

Xiaochun LUO, Heng LI, SangHyun LEE

Bridging the gap: Neuro-Symbolic Computing for advanced AI applications in construction

© Higher Education Press 2023

Abstract Deep Learning (DL) has revolutionized the field of Artificial Intelligence (AI) in various domains such as computer vision (CV) and natural language processing. However, DL models have limitations including the need for large labeled datasets, lack of interpretability and explainability, potential bias and fairness issues, and limitations in common sense reasoning and contextual understanding. On the other side, DL has shown significant potential in construction for safety and quality inspection tasks using CV models. However, current CV approaches may lack spatial context and measurement capabilities, and struggle with complex safety and quality requirements. The integration of Neuro-Symbolic Computing (NSC), an emerging field that combines DL and symbolic reasoning, has been proposed as a potential solution to address these limitations. NSC has the potential to enable more robust, interpretable, and accurate AI systems in construction by harnessing the strengths of DL and symbolic reasoning. The combination of symbolism and connectionism in NSC can lead to more efficient data usage, improved generalization ability, and enhanced interpretability. Further research and experimentation are needed to effectively integrate NSC with large models and advance CV technologies for precise reporting of safety and quality inspection results in construction.

Keywords advanced AI in construction, safety and quality inspection, Neuro-Symbolic Computing, Deep Learning, computer vision

Received May 10, 2023; revised Jul. 18, 2023; accepted Jul. 25, 2023

Xiaochun LUO
Department of Construction Management, South China University of Technology, Guangzhou 510641, China

Heng LI
Department of Building and Real Estate, Hong Kong Polytechnic University, Hong Kong, China

SangHyun LEE (✉)
Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109-1316, USA
E-mail: shdpm@umich.edu

1 Introduction

Deep Learning (DL) has received great attention from academia and industry in the past decade and become the focus of Artificial Intelligence (AI) research (d'Avila Garcez and Lamb, 2023). DL has not only changed the research paradigm in the field of AI, but also directly affected many fields of computer science, including computer vision (CV) and natural language processing (LeCun et al., 2015). However, DL is still being criticized due to the following main weaknesses, including the dependency on large labeled datasets, lack of interpretability and explainability, potential bias and fairness issues, and limitations in common sense reasoning and contextual understanding. These limitations can impact the performance, robustness, interpretability, and fairness of DL models in certain situations.

DL applications in construction are still in the early stages of development, but they show significant potential in various areas. For example, DL-based CV models can analyze images and videos to assist with construction site monitoring (Luo et al., 2019; 2020), progress tracking (Lei et al., 2019; Braun et al., 2020), and quality control (Rodríguez-González et al., 2017; Yang et al., 2018). They can identify objects, detect safety hazards, and monitor construction activities in real-time, improving safety, efficiency, and quality. However, the current CV applications in construction may have limitations in accurately evaluating specific safety or quality requirements due to their reliance on basic object detection networks that may lack spatial context and measurement capabilities. Furthermore, these networks may struggle to accurately assess spatial relations, measure distances, and evaluate complex safety and quality requirements in dynamic construction environments.

• In terms of safety management, the previous research on unsafe behavior analysis mainly focuses on the topological relationship on the two-dimensional image plane, without considering three-dimensional spatial factors such as distances and angles, and it is difficult to support

the reasoning of quantitative safety rules. Vision-based construction safety risk monitoring is mainly aimed at the unsafe behavior of workers and construction machinery (Seo et al., 2015). Although logical reasoning technology has been introduced to improve the flexibility of DL methods for identifying human–machine risk behaviors (Zhong et al., 2020; Wu et al., 2021), the existing research mainly focuses on the topological relationship on the two-dimensional image plane, without considering the spatial distance and direction relationship.

- In terms of qualitative quality inspection, vision-based quality defect detection is widely investigated in the past decade. Quantitative quality inspection adopts point clouds as the main data form, which is organized by irregular points with characteristics such as inhomogeneity and disorder. The previous research focuses on scanning to building information modeling (BIM) and evaluates the results against quality rules (Wang, 2019). The various shapes formed by the abstraction of these points are usually difficult to distinguish, making it still challenging to segment fine-grained semantic objects from point clouds. Hard coded spatial rules have good efficiency and accuracy in specific situations, but it will lose generality and reduce the generalization ability of the technology.

The integration of symbolism and connectionism has become the main research trend in the realization of advanced AI in recent years (d'Avila Garcez and Lamb, 2023). Symbolism postulates physical symbol systems and the principle of bounded rationality, while the main principles of connectionism are the neural network and the connection mechanism and learning algorithm between neural networks. High-level cognitive tasks, such as abstraction, reasoning, and interpretation, are closely related to symbolic systems, but often cannot adapt to complex high-dimensional spaces. Neuro-Symbolic Computing (NSC) combines the strengths of DL models with symbolic methods, thereby significantly reducing the search space of symbolic methods (Mao et al., 2019).

The strengths of symbolic logic and neural networks can complement each other (Garnelo and Shanahan, 2019). First, the declarative properties of symbols enable symbolic representations to be reused in different tasks, thus improving data usage efficiency. Second, symbolic representations are usually abstract and thus help to improve the generalization ability of the model. Finally, symbolic representations use natural language-like propositional notations and are thus more understandable to humans, improving model interpretability. Kahneman et al. (2020) believed that combining neural networks with symbolic reasoning, that is, constructing neuro-symbolic AI, is the key to constructing rich AI systems (semantics are reasonable, explainable, and credible AI systems).

In this paper, we emphasize that current CV applications, based on basic object detection and semantic

segmentation networks, may struggle to accurately evaluate specific safety or quality requirements due to the lack of spatial context, measurement capabilities, and domain-specific knowledge. We propose the use of NSC as the methodology that integrates construction engineering knowledge with DL to address these limitations. We argue that NSC, which combines the strengths of DL and symbolic reasoning, has the potential to provide more robust, interpretable, and accurate AI systems for safety and quality inspection in construction. In the long run, the highly intelligent and autonomous safety and quality management could be achieved.

2 Computer vision for construction safety and quality management

This paper focuses on safety and quality management in construction. Therefore, the literature on these two aspects is selected. We trust that the cognitive statements about them can be generalized to other project management areas requiring accurate spatial rules.

2.1 Vision-based unsafe behavior detection

Vision-based construction safety risk monitoring is mainly carried out around unsafe behaviors of workers and construction machinery (Seo et al., 2015). Before DL was widely recognized and applied, visual surveillance research mainly used CV techniques based on handcrafted features (LeCun et al., 2015). Numerous studies (Chi et al., 2009; Brilakis et al., 2011) have focused on the task of identifying and tracking workers and machinery on construction sites for service productivity analysis (Gong and Caldas, 2010; Teizer, 2015; Yang et al., 2015) and safety monitoring (Teizer et al., 2010; Vahdatikhaki and Hammad, 2015). There are also studies (Chi and Caldas, 2011; 2012) using stereo vision cameras to obtain depth data in earthwork scenes, estimate the moving speed and mutual distance of workers and equipment, and then implement safety risk monitoring based on speed and distance.

Since DL outperformed manual feature-based machine learning techniques in image classification tasks in 2012 (Krizhevsky et al., 2012), it has gradually gained the attention of academics and the industry, and has achieved breakthroughs in object recognition (Girshick, 2015), object tracking (Girdhar et al., 2018), and activity analysis (Kim et al., 2018) tasks. It has greatly promoted its application in the field of engineering management. In terms of unsafe behavior analysis tasks, using convolutional neural networks (CNNs) (Ren et al., 2017) to identify whether workers wear hard hats is a common research pattern (Fang et al., 2018a; Wu et al., 2019; Li et al., 2020). Fang et al. (2018c) also utilized CNNs to identify

the use of safety ropes. Ding et al. (2018) combined CNNs with recurrent neural networks (RNNs) (Hochreiter and Schmidhuber, 1997) to identify unsafe behaviors associated with work ladder use by means of activity classification. In addition, Fang et al. (2018b) combines activity recognition technology (Luo et al., 2018) with face recognition technology (Zhang et al., 2016) to check whether special operation workers are licensed to operate. Research on unsafe behavior of machinery includes risk identification and early warning of strikes by lifting machinery. DL techniques are widely used in the study of unsafe behavior of workers and machinery, which further demonstrates their superior performance in low-level perception tasks (such as object recognition, semantic segmentation, and activity recognition), but methods using only DL techniques are only suitable for specific behavioral risk identification.

To solve the above-mentioned flexibility problem, logical reasoning technology is introduced into the analysis of unsafe behavior. Studies (Zhong et al., 2020; Wu et al., 2021) use semantic networks to model safety rules, use semantic segmentation techniques (He et al., 2017) to obtain visual information about workers' location, morphology, and their environment, and finally use safety rules grounded in visual information to make inferences to determine the risks of worker behavior. Fang et al. (2021) adopted another approach. After extracting the visual features of workers with DL methods, the visual features were matched with safety rule texts (Huang et al., 2020) to identify unsafe behaviors of workers in on-site photos. The advantage of the method matching natural language and visual features is that the data is easier to obtain, but it is not as strict as the description logic and therefore maintains the flexibility. Although logical reasoning technology has been introduced to improve the flexibility, the methods mainly focus on the topological relationship on the two-dimensional image plane, without considering the spatial distance and direction relationship.

In conclusion, the use of DL techniques in combination with CV methods has significantly advanced the field of construction safety risk monitoring, particularly in analyzing unsafe behaviors of workers and machinery. However, there are limitations in using only DL techniques for specific behavioral risk identification, and the introduction of logical reasoning technology has been explored to address the flexibility problem. Further research is needed to develop comprehensive and robust methods for construction safety risk monitoring.

2.2 Infrastructure 3D reconstruction for quality and safety checking

Infrastructure 3D reconstruction is usually carried out without a design model. The typical technical route is to apply or propose a spatial clustering algorithm to process

different views of a point cloud, obtain the position of a specific type of components according to the projection point density or clustering, and realize the detection and segmentation of a specific component, and then generate mesh surfaces or fit solid geometry based on the segmentation results, and finally carry out downstream quality and safety checking tasks on this basis.

Wang (2019) used point clouds for safety rule checking on scaffolding platforms. The study employed a clustering method on slices from the top view of the point cloud to determine the positions of the upright poles and identify the scaffold's location. Scaffolding platforms, including scaffolding boards and horizontal members, were then identified based on the histogram of the top view of the point cloud. Toe guards and guardrails were identified based on the side view. Finally, qualitative or quantitative rule metric values were extracted manually or through hardcoding and compared with rule requirements to assess the compliance of the scaffolding placement.

Similarly, Kim et al. (2022) involved fine-tuning the point cloud segmentation network RandLA-Net (Hu et al., 2020) using a transfer learning strategy for the segmentation of scaffold and background point clouds. The segmentation results were further processed using multi-view projection, with outliers cleaned up through spatial clustering. Different views were selected for projection processing based on the direction characteristics of different scaffold components. Thermal maps were used to identify the position, and the side view was used for the support and horizontal rod. Linear fitting was employed to determine the position, and the working platform was established based on the horizontal view after removing the vertical, support, and horizontal rods. However, it's worth noting that these methods rely on manual experience for view selection, component location, and segmentation, which may limit their generality and flexibility.

In summary, both studies utilize point clouds for safety rule checking on scaffolding platforms. Wang (2019)'s study employs clustering methods and manual extraction of rule metric values, while Kim et al. (2022)'s study involves fine-tuning a point cloud segmentation network and uses multi-view projection with manual component location. However, both methods may have limitations in terms of generality and flexibility due to reliance on manual experience for certain tasks.

3 Neuro-Symbolic Computing

According to the form and combination of neural architecture and symbolic architecture, NSC architecture can be divided into three categories (Garnelo and Shanahan, 2019). Based on these categories, we will discuss the potential of NSC to address some of the limitations of current CV applications in construction in Section 4.

3.1 Neural networks performing logical reasoning

The first class of architectures uses neural networks to perform logical reasoning. TensorLog (Cohen et al., 2017) first converts the first-order logic knowledge base into a factor graph called random logic program (Cussens, 2001), whose random variables correspond to the logical variables in the logic proof, and the factors correspond to the predicates in the knowledge base. Subsequently, constants in the knowledge base are converted into one-hot vectors, binary predicates are encoded into sparse matrices, unary predicates are encoded into sparse vectors, and logical queries are compiled into differentiable functions in the neural network structure. Ultimately, the logical reasoning process can be seen as a belief propagation process on a factor graph on a DL architecture, realizing the integration of logical reasoning and DL architecture. pLogicNet (Qu and Tang, 2019) combines the Markov logic network (Richardson and Domingos, 2006) with the knowledge graph embedding (Wang et al., 2017), uses the domain knowledge defined by the first-order predicate logic, constructs the joint distribution of knowledge triples based on a Markov logic network, and uses the expectation maximization algorithm to perform optimization. Among them, the parameter expectation estimation stage uses the knowledge graph embedding method to infer the missing triplets, and the likelihood maximization stage updates the logical rule weights based on the observed and predicted triplets.

3.2 Logical language defining neural networks

The goal of the second class of architectures is to use logical language to define neural networks. Lifting models (aka templated models) have received a lot of attention in the field of statistical relational learning in recent years, such models derived from predefined schemas (Kimmig et al., 2015). For example, a Markov logic network model can express logic such as “friends of smokers are probably also smokers”. The idea of defining models with patterns in the lift model was also developed in the framework of neuro-symbolic computation. For example, in Lifting Relational Neural Network (LRNN) (Sourek et al., 2018), the rules of first-order predicate logic are used to describe the structure of the question and construct the template of the neural network. Different neural networks are constructed for different query questions, and the weights of neurons in the neural network (corresponding to logic rules) are shared across networks. LRNN is a one-way reasoning, so it is necessary to build a separate neural network for each logical query problem with a new structure, which has the problem of high computational cost. Logical Neural Networks (LNNs) (Riegel et al., 2020) transform weighted real-valued logic programs into syntactic graph structures, and syntactic graphs further normalize neural network structures,

where each neuron represents an element (atomic concept, predicate, or connection word). The reasoning of LNN is omnidirectional and does not focus on predefined target variables. Therefore, building a network model for a knowledge base can be applied to different logical queries.

3.3 Integrating inductive learning and deductive reasoning

A third class of architectures is used to integrate inductive learning and deductive reasoning. This type of architecture can integrate inductive and deductive reasoning under a differentiable framework. Systems using this type of structures usually include two types of components: Neural components, which consist of one or more neural networks; and logical components, which provide a set of algorithms to complete logical reasoning mission (Badreddine et al., 2022). There are three common types of integration mechanisms.

- One is to introduce an additional layer in the neural network to encode logical constraints and modify the prediction results of the network through logical constraints. Representative works are Deep Logical Model (DLM) (Tran and d’Avila Garcez, 2018) and Knowledge Enhancement Neural Network (KENN) (Daniele and Serafini, 2019).

- The second is that differentiable logical reasoning is used to calculate the logical results predicted by the underlying neural network. The representative work is DeepProbLog (Manhaeve et al., 2021), whose differentiable logical reasoning is realized based on algebraic probabilistic logic aProbLog (Kimmig et al., 2011), which is the basis for guaranteeing the integration of logical reasoning and neural network computation.

- The third is to integrate logical knowledge as an additional constraint into the objective function or loss function of the neural network to train the neural network. A representative work is the Logical Tensor Network (LTN) (Badreddine et al., 2022), which defines concept grounding based on real-valued logic and logical domains, and maps them to real-field tensors or tensor-based operations, including functions and predicates. The reasoning process of LTN consists of three stages: 1) grounding stage, which uses logical query data to initialize the relevant logical axioms in the knowledge base; 2) feedforward stage, which calculates the true score of each logical axiom based on the given grounding results; and 3) aggregation stage, which aggregates the true score of the axioms, and calculates the satisfaction level of the data on the entire knowledge base.

4 NSC as an engineering methodology

NSC has the potential to address some of the limitations of current CV applications in construction. Here, we pick

out six critical limitations according to our knowledge.

(1) Incorporating spatial context

One of the limitations of current CV applications in construction is the difficulty in incorporating spatial context, such as understanding the relationships between different objects or elements on construction sites. NSC architectures, specifically those that use neural networks for logical reasoning, can leverage their ability to capture spatial relationships and contextual information through their layered representations and inference mechanisms. By incorporating symbolic reasoning capabilities, these architectures can better understand the spatial context and relationships among objects, enabling more accurate and context-aware CV applications in construction. For example:

- Using neural networks for logical reasoning to capture spatial relationships, such as understanding the relative positions of different construction elements, and reasoning about their interactions and dependencies.
- Employing CNNs in combination with symbolic reasoning mechanisms to capture local and global spatial context, enabling more accurate object detection, segmentation, and tracking on construction sites.

(2) Enhancing spatial measurement capabilities

Spatial measurement is an important task in construction, and accurate measurements are crucial for various applications, such as progress tracking and quality control. NSC architectures, particularly those that allow for logical language-defined neural networks, can provide a framework to define rules and constraints for spatial measurement tasks using logical language. This can enable more precise and consistent spatial measurements, improving the accuracy and reliability of CV applications in construction. For example:

- Defining logical rules in a symbolic language to specify precise measurement requirements, such as tolerances, alignments, and distances, and using these rules to guide the measurement process in CV applications for progress tracking and quality control.
- Incorporating geometric constraints and symbolic reasoning mechanisms into neural networks to improve the accuracy and consistency of spatial measurements, such as measuring the dimensions of building elements or checking the alignment of structural components.

(3) Handling complex safety and quality requirements

Construction sites are often subject to complex safety and quality requirements, and CV applications need to be able to understand and enforce these requirements. NSC architectures that integrate inductive learning and deductive reasoning can facilitate the integration of safety and quality rules into the learning and inference processes. This can help in better handling complex safety and quality requirements in CV applications, ensuring compliance with regulations and standards. For example:

- Integrating symbolic reasoning mechanisms into neural networks to enforce complex safety and quality

rules, such as checking for compliance with building codes, safety regulations, and quality standards, in CV applications for site monitoring, inspection, and compliance assessment.

- Using logical language-defined neural networks to specify safety and quality requirements in a symbolic form and leveraging the reasoning capabilities of these architectures to dynamically adapt the CV applications to complex safety and quality requirements on construction sites.

(4) Adapting to variable construction environments

Construction sites can vary greatly in terms of environmental conditions, lighting, weather, and other factors, which can affect the performance of CV applications. NSC architectures that use neural networks for logical reasoning can leverage their ability to adapt to variable inputs and conditions through their learning and inference mechanisms. This can enhance the robustness and adaptability of CV applications to different construction environments. For example:

- Employing NSC architectures that use RNNs or attention mechanisms to capture temporal dependencies and adapt to changing environmental conditions, such as lighting changes, weather variations, and seasonal effects, in CV applications for construction site monitoring, progress tracking, and change detection.
- Leveraging symbolic reasoning mechanisms in neural networks to handle variations in construction site layouts, equipment configurations, and material conditions, and adapt the CV applications accordingly, for example, in applications like site layout optimization or material tracking.

(5) Improving interpretability and explainability

Interpretability and explainability are important aspects of CV applications, especially in safety-critical domains like construction. NSC architectures that use logical language-defined neural networks or allow for symbolic reasoning can provide interpretable and explainable models. These architectures can use symbolic rules and constraints that can be easily understood and explained, facilitating the adoption of CV applications in construction. For example:

- Using symbolic reasoning mechanisms in neural networks to generate explanations or justifications for the decisions made by the CV applications, such as providing logical proofs or rule-based explanations for object detections, measurements, or quality assessments.

• Employing logical language-defined neural networks to generate interpretable and explainable models that can be easily understood and verified by construction stakeholders, such as project managers, engineers, and inspectors, facilitating the trust and adoption of CV applications in the industry.

(6) Handling large dataset problem

Construction projects generate large amounts of data, and handling such large datasets can be challenging for

CV applications. NSC architectures that use neural networks for logical reasoning or integrate inductive learning can leverage their ability to handle large datasets through their training and inference processes. This can enable more efficient and scalable CV applications in construction. For example:

- Leveraging the inductive learning capabilities of NSC architectures to handle large and diverse construction datasets for training CV applications, such as using symbolic rules and constraints to guide the learning process and improve the efficiency and scalability of training.
- Employing transfer learning techniques in combination with symbolic reasoning mechanisms to leverage knowledge learned from related domains or datasets, and adapt CV applications to construction-specific datasets with limited data availability, reducing the data requirements and overcoming the challenges of large dataset problem in construction CV applications.

5 NSC and large models

Apart from addressing issues related to spatial context, spatial measurement, safety and quality requirements, variable construction environments, interpretability, and handling large datasets, NSC can also tackle input and output-related challenges. The input-related challenge, known as abstract concept grounding (Du et al., 2022; Jain et al., 2022), involves mapping natural language specifications of safety and quality requirements to concrete measurements. On the other hand, the output-related challenge, referred to as professional inspection reporting, can also be addressed using NSC. By combining symbolic reasoning with large models, a comprehensive and reliable approach can be achieved for evaluating safety and quality requirements.

(1) Abstract concept grounding

Construction safety and quality requirements are specified in natural language. Large language models like Generative Pre-trained Transformer (GPT) series (Radford et al., 2018; Brown et al., 2020) and CV models like Detection Transformer (DETR) series (Zhu et al., 2020; Carion et al., 2020) can potentially be used in abstract requirement concept grounding for construction safety and quality requirements, but with certain limitations. Abstract requirement concept grounding involves mapping natural language specifications of safety and quality requirements to concrete measurements that can be used for evaluation or inspection purposes in construction.

Large language models can be utilized in this process by leveraging their language understanding and generation capabilities to interpret and generate text-based representations of safety and quality requirements. For example,

large models can be trained to analyze and understand natural language specifications of safety and quality requirements, extract relevant information, and generate structured representations or annotations that capture the abstract concepts involved. These structured representations can then be used for further processing, such as mapping to specific actions or measurements for safety and quality inspections.

However, there are limitations to using large models for abstract concept grounding in construction. One limitation is that the language understanding capabilities of large models may not always be accurate or comprehensive, especially when dealing with domain-specific jargon or technical terminology used in construction. Another limitation is that large models may lack the ability to fully understand the contextual nuances and domain-specific knowledge required for accurate grounding of abstract requirements to concrete actions or measurements. Additionally, large models may not have the necessary domain-specific expertise to capture the intricacies and complexities of safety and quality requirements in construction accurately.

(2) Professional inspection reporting

NSC can potentially be combined with large models to report safety and quality inspection results accurately and professionally in construction. The combination of symbolic reasoning and large models can provide a comprehensive and reliable approach for evaluating safety and quality requirements.

One possible approach is to use symbolic reasoning to define rules or knowledge representations that capture safety and quality requirements in the construction domain. These rules can be based on industry standards, regulations, guidelines, and best practices. The symbolic reasoning component can then analyze the output of the large models, which may include detected objects, scene features, or event recognition, and compare it with the predefined rules or knowledge representations to evaluate if the safety and quality requirements are met. The large models, such as deep neural networks, can be trained on a large dataset of construction images or videos to learn visual features, object detection, scene understanding, and other relevant tasks. The output of the large models can serve as input to the symbolic reasoning component, which can interpret the results in the context of safety and quality requirements and generate professional reports accordingly.

Additionally, NSC can also provide explainable AI capabilities for reporting safety and quality inspection results. The symbolic reasoning component can generate explanations and justifications for the decisions made by the large models, providing transparency and interpretability in the reporting process. This can help in understanding the reasons behind the inspection results and provide professional and reliable reports for safety and quality assessments in construction.

However, it is important to note that using NSC to combine engineering knowledge with large models for safety and quality inspection in construction may have challenges, such as the integration of different reasoning techniques, potential conflicts between symbolic and neural components, and the need for accurate representation and interpretation of safety and quality requirements. Careful consideration, experimentation, and research would be required to effectively combine NSC with large models for accurate and professional reporting of safety and quality inspection results.

6 Conclusions

NSC is an emerging field that combines the strengths of DL and symbolic reasoning (symbolic logic and knowledge representation) to enable more robust and interpretable AI systems. NSC has the potential to address the limitations of current CV applications in construction by combining DL with symbolic reasoning to enable more robust, interpretable, and accurate AI systems. Further research and development in this field could lead to advancements in CV technologies that better meet the specific safety and quality requirements of the construction industry.

Competing Interests The authors declare that they have no competing interests.

References

- Badreddine S, d'Avila Garcez A, Serafini L, Spranger M (2022). Logic tensor networks. *Artificial Intelligence*, 303: 103649
- Braun A, Tuttas S, Borrmann A, Stilla U (2020). Improving progress monitoring by fusing point clouds, semantic data and computer vision. *Automation in Construction*, 116: 103210
- Brilakis I, Park M W, Jog G (2011). Automated vision tracking of project related entities. *Advanced Engineering Informatics*, 25(4): 713–724
- Brown T B, Mann B, Ryder N, Subbiah M, Kaplan J D, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, et al. (2020). Language models are few-shot learners. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver, BC: Curran Associates Inc., 1877–1901
- Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S (2020). End-to-end object detection with transformers. In: *Proceedings of the 16th European Conference*. Glasgow: Springer-Verlag, 213–229
- Chi S, Caldas C H (2011). Automated object identification using optical video cameras on construction sites. *Computer-Aided Civil and Infrastructure Engineering*, 26(5): 368–380
- Chi S, Caldas C (2012). Image-based safety assessment: Automated spatial safety risk identification of earthmoving and surface mining activities. *Journal of Construction Engineering and Management*, 138(3): 341–351
- Chi S, Caldas C H, Kim D Y (2009). A methodology for object identification and tracking in construction based on spatial modeling and image matching techniques. *Computer-Aided Civil and Infrastructure Engineering*, 24(3): 199–211
- Cohen W W, Yang F, Mazaitis K R (2017). TensorLog: Deep learning meets probabilistic DBs. arXiv preprint, arXiv:1707.05390
- Cussens J (2001). Parameter estimation in stochastic logic programs. *Machine Learning*, 44(3): 245–271
- d'Avila Garcez A, Lamb L C (2023). Neurosymbolic AI: The 3rd wave. *Artificial Intelligence Review*, 56(11): 12387–12406
- Daniele A, Serafini L (2019). Knowledge enhanced neural networks. In: *Proceedings of the 16th Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence*. Cuvu, Yanuca Island: Springer Cham, 542–554
- Ding L, Fang W, Luo H, Love P E D, Zhong B, Ouyang X (2018). A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Automation in Construction*, 86: 118–124
- Du Y, Fu Z, Liu Q, Wang Y (2022). Visual grounding with transformers. In: *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*. Taipei: IEEE, 1–6
- Fang Q, Li H, Luo X, Ding L, Luo H, Rose T M, An W (2018a). Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in Construction*, 85: 1–9
- Fang Q, Li H, Luo X, Ding L, Rose T M, An W, Yu Y (2018b). A deep learning-based method for detecting non-certified work on construction sites. *Advanced Engineering Informatics*, 35: 56–68
- Fang W, Ding L, Luo H, Love P E D (2018c). Falls from heights: A computer vision-based approach for safety harness detection. *Automation in Construction*, 91: 53–61
- Fang W, Love P E D, Ding L, Xu S, Kong T, Li H (2021). Computer vision and deep learning to manage safety in construction: Matching images of unsafe behavior and semantic rules. *IEEE Transactions on Engineering Management*, 70(12): 4120–4132
- Garnelo M, Shanahan M (2019). Reconciling deep learning with symbolic artificial intelligence: Representing objects and relations. *Current Opinion in Behavioral Sciences*, 29: 17–23
- Girdhar R, Gkioxari G, Torresani L, Paluri M, Tran D (2018). Detect-and-track: Efficient pose estimation in videos. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT: IEEE, 350–359
- Girshick R (2015). Fast R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision*. Santiago: IEEE, 1440–1448
- Gong J, Caldas C H (2010). Computer vision-based video interpretation model for automated productivity analysis of construction operations. *Journal of Computing in Civil Engineering*, 24(3): 252–263
- He K, Gkioxari G, Dollár P, Girshick R (2017). Mask R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Venice: IEEE, 2980–2988
- Hochreiter S, Schmidhuber J (1997). Long short-term memory. *Neural Computation*, 9(8): 1735–1780
- Hu Q, Yang B, Xie L, Rosa S, Guo Y, Wang Z, Trigoni N, Markham A (2020). Randa-net: Efficient semantic segmentation of large-scale point clouds. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA:

- IEEE, 11105–11114
- Huang Y, Wu Q, Wang W, Wang L (2020). Image and sentence matching via semantic concepts and order learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(3): 636–650
- Jain A, Gkanatsios N, Mediratta I, Fragkiadaki K (2022). Bottom up top down detection transformers for language grounding in images and point clouds. In: *Proceedings of the 17th European Conference on Computer Vision*. Tel Aviv: Springer-Verlag, 417–433
- Kahneman D, Hinton G, Bengio Y, LeCun Y, Rossi F (2020). AAAI-20 fireside chat with Daniel Kahneman
- Kim J, Chung D, Kim Y, Kim H (2022). Deep learning-based 3D reconstruction of scaffolds using a robot dog. *Automation in Construction*, 134: 104092
- Kim P S, Lee D G, Lee S W (2018). Discriminative context learning with gated recurrent unit for group activity recognition. *Pattern Recognition*, 76: 149–161
- Kimmig A, Mihalkova L, Getoor L (2015). Lifted graphical models: A survey. *Machine Learning*, 99(1): 1–45
- Kimmig A, van den Broeck G, de Raedt L (2011). An algebraic prolog for reasoning about possible worlds. In: *Proceedings of the 25th AAAI Conference on Artificial Intelligence*. San Francisco, CA: AAAI Press, 209–214
- Krizhevsky A, Sutskever I, Hinton G E (2012). ImageNet classification with deep convolutional neural networks. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe, NV: Curran Associates Inc., 1097–1105
- LeCun Y, Bengio Y, Hinton G (2015). Deep learning. *Nature*, 521(7553): 436–444
- Lei L, Zhou Y, Luo H, Love P E D (2019). A CNN-based 3D patch registration approach for integrating sequential models in support of progress monitoring. *Advanced Engineering Informatics*, 41: 100923
- Li Y, Wei H, Han Z, Huang J, Wang W (2020). Deep learning-based safety helmet detection in engineering management based on convolutional neural networks. *Advances in Civil Engineering*, 9703560
- Luo X, Li H, Cao D, Dai F, Seo J, Lee S (2018). Recognizing diverse construction activities in site images via relevance networks of construction-related objects detected by convolutional neural networks. *Journal of Computing in Civil Engineering*, 32(3): 04018012
- Luo X, Li H, Yang X, Yu Y, Cao D (2019). Capturing and understanding workers' activities in far-field surveillance videos with deep action recognition and Bayesian nonparametric learning. *Computer-Aided Civil and Infrastructure Engineering*, 34(4): 333–351
- Luo X, Li H, Yu Y, Zhou C, Cao D (2020). Combining deep features and activity context to improve recognition of activities of workers in groups. *Computer-Aided Civil and Infrastructure Engineering*, 35(9): 965–978
- Manhaeve R, Dumančić S, Kimmig A, Demeester T, de Raedt L (2021). Neural probabilistic logic programming in DeepProbLog. *Artificial Intelligence*, 298: 103504
- Mao J, Gan C, Kohli P, Tenenbaum J B, Wu J (2019). The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. In: *Proceedings of the 7th International Conference on Learning Representations*. New Orleans, LA: OpenReview.net
- Qu M, Tang J (2019). Probabilistic logic neural networks for reasoning. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Vancouver, BC: Curran Associates Inc., 7712–7722
- Radford A, Narasimhan K, Salimans T, Sutskever I (2018). Improving language understanding by generative pre-training. *OpenAI*
- Ren S, He K, Girshick R, Sun J (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137–1149
- Richardson M, Domingos P (2006). Markov logic networks. *Machine Learning*, 62(1–2): 107–136
- Riegel R, Gray A, Luus F, Khan N, Makondo N, Yunus Akhalwaya I, Qian H, Fagin R, Barahona F, Sharma U, Ikbal S, Karanam H, Neelam S, Likhyan A, Srivastava S (2020). Logical neural networks. *arXiv preprint, arXiv:2006.13155*
- Rodríguez-González P, Rodríguez-Martín M, Ramos L F, González-Aguilera D (2017). 3D reconstruction methods and quality assessment for visual inspection of welds. *Automation in Construction*, 79: 49–58
- Seo J, Han S, Lee S, Kim H (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, 29(2): 239–251
- Sourek G, Aschenbrenner V, Zelezny F, Schockaert S, Kuzelka O (2018). Lifted relational neural networks: Efficient learning of latent relational structures. *Journal of Artificial Intelligence Research*, 62: 69–100
- Teizer J (2015). Status quo and open challenges in vision-based sensing and tracking of temporary resources on infrastructure construction sites. *Advanced Engineering Informatics*, 29(2): 225–238
- Teizer J, Allread B S, Fullerton C E, Hinze J (2010). Autonomous proactive real-time construction worker and equipment operator proximity safety alert system. *Automation in Construction*, 19(5): 630–640
- Tran S N, d'Avila Garcez A S (2018). Deep logic networks: Inserting and extracting knowledge from deep belief networks. *IEEE Transactions on Neural Networks and Learning Systems*, 29(2): 246–258
- Vahdatikhaki F, Hammad A (2015). Risk-based look-ahead workspace generation for earthwork equipment using near real-time simulation. *Automation in Construction*, 58: 207–220
- Wang Q (2019). Automatic checks from 3D point cloud data for safety regulation compliance for scaffold work platforms. *Automation in Construction*, 104: 38–51
- Wang Q, Mao Z, Wang B, Guo L (2017). Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12): 2724–2743
- Wu F, Jin G, Gao M, He Z, Yang Y (2019). Helmet detection based on improved YOLO v3 deep model. In: *Proceedings of the 16th International Conference on Networking, Sensing and Control (ICNSC)*. Banff, AB: IEEE, 363–368
- Wu H, Zhong B, Li H, Love P E D, Pan X, Zhao N (2021). Combining computer vision with semantic reasoning for on-site safety management in construction. *Journal of Building Engineering*, 42: 103036
- Yang J, Park M W, Vela P A, Golparvar-Fard M (2015). Construction performance monitoring via still images, time-lapse photos, and

- video streams: Now, tomorrow, and the future. *Advanced Engineering Informatics*, 29(2): 211–224
- Yang X, Li H, Yu Y, Luo X, Huang T, Yang X (2018). Automatic pixel-level crack detection and measurement using fully convolutional network. *Computer-Aided Civil and Infrastructure Engineering*, 33(12): 1090–1109
- Zhang K, Zhang Z, Li Z, Qiao Y (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10): 1499–1503
- Zhong B, Li H, Luo H, Zhou J, Fang W, Xing X (2020). Ontology-based semantic modeling of knowledge in construction: Classification and identification of hazards implied in images. *Journal of Construction Engineering and Management*, 146(4): 04020013
- Zhu X, Su W, Lu L, Li B, Wang X, Dai J (2020). Deformable DETR: Deformable transformers for end-to-end object detection. *arXiv preprint, arXiv:2010.04159*