

# 거푸집 부재 인식을 위한 인공지능 이미지 분할

## Artificial Intelligence Image Segmentation for Extracting Construction Formwork Elements

아이샤 무니라 초드리<sup>1)</sup>, 문성우<sup>2)</sup>

Chowdhury, Ayesha Munira<sup>1)</sup> · Moon, Sung-Woo<sup>2)</sup>

Received February 10, 2022; Received March 29, 2022 / Accepted March 30, 2022

**ABSTRACT:** Concrete formwork is a crucial component for any construction project. Artificial intelligence offers great potential to automate formwork design by offering various design options and under different criteria depending on the requirements. This study applied image segmentation in 2D formwork drawings to extract sheathing, strut and pipe support formwork elements. The proposed artificial intelligence model can recognize, classify, and extract formwork elements from 2D CAD drawing image and training and test results confirmed the model performed very well at formwork element recognition with average precision and recall better than 80%. Recognition systems for each formwork element can be implemented later to generate 3D BIM models.

**KEYWORDS:** Construction Formwork, Safety Management, Artificial Intelligence Model, Mask R-CNN, Image Segmentation

**키워드:** 거푸집, 안전관리, 인공지능모델, 마스크 R-CNN, 이미지 분할

### 1. Introduction

#### 1.1 Background

Formwork is a temporary structure erected in advance to help with important concrete construction such as slabs, columns crossbeams, longitudinal beams, and walls when constructing a permanent structure. The formwork comprises joists, struts, pipe supports to form decks or sheathing to create a mold during construction where wet concrete is placed and compacted. Planning and management formwork has considerable impact on the entire project and depending on the environment and affects formwork planning depending on site conditions (Lee et al., 2020).

Constructability, safety, and economic feasibility must be carefully considered when designing formwork. If a proper plan is not established in design, it can lead to poor

construction, and potentially to serious problems in terms of construction management such as accidents, structural failures and cost overruns.

Serious social and economic losses can occur should accidents occur during formwork construction; and can lead to significant project schedule delays. Since construction projects need to meet preset schedules, any incidents that lead to delays are unacceptable (Moon and Chowdhury 2021). However, often sufficient consideration is not given to formwork design due to the perception it is a temporary structure.

Three-dimensional (3D) building information management (BIM) modelling has gained popularity recently and has also been applied to formwork design and planning. However, 3D modelling requires considerable manual operation and specialized manpower (Horna et al., 2009; Sigalov and König,

<sup>1)</sup> 학생회원, 부산대학교 석사과정 (cuijinrui@pusan.ac.kr)

<sup>2)</sup> 정회원, 부산대학교 토목공학과 교수 (sngwmoon@pusan.ac.kr) (교신저자)

2017) either from 2D digital data, such as CAD, PDF or image file; or documents and specification for new or old buildings. Therefore, it is essential to introduce an artificial intelligence model to properly recognize forms in a construction drawing and assist the engineer/contractor to prepare 3D models.

Image segmentation classifies and localizes each objects within images, producing a segmentation mask for each object in the image (Ahn and Kwak, 2018). Some recent studies applied image segmentation to construction and architectural engineering drawings. Seo et al. (2020) used image segmentation with GOOGLE DeepLabV3+ to automatically generate labeled components for floor plans. Xiao et al. (2020) used fully convolutional networks (FCNs) for architectural component segmentation and categorization to automatically generate 3D models. Zhao et al. (2020) used YOLO CNN to recognize structural components.

The present study applies image segmentation on 2D formwork drawing images to extract formwork elements in an attempt to automate 3D model generation without requiring extracting additional features such as grid lines or dimensions and better distinguish elements from the background automatically.

## 1.2 Research Objective and Procedure

The purpose for the current study was to provide an artificial intelligence model to extract temporary formwork elements by applying Mask R-CNN, a popular convolutional neural network (CNN) model. We provide an artificial intelligence procedure to extract formwork components such as joists, struts, pipe supports, and sheathing from 2D formwork drawings.

Although the motivation for this research was to provide a feasible method to extract required features for 3D modelling, the scope was limited to extracting formwork elements using image segmentation on 2D formwork drawings. To the authors knowledge, this is the first study that applies image segmentation on 2D formwork drawings. This study is unique, since the proposed system also helps engineers or contractors-in-charge to inspection of formwork drawings. Based on the inspection the engineer/contractor-in-charge can take appropriate steps if necessary.

Three-dimensional formwork models are required for AutoCAD to perform image segmentation. Two views are

collected at front and side in 2D wireframe format from the 3D model, preparing the from drawings with different perspectives. Each drawing image is manually labelled and post processed for image segmentation training. After training, the model is validated with appropriate evaluation metrics. We also conducted a case study, as shown in Fig. 1, preparing a test database separately to test how well the trained model learned image segmentation. The following outputs are produced by image segmentation for each drawing image: classified formwork components, localized formwork components and segmented formwork components.

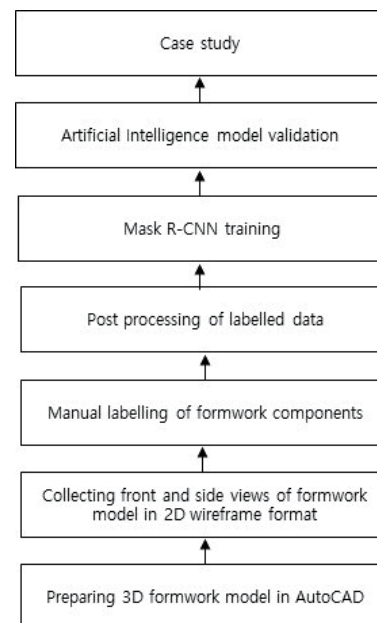


Figure 1. Research procedure

The remainder of this paper is organized as follows. Section 2 briefly describes the CNN based AI system, including literatures and Mask R-CNN. Section 3 details the AI model structure for formwork drawing inspection, including data collection, segmentation, and training procedures. Section 4 provides results and discussion, and Section 5 summarizes and concludes the paper.

## 2. CNN model overview

### 2.1 CNN Model

Convolution neural networks are powerful image

processing tools, comprising an artificial intelligence (AI) system that uses deep learning to perform both generative and descriptive tasks. CNNs are usually deep feedforward neural networks where the knowledge goes in one direction from input to output layer(s) (Rawat and Wang, 2017), and interconnections between the layers do not assemble a cycle (Habibi and Heravi, 2017).

A CNN can comprise single or multiple convolutional and pooling layer(s) arranged in blocks, followed by single or a few fully connected layers. Images input to the model passes through several convolutional and pooling layer stages and finally fully connected layer(s) (usually followed by a SoftMax layer) to delivers outputs (Rawat and Wang, 2017). Fig. 2 shows a typical image classification procedure for a deep CNN, where input 'x' is labeled to be 'class 1' as output.

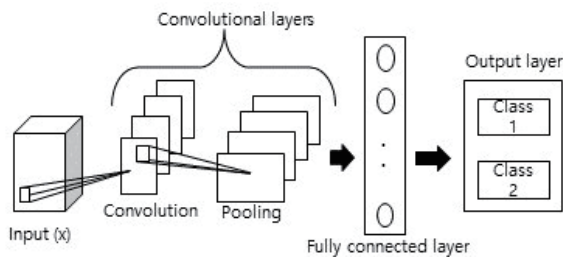


Figure 2. Deep CNN architecture

Other than classification, CNN can also perform various tasks such as classification and localization (He et al. 2016), or semantic segmentation (Girshick et al. 2014), and instance segmentation (He et al. 2018). Dai et al. (2016) presented region based fully convolutional networks (FCNs) for accurate and efficient object detection. Ren et al. (2017) proposed Faster R-CNN, an object detection model that depends on region proposal algorithms to share full-image convolutional features with a detection network to enable region proposals for objects of interest.

## 2.2 Mask R-CNN Model

Region based convolutional neural networks (R-CNNs) are deep learning models for object detection. Mask R-CNN (He et al., 2017) is a popular model for image segmentation. Mask R-CNN is the same as Faster R-CNN (Ren et al., 2017), expect Mask R-CNN has an additional branch for predicting segmentation masks on each region of interest

(RoI).

The Mask R-CNN architecture comprises two phases. A region proposal network (RPN) submits bounding boxes for the targets, i.e., Rols, the same as Faster R-CNN phase one. However, Faster R-CNN extracts features using RoIPool from each target bounding box in the second phase, and classifies object labels and performs bounding-box regression; whereas Mask R-CNN also provides a binary mask for each RoI.

Mask R-CNN has been used recently in various construction engineering applications. Wang et al. (2020) combined Faster R-CNN and Mask R-CNN models to detect morphological features for damage in historic glazed tiles. Ying and Lee (2019) utilized Mask R-CNN for building object automatic recognition and segmentation for constructing as-is BIM objects from building images. Kim and Cho (2020) used Mask R-CNN for damage detection in concrete, such as cracks, efflorescence, rebar exposure, and spalling; and Li et al. (2021) used histogram thresholding Mask R-CNN for mapping new and old buildings.

Fig. 3 shows Mask R-CNN proposal and segmentation phases. The proposal phase generates proposals as probable object locations. The feature pyramid network (FPN) (Lin et al., 2017) architecture has been used to feature maps extraction. Region proposal networks (RPNs) (Ren et al., 2017) generate object proposals with objectness scores for each object. The segmentation stage generates object masks, binary boxes and predicts formwork element object classes after receiving proposals from the first phase.

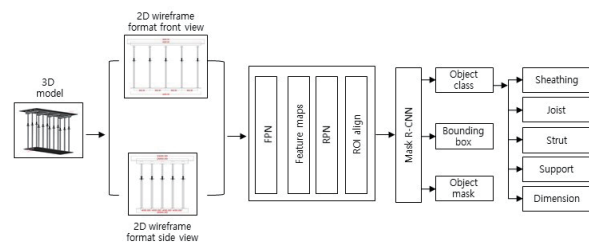


Figure 3. Mask R-CNN structure

## 3. Formwork AI model structure

### 3.1 Training Data Collection

Large datasets are usually preferable for training full-

scale CNNs. However, smaller datasets are okay if transfer learning can be applied (Zhang et al., 2019). In transfer learning, knowledge learned in previous task is transferred from a pretrained CNN architecture to perform a new similar task. Usually, in transfer learning, feature extractor uses the pre-trained weights and bias, while the later layers are retrained for the new task. Here, pretrained Mask R-CNN is used, which was trained on the COCO dataset (He et al., 2017). When transfer learning is applied, feature extractor part of pretrained Mask R-CNN is kept frozen, while the classifier part is retrained. This study focuses only on slab formwork. Three-dimensional slab formwork models were prepared in AutoCAD to train the CNN model, since 3D models provide detailed visualization, and views from two sides in 2D wireframe format (front and side) were collected. We prepared 93 3D models, comprising 186 images, were collected for training.

The models were prepared at different lengths for all horizontal elements, i.e., sheathing, joists, and struts, at different joist, strut, and pipe support spacings; and finally at different sheathing widths. 2D AutoCAD drawings in .pdf format were converted to .jpeg using python PyPDF and pdf2image libraries.

Since the actual formwork model was drawn in 3D, it provides better accuracy than traditional 2D formwork drawings, and better visualization and accuracy for 3D models in BIM. Fig. 4 shows the original 3D model South-East view and Fig. 5 shows the front view in 2D wireframe format.



Figure 4. Three-dimensional formwork model in AutoCAD

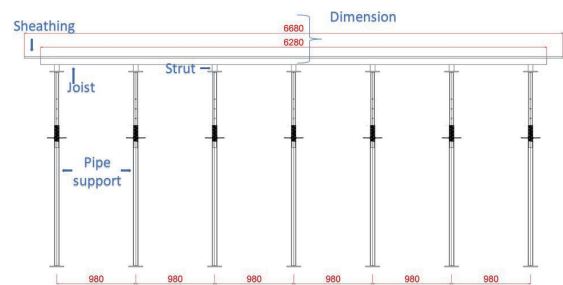


Figure 5. Typical two-dimensional formwork model and its elements (front view)

### 3.2 Formwork Element Image Segmentation

Image segmentation is a popular technique used to segment an image into meaningful parts. The first stage in any image analysis task is image segmentation to evaluate the analysis from all objects of interest in the corresponding image. Image segmentation determines exact object boundaries (Sahu et al., 2018) and provides more details about target features than traditional object detection and classification models (Ker et al., 2017). Various segmentation techniques have been developed for object segmentation (Sahu et al., 2018) in many fields.

Concrete slab formwork comprises four elements: sheathing, joists, struts, and pipe support. Sheathing is a plywood sheet supported by joists and struts that transfers load from joists to the pipe supports, which finally transfer the load to the attached ground. Typical 2-D slab formwork drawings contain one additional drawing element: dimension, or numerical specification. When an image such as Fig. 5 is given to the trained physical model, the model first detects each element, creates segmentation masks around each element, and finally predicts the label or classes for each of the element (fig. 3) This is image segmentation, with three outputs (Fig. 3):

1. identify the location for each element,
2. classify the label (i.e., sheathing, joist, strut, pipe support, and dimension) for each element, and
3. generate binary masks around each object of interest.

Deep learning is one of the most popular image segmentation methods. Long et al. (2015) proposed an FCN using fully convolutional layers to output feature maps for the object of interest in the given image rather than using

a fully connected layer. The CNN usually outputs object classifications from an input image in the fully connected layer, as discussed in section 2.1. In contrast, FCN outputs the feature map using convolution and pooling layers and the feature map becomes the probability map for objects to be classified and labeled for each pixel. However, feature map size for the input image is reduced every time it passes through a convolutional layer, eventually becoming very small and merges with the original image, enlarging the generated output feature map, and hence creating high memory usage issues.

SegNet (Badrinarayanan et al., 2017) fixed the memory usage issue by adopting an encoder–decoder structure. Convolution and pooling is achieved during the encoder stage, and the feature map is subsequently enlarged in the decoder stage.

Mask R–CNN has become very popular for accuracy and speed in image segmentation tasks. Since Mask R–CNN is based on Faster R–CNN, it extracts input image features using the backbone CNN, then the RPN chooses object regional areas and finally segmentation is performed within the proposed bounding box area (Ren et al., 2017; He et al., 2017).

Image segmentation with deep learning has been used in various fields, including automated driving, medical science, factory inspection, etc. Unfortunately, image segmentation applications in civil engineering remain in their early stage. Since construction drawing elements are very complicated, we chose image segmentation to ensure the best performance for formwork drawing element extraction for future applications in 3D modelling.

### 3.3 Artificial Intelligence Training

Postprocessing is mandatory after data collection to launch CNN training, particularly when segmentation is the primary objective using the TensorFlow Object Detection API. Labelling is crucial for image segmentation, where labelling refers to putting a label or name for each object in an image (Bhavsar et al., 2020). Each formwork element was manually labelled in all 186 images. However, manual data labelling is a tiresome process, and data labeling is the first and essential step for CNN training. Polygon object masks were created in the labeling process for pipe supports.

Fig. 6 shows an example of labelling procedure on 2D formwork drawing elements (front view) on Labelbox, an online platform for image annotation. The labelled data was received in .json format, then used to create object masks and XML files for each labelled image. Finally, TFRecord is created containing all the RGB images, PNG mask images, data from XML files, image index and class index.

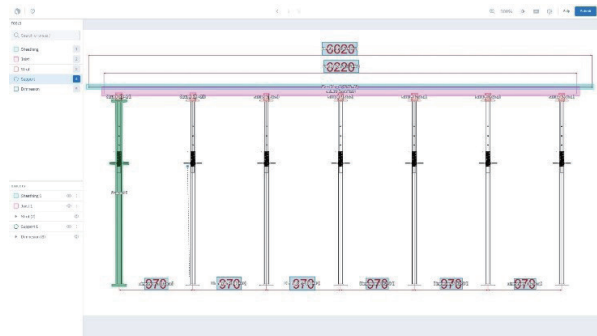
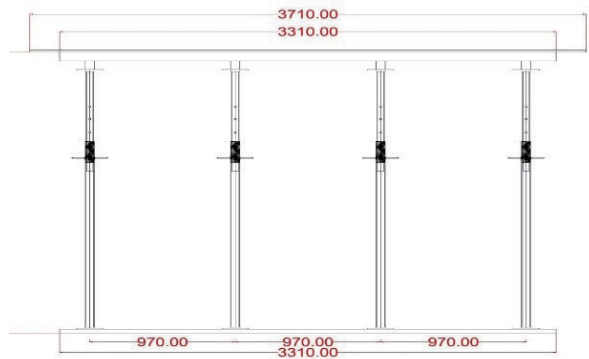
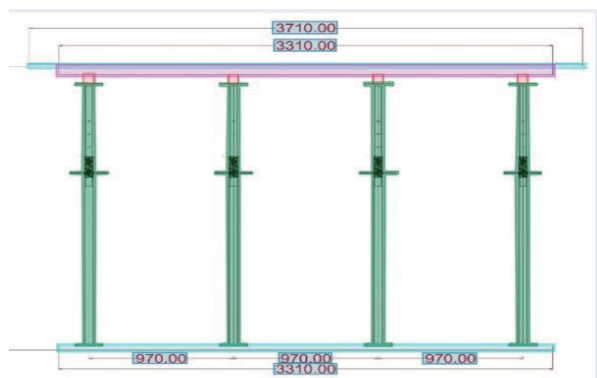


Figure 6. Labelling on 2D formwork image from Labelbox

Fig. 7 shows an example of the training data, an original image (Fig. 7 (a)), and the same image with labelled object (Fig. 7 (b)):



(a) Original image



(b) Labeled image

Figure 7. Example of training data

The TFRecord file contains this information together for all the training data, i.e., the original image and the labelled image information.

Training was launched in a local system using the TensorFlow Object Detection API training module. The datasets were converted into two distinct TFRecord files: training (70%) and validation (30%) subsets. Image size was approximately 1600×1200 pixels. Mask R-CNN was trained with basic learning rate = 0.003, momentum = 0.9, and weight decay = 0.0005 for 5,000 and 10,000 iterations.

The TensorFlow Object Detection API training platform was launched on a local system comprising Intel CITM) i5–9600K @ 3.70GHz CPU, 48.0 GB RAM, and GEFORCE GTX 1660 GPU, on a Windows system; running python V3.6 and TensorFlow V1.15. Training took approximately 15 days to complete (50,000 epochs) and model hyperparameters were tuned according to the Faster R-CNN model.

Fig. 8 shows total training loss during training, including classification, localization, and mask loss for box classifiers, localization, and objectness loss for the RPN. The loss curve reduced 10 fold by the end of training, implying the model learned to predict well during training.

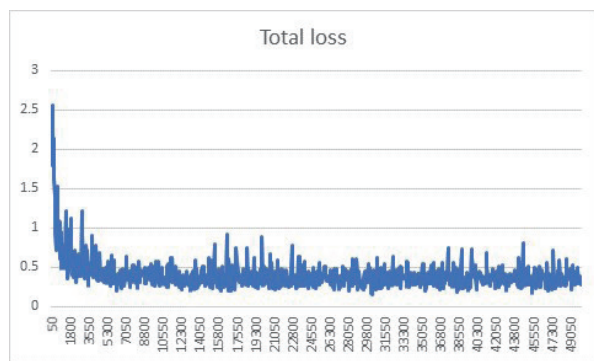


Figure 8. Total loss with respect to training steps

#### 4. Result and Discussion

Two evaluation metrics were utilized to validate the trained CNN model: mean average precision and mean average recall, as shown in Table 1. Precision measures prediction accuracy, defined as the number of true positives (TP) over the number of true positives plus the number of false positives (FP),

$$\text{Precision} = \frac{TP}{TP+FP} \tag{1}$$

whereas Recall (R) measures TP accuracy, defined as the number of true positives (TP) over the number of true positives plus the number of false negatives (FN),

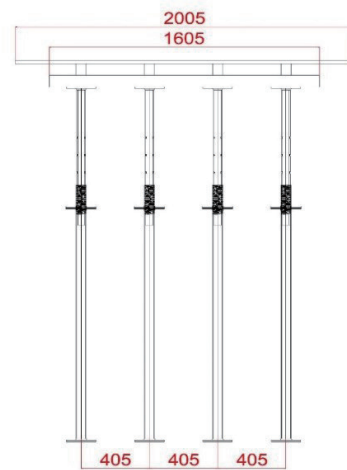
$$\text{Recall} = \frac{TP}{TP+FN} \tag{2}$$

Both evaluation metrics > 80%, confirming the model is trained well.

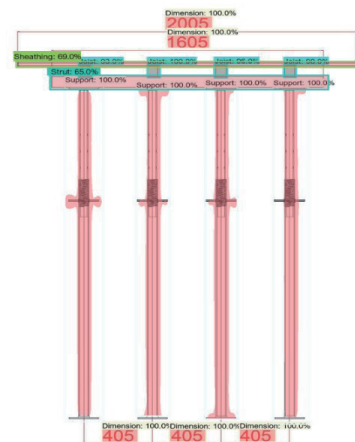
Table 1. Evaluation metrics for the validation set

Serial	Metric	Value
1	mAP	0.826
2	mAR	0.817

Finally, we performed a case study is on 10 new test data that were not used during training. The model performed accurate image segmentation for all cases considered.

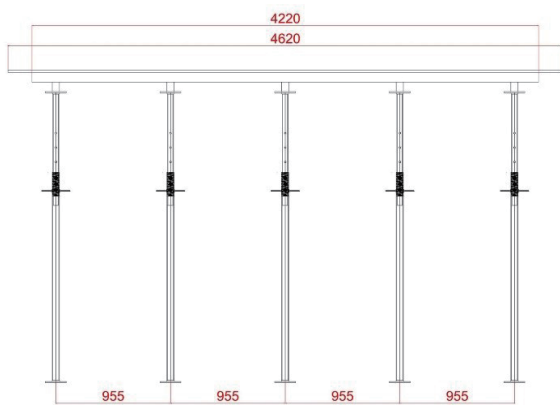


(a) Test data (before image segmentation)

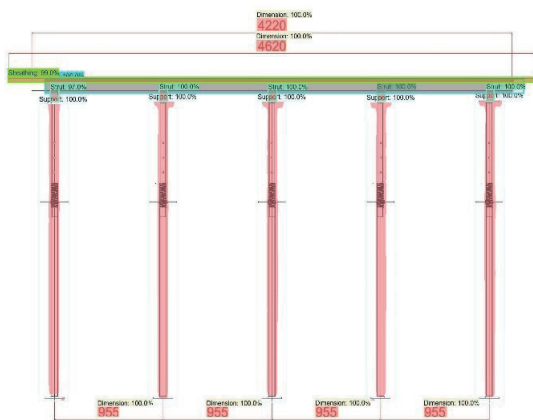


(b) Test data (after image segmentation)

Figure 9. Image segmentation by trained Mask R-CNN in a 2D formwork model (side view)



(a) Test data (before image segmentation)



(b) Test data (after image segmentation)

Figure 10. Image segmentation by trained Mask R-CNN in a 2D formwork model (front view)

Fig. 9 shows a typical test image before and after image segmentation. The test image included 15 elements: 1 sheathing, 1 strut, 4 joists, 4 pipe supports, and 5 dimensions. Similarly, Fig. 10 shows a 2D formwork drawing image in front view, before and after segmentation.

The trained AI model provided all three outputs as discussed in Section 3.2, i.e., the class label, bounding box (location), and segmentation mask around each object of interest. Thus, the results confirm that the trained AI image segmentation model learned adequately and was ready for formwork element extraction during 3D BIM formwork model construction.

## 5. Conclusion

The objective for this study was to develop a unique

system to segment and categorize construction formwork drawing elements. To achieve that goal, we proposed an image segmentation model based on the popular Mask R-CNN model. Manual drafting and extracting drawing elements from 2D CAD/.pdf file is time-consuming, and the proposed system would be useful for automatic drawing element recognition and hence 3D automation model.

However, although the proposed model performed well, there are some limitations. We only used a small number of training images. The model is particularly developed for slab formwork with pipe supports. Further study needs to be done for various types of concrete formwork structures. Future work should investigate overcoming these limitations.

## Acknowledgments

This work was supported by a 2-Year Research Grant of Pusan National University.

## References

- Ahn, J., Kwak, S. (2018). "Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June, pp. 4981–4990.
- Badrinarayanan, V., Kendall, A., Cipolla, R. (2017). "Segnet: A deep convolutional encoder-decoder architecture for image segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), pp. 2481–2495.
- Bhavsar, K., Jani K., Vanzara, R. (2020). "Indian currency recognition from live video using deep learning". In Chaubey N., Parikh S., Amin K. (eds) Computing Science, Communication and Security. COMS2 2020, Communications in Computer and Information Science, 1235, Springer, Singapore, pp. 70–81.
- Dai, J., Li, Y., He, K., Sun, J. (2016). "R-FCN: Object detection via region-based fully convolutional networks", Advances in Neural Information Processing Systems 29 (NIPS 2016), pp. 379–387.

- Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation", *Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 23–28 June, pp. 580–587.
- Habibi, H., Heravi, E. J. (2017). "Guide to convolutional neural networks: A practical application to traffic sign detection and classification, Springer International Publishing AG, Switzerland, pp. 63.
- He, K., Gkioxari, G., Dollár, P., Girshick, R. (2018). "Mask R-CNN", *IEEE transactions on pattern analysis and machine intelligence*, 42 (2), pp. 386–397.
- He, K., Gkioxari, G., Dollár, P., Girshick, R. (2017). "Mask R-CNN", In *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 22–29 October, pp. 2980–2988.
- He, K., Zhang, X., Ren, S., Sun, J. (2016). "Deep residual learning for image recognition", *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June, pp. 770–778.
- Horna, S., Méneveaux, D., Damiand, G., Bertrand, Y. (2009). "Consistency constraints and 3D building reconstruction", *Computer-Aided Design*, 41 (1), pp. 13–27 2009.
- Ker, J., Wang, L., Rao, J., Lim, T. (2017). "Deep learning applications in medical image analysis", *IEEE Access*, 6, pp. 9375–9379.
- Kim, B., Cho, S. (2020). "Automated multiple concrete damage detection using instance segmentation deep learning model", *Appl. Sci.*, 10 (22), pp. 8008.
- Lee, B., Choi, H., Min, B., Lee, D.-E. (2020). "Applicability of formwork automation design software for aluminium formwork", *Applied Sciences*, 10 (24), pp. 9029.
- Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S. (2017). "Feature pyramid networks for object detection", *Proceedings of the IEEE conference on computer vision and pattern recognition*, 21–26 July, Honolulu, HI, USA, pp. 2117–2125.
- Li, Y., Xu, W., Chen, H., Jiang, J., Li, X. (2021). "A novel framework based on Mask R-CNN and histogram thresholding for scalable segmentation of new and old rural buildings", *Remote Sens*, 13 (6), pp. 1070.
- Long, J., Shelhamer, E., Darrell, T. (2015). "Fully convolutional networks for semantic segmentation." In: S. Lakušić, Ed. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 8–10 June, New Jersey, USA: Institute of Electrical and Electronics Engineers, pp. 3431–3440.
- Moon, S., Cowdhury, A. M. (2021). "Utilization of prior information in neural network training for improving 28-day concrete strength prediction", *Journal of Construction Engineering and Management*, 147 (5), pp. 04021028.
- Rawat, W., Wang, Z. (2017). "Deep convolutional neural networks for image classification: A comprehensive review", *Neural Computation*, 29 (9), pp. 62352–2449.
- 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, pp.1137–1149.
- Sahu, S., Sarma, S., Bora, D. J. (2018). "Image segmentation and its different techniques: An in-depth analysis", *Proceedings of the 3rd IEEE International Conference on Research in Intelligent and Computing in Engineering*, RICE 2018, 22–24 August, Ciudadela Don Bosco, San Salvador, El Salvador, pp. 1–7.
- Seo, J., Park, H., Choo S. (2020). "Inference of drawing elements and space usage on architectural drawings using semantic segmentation", *Applied Science*, 10 (20), pp. 7347.
- Sigalov, K., König, M. (2017). "Recognition of process patterns for BIM-based construction schedules", *Advanced Engineering Informatics*, 33, pp. 456–472.
- Wang, N., Zhao, X., Zou, Z., Zhao, P., Qi, F. (2020). "Autonomous damage segmentation and measurement of glazed tiles in historic buildings via deep learning", *Computer-Aided Civil and Infrastructure Engineering*, 35 (3), pp. 277–291.
- Xiao, Y., Chen, S., Ikeda, Y., Hotta, K. (2020). "Automatic recognition and segmentation of architectural elements from 2D drawings by convolutional neural network", *Proceedings of The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA)*, pp. 843–852.



- Ying, H., Lee, S. (2019). "A Mask R-CNN based approach to automatically construct as-is IFC BIM objects from digital images," ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, 36. IAARC Publications, Banff, Canada, pp. 764–771.
- Zhang, Z., Pan, W., Zheng, Z. J. (2019). "Transfer learning enabled process recognition for module installation of high-rise modular buildings", Proceedings of the 2019 Modular and Offsite Construction (MOC) Summit, 21–24 May, Banff, Alberta, Canada, pp. 268–275.
- Zhao, Y., Deng, X., Lai, H. (2020). "A deep learning-based method to detect components from scanned structural drawings for reconstructing 3D models", Applied Science, 10 (6), pp. 2066.
- V. Badrinarayanan, A. Kendall, R. Cipolla, "Segnet: A deep convolutional encoder–decoder architecture for image segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481–2495, 2017.
- K. Bhavsar, K. Jani, R. Vanzara, "Indian currency recognition from live video using deep learning". In Chaubey N., Parikh S., Amin K. (eds) Computing Science, Communication and Security. COMS2 2020, Communications in Computer and Information Science, 1235, Springer, Singapore, pp. 70–81, 2020.