

AUTOMATIC ROAD CRACK RECOGNITION BASED ON DEEP LEARNING NETWORKS FROM UAV IMAGERY

F. Samadzadegan^{1*}, F. Dadrass Javan², M. Hasanlou¹, M. Gholamshahi³, F. Ashtari Mahini¹

¹ School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran; (samadz, hasanlou, f.ashtari)@ut.ac.ir

² Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7522 NB Enschede, the Netherlands; f.dadrassjavan@utwente.nl

³ Department of Electrical and Computer Engineering, Faculty of Engineering, Kharazmi University, Tehran, Iran, mehrnazgholamshahi.khu@gmail.com

* Correspondence: samadz@ut.ac.ir

Commission IV, WG IV/3

KEY WORDS: Deep Learning, Convolutional Neural Network, Road Cracks, Road Crack Recognition, YOLO Deep Learning Network, YOLOv4 Deep Learning Network

ABSTRACT:

Roads are one of the essential transportation infrastructures that get damaged over time and affect economic development and social activities. Therefore, accurate and rapid recognition of road damage such as cracks is necessary to prevent further damage and repair it in time. The traditional methods for recognizing cracks are using survey vehicles equipped with various sensors, visual inspection of the road surface, and recognition algorithms in image processing. However, performing recognition operations using these methods is associated with high costs and low accuracy and speed. In recent years, the use of deep learning networks in object recognition and visual applications has increased, and these networks have become a suitable alternative to traditional methods. In this paper, the YOLOv4 deep learning network is used to recognize four types of cracks transverse, longitudinal, alligator, and oblique cracks utilizing a set of 2000 RGB visible images. The proposed network with multiple convolutional layers extracts accurate semantic feature maps from input images and classifies road cracks into four classes. This network performs the recognition process with an error of 1% in the training phase and 77% F1-Score, 80% precision, 80% mean average precision (mAP), 77% recall, and 81% intersection over union (IoU) in the testing phase. These results demonstrate the acceptable accuracy and appropriate performance of the model in road crack recognition.

1. INTRODUCTION

Road infrastructures play an important role in economic development and growth and are widely considered the most important platform for transportation (Doshi et al., 2020; Shu et al., 2021). Roads crack over time due to various reasons such as heavy vehicles, changing weather conditions, human activity, and the use of inferior materials. Delayed repair of road cracks leads to problems such as reduced serviceability of roads, rise in pavement collapse, increased traffic accidents, excessive damages, and increased repair costs (Guo et al.; Yan et al., 2021a). Therefore, it is necessary to recognize different types of road cracks to regularly repair the road and maintain them to carry out smooth daily activities in society (Doshi et al., 2020; Guo et al.).

There are different traditional methods to recognize types of road cracks (Zhu et al., 2021). The first method is using road survey vehicles equipped with various sensors; However, the application of this method is expensive for multiple organizations. The second method is entirely manual recognition, performed in some developing countries and requires hours of visual inspection of the road surface by experts. However, this method is time-consuming, and the accuracy in determining the severity and type of the crack depends on the accuracy of the expert (Guo et al.; Zhu et al., 2021). The third method for recognizing road cracks includes

Gabor filtering (Salman et al., 2013), edge detection (Chambon et al., 2011), and intensity thresholding (Ayenu-Prah et al., 2008), and texture analysis. These methods are slow and do not work well on a large scale because they rely on manipulating image pixels. These methods can also accurately recognize road cracks when the image configuration is static and do not work precisely when the camera configurations are different and their widespread use of them is impractical (Majidifard et al., 2020); Because of the limitations mentioned in these three traditional methods, it is important to develop a cost-effective, accurate, fast, and independent method for recognizing road cracks.

Automatic deep learning methods have been introduced as an accurate alternative to traditional object recognition methods and have great potential in visual applications and image analysis (Majidifard et al., 2020). These methods can not only detect the category of an object but also determine the object's location in the image (Yan et al., 2021b). The use of deep learning methods can reduce labor costs and improve work efficiency and intelligence in recognizing road cracks. Among these networks, the YOLO Network is a widely used one-stage detector that determines the coordinates of the bounding box and the object class (Redmon et al., 2016). This network is more accurate and efficient than two-stage detection models such as SPP-Net (He et al., 2015), Fast-RCNN, and Faster RCNN (Ren et al., 2015).

* Corresponding author

Recently, research has been conducted on crack detection using deep learning networks. In 2020, Silva et al., used a deep learning network training for real-time detection of cracks and potholes by drones, and the results of this model reached 95% accuracy (Silva et al., 2020). In 2021, Teng et al., detected cracks using a deep learning network and 11 feature extractors. In this paper, the YOLOv2 object detector achieved better crack detection results (Teng et al., 2021). Also, in this year, Yang et al., used AlexNet, VGGNet13, and ResNet18 neural networks to detect and classify cracks in an image collection. Among them, the ResNet18 network performed better. This paper also performed crack detection using an accurate and quick deep learning network and concluded that this network could be an accurate and efficient tool for crack detection compared to conventional methods (Yang et al., 2021). In 2022, Xu et al., detected single and bifurcation cracks in roads using two R-CNN networks (Faster R-CNN and Mask R-CNN). They also investigated the performance of these networks in detecting cracks with or without sunlight interference, straight and Bending Cracks, and deep and shallow cracks. Both networks performed well in detecting single cracks

and required more datasets to train complex cracks (Xu et al., 2022). In 2022, Fan et al., proposed a novel automatic method for detecting and measuring pavement cracks using the parallel ResNet module and a skeleton. Mathematical results show that the new method performs well compared to several competing methods (Fan et al., 2022).

YOLOv4 Deep Learning Network is one of the versions of the YOLO network, which has remarkable accuracy and speed in object recognition and has attracted much attention due to its high computational capabilities (Bochkovskiy et al., 2020). Due to the importance of accuracy and speed in road crack recognition, this network has been used to recognize four types of road cracks: transverse, longitudinal, alligator, and oblique cracks.

When asphalt pavement is appropriately designed and constructed, it can function properly for many years. However, asphalt pavements crack over time after use and require maintenance. Road pavement is exposed to cracks due to traffic loading, temperature, humidity, and subsoil movement. According to Figure 1, crack images are shown along with their causes.

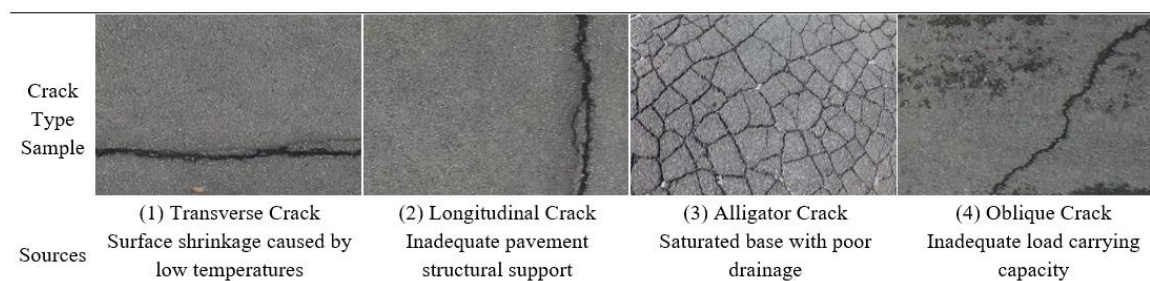


Figure 1. Some samples of common pavement cracks.

This paper is organized as follows: In the second section, we discuss the proposed method; in the third section, we explain the network training and the evaluation criteria; in the fourth section, we perform the results and evaluation of the model, and in the fifth section, we discuss the future area and propose improvements.

2. PROPOSED METHOD

In this study, the YOLOv4 deep learning network is used to recognize four types of road cracks, including transverse, longitudinal, alligator, and oblique cracks. The proposed method is performed in four phases:

- Input Preparation,
- Training the Network,
- Testing the Network,
- Model Evaluation.

First, the input images are divided into two categories: training data and testing data, and the collected dataset, which includes four types of cracks, is labeled by drawing a bounding box around the object. In step 2, the adjustment of the proposed network parameters is made to start the network training process. In step 3, the test images are examined with the obtained model weight files and the Non-Maximum Suppression (NMS) algorithm. Finally, the network's performance in the recognition process is evaluated with evaluation metrics.

2.1 Input Preparation

There are several types of road cracks, and each is important because of the extent of the damage it causes to the road. In this

study, a dataset of four types of cracks, including transverse, longitudinal, alligator, and oblique cracks, is collected and labeled for road crack recognition. For labeling the input data, a bounding box is drawn around each target object, and the information about each bounding box, including the class name, the center coordinates of the bounding box, and its width and height, is stored. Finally, 70% of these images are selected for training and the remaining 30% for network testing.

2.2 Network Training

In this stage, a fast and accurate deep convolutional network is selected for training the network to recognize four types of road cracks. Then, 70% of the collected road crack dataset is used for executing this process.

2.2.1 Network Architecture

The architecture of proposed deep learning network consists of four stages, including the input, the backbone, the neck, and the head. CSPDarknet53 is used in the backbone of the network responsible for extracting the feature and creating the feature map from the input dataset. CSPDarknet53 is a convolutional neural network for object recognition and uses DarkNet-53 as its base network. The neck consists of layers between the backbone and the head, which receives feature maps from the backbone and uses Spatial Pyramid Pooling (SPP) and the Path Aggregation Network (PAN) to increase accuracy (Liu et al., 2018; Ren et al., 2015). Using the SPP network, feature maps are created from the entire image only once, and then, the features are collected in these areas, and fixed-length images are created to train the detectors. PAN network also improves the

input segmentation process with the ability to accurately store spatial information. Finally, in the head, the classification and positioning of the objects are performed using the YOLOv3

network, and the probabilities and the bounding box coordinates (x, y, height, and width) are given (Bochkovskiy et al., 2020).

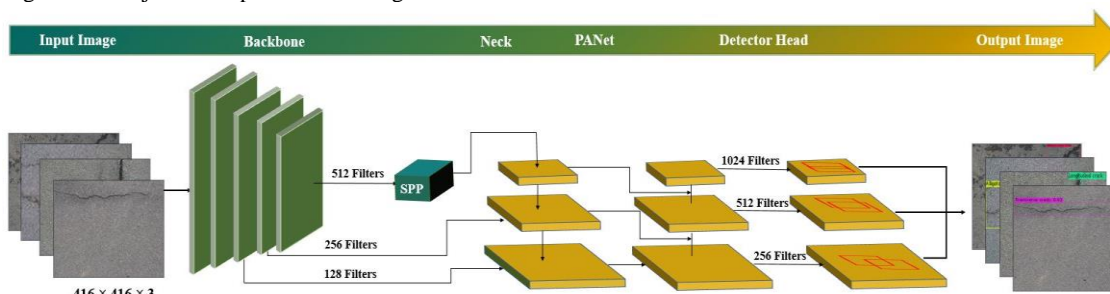


Figure 2. The proposed automatic crack recognition deep learning network architecture.

2.3 Network Testing

After completing the network training, the network testing process begins with 30% of the entire dataset. Object Recognition Algorithms use the Non-Maximum Suppression (NMS) algorithm to select the best bounding box containing an object from several predicted bounding boxes. This method "removes" possible bounding boxes and selects the best bounding box that contains the object (Figure 3).

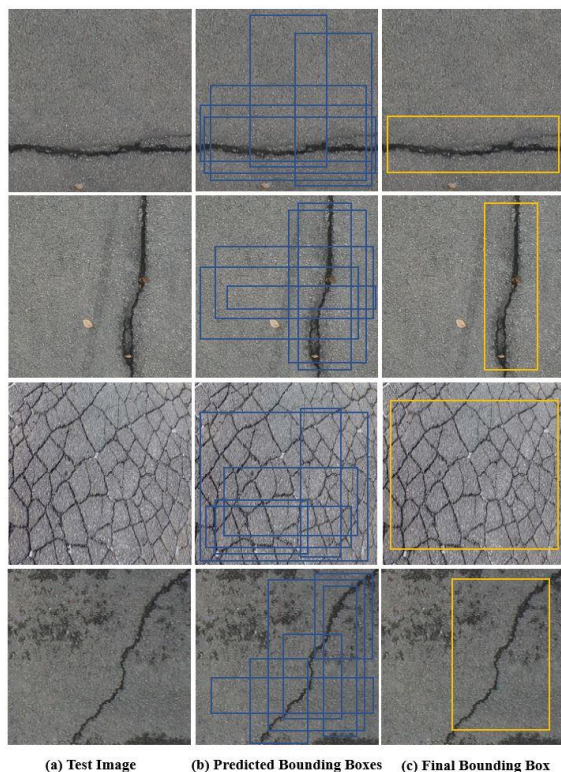


Figure 3. Non-Maximum Suppression (NMS) algorithm.

In this algorithm, the bounding box with the highest confidence score is selected first, and then the parameter IoU between the selected box and the other predicted boxes is calculated. If this value exceeds the IoU threshold of 0.7, the box is deleted and the process continues until only one bounding box remains. Figure 3 shows the performance of the Non-Maximum Suppression (NMS) algorithm. The predicted bounding boxes are initially shown in blue. After applying the algorithm, the yellow box is selected as the bounding box containing the object.

2.4 Model Evaluation

To evaluate the performance of the proposed model, evaluation metrics such as F1-score, precision, recall, mAP, and IoU are used, defined as follows (Simonyan et al., 2014):

$$F1_Score = \frac{2TP}{2TP + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

where TP = True Positive
FP = False Negative

TP (True Positive) means that the input is predicted to be positive and is actually positive, and FP (False Positive) means that the input is predicted to be positive and is not actually positive. TN (True Negative) means that the input is predicted to be negative and is actually negative, and FN (False Negative) means that the input is predicted to be negative and is actually positive (Dadrass Javan et al., 2022). In addition, other evaluation metrics are defined as follows (Simonyan et al., 2014):

$$mAP = \frac{1}{|Classes|} \sum_{c \in Classes} \frac{TP(c)}{TP(c) + FP(c)} \quad (4)$$

$$IoU = \frac{Area\ of\ overlap}{Area\ of\ Union} \quad (5)$$

where Classes = Number of Classes

3. EXPERIMENTAL RESULTS

In this section, the results of training and testing the proposed deep learning network in recognizing four types of road cracks are discussed. The dataset used, the network implementation and the system required to operate the network are also investigated.

3.1 Data preparation

To begin the network training process, a set of 2000 RGB images with different road cracks is collected. Figure 1 shows

four types of road cracks in this study, including transverse, longitudinal, alligator, and oblique cracks. To assess the performance of the proposed model and increase the reliability, and generalizability of the network in recognizing small cracks and distinguishing them from each other, a comprehensive and challenging dataset is used. To collect this dataset, public images and videos from UAV are used, 70% selected for network training and 30% for testing. Common to all these images is the use of a visible sensor with a resolution between 100 dpi and 300 dpi. In addition, the collection of videos is converted into images with a frame rate of 2 FPS. Then, the labelling tool is used to label each image and draw a rectangular bounding box around the objects. In this sketching tool, transverse cracks are classified as class 0, longitudinal cracks as class 1, alligator cracks as class 2, and oblique cracks as class 3.

3.2 Model Implementation and Training Result

The proposed network is trained on NVIDIA GeForce RTX 3050 Ti Graphics Processing Unit (GPU) hardware using CUDA (Compute Unified Device Architecture) toolkit version 11.1, CUDNN (CUDA Deep Neural Network library) version 8.1, and OpenCV 4.5. For the network training, the configuration file settings are changed as follows:

- The number of iterations is changed to 40k.
- The size of the input image is set to 160 × 160.
- The subdivision value is changed to 32.
- The batch value is changed to 32.
- The step parameter is changed to 32000 and 36000 (80% and 90% of the number of iterations).
- The size of the convolution layer filters before the YOLO network is changed to 27 according to the number of classes.

Figure 4 shows the change in the values of loss and mAP after 40k iterations and 24 hours. It can be seen that the proposed network achieves 1% error and 80% mAP during training. These results show that recognizing four types of road cracks with the proposed network has acceptable performance.

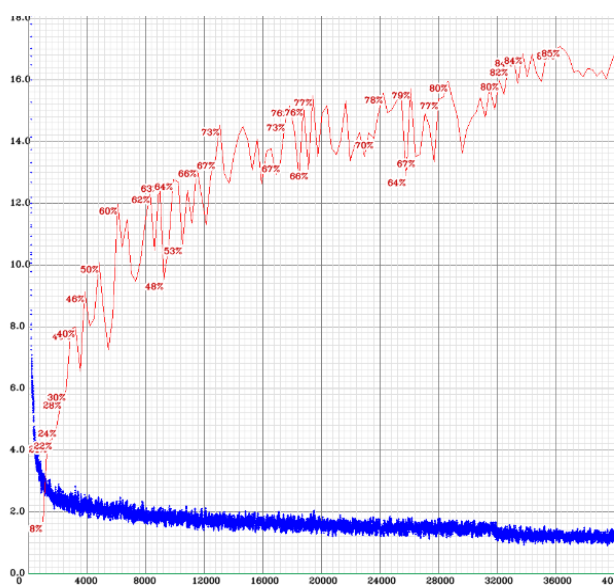


Figure 4. The loss graph in the training process of the network.

3.3 Implemented Result

After the training process, the proposed deep learning network is tested, and the performance of the network is evaluated using the evaluation metrics such as F1 score, Precision, Recall, mAP, and IoU. Table 1 shows the evaluation results of this network.

Class	Precision %	Recall %	F1-Score %	mAP %	IoU %
Transverse Crack	79	73	75	-	-
Longitudinal Crack	82	80	80	-	-
Alligator Crack	88	75	80	-	-
Oblique Crack	71	80	75	-	-
Total	80	77	77	80	81

Table 1. Evaluation results of the proposed network.

According to this table, the general evaluation metrics of the model, such as F1-Score, IoU, precision, recall, and mAP, reach 77%, 81%, 80%, 77%, and 80%, respectively. These values show high network performance and the probability of a lower error rate in road crack recognition.

Figure 5 shows examples of road crack recognition using the proposed deep learning network. In this Figure, the recognition of road cracks is indicated with bounding boxes and class probabilities. As it appears, the trained model has good performance in recognizing four types of road cracks, including transverse, longitudinal, alligator, and oblique cracks.



Figure 5. Some samples of pavement distress recognition results using proposed deep learning network.

4. DISCUSSION

The evaluation metrics in this study include F1-score, IoU, precision, recall, and mAP, and their corresponding values in each class are shown in Table 1. The values of precision, recall, and F1-score show the low error rate of the model and the high classification accuracy in recognizing different types of cracks in different sizes and backgrounds. In addition, the IOU metric indicates the acceptable overlap of the predicted bounding boxes with the ground truth boxes.

Figure 6 shows some examples of the success and failure of the network in recognizing four types of cracks. In the images in the first column, the model can recognize small cracks in different backgrounds and assign them to the correct class with high class probability. Even in the images where multiple types of cracks are present, the model can recognize the types of cracks simultaneously and distinguish between them. However, in the images of the second column, the network may incorrectly assign two classes to the same crack due to the similarity of the behavior of some crack types, and it may even fail to recognize a number of tiny cracks in the image.



Figure 6. Some samples of the model's (a) ability and (b) inability in crack recognition.

5. CONCLUSION

Automatic road crack recognition and its on-time repairment are critical to avoid further damages and to carry out daily transportation activities smoothly. This study presented the YOLOv4 deep learning network to recognize four types of road cracks automatically. To train and test the network, a set composing of 2000 images covering four types of road cracks, as transverse, longitudinal, alligator, and oblique cracks were collected. Evaluation of the model based on f1-score, IoU, precision, recall, and mAP metrics reached 77%, 81%, 80%,

77%, and 80% successfulness respectively, which shows the acceptable performance of the proposed model in road crack recognition. For future studies, it is also possible to use modified YOLOv4 network, YOLOv5 network and semantic segmentation methods such as U-Net network and deep learning methods based on edge detection and compare their performance with the network used. Horizontal images can also be added to the dataset to allow the model to recognize different types of cracks at different angles to the camera.

6. REFERENCES

- Ayenu-Prah, A., & Attoh-Okine, N. 2008. Evaluating pavement cracks with bidimensional empirical mode decomposition. *EURASIP Journal on Advances in Signal Processing*, 2008, 1-7.
- Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. 2020. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
- Chambon, S., & Moliard, J.-M. 2011. Automatic road pavement assessment with image processing: Review and comparison. *International Journal of Geophysics*, 2011.
- Dadrass Javan, F., Samadzadegan, F., Gholamshahi, M., & Ashotari Mahini, F. 2022. A Modified YOLOv4 Deep Learning Network for Vision-Based UAV Recognition. *Drones*, 6(7). <https://doi.org/10.3390/drones6070160>
- Doshi, K., & Yilmaz, Y. 2020. Road damage detection using deep ensemble learning. 2020 IEEE International Conference on Big Data (Big Data),
- Fan, Z., Lin, H., Li, C., Su, J., Bruno, S., & Loprencipe, G. 2022. Use of Parallel ResNet for High-Performance Pavement Crack Detection and Measurement. *Sustainability*, 14, 1825. <https://doi.org/10.3390/su14031825>
- Guo, L., Li, R., & Jiang, B. A ROAD SURFACE DAMAGE DETECTION METHOD USING YOLOV4 WITH PID OPTIMIZER.
- He, K., Zhang, X., Ren, S., & Sun, J. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9), 1904-1916.
- Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. 2018. Path aggregation network for instance segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition,
- Majidifard, H., Adu-Gyamfi, Y., & Buttlar, W. G. 2020. Deep machine learning approach to develop a new asphalt pavement condition index. *Construction and Building Materials*, 247, 118513.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. 2016. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition,
- Ren, S., He, K., Girshick, R., & Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

Salman, M., Mathavan, S., Kamal, K., & Rahman, M. 2013. Pavement crack detection using the Gabor filter. 16th international IEEE conference on intelligent transportation systems (ITSC 2013),

Shu, Z., Yan, Z., & Xu, X. 2021. Pavement Crack Detection Method of Street View Images Based on Deep Learning. *Journal of Physics: Conference Series*,

Silva, L. A., Sanchez San Blas, H., Peral García, D., Sales Mendes, A., & Villarubia González, G. 2020. An architectural multi-agent system for a pavement monitoring system with pothole recognition in UAV images. *Sensors*, 20(21), 6205.

Simonyan, K., & Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Teng, S., Liu, Z., Chen, G., & Cheng, L. 2021. Concrete crack detection based on well-known feature extractor model and the YOLO_v2 network. *Applied Sciences*, 11(2), 813.

Xu, X., Zhao, M., Shi, P., Ren, R., He, X., Wei, X., & Yang, H. 2022. Crack Detection and Comparison Study Based on Faster R-CNN and Mask R-CNN. *Sensors*, 22, 1215. <https://doi.org/10.3390/s22031215>

Yan, K., & Zhang, Z. 2021a. Automated Asphalt highway pavement crack detection based on deformable single shot multi-box detector under a complex environment. *IEEE Access*, 9, 150925-150938.

Yan, K., & Zhang, Z. 2021b. Automated Asphalt Highway Pavement Crack Detection Based on Deformable Single Shot Multi-Box Detector Under a Complex Environment. *IEEE Access*, PP,1-1. <https://doi.org/10.1109/ACCESS.2021.3125703>

Yang, C., Chen, J., Li, Z., & Huang, Y. 2021. Structural Crack Detection and Recognition Based on Deep Learning. *Applied Sciences*, 11(6). <https://doi.org/10.3390/app11062868>

Zhu, J., Zhong, J., Ma, T., Huang, X., Zhang, W., & Zhou, Y. 2021. Pavement distress detection using convolutional neural networks with images captured via UAV. *Automation in Construction*, 133, 103991. <https://doi.org/10.1016/j.autcon.2021.103991>