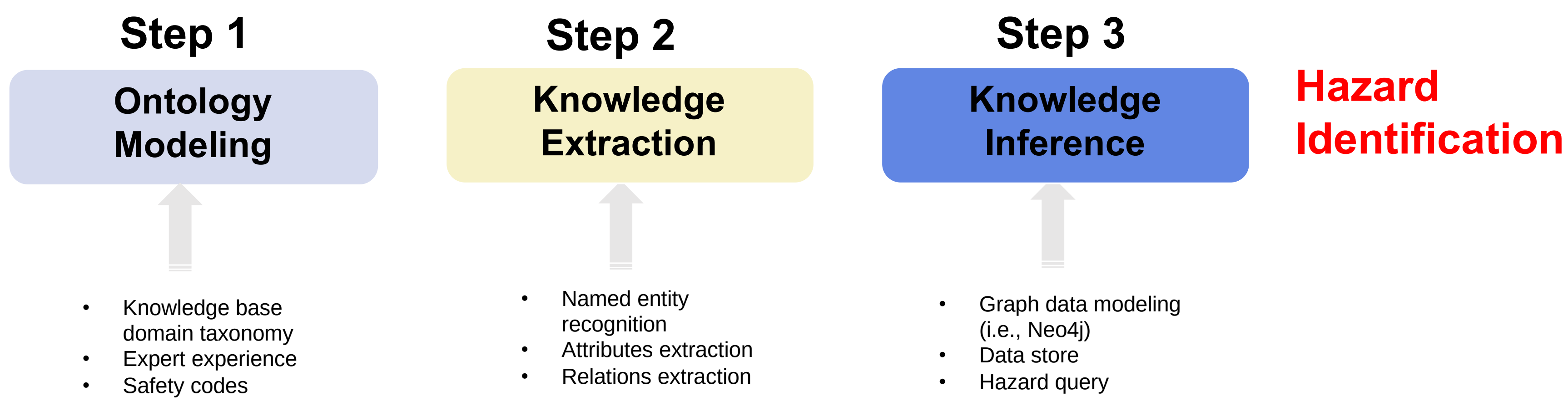


Figure 1. The workflow of the proposed hybrid semantic computer vision approach

3. *Knowledge inference*: A reasoning model for hazard identification was developed using the Neo4j database to create nodes, relationships, and their properties for modelling. The Neo4j database stores and records all types of objects, their attributes, and the relationship of objects, which were obtained from step two. Thus, hazards in the images are automatically identified by querying the created Neo4j database.

Each of these steps is examined in further detail below.

3.1 Ontology Modelling

The initial process for implementing our semantic computer vision-based hazard identification model was to develop an ontology of a construction site. The ontology was developed using the Graph Database Language instead of the traditional RDF mapping models. The Chinese code for 'Quality and Safety Inspection Guide of Urban Rail Transit Engineering,' for example, was selected as a point of reference to examine hazards that were incurred during the construction of a metro-rail project in Wuhan, China. In our ontological model, the information is categorised into seven classes: (1) thing; (2) part; (3) attribute; (4) time; (5) space; (6) event; and (7) attribute-value. Within the context of construction, a hazard can be defined by its given *time* and *space*, and *entities* (with specific attributes), which perform certain activities [12,14]. Thus, a hazard event consists of semantic information that specifies its:

1. *Entity*: The entities that are the objective existence. In this research, the entities are classified into four categories: (1) people; (2) equipment; (3) materials: and (4) environment. An example of taxonomy entities is presented in Figure 2.
2. *Activity*: A change that is caused by a hazard, such as its attributes, states, and relations, which contain static and dynamic subclasses. For example, "more than two workers standing in a basket". Here, "standing" represents the activity.

3. *Location*: Specific location and the interface with concepts, such as working "in height".
4. *Time*: The specific time involved with hazards, such as their duration on a timeline.
5. *Attribute*: Specific description of properties. For example, distance, colour, height, and speed.

Examples of the entities in the ontology model are shown in Figure 2.

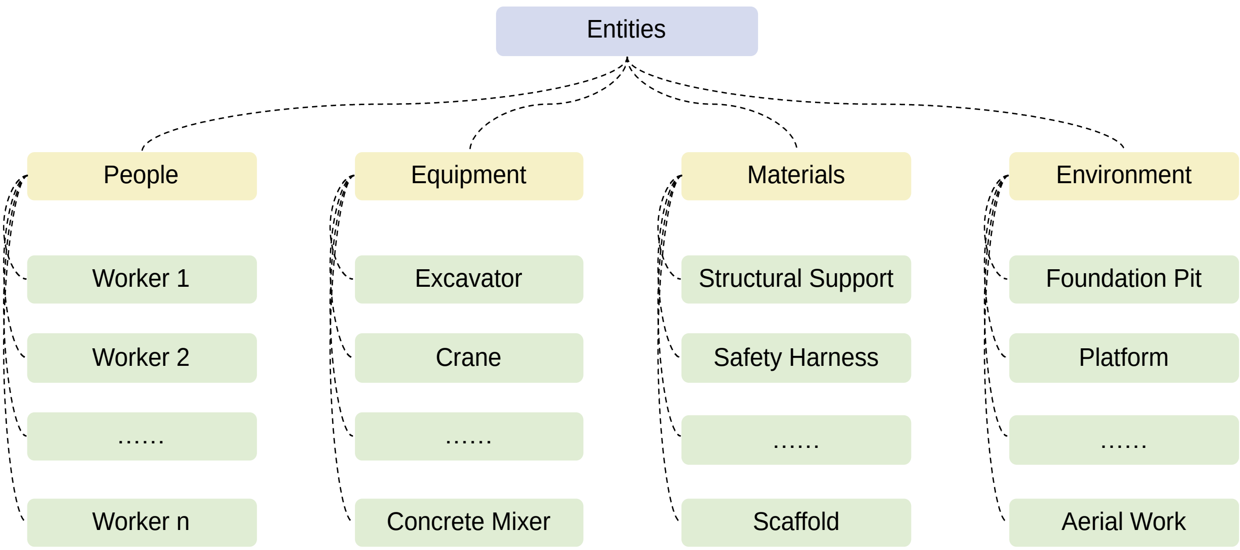


Figure 2. Examples of the entities in the ontology model

Figure 3 shows an example relationship – 'Spatial relationship' between entities. The relationship exists between people, between people and a safety helmet, and between people and machinery. The model will be able to answer the following queries:

- Who is behind 'John'
- Is there anyone who stands close to 'John' not wearing a safety helmet?
- Who is driving the excavator?
- Is there any worker stands outside of the excavator driver's view range?

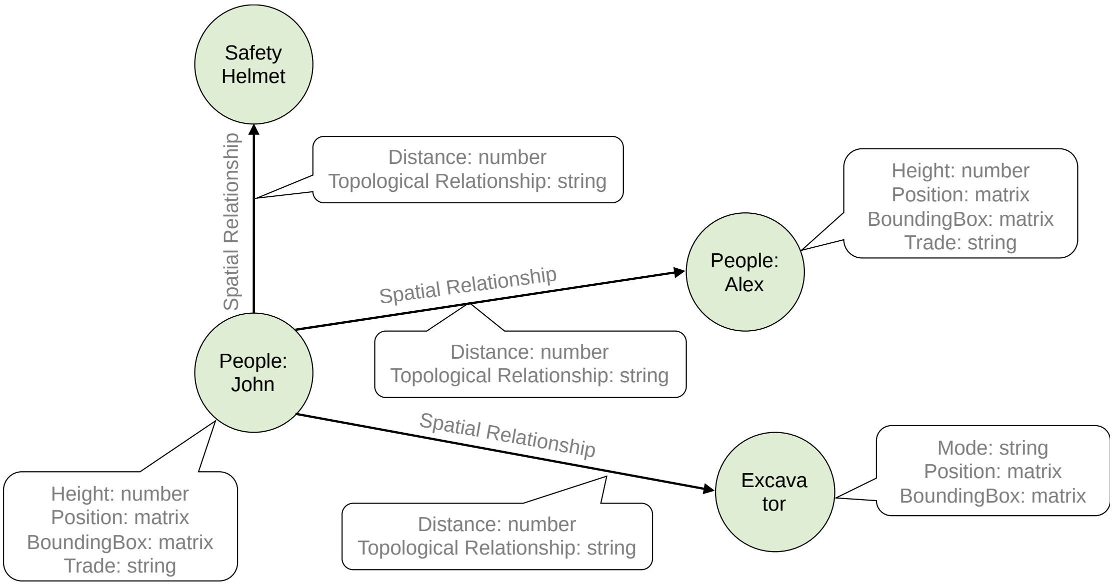


Figure 3. Examples of the entity relationships in the ontology model

3.2 Knowledge Extraction

Knowledge extraction is a vital step in the construction of a knowledge graph, which includes the detection of and the relationship between entities.

3.2.1 Computer Vision-based Entity Detection

The aim of our research is to develop a computer vision approach that can be used to identify and warn people about the likelihood of hazards. For example, if a person is entering an area where machinery is present, regardless if it is moving or static, our model, will identify the action as being 'unsafe'. Our research solely considers the extraction of attributes by using a computer vision approach, which was used to explore the development of a knowledge graph. To this end, we use computer vision to determine contextual information from a construction site by:

- *Entity Recognition*: As shown in Figure 2, entities can be divided into four types of objects: (1) people; (2) equipment; (3) materials; and (4) environment. In this research, two detection approaches are used: (1) object; and (2) scene recognition. Here, object detection is used to identify people, equipment (i.e., excavator), and materials (e.g., structural support). The scene recognition approach, one of the hallmark tasks of computer vision, enables us to define a context for given object recognition. The Mask R-CNN developed by He *et al*. [30] adopts a two-stage procedure whereby:

1. Images are taken as input for the ResNet network to obtain feature maps. Then candidates of object bounding boxes are obtained by using the Region Proposal Network (RPN); and
2. RoiAlign is used to preserve and extract spatial locations from each candidate box and perform classification, bounding box regression, and mask generation.

The Mask R-CNN has achieved higher levels of detection accuracy for objects than other approaches [30]. With this in mind, we adopted the Mask R-CNN in our research for entity (i.e., people, equipment) detection. We assume that this approach can be expanded to identify several types of objects (i.e., people, equipment, materials) in construction through a process of training. Specific details about the Mask R-CNN can be found in Fang *et al*. [22].

To understand and accurately recognise scenes (e.g., people working at a height), a Unified Perceptual Parsing approach (UPP) based on a feature pyramid network (FPN) is used to segment concepts from images effectively. The UPP approach was developed by Xiao *et al*. [60] and can infer and discover rich visual knowledge from images. The UPP performs better than prevailing state-of-the-art machine learning tools that can be used for segmentation (e.g., fully convolutional network (FCN), SegNet, and DilatedNet). A detailed description of the UPP can be found in Xiao *et al*. [60].

- *Attributes Extraction:* As our research focuses on identifying hazards based on distance and spatial features, as we only need to extract two types of attributes: (1) the coordinates of each object in the image; and (2) distance among objects detected by Mask R-CNN. We, therefore, utilised the transformation matrix [36] within our hybrid semantic computer vision model to compute distances between objects.`

3.2.2 Extraction of Spatial-Relationships from Images

After identifying the types of objects and their attributes, three spatial relationships between them can be computed: (1) within; (2) overlap; and (3) away. An example of a spatial relationship is presented in Figure 4. In this research, the choice of terminology and semantics for the spatial relationships is based on the distance between objects (i.e., between two geometries A and B) and rules specified by Chinese safety codes (Section 4.1).

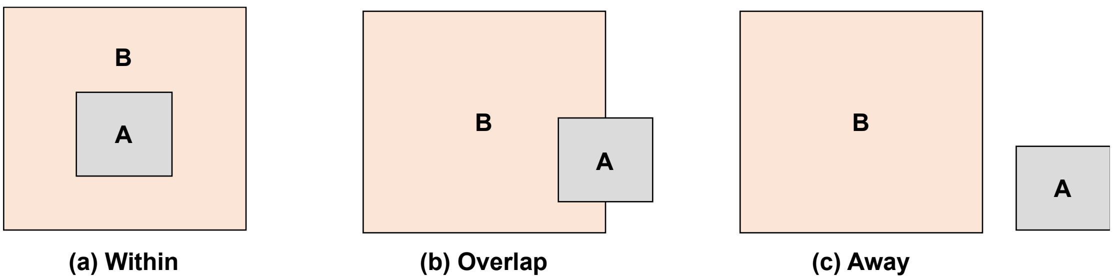
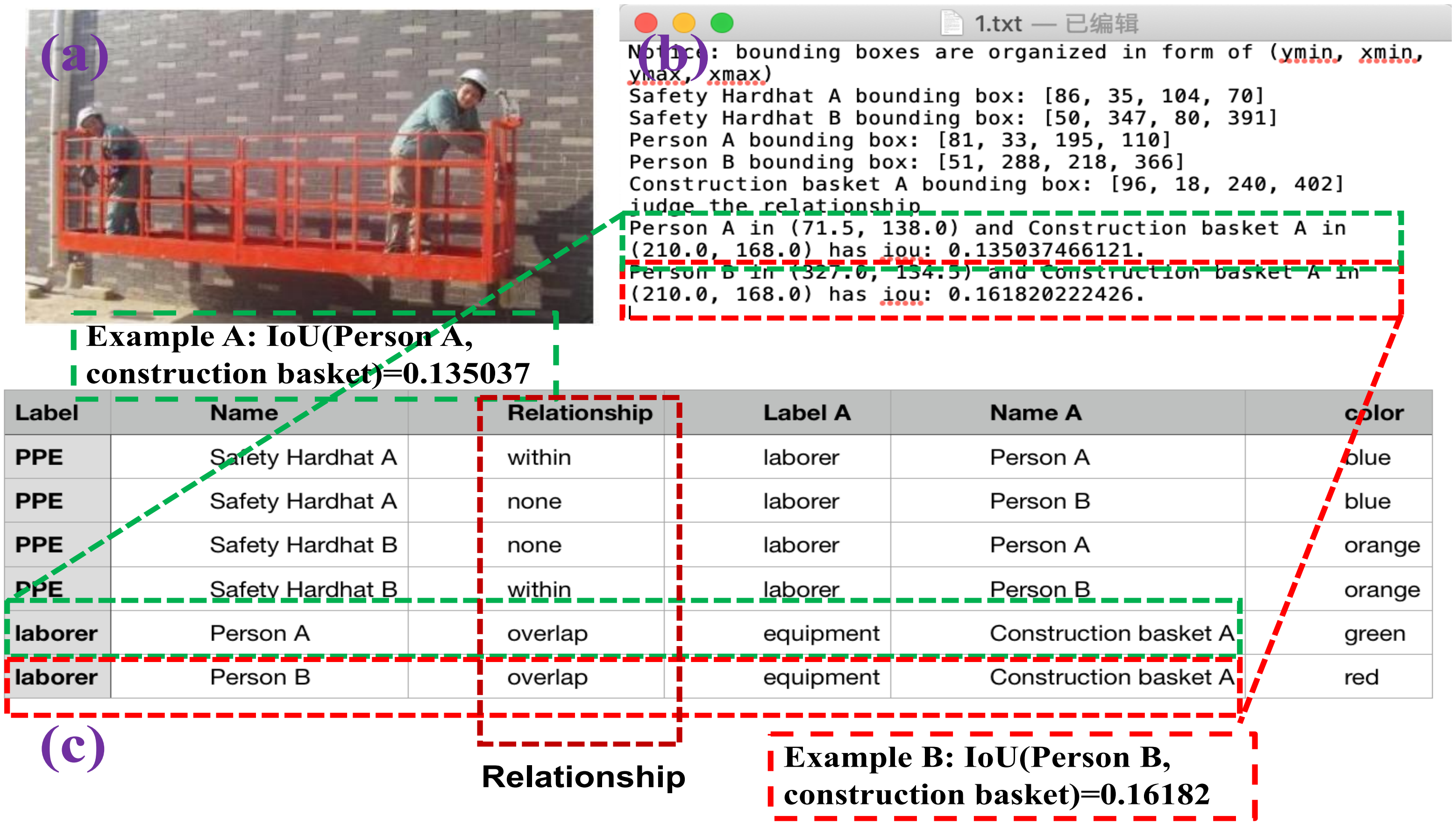


Figure 4. Examples of spatial relationship

The spatial relationship between object A and object B is defined as the IoU of the bounding box A and B, as shown in Eq. [1]:

$$IoU(A,B) = \frac{area(A) \cap area(B)}{\min\{area(A), area(B)\}} = \begin{cases} 1 & within \\ [0,1] & overlap \\ 0 & away \end{cases} \qquad \text{Eq. [1]}$$

For the conditions of within and overlap, we can use the IoU to identify the spatial relationships between objects. If the IoU of two objects is 0, we then compute the distance between them by using the transformation matrix approach (Section 4.2.2). Figure 5 presents an example of a spatial relationship using the IoU and where distance is extracted.



(a) Original image (b) Attributes extraction (i.e., IoU, coordinate) (c) Extraction of spatial relationship

Figure 5. Extraction of spatial relationship

3.3 Knowledge Inference for Hazard Identification with Graph Database

We use a graph database to present the knowledge needed to infer hazards in a highly accessible way. A graph structure is used to represent semantic queries with nodes, relationships and properties, and store data. Due to its ability to present data in a robust and scalable way, we use the Neo4j graph database management system so that queries with multiple relationships can be identified [13,34]. To automatically identify hazards, we perform the following tasks: (1) data modelling; and (2) automated reasoning and query.

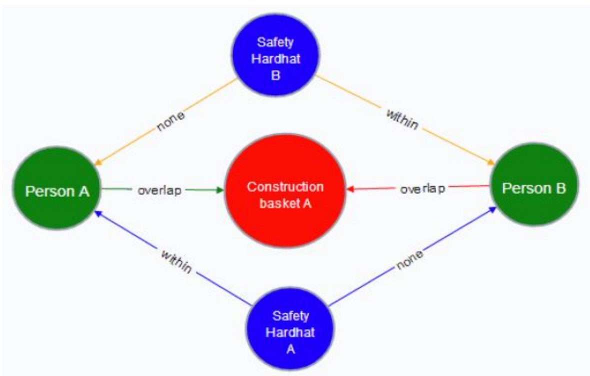
3.3.1 Data Modelling

The procedure to extract object classes and their spatial relationships have been described above. The outputs from these procedures are saved as a '.csv' file and loaded into the Neo4j database. The Neo4j database automatically processes the data and then provides an output. An example of the detection output is presented in Figure 6.

(a) Computer Vision detection

Label	Name	Relationship	Label A	Name A	Color
Material	Safety hardhat A	Within	people	Worker A	blue
Material	Safety hardhat A	none	people	Worker B	blue
Material	Safety hardhat B	none	people	Worker A	orange
Material	Safety hardhat B	within	people	Worker B	orange
Material	Worker A	overlap	equipment	Construction basket A	green
Material	Worker B	overlap	equipment	Construction basket A	red

(b) Output information from computer vision system

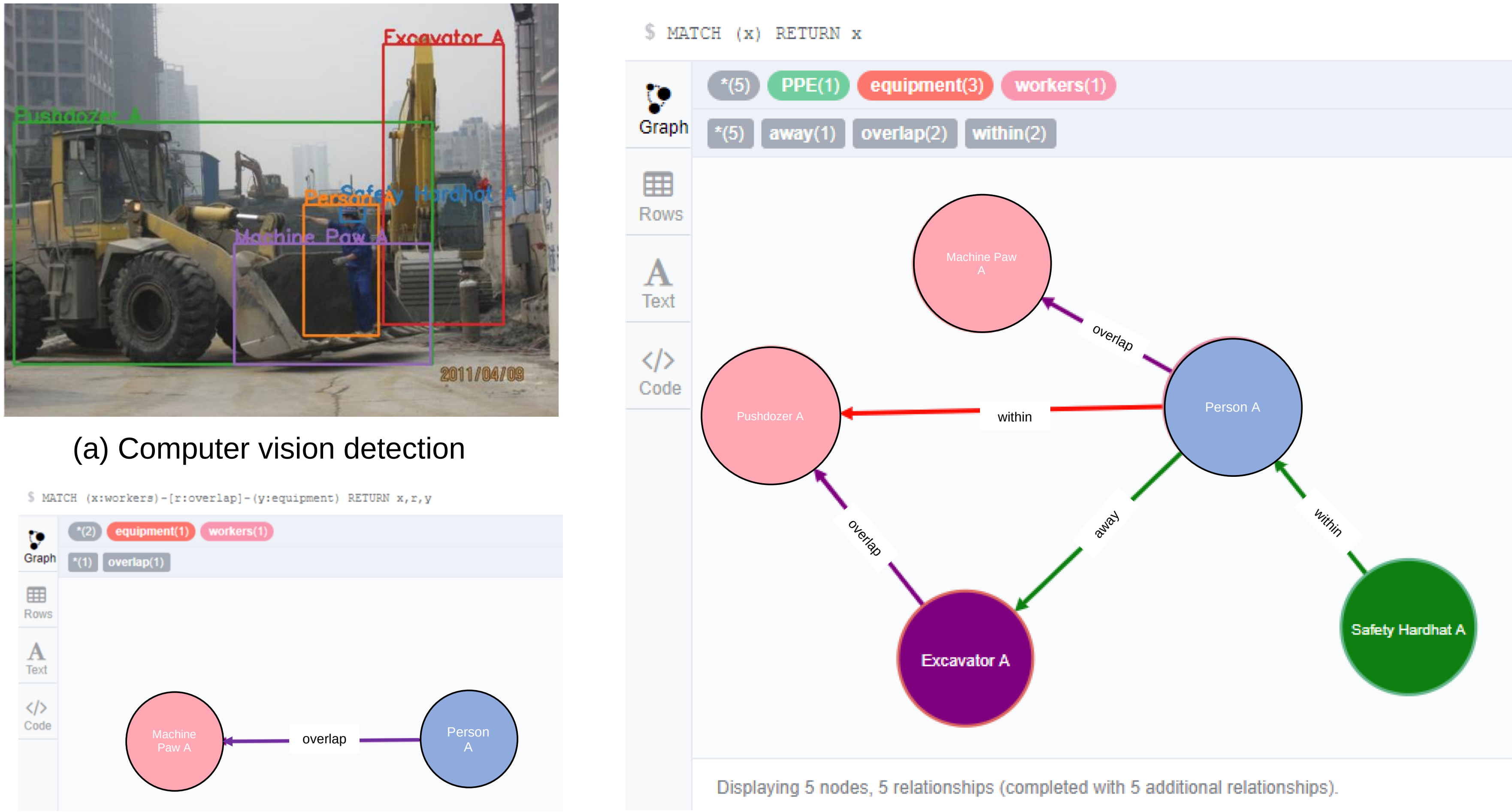


(c) Data modelling with Neo4j

Figure 6. An example of computer vision detection results and the output information

3.3.2 Automated Reasoning and Query

The final step of the modelling process is to identify hazards by querying the unsafe behaviour rules that had been defined in the model. The as-built graph database (Section 4.4.1) is constructed based on the objects and their spatial relationship; unsafe rules are derived from the safety codes, which were re-defined as queries. An unsafe behaviour, for example, occurs when "people stand on machinery when hoisting". Then, we can identify the unsafe behaviour by searching for the people (i.e. worker) "whose bounding box is within a machinery's bounding box". Figure 7 shows that an unsafe condition, in which a person is standing in a machine paw, is identified by using the rule: "MATCH (x: worker) – [r: overlap] – (y: equipment) RETURN x,r,y".



(a) Computer vision detection

(c) Hazard identification by reasoning

(b) Data modelling

Figure 7. The reasoning of unsafe conditions by querying in the graph database

4.0 Case Study

To demonstrate and test the validity of our developed semantic model, we can focus on identifying the unsafe condition that may lead to FFH (Table 2). We have selected an urban metro system under construction in Wuhan China to evaluate the effectiveness of detection for the developed semantic approach. Working in collaboration with a contractor who is involved with constructing the metro system in Wuhan (China) we were provided safety data from nearly 120 sites and images from a Web-based near-miss management system that had been installed on their sites. In sum, we had access to more than 3000 near-miss reports and over 40,000 related images (Figure 8).

The Web-based near-miss management system contains information about hazards, which includes their code, line, location, name, area, and description. We present an example of the hazard code in Figure 8: report number: No0000087; Lines: 2; Hazard name: adjacent edges and other protections do not meet requirements; hazard description: missing neighbour protection net. We individually examine FFH as they account for a high proportion (over 30%) of fatalities in construction [42,46]. By being able to detect of FFH hazards and mitigate their adverse consequences, we can make headway toward reducing safety incidents [41]. To validate our approach, we focus on identifying six types of unsafe behaviour that were selected from the near-miss accident reports (Table 2).

4.1 Development of Ontology for FFH

A taxonomy of hazards related to FFH was developed based on the checklist in Table 2. The core concepts identified are analysed and classified, which can be seen in Table 3 and serve as an extension to the taxonomy.

Table 2. Checklist of unsafe behaviour related to FFH

Number	Unsafe Behavior Description
1	There should be no more than two people in a lift's basket
2	People should not walk on the support of excavation if there has no guardrail
3	Edges of excavations (over 2m deep) should be protected with a guardrail
4	People should not stand on machinery when hoisting
5	People should wear a safety harness when working above a certain height
6	It is not allowed to use car hopper to pick up people

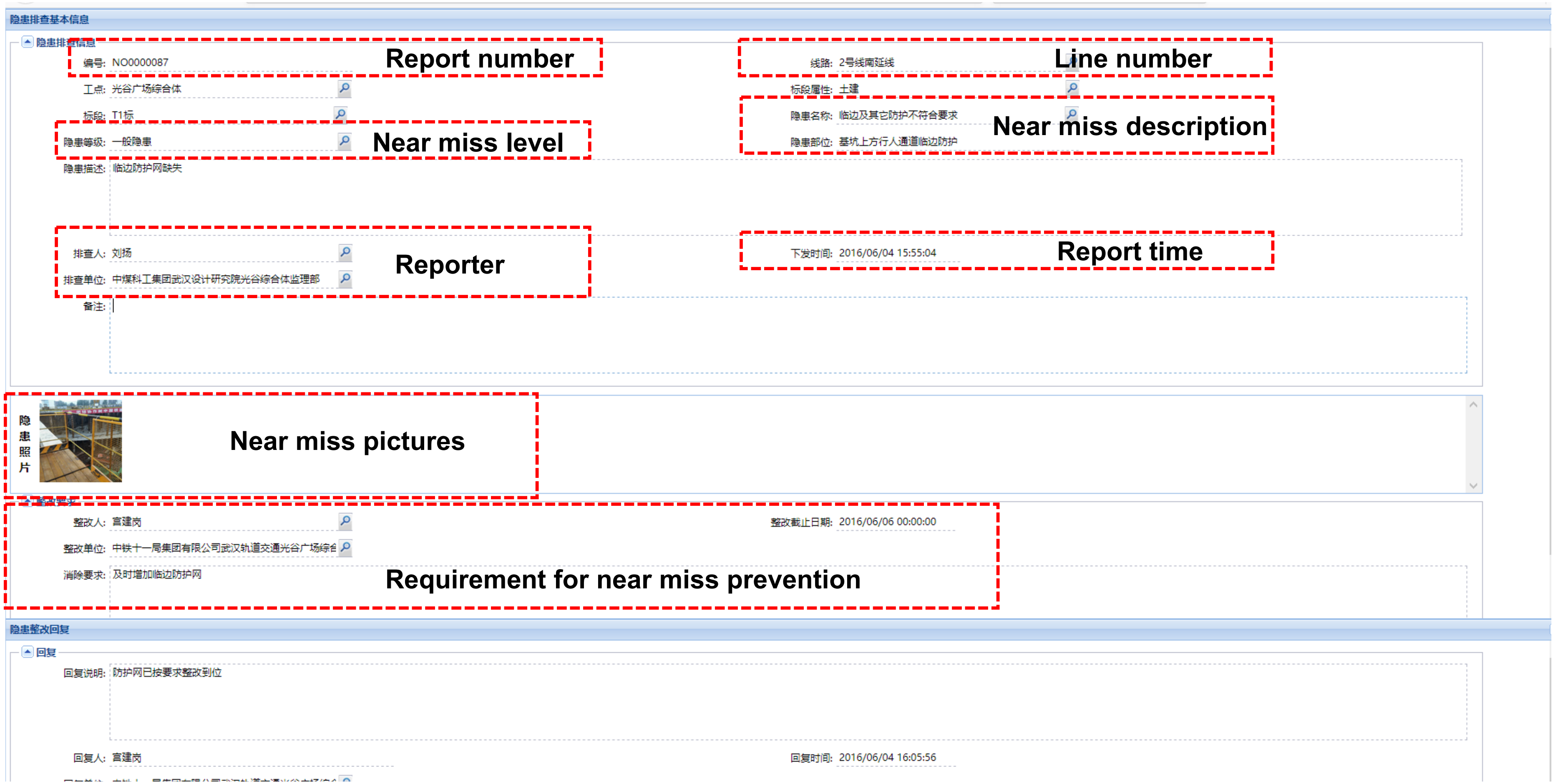


Figure 8. A web-based near-miss management system

Table 3. Concept identification of hazard information in FFH

Number	Images of hazards	Description of hazards	Hazard entity	Activity type	Location	Attribute	Relationship
1		There should be no more than two people in a lift's basket	People, lift basket	Stand		Number, coordinate	Overlapped/Within
2		People should not walk on the support of excavation if there has no guardrail	People, support, excavation, guardrail	Stand		coordinate	Touch/overlap

3		Edges of excavations (over 2m deep) should be protected with a guardrail	People, excavation, over 2m,	stand	Coordinate	Near/overlap
4		people should not stand on machinery when hoisting	People, machinery	Stand	Coordinate	Overlap/within

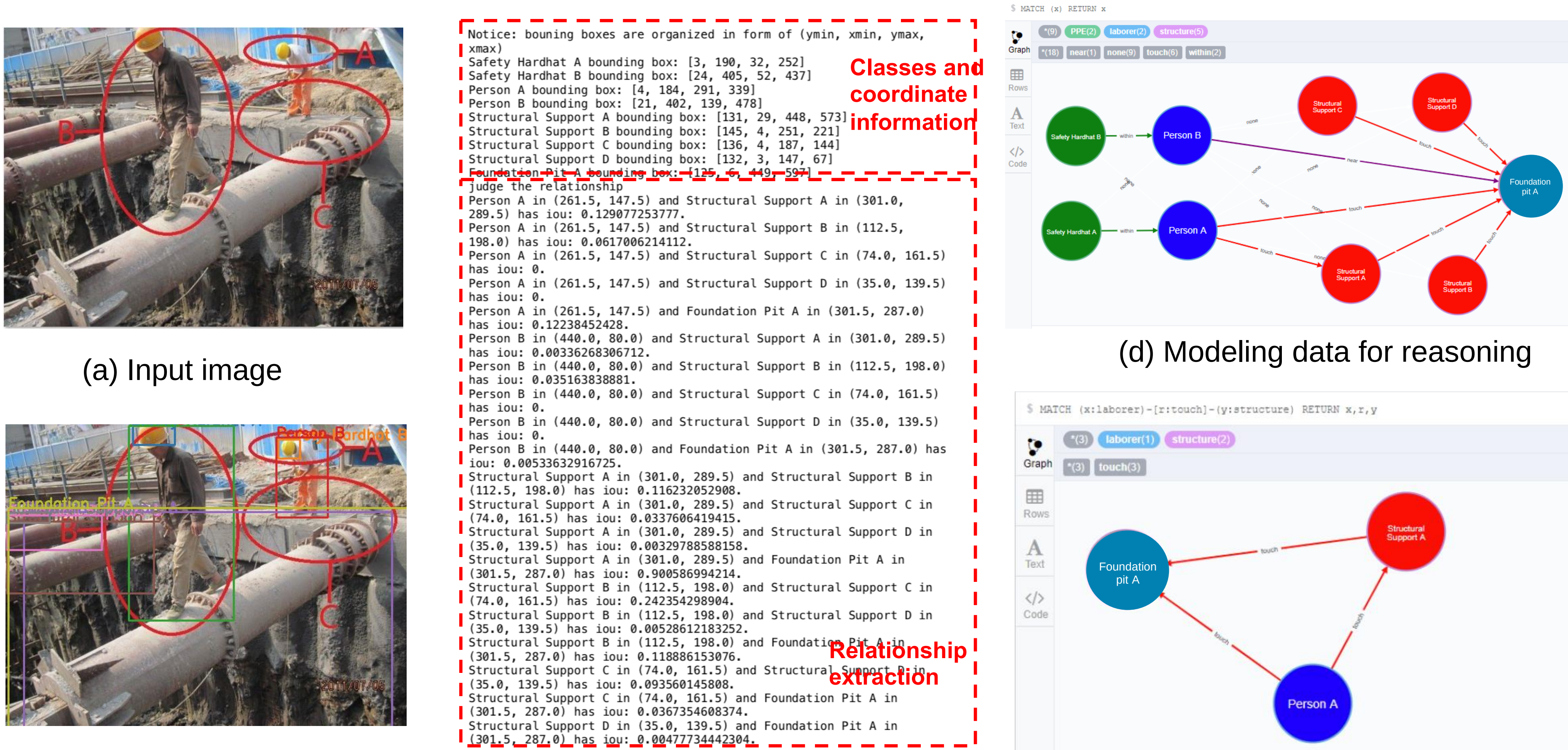
5		People should wear a safety harness when working above a certain height	People, safety harness	Wear	Working at heights	Coordinate	Overlap/within
6		There should not use car hopper to pick up people	People, car hopper	Pick-up		Coordinate	Within/overlap

4.2 Hazard Identification Results

We initially used computer vision to detect objects and their attributes with individuals, structural supports, and the foundation pit, as identified in Figure 8. The spatial relationships between objects are recognised using the IoU and determining the distance between them. As previously mentioned, the results are stored in the Neo4j database to identify unsafe conditions using rule the "MATCH (x: labourer)-[r: touch]-(y: structure) RETURN x,r,y" (Figure 9e).

The performance of our research results is based on two aspects: (1) entity detection; and (2) attributes detection. The precision and recall are selected as a critical evaluation metric for object detection. Our developed object detection approach is based on the previous work of Fang *et al*. (2019). Also, two key evaluation metrics are used for scene recognition: (1) pixel accuracy (PA); and (2) mean IoU (mIoU). The applied UPP achieved mIoU and PA of 41.22 and 79.98 on ADE20K dataset, respectively [60].

The performance of attributes detection relies on the extraction of coordinates and the computation of distance from images. Previous studies have demonstrated that the transformation matrix can be used for distance computation for objects [36-38]. Based on these performance metrics, our developed semantic computer vision approach achieves an acceptable level of accuracy for identifying unsafe behaviour.



(a) Input image

(d) Modeling data for reasoning

(b) Objects detection

(c) Attributes and relationships extraction

(e) Hazard identification

Figure 9. Semantic computer vision detection results

5.0 Discussion

To improve the efficiency and effectiveness of the safety inspection process and mitigate unsafe behaviour that occurs on construction sites, a semantic computer vision-based approach that integrates computer vision algorithms with ontologies was developed to identify hazards from images automatically. This approach provides site management with a mechanism to proactively identify, record, and analyse unsafe behaviours and therefore enable appropriate action to be undertaken to reduce and mitigate the likelihood of FFH. It can also be used for safety intervention by site management as a means to highlight potential hazards and the possible consequences that may materialise from peoples unsafe actions. If people are made aware that their actions are being monitored, then there will be a greater tendency for them to abide by safety rules.

In comparison with previous studies that have utilised computer vision to identify hazards, our study has the following advantages:

- We provide an integrated semantic model that can be used for training even when data is scarce. The unavailability of unsafe behaviour databases, especially for specific tasks, has hindered the development of deep learning applications in construction. Our approach not only relies on accurately detecting objects, but also the use of the spatial relationship between objects to reason hazards. Studies have demonstrated that prevailing computer-vision based approaches have achieved a satisfying performance to detect a variety of objects, which renders our semantic approach to be useful [20-22]. Thus, we have combined graph database to model data obtained from computer vision detection results to identify hazards, which makes our approach useable without a specific database for training; and

- The integrated approach is more generalizable than data training-based approaches due to its excellent performance (i.e., high accuracy on object detection in the cross-database) on object detection.

Our knowledge-based graph uses the output (e.g., the location of a person or a basket, computed by CV and machine learning as the input of the graph database (Neo4j)) to detect hazards. The knowledge graph can detect hazards which single computer-vision algorithms unable to do due to the complexity of the rules that need to be considered to define them. Improving the accuracy of computer vision algorithms and determining how to extract knowledge (i.e., entity detection) has not been the focus of our paper. Instead, we have built on the previous work of Fang *et al*. [22] who used deep learning to detect FFH by integrating a Mask R-CNN with ontology. As a result, there was no requirement to develop new algorithms. We acknowledge an array of robust vision-based algorithms are available, but undertaking a comparison between them, however, is outside the remit of this paper.

6.0 Limitation

Despite the novelty of the research presented, we need to acknowledge that it has several limitations. Firstly, our research relied on distance and coordinate information to extract spatial relationship for reasoning hazards. Many hazards comprise safety rules with specific features. For example, due to the presence of apanage management, persons on-site may be prohibited from entering a specific working area. In this case, computer vision cannot be used to extract the attributes and individuals and the area where they are performing their tasks. Our future work will need to integrate other technologies such as Radio Frequency Identification, to extract additional information to address this limitation, (e.g., identity).

Secondly, our research extracts the coordinates and the distance between objects from 2D images and then obtains spatial-relationship following the information obtained (i.e., coordinate, distance). Mistakes can be made when using the transformation matrix to compute the distance of objects from single images. Therefore, we suggest that future research will need to use stereo cameras to collect data and compute depth information to improve the accuracy of calculating spatial relationships.

Thirdly, our research solely considers the attribute (i.e., the distance between entities) in an as-built ontological model to determine whether hazards with complex rules are identifiable. A hazard is determined by combinations of semantic information (i.e., activity, time, and location). For example, an individual is not allowed to approach the working area of a piece of machinery. In this case, we should detect the machinery's working status (static or moving). We suggest that our approach can be expanded with consideration of other semantic information according to the as-built ontological model.

Fourthly we should acknowledge there have been a limited number of examples that have been able to integrate computer vision with ontology to identify hazards as data is scarce. Thus, our future research will focus on creating a database with a significant number of images in order to validate further and improve the reliability of our proposed approach.

Finally, we have also assumed that Mask R-CNN can accurately detect a variety of objects. However, if an object is occluded or there are unavailable images in the database for training, then the error rate for object detection may be high. We, therefore, intend to integrate ontology with the object's features to identify them in the future. For example, if an object partly occludes an individual, we may infer their presence using other features, such as shape, size, colour, and clothes.

7.0 Conclusion

We have introduced a novel semantic model that integrates computer vision and ontology to identify hazards from images automatically. We utilised the following tools to develop our model: (1) computer vision algorithms, which were used to extract implied knowledge from images (i.e., objects detection and attributes extraction); and (2) ontological reasoning to identify unsafe conditions based on their identified distance and spatial information. To validate our approach, we created a database of individuals unsafe behaviour related to FFH from several construction sites. We reveal that our semantic model can accurately recognise hazards from images with complex rules. We also suggest that our proposed semantic model can be used by site management to automatically identify potential hazards and therefore put in place strategies to mitigate potential injuries and accidents.

Our future research will focus on (1) combining temporal and spatial information to identify hazards from video streaming; (2) using stereo a camera to collect data, and then compute the 3D depth information from stereo videos; (3) combining other information techniques and computer vision to extract additional features, such as the size of the foundation, and colour of a hardhat, to identify additional hazard types; and (4) expanding our approach to integrate semantic information in accordance to our as-built ontological model.

Acknowledgements: The authors would like to acknowledge the financial support provided by the National Natural Science Foundation of China (Grant No. 51978302, No.71732001, No.51678265, No.51878311) and the China Scholarship Council (CSC). The authors would also like to thank the three anonymous reviewers for their constructive and insightful comments that have helped improve this manuscript.

References

[1] C.J. Anumba, R.R.A. Issa, J. Pan, I. Mutis, Ontology-based information and knowledge management in construction, Construction Innovation, 8(2008), pp. 218-239. https://doi.org/10.5220/0003053903310334.

[2] E.R. Azar, B. McCabe, Part based model and spatial-temporal reasoning to recognise hydraulic excavators in construction images and videos, Automation in Construction, 24(2012), pp.194-202. https://doi.org/10.1016/j.autcon.2012.03.003.

[3] E.R. Azar, B. McCabe, Automated visual recognition of dump trucks in construction videos, ASCE Journal of Computing in Civil Engineering, 26(2012), pp.769-781. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000179.

[4] CDC. (2012). "*Occupational Injury and Illness Classification Manual". US Department of Labor*.

[5] N. Chi, K. Lin, S. Hsieh, Using ontology-based text classification to assist Job Hazard Analysis, Advanced Engineering Informatics, 28(2014), pp. 381-394. https://doi.org/10.1016/j.aei.2014.05.001.

[6] S. Chi, C.H. Caldas, Automated Object Identification Using Optical Video Cameras on Construction Sites, Computer-Aided Civil and Infrastructure Engineering, *26*(2011), pp. 368-380. https://doi.org/10.1111/j.1467-8667.2010.00690.x.

[7] E. Corry, P. Pauwels, S.S. Hu, M. Keane, J. O'Donnell, A performance assessment ontology for the environmental and energy management of buildings", Automation in Construction, 57(2015), pp. 249-259. http://dx.doi.org/10.1016/j.autcon.2015.05.002

[8] Dalal, N., and Triggs, B. (2005), Histograms of Oriented Gradients for Human Detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, June 20-25. pp. 886-893. https://doi.org/10.1109/cvpr.2005.177.

[9] Deng, J., Dong, W., Socher, R., Li, LJ, Li, FF, (2009, June). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 248-255. https://doi.org/10.1109/CVPR.2009.5206848.

[10] L. Ding, W. Fang, H. Luo, P.E.D. Love, B. Zhong, X. Ouyang, X, A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory, Automation in Construction, 86(2018), pp. 118-124. https://doi.org/10.1016/j.autcon.2017.11.002.

[11] L. Ding, B. Zhong, S. Wu, H. Luo, Construction risk knowledge management in BIM using ontology and semantic web technology, Safety Science, 87(2016), pp. 202-213. https://doi.org/10.1016/j.ssci.2016.04.008.

[12] Y.M. Goh, D.K.H. Chua, Case-Based Reasoning Approach to Construction Safety Hazard Identification: Adaptation and Utilisation, ASCE Journal of Construction Engineering and Management, 136(2)(2010), pp. 170-178. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000116.

[13] J. Guia, v.G. Soares, S. Bernardina, Graph Database: Neo4j Analysis, in: Proc. of the 19th International Conference on Enterprise Information System, Porta, Portugal, 2017, pp. 351-356. https://doi.org/10.5220/0006356003510356.

[14] BHW. Guo, Y.M. Goh, Ontology for the design of active fall protection systems, Automation in Construction, 82(2017), pp. 138-153. https://doi.org/10.1016/j.autcon.2017.02.009.

[15] Gruber, T, A translation approach to portable ontologies, Knowledge Acquisition, 5(2)(1993), pp. 199-220. https://tomgruber.org/writing/ontolingua-kaj-1993.pdf.

[16] Edwards, H. (2017). "Fatal injuries arising from accidents at work in Great Britain 2017 to 2018." http://www.hse.gov.uk/statistics/fatals.htm.

[17] T.E. El-Diraby, Validating ontologies in informatics systems: Approaches and lessons learned for AEC, *Journal of Information Technology in Construction, 19*(2014), pp. 474-493.

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.663.7511&rep=rep1&type=pdf.

[18] T. Elghamrawy, F. Boukamp, H.S. Kim, Ontology-Based, Semi-Automatic Framework for Storing and Retrieving On-Site Construction Problem Information — An RFID-Based Case Study, in: Construction Research Congress, Seattle, Washington, United States, 2009, April5-7, pp. 457-466. https://doi.org/10.1061/41020(339)47.

[19] S. Executive, Health and safety in construction, 2nd ed., London, 2015. ISBN: 9781317423980.

[20] W. Fang, L. Ding, H. Luo, P.E.D. Love,Falls from Heights: A Computer Vision-based Approach for Safety Harness Detection, Automation in Construction. 91(2018), pp. 53-61. https://doi.org/10.1016/j.autcon.2018.02.018.

[21] W. Fang, L. Ding, B. Zhong, P.E.D. Love, H. Luo, Automated Detection of Workers and Heavy Equipment on Construction Sites: A Convolutional Neural Network Approach, Advanced Engineering Informatics, 37(2018), pp. 139-149. https://doi.org/10.1016/j.aei.2018.05.003.

[22] W. Fang, B. Zhong, N. Zhao, P.E.D. Love, H. Luo, J. Xue, S. Xu, A deep learning-based approach for mitigating falls from a height with computer vision: Convolutional neural network, Advanced Engineering Informatics. 39(2019), pp. 170-177. https://doi.org/10.1016/j.aei.2018.12.005.

[23] W. Fang, L.Ding, P.E.D. Love, H. Luo, H. Li, F. Peña-Mora, B. Zhong, C. Zhou, Computer vision applications in construction safety assurance, Automation in Construction, 110(2020), pp. 103013. https://doi.org/10.1016/j.autcon.2019.103013.

[24] W. Fang, P.E.D. Love, H. Luo, L. Ding, Computer vision for behavior-based safety in construction: A review and future directions, Advanced Engineering Informatics. 43(2020), pp. 100980. https://doi.org/10.1016/j.aei.2019.100980.

[25] M. Golparvar-Fard, A. Heydarian, J.C. Niebles, Vision-based action recognition of earthmoving equipment using spatiotemporal features and support vector machine classifiers, Advanced Engineering Informatics, *27*(2013), pp. 652-663. https://doi.org/10.1016/j.aei.2013.09.001.

[26] J. Gong, C.H. Caldas, Computer vision-based video interpretation model for automated productivity analysis of construction operations, ASCE Journal of Computing in Civil Engineering, *24*(2009), pp. 252-263. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000027.

[27] BHW. Hadikusumo, S. Rowlinson, Integration of virtually real construction model and design-for-safety-process database, Automation in Construction, 11(5)(2002), pp. 501-509. https://doi.org/10.1016/S0926-5805(01)00061-9.

[28] PD Haghighi, F. Burstein, A. Zaslavsky, P. Arbon, P, Development and evaluation of ontology for intelligent decision support in medical emergency management for mass gatherings, Decision Support Systems, 54(2013), pp. 1192-1204. https://doi.org/10.1016/j.dss.2012.11.013.

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

[30] K. He, G. Gkioxari, P. Dollar, R. Girshick, R, Mask R-CNN, (2017) in: IEEE International Conference on Computer Vision. Venice, Italy, 2017, pp.22-29 october. https://doi.org/10.1109/ICCV.2017.322.

[31] JL Hippolyte, Y.Rezgui, H. Li, B. Jayan, S. Howell, Ontology-driven development of web services to support district energy applications, Automation in Construction, 86(2018), pp. 210-225. http://dx.doi.org/10.1016/j.autcon.2017.10.004.

[32] N. Jia, R.R.A. Issa, Developing taxonomy for the domain ontology of construction contractual semantics: A case study on the AIA A201 document, Advanced

Engineering Informatics, 29(2015), pp. 472-482. https://doi.org/10.1016/j.aei.2015.03.009.

[33] S. Ji, W. Xu, M. Yang, K. Yu, 3D convolutional neural networks for human action recognition, IEEE Transactions on pattern analysis and machine intelligence. 35(1)(2013), pp. 221-231. https://doi.org/10.1109/TPAMI.2012.59.

[34] CI. Johnpaul, T. Mathew, A cypher query based NoSQL data mining on protein databases using Neo4j graph database, in 4th International Conference on Advanced Computing and Communication Systems. Coimbatore, India, 2017, Janurary 6-7. https://doi.org/10.1109/ICACCS.2017.8014558.

[35] H. Kim, H., Kim, 3D reconstruction of a concrete mixer truck for training object detectors, Automation in Construction, 88(2018), pp. 23-30. https://doi.org/10.1016/j.autcon.2017.12.034.

[36] K. Kim, K. Kim, K. Kim, Image-based construction hazard avoidance system using augmented reality in a wearable device, Automation in construction, 83(2017), pp. 390-403. https://doi.org/10.1016/j.autcon.2017.06.014.

[37] H. Kim, K. Kim, K. Kim, Vision-based object-centric safety assessment using fuzzy inference: Monitoring struck-by accidents with moving objects, Journal of Computing in Civil Engineering, 30(2016), pp. 04015075. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000562.

[38] D. Kim, M. Liu, S. Lee, V.R. Kamat, Remote proximity monitoring between mobile construction resources using camera-mounted UAVs, Automation in Construction. 99(2019), pp. 168-182. https://doi.org/10.1016/j.autcon.2018.12.014.

[39] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature, 521(2015)436-444. https://xs.scihub.ltd/https://doi.org/10.1038/nature14539.

[40] S.K. Lee, K.R. Kim, J.H. Yu, BIM and ontology-based approach for building cost estimation, Automation in Construction, 41(2014), pp. 94-105. http://dx.doi.org/10.1016/j.autcon.2013.10.020.

[41] P.E.D. Love, J. Smith, P. Teo, Putting into practice error management theory: Unlearning and learning to manage action errors in construction, Applied Ergonomics, 69(2018), pp. 104-114. doi.org/10.1016/j.apergo.2018.01.007

[42] 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), pp. 1273-1288. https://doi.org/10.1080/00140139.2019.1644379.

[43] Z. Lin, J. Jin, H. Talbot, Unseeded region growing for 3D image segmentation, Proc. VIP 2000, CRPIT, Sydney, 2000, pp. 31–37. https://dl.acm.org/doi/10.5555/563752.563757.

[44] H. Lingard, Occupational health and safety in the construction industry, Construction Management and Economics, 31(2013), pp. 505-514. https://doi.org/10.1080/01446193.2013.816435.

[45] D.G. Lowe, Distinctive image features from scale-Invariant keypoints, International Journal of Computer Vision, 60(2004), pp. 91-110. https://doi.org/10.1023/b:visi.0000029664.99615.94.

[46] E.A. Nadhim, C. Hon, B. Xia, I. Stewart, Falls from Height in the Construction Industry: A Critical Review of the Scientific Literature, International Journal of Environmental Research and Public Health. 13 (7)(2016), pp. 638. https://doi.org/10.3390/ijerph13070638.

[47] L. Ma, H. Luo, H. Chen, Safety risk analysis based on a geotechnical instrumentation data warehouse in metro tunnel project, Automation in Construction, 34(2013), pp. 75-84. https://doi.org/10.1016/j.autcon.2012.10.009.

[48] L. Ma, R. Sacks, U. Kattel, Building Model Object Classification for Semantic Enrichment Using Geometric Features and Pairwise Spatial Relationships, *Lean and Computing in Construction Congress - Joint Conference on Computing in Construction*. 2017, pp. 373-380. https://doi.org/10.24928/jc3-2017/0044.

[49] L. Ma, R. Sacks, U. Kattel, T. Bloch, 3D Object Classification Using Geometric Features and Pairwise Relationships, Computer-aided Civil and Infrastructure Engineering, 33(2)(2018), pp. 152-164. https://doi.org/10.1111/mice.12336.

[50] M. Martinez-Aires, M. Lopez-Alonso, M. Martinez-Rojas, Building information modeling and safety management: A systematic review, Safety Science, 101(2018), pp. 11-18. https://doi.org/10.1016/j.ssci.2017.08.015.

[51] M. Memarzadeh, A. Heydarian, M. Golparvarfard, J.C. Niebles, Real-Time and Automated Recognition and 2D Tracking of Construction Workers and Equipment from Site Video Streams, in: International Conference on Computing in Civil Engineering, Florida, United States, 2012, June 17-20, pp. 429-436,. https://doi.org/10.1061/9780784412343.0054.

[52] L. Nanni, S. Ghidoni, S. Brahnam, Handcrafted vs. non-handcrafted features for computer vision classification. Pattern Recognition. 71(2017), pp. 158-172. http://dx.doi.org/10.1016/j.patcog.2017.05.025.

[53] OSHA, Commonly Used Statistics, https://www.osha.gov/oshstats/commonstats. html2014 (accessed January 17, 2015)

[54] MW. Park, I. Brilakis, Continuous localisation of construction workers via integration of detection and tracking". Automation in construction, 72(2016), pp. 129-142. https://doi.org/10.1016/j.autcon.2016.08.039.

[55] MW. Park, I. Brilakis, Construction worker detection in video frames for initialising vision trackers, Automation in Construction, 28(2012), pp. 15-25. https://doi.org/10.1016/j.autcon.2012.06.001.

[56] K. Simonyan, A. Zisserman, Two-stream convolutional networks for action recognition in videos". Part of Advances in Neural Information Processing Systems (NIPS 2014). http://papers.nips.cc/paper/5353-two-stream-convolutional-networks-for-action-recognition-in-videos.pdf.

[57] H. Wang, C. Schmid, C, Action Recognition with Improved Trajectories, In: IEEE International Conference on Computer Vision, 2013, pp. 3551-3558. https://doi.org/10.1109/iccv.2013.441.

[58] H.H. Wang, F. Boukamp, T. Elghamrawy, Ontology-based approach to context representation and reasoning for managing context-sensitive construction information, Journal of Computing in Civil Engineering, 25(2011), pp. 331-346. https://doi.org/10.1061/(ASCE)cp.1943-5487.0000094.

[59] H.H. Wang, F. Boukamp, Ontology-Based Representation and Reasoning Framework for Supporting Job Hazard Analysis, Journal of Computing in Civil Engineering, 25(2011), pp. 442-456. https://doi.org/10.1061/(ASCE)cp.1943-5487.0000125.

[60] T. Xiao, Y.C. Liu, B.L. Zhou, Y.N. Jiang, J. Sun, Unified Perceptual Parsing for Scene Understanding, Lecture Notes in Computer Science, 2018, pp.432-448. https://doi.org/10.1007/978-3-030-01228-1_26.

[61] Y. Yu, H. Guo, Q. Ding, H. Li, M. Skitmore, An experimental study of real-time identification of construction workers' unsafe behaviors, Automation in Construction. 82(2017), pp. 193-206. https://doi.org/10.1016/j.autcon.2017.05.002.

[62] M. Zhang, T. Cao, X. Zhao, Applying sensor-based technology to improve construction safety management." Sensors, 17(8)(2017), pp. 1481. https://doi.org/10.3390/s17081841.

[63] Z. Zhou, Y.M. Goh, L. Shen, Overview and Analysis of Ontology Studies Supporting Development of the Construction Industry, Journal of Computing in Civil Engineering, 30(6)(2016), pp. 0406026. 10.1061/(ASCE)CP.1943-5487.0000594.