Automatic Road Surface Defect Detection Using Machine Learning Methods

Ian Furman[a] and Dmitry Skorobogatchenko[b]
Volgograd State Technical University, Lenina 28, 400064, Volgograd, Russia
furmanyan@gmail.com, dmitryskor2004@gmail.com

Keywords: Convolutional neural networks, CNN, Instance segmentation, Machine learning, Defect recognition, Road surface defects detection.

Abstract: The study aimed to develop software for recognizing road surface defects from an image. The article analyzes the main tools and methods for recognizing road surface defects. It is concluded that the most effective is the use of convolutional neural networks (CNNs) in the problem of instance segmentation, in particular, the Mask R-CNN and YOLOv8 models. During the research, these networks were used in training on a dataset containing selected road damage in the form of potholes and cracks. It was concluded that the most effective algorithms to solve instance segmentation tasks are convolutional neural networks, including the Mask R-CNN and YOLOv8 models. During the work, those models were used and learned with the dataset that contains marked road damages, such as potholes and cracks. Results of the neural networks work after learning on a real data are presented.

1 INTRODUCTION

The road network performs essential functions in the economy of any country. However, this resource wears out over time and requires maintenance. Without timely repairs, damage progresses, leading to significant costs for operating organizations and end-users. To maintain the operability of roads, it is necessary to have an accurate assessment of the condition of their structural elements (Skorobogatchenko, 2016). For an objective assessment of the condition of the road surface, the use of technologies based on computer vision is promising. (Farhadmanesh, 2021). Consequently, nowadays the use of new image processing technologies using artificial intelligence and deep learning technology is relevant (Majidifard, 2020).

The purpose of the research is to develop an AI-based solution in the field of recognition of road surface defects based on image processing.

2 RELATED WORKS

As an analysis of the literature shows, convolutional neural networks (CNNs) and models that implement their various modifications are most effective for recognizing surface defects. (Seongdeok, 2020). A feature of CNNs that has determined their effective use for solving this class of problems is their unique feature of local connection and weight distribution (Hu, 2022). CNN can efficiently extract objects from the input image through basic procedures including convolution, pooling, and activation, which greatly improves the accuracy and efficiency of object detection and classification (Shim, 2021). Following the basic structure of CNN, a growing number of advanced CNN-based networks have been created, for example, AlexNet (Liu, 2022), VGG (Guan, 2021), GoogLeNet (Zhong, 2022), ResNet (Wang, 2021), and DenseNet (Weng, 2022), to continually expand the types of identification targets and improve identification accuracy. The development of deep convolutional neural network architectures, such as Faster R-CNN, YOLO, SSD, MobileNet, and Inception, has made it possible to reduce

[a] https://orcid.org/0009-0007-3090-5782
[b] https://orcid.org/0000-0003-0280-0020

computational and time costs when determining the location of a defect in an image and its classification.

In the paper (Deru, 2023) tools for increasing the efficiency of training CNNs to recognize coating damage are presented.

3 METHODOLOGY

3.1 CNNs Libraries Used in Work

When choosing a network for use in the training model, the authors settled on Mask R-CNN, as it has proven itself in tasks of road damage instances segmentation (He, 2017), and the 8th version of YOLO (YOLOv8) (Terven, 2023) as having a SOTA algorithm, the previous version of which was actively used in the task of road damage's objects detection. No scientific articles from its developers have been published for YOLOv8. The architecture of the YOLOv8 model was studied based on console output.

The Mask R-CNN neural network model is a development of Faster R-CNN, designed for object recognition. Both algorithms first obtain a feature map from a CNN ResNet-50 or ResNet-101. Then recognition is made. In Faster R-CNN, the estimated regions of the object's location (Region of Interest, RoI) are transferred to the RoIPool algorithm (Girshick, 2015), (He, 2015). The Mask R-CNN network uses the RoIAlign algorithm, to which a fully convolutional network (FCN) is added, which simultaneously allocates segmentation masks (Shelhamer, 2018).

3.2 Dataset

Currently, there are a large number of datasets that can be used to train neural networks. In addition, some researchers offer their images and binary masks (Vasil'ev, 2020). Thus, the paper (Kanaeva, 2021) presents a broad analytical review of the most popular and open datasets for solving problems of identifying roadway defects.

One of the largest and most popular datasets containing road damage is RDD2022 (Arya, 2022). This dataset, developing since 2018, contains more than 47,000 road images, as well as more than 55,000 damage bounding boxes, divided into 8 classes.

The «Cracks and Potholes in Road Images Dataset» dataset was chosen for the study. (Passos, 2020). It consists of images of road surfaces and bitwise masks that define the road, cracks, and potholes. The dataset contains 2235 images. Instances are not marked, so the dataset is initially suitable only for semantic segmentation.

Next, the authors prepared the data. Two types of defects were used as defects that could be fixed by road surface maintenance work: potholes and cracks, following (FDS Rossii, 1998). Thus, the work distinguishes two classes of objects: pothole and crack. The dataset size is defined as 1200 images. The dataset was divided into training, validation, and test samples in proportions of 65%, 20%, and 15%, respectively.

For each semantic mask corresponding to a specific image, segmentation masks and boundary boxes for individual object instances were extracted using the «findCoutours» method from the OpenCV library. As a result, annotation files were obtained. The number of crack annotations turned out to be several orders of magnitude greater than the number of pothole annotations, which is why the dataset turned out to be heterogeneous and overloaded. To eliminate disproportions, a restriction was imposed on the minimum segmentation area of a particular crack instance. As a result, the number of cracks began to exceed the number of holes by only 1.7 times, and the size of the annotation file was reduced by approximately 6.5 times.

3.3 CNNs Learning Frameworks

To train the Mask R-CNN network, the MMDetection framework was used, which allows the implementation of various models of instance segmentation and object detection, containing tools for visualization of evaluation metrics of training. The best results were shown with two sets of parameters, with the code names 3_epoch_9 and 5_epoch_25. Both models were based on the Mask R-CNN model with the backbone network ResNet-50-FPN.

To train the YOLOv8 network, the Ultralytics framework was used, which is the official implementation of the YOLOv8 model tools, including training, evaluation, and visualization of learning results. The best results were shown with two sets of parameters, with the code names train20 and train22. Both models were based on the YOLOv8m-seg model.

3.4 Evaluation Metrics

To compare the training results, the mAP (mean average precision) metric was used, and the average precision value averaged over all categories:

$$\mathrm{mAP} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{AP_i} \quad (1)$$

where n is the number of classes. Average Precision – the area under the precision versus recall graph:

$$\mathrm{AP} = \int_0^1 \mathrm{P(R)dR} \quad (2)$$

where P is precision, and R is recall based on the assumption that the segmentation mask is determined correctly if IoU≥0.5.

IoU (Intersection over Union) – the ratio of the areas of bounding boxes or segmentation masks:

$$\mathrm{IoU} = \frac{\mathrm{A} \cap \mathrm{B}}{\mathrm{A} \cup \mathrm{B}} = \frac{\mathrm{TP}}{\mathrm{TP + FP + FN}} \quad (3)$$

where A and B are the predicted and real segmentation masks, respectively; TP – true positive prediction (True Positive).

Precision – the ratio of true positive prediction to predicted mask:

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP + FP}} \quad (4)$$

Recall – the ratio of the true-positive prediction to the true segmentation mask:

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP + FN}} \quad (5)$$

4 RESULTS

4.1 Mask R-CNN Results

For the Mask R-CNN network, compare the parameters of mAP segmentations with IoU=50%, IoU=75%, and mAP for small segmentations. The best parameters for each model are shown in Table 1.

Table 1: Parameters of Mask R-CNN model training results.

Model	mAP(IoU =0.5)	mAP(IoU =0.75)	mAP(small)	Epoch
3_epoch_9	0.3198	0.1372	0.0703	9
5_epoch_25	0.3175	0.1307	0.0721	13
	0.3171	0.1423	0.0733	25
	0.3152	0.1289	0.0759	11

Model 3_epoch_9 showed the best results for the specified parameters at epoch 9. Model 5_epoch_25 showed the best parameter mAP(0.5) at epoch 9. mAP(0.75) is the best at epoch 25. The best parameter for both models, mAP(s), was shown at epoch 11 of this model, but at the same time, at this epoch, the worst parameters mAP(0.5) and mAP(0.75) were shown. It was decided to choose epoch 25 for the 5_epoch_25 model since the value of mAP(0.5) is slightly inferior to the results of epoch 13.

A visual comparison of training results is presented in Figure 1. From left to right are shown a clean original image, an image with overlaid segmentations, the result of interaction with model 3_epoch_9, and the result of interaction with model 5_epoch_25.

4.2 YOLOv8 Results

For the YOLOv8 model, built-in analysis tools allow to calculate the parameters mAP(IoU=50%) and mAP(IoU=50-95%) for different classes. Compared with the training results of Mask R-CNN, one can compare mAP50 for all classes with mAP(0.5), mAP50-95 for all classes with mAP(0.75), mAP50 for cracks with mAP for small segmentations.

As the best training result, the framework saves the results of the iteration in which the maximum parameter mAP50-95 was achieved for all classes. Results are shown in Table 2.

Table 2: Parameters of YOLOv8 model training results.

Model	Class	mAP50	mAP50-95
train20	all	0.389	0.177
	crack	0.0875	0.0189
	pothole	0.69	0.336
train22	all	0.36	0.162
	crack	0.0467	0.0103
	pothole	0.673	0.314

A visual comparison of the training results is presented in Figure 2. The same images were selected on which Mask R-CNN was tested

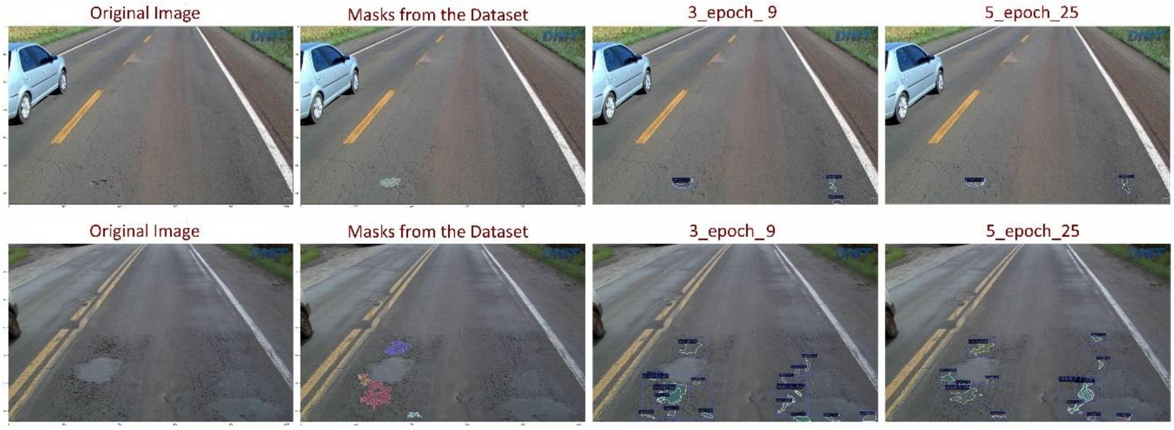

Figure 1: A visual comparison of Mask R-CNN training results.

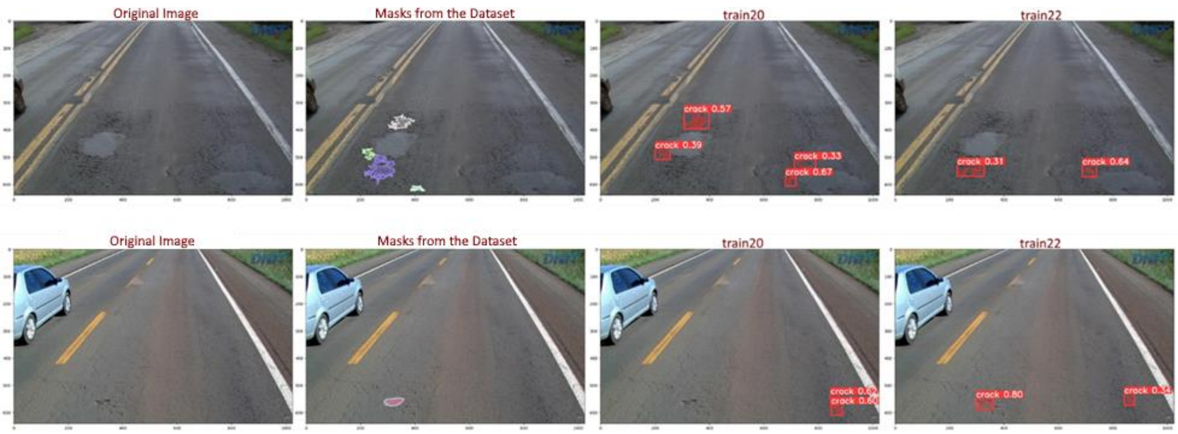

Figure 2: A visual comparison of YOLOv8 training results.

5 CONCLUSION

As a result of the work, the authors obtained a solution for recognizing road surface defects, which allows identifying potholes and cracks on road surfaces from an image, by using machine learning algorithms that implement the task of instance segmentation, based on the use of Mask R-CNN and YOLOv8 models. During the work, the authors established:

1) The use of Mask R-CNN networks is most appropriate when using the 3_epoch_9 model. This combination gives the result mAP(IoU=0.5) = 0.3198.

2) According to the estimated mAP parameters, the best result is shown by YOLOv8 on the train20 model mAP(IoU=0.5) = 0.389. However, according to the results of a visual comparison of the results, Mask R-CNN on the 3_epoch_9 model shows higher recognition quality.

3) The authors conclude that this circumstance is because the dataset was created with ground-truth bitwise masks, which is why, when selecting individual segmentations, the number of crack annotations turned out to be several orders of magnitude greater than the hole annotations, which is why the dataset turned out to be heterogeneous and overloaded. Some of the cracks were missed because of the elimination of imbalances, and the YOLOv8 network was overfitted using this approach. In the future, it is planned to work on optimizing the dataset, or moving to another dataset, or creating synthetic

samples with images of road surface areas and overlaying cracks and potholes based on datasets with semantic masks of individual damage.

In conclusion, we note that it is planned to scale up the technology used based on the recognition of such defects as rutting, a network of cracks, as well as places of bitumen sweating. The solution proposed by the authors can be used in systems for operational diagnostics of the condition of highways when receiving road surface images from streaming video.

REFERENCES

Skorobogatchenko, D.A., Kamaev, V.A., Finogeev, A.G., Trung, D.T., 2016. Application of indistinct neural networks for solving forecasting problems in the road complex. *ARPN Journal of Engineering and Applied Sciences*, 11(16), pp. 9646-9654.

Farhadmanesh, M., Cross, C., Mashhadi, A.H., Rashidi, A., Wempen, J., 2021. Highway Asset and Pavement Condition Management using Mobile Photogrammetry. *Transportation Research Record*, 2675(9), pp. 296–307.

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.

Seongdeok, B., Baek, F., Park, S., Kim, W., Kim, H., 2020. Image augmentation to improve construction resource detection using generative adversarial networks, cut-and-paste, and image transformation techniques. *Automation in Construction*, 115.

Hu, Da, Chen, J., Li, S., 2022. Reconstructing unseen spaces in collapsed structures for search and rescue via deep learning-based radargram inversion. Automation in Construction, 140.

Shim, S., Kim, J., Lee, S.-W., Cho, G.-C., 2021. Road surface damage detection based on hierarchical architecture using lightweight auto-encoder network. *Automation in Construction*, 130.

Liu, F., Liu, J., Wang, L., 2022. Deep learning and infrared thermography for asphalt pavement crack severity classification. *Automation in Construction*, 140.

Guan, J., Yang, X., Ding, L., Cheng, X., Lee, V. C.S., 2021. Automated pixel-level pavement distress detection based on stereo vision and deep learning. *Automation in Construction*, 129.

Zhong, J., Zhu, J., Huyan, J., Ma, T., Zhang, W., 2022. Multi-scale feature fusion network for pixel-level pavement distress detection. *Automation in Construction*, 141.

Wang, D., Ren, B., Cui, B., Wang, J., Wang, X., 2021. Real-time monitoring for vibration quality of fresh concrete using convolutional neural networks and IoT technology. *Automation in Construction*, 123.

Weng, Z., Gulnigar, A., Wu, D., Liu, C., Li, F., 2022. Rapid pavement aggregate gradation estimation based on 3D data using a multi-feature fusion network. *Automation in Construction*, 134.

Deru, Li., Duan, Z., Hu, X., Zhang, D., Zhang, Y., 2023. Automated classification and detection of multiple pavement distress images based on deep learning. *Journal of Traffic and Transportation Engineering (English Edition)*, 10(2), pp. 276-290.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. 2017 *IEEE International Conference on Computer Vision (ICCV), IEEE*, pp: 2980-2988.

Terven, J., Cordova-Esparza, D., 2023. A Comprehensive Review of YOLO: From YOLOv1 and Beyond. *arXiv e-print*.

Girshick, R., 2015. Fast R-CNN. 2015 *IEEE International Conference on Computer Vision (ICCV), IEEE*, pp: 1440-1448.

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), pp. 1904-1916.

Shelhamer, E., Long, J., Darrell, T., 2017. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(4), pp. 640-651.

Vasil'ev, P.V., Senichev, A.V., 2020. Application of neural network technologies in the task of surface defects monitoring. *Izvestija vysshih uchebnyh zavedenij. Severo-Kavkazskij region. Tehnicheskie nauki*. 2020. no. 1 (205). pp. 33-40 (in Russ.).

Kanaeva, I.A., Ivanova, Ju.A., Spicyn, V.G., 2021. Segmentation of road surface defects based on the formation of synthetic samples using deep generative adversarial convolutional networks, *Komp'juternaja optika*. no. 6, pp. 907-916 (in Russ.).

Arya, D., Maeda, H., Ghosh, S.K., Durga, T., Sekimoto, Y., 2022. RDD2022: A multi-national image dataset for automatic Road Damage Detection. *arXiv e-print*.

Passos, B.T., Cassaniga, M.J., Fernandes, A.M., Medeiros, K.B., Comunello, E., 2020. Cracks and Potholes in Road Images. *Mendeley Data*.

Katalog tipichnyh defektov soderzhanija konstruktivnyh jelementov avtomobil'nyh dorog. Tom 1. Defekty vesenne-letne-osennego perioda tipichnyh defektov soderzhanija konstruktivnyh jelementov avtomobil'nyh dorog [Catalog of typical defects in the maintenance of structural elements of highways. Volume 1. Defects in the spring-summer-autumn period of typical defects in the maintenance of structural elements of highways]. FDS Rossii. 1998 (in Russ.).