International Journal of Artificial Intelligence,
ISSN 0974-0635; Int. J. Artif. Intell.
2024, Volume 22, Issue Number 2

Optimization of A Pavement Defects Detection System Using A Deep Learning Algorithm with Pre-train

Nguyen T. Long[1,2], **Nguyen T. Huong**[1,2,3], **Shmeleva G. Anna**[1] **and Afanasiev D. Alexander**[2]

[1]Department of Informatics, Institute for Cybersecurity and Digital Technologies,
MIREA Russian Technological University, Moscow, Russia,
nguen@mirea.ru, nguen_t@mirea.ru, ann_shmeleva@mail.ru

[2]Laboratory of Artificial Intelligence and Machine Learning
Institute of Information Technology and Data Science
Irkutsk National Research Technical University, 664074, Irkutsk, Russia
aad@ex.istu.edu

[3]University of Information and Communication Technology
Thai Nguyen University, 70000, Thai Nguyen, Viet Nam
nthuongcntt@ictu.edu.vn

ABSTRACT

Detecting defects in road pavement is a highly practical problem and of great research interest in the field of computer vision. In recent years, Deep Learning has made many important breakthroughs in object detection and recognition. In this paper, we apply a Deep Learning method using VGG16 and R-CNN Region Convolution Neural Network architectures trained on the individual HL88 dataset. The performance of the network is evaluated on the Portuguese and Russia datasets to detect crack types. The Russia dataset provides the best results for the R-CNN architecture using weights trained on a large dataset of sidewalk defects, achieving an accuracy of 96.5% and an F1 score of 97.9% in the test set.

Keywords: artificial intelligently, deep learning, road pavement defects, VGG -16, R-CNN, pre-train weight.

Mathematics Subject Classification: 68U10, 68T45.
Computing Classification System: 10010147.10010257, 10010147.10010178, 10010147.10010371.10010382.10010383, 10010147.10010178.10010224.10010245.10010247, 10010520.10010521.10010542.10010294.

1 Introduction

Currently, artificial intelligence applications are being deployed around the world to simplify work, reduce labor costs, and improve the ability for people and machines to interact and communicate. This has proven to be very useful in human life, such as object recognition and human behavior identification. One of the most important recent research trends is the automatic detection of road pavement defects by machines. In daily life, people's road safety is greatly affected by accidents or vehicle breakdowns due to road damage. Monitoring and updating the condition of roads is important and practical for engineers and construction workers. In research communities such as computer vision and computer science, scientists have conducted extensive research to enable computers to remotely monitor and detect pavement defects. This helps people plan for assessment and repair as quickly as possible. However, there are always many challenges associated with detecting pavement defects. In addition to objective factors such as weather, equipment resolution, and monitoring time, the algorithms and methods used to build the system are one of the critical factors in the accuracy of defect detection results.

In this paper, we use the VGG16 (Huong and Long, 2021) and R-CNN (Wei and Yunfei, 2022) (Yamashkina, Yamashkin and Platonova, 2022) neural network architectures and two datasets from Portugal and Moscow to train and build a system that automatically detects two types of defects: longitudinal cracks and alligator cracks. Then, the results of the meshes are compared with the different datasets. This paper is organized as follows: the first part is the introduction, part 2 is related works, part 3 is the proposed method, part 4 is the comparison of results, and the last part is the summary.

2 Related works

With the development of artificial intelligent field, many researchers have proposed a variety of detection methods for concrete pavement crack detection. A stereo vision-based crack width detection method was proposed to evaluate the crack width of concrete structures surfaces quantitatively (Baohua and Shijie, 2015). Edge-based is used to build crack detection method, which used a width map to remove noise and residual noise can be removed by reclassification of crack regions (Cho, Yoon and Jung, 2018). The concrete pavement under different noises is processed by genetic algorithm based on genetic programming (GP) and percolation model (Zhong and Chen, 2019). In (Mir and Sasaki, 2022) presented method can accurately identify the concrete surface cracks from several types of disturbance characteristics by training the classification model. Then, in order to improve the efficiency of crack detection, the algorithm scales the network model horizontally and accesses the convolution layer with the kernel size. Classification in terms of multiple features is inherently a multidimensional problem. For visual data acquired using cameras, the raw data contains a large amount of data represented as pixels. Since the computational cost and memory constraints are directly related to the number of features, it is not practical to treat each pixel as one feature when using machine learning classifiers. Therefore, several steps are taken to reduce the dimensionality of the data. These steps are defined and can be abstracted in common practice as image segmentation and feature extraction (Laith and Zhang, 2021), (Kramarov, Popov and Dzhariev, 2023). Consequently, the work flow for image classification can be represented as three steps, namely image segmentation, feature extraction, and feature classification (H. and D., 2021).

Among these steps, image segmentation is used as a tool to perform feature extraction and is generally perceived as part of the feature extraction step. The image segmentation and feature extraction steps are often performed using manual or semi-autonomous/adaptive methods that require user input. Although there are studies that aim to perform image segmentation and feature extraction, performing the steps separately leads to an accumulation of errors in classification (Ferjani and Alsaif, 2022). As a result, either the overall process involves manual decisions or the process is performed in a series of autonomous algorithms, which cumulatively increases the error. In such cases, the obtained framework is often applicable only to specific cases and loses performance as conditions change. Studies that consider the crack detection workflow as a series of operations focus on selecting and finding the optimal combination of methods.

Convolutional neural networks (CNNs) represent a different approach to the fragmentary performance of classification tasks. CNNs perform the feature extraction and classification tasks in a single frame without the need for image segmentation preprocessing. Therefore, the case-specific bias due to manual feature extraction decisions and cumulative error accumulation in CNNs is avoided. Deep learning methods based on CNN are used for crack detection in several publications (Ali and Jassmi, 2021). Two crack datasets without preprocessing (Amhaz and Chambon, 2016) were used to solve the unbalanced data problem (Wu, Zhao and Sun, 2021). The training phase uses crack images with labels, and the experiments show that training the ground truth with sparser crack labels leads to sparser initial cracks. The small sample size cannot be applied to low spatial resolution images that contain a lot of noise and artefacts, such as the 2StageCrack dataset. In addition, the method obtained a high F1 score using an unbalanced training set with twice as many negative samples as positive samples. CNN-based methods have also been investigated for road damage detection (Rezaie, Achanta and Godio, 2020). They detect road surface damage using object detection. Object detection involves finding the position of an object on the image in the form of bounding boxes and determining which class the object belongs to (Liu, Yao, Lu and Xie, 2019). In object detection, the damaged parts of the road surface are not accurately segmented. In this work, we focus on road damage detection in the form of semantic segmentation. The work in (Chun and Ryu, 2019) proposes a system architecture based on a deep convolutional neural network (CNN) for road detection and segmentation from aerial images. These images are captured by an unmanned aerial vehicle implemented by the authors. The CNN was designed using MatConvNet and has the following structure: four convolutional layers, four pooling layers, one ReLu layer, one fully linked layer, and one softmax layer. The approach can lead to a high. To improve the robustness of unstructured road detection, a new algorithm is proposed in (Liu, Xu and Hu, 2022). The work (Eduardo, Margarita and Jose, 2018) presented an implementation of short-time Fourier transform (STFT) in feature extraction and artificial neural networks (ANN) in classification for acoustic fingerprint recognition used in ship identification. The experimental results show that the proposed acoustic fingerprint based ship detection system is accurate and robust to signal variations such as noise, speed and position changes. The algorithm EMODS (Evolutionary Metaheuristic of Deterministic Swapping) was introduced by the authors in the work (Jonathan, Carlos and Luis, 2018) at the level of its mutation stage to train algorithms for each problem. The application of the method in the neural networks will generate sets of networks with optimal weights for a given problem. Training a neural network consists of estimating the parameters of the neural network by minimizing a cost function. In

this paper (Miguel, 2022), the authors offer an alternative in the context of costing in normalized spaces. The Frechet derivation and generalizations of the chain rule and the Leibniz formula were used. Backpropagation for the gradient of an arbitrary cost function results from a simple application of the chain rule. RanSac is used to different image regions, and the features are automatically extracted using convolutional neural network (CNN), and the corresponding Gaussian mixture models (GMMs) are constructed based on Orchard-Bouman, and then the Gibbs energy function is used to achieve the road recognition in subsequent images. Mathematical optimization and the theory of operations research form the core of the modern theory of artificial intelligence and machine learning (Denis, Igor and Dong, 2021).
In the paper (Deeksha and Hiroya, 2021), several Convolutional Neural Network (CNN) algorithms were used to classify road damage, and it was determined which algorithm performs better in detecting and classifying road damage. The damages are classified into three categories: Pothole, Crack and Exposure. deep education in the field of computer vision has achieved remarkable results and shown great effectiveness in many research areas. In (Sadia and Khan, 2020), evolutionary neural network algorithms for road damage detection and SVM for object classification are proposed. In previous work on road damage detection using image processing, various algorithms have been used. Here, R-CNN and faster R-CNN are used for this work to compare which algorithm is more suitable for road damage detection.

3 Methodology

The main method in this article is the use of the pre-trained model VGG16 and R-CNN (fig. 1). This method is also known as transfer learning. These are models that have been previously trained on a large dataset using state-of-the-art methods, which reduces the training effort from scratch. The problem is given an image X with an input of 128×128. The images can come from still or moving images from the camera and are then normalized before being included in the model. The output of the model consists of 2 types of defects: longitudinal crack and alligator crack. After training, real-time experiments are performed. Images from the video are used as input to the model.

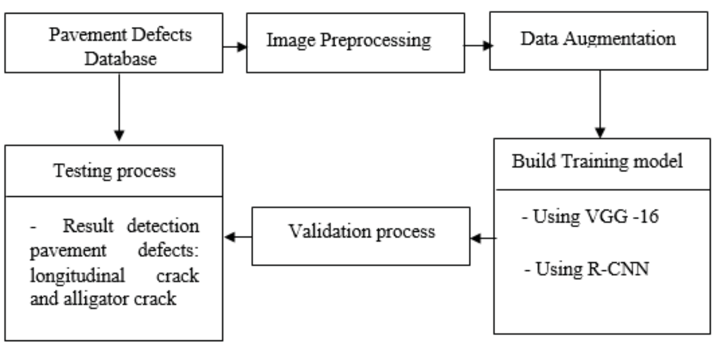

Figure 1: Data-flow of proposed method based on VGG-16 and R-CNN to detection pavement defects.

3.1 Architecture of VGG 16 and R-CNN for pavement defects detection system

The two models VGG16 and R-CNN were both trained on the dataset HL-88, with the input and output of both modified to match the output of the problem. After the feature extractor, the networks are fed into the classifier. Then it goes through the softmax activation layer to output the probability distribution of the labels.
VGG16's network architecture includes 17 layers: of which 13 Convolutional layers all have 3×3 kernels and 4 Fully connected layers are added to suit the problem. In VGG16, convolutional blocks appear, which is an architecture consisting of a set of similar repeating convolutional layers and is widely applied to later architectures. VGG16 has also changed the order of blocks when arranged into multiple layers convolutional + max pooling instead of alternating one layer convolutional + max pooling.
The R-CNN network is designed to work with hundreds of layers. A problem with many convolutional networks is the phenomenon of gradient vanishing, leading to poor training results. Therefore, the solution that R-CNN uses is shortcut connections - directly connecting the input of layer n to layer $n+1$. The architecture of R-CNN includes 3 components. Region proposal has the effect of creating and extracting proposed regions containing objects surrounded by bounding boxes. Extracts features to help identify images from proposal regions through deep convolutional neural networks. Classifier based on the input of the features in the previous section to classify images contained in the region proposal to the correct label.

3.2 Enhance datasets using Data Augmentation techniques

Data Augmentation is a technique for expanding a data set by creating additional transformed data from the original data. There are many ways to augment data, such as collecting data or synthesizing data, but data augmentation is a relatively simple approach and is strongly supported when using Deep Learning. Data augmentation not only solves the problem of insufficient data, but also solves problems such as overfitting and helps the model perform better with a variety of data samples. The data used in this article is quite imbalanced. Figure 2 depict the expected distribution of data on each label of the Train set of the data sets before image enhancement and after image enhancement.

3.3 Optimize network structure using loss-function and Adam algorithm

Loss function, also known as loss function, represents a relationship between $\widehat{y}$ (the model's predicted result) and y (the actual value). The purpose of the loss function is to optimize your model as best as possible and use it to evaluate the goodness of the model. The closer $\widehat{y}$ is to y, the better. That is, when we rely on the loss function, we can calculate gradient descent to optimize the loss function as close to 0 as possible. Categorical_crossentropy is the error function for a multi-class classification model where there are two or more output labels. The output label is converted to a one-hot encoding vector before applying this function (eq. 3.1):

$$Loss = -\sum_{i=1}^{10} y_i * \log(\hat{y}_i) \tag{3.1}$$

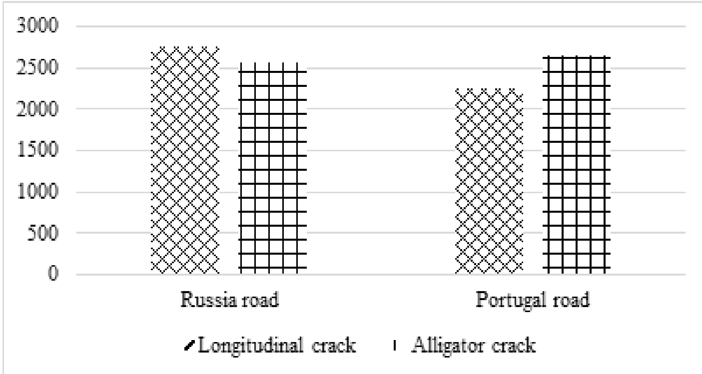

Figure 2: Result of distribution label of defect pavement after augmentation process.

Optimizer (optimization algorithm) is the basis for building models with the purpose of learning the features of input data, from which appropriate weights and biases could found to optimize the model. A method to optimize the objective function in deep learning, Adam applies different learning rates to each parameter based on the parameters β_1 and β_2. Adam optimizer is an algorithm that combines the technique of RMS prop and momentum. The algorithm uses two internal states momentum (m) and squared momentum (v) of the gradient for parameters. After each training batch, the values of m and v are updated using exponential weighted averaging (eq.3.2):

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\, g_t; v_t = \beta_2 v_{t-1} + (1 - \beta_2)\, g_t^2 \tag{3.2}$$

where β is considered a hyperparameter. The formula for updating θ is as follows (eq. 3.3):

$$\theta_t = theta_{t-1} - \alpha \frac{m_t}{\sqrt{v_t + \epsilon}} \tag{3.3}$$

where α is the learning rate, ϵ is the value added to prevent division by zero.

4 Analysis Result

4.1 Experimental environment and datasets description

In this research, Google Colab (Colaboratory) is used as a product of google research. Colab supports Python platform in Notebook form through the browser, very suitable for machine learning and data analysis tasks. data and deploy deep learning models. In addition, Colab provides powerful GPUs and TPUs to train models more easily. Currently, there are many libraries that support model deployment. In this report, Tensorflow is used as the main library to perform model training on data sets.

A dataset of 5000 images was developed by us. In this work, using smartphone cameras to collect defect datasets from the streets of Russia and and Porgutal, then the images were corrected by white balance and contrast transformation. Then the image size is changed to 128×128 pixels. After that, the images are assigned with two categories: longitudinal crack and alligator crack.

4.2 Evaluate and analyze the result of system

The CNN is used network baseline to bootstrap from scratch to compare more intuitively before using pre-trained networks. After experimenting many times on the data, the performance of the two networks VGG16 and R-CNN will be evaluated through a set of tests. The test is evaluated after the training process ends, after many executions with the data sets that have given the best results as shown in table 1. In addition, the accuracy for overall model evaluation is good. Critial F1- Score is used to evaluate the performance of the model because this is data imbalance.

Table 1: The comparision between CNN, VGG-16, R-CNN for two dataset of accuracy and F1-Score.

Database	Critial	CNN	VGG-16	R-CNN
RPD	Accuracy	94.05	96.24	95.97
	F1-Score	0.72	0.89	0.94
PPD	Accuracy	93.78	94.67	97.52
	F1-Score	0.70	0.92	0.96

It can be seen that VGG16 and R-CNN perform much better than a simple CNN network. For the Portugal set, the accuracy and F1 score reach the lowest and highest accuracy for the Moscow set, respectively. When using pre-trained weights on the individual dataset, R-CNN outperforms VGG-16 on both datasets. R-CNN achieved the highest accuracy on the Portugal dataset with an F1 score of 96.75%. This shows the importance of using the pre-trained model. If the image data does not match the data from the pre-trained model, there is a high probability that the model will converge very poorly for that data set. To get a more general overview of each data class, the confusion matrix shows how many data points actually belong to that label (table 2, 3.

Table 2: The confusion matrix of Russian pavement defects dataset.

	LC	AC
LC	1702	60
AC	135	1105

In the experiments with both the Portugal and Moscow datasets, the results were most accurately predicted for longitudinal cracking on the test set with over 97% accuracy in the two confusion matrices described in table 2, 3. Through the two tables the confusion matrix of two identical datasets, but labeled differently, has quite different results, proving that correct labeling plays an important role in data generation and training of deep learning models.

Table 3: The confusion matrix of Portugal pavement defects dataset.

	LC	AC
LC	1811	98
AC	120	976

Comparison between VGG16 and R-CNN using the individual dataset as starting weights on the Russia and Portugal pavement defects to research how the two models differ. For VGG16, the label with the highest prediction accuracy is vertical crack with 96.21% accuracy and alligator crack with 94.67% accuracy, as shown infigure 3. For R-CNN, the highest prediction

Figure 3: Result of training and testing using VGG-16 for Russia pavement defect dataset.

accuracy is for longitudinal cracks with an accuracy of 95.97% and the lowest is for alligator cracks with an accuracy of 97.52% as shown in figure 4.

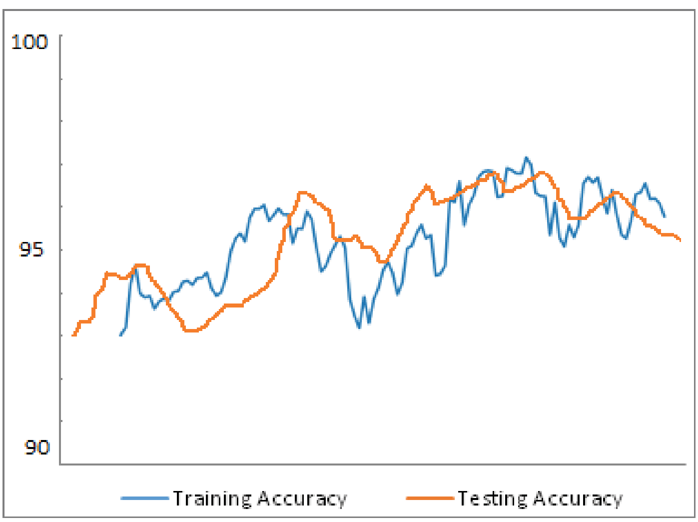

Figure 4: Result of training and testing using R-CNN for Portugal pavement defect dataset.

5 Conclusions and future work

In this study, CNN models and VGG16 and R-CNN network architectures were tested on road pavement defect detection datasets. Here, the comparison of the importance of labeling between the Portugal and Moscow datasets has a strong effect on model training. The advantage of using pre-trained models is that the model converges more easily and quickly than a pure CNN, also the initial weighting is essential for improving model training. It is a very practical problem that can be applied to many real-world problems, especially in the context of monitoring and scheduling repairs and applying safety warnings when participating in traffic. for the community. In the near future, we will focus on improving the quality of the model with new methods such as Attention and use the pytorch library to experiment and compare the results.

References

Alam, H. and Valles, D. 2021. Debris object detection caused by vehicle accidents using uav and deep learning techniques, *In Proceedings of the IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* pp. 27–30.

Ali, L. and Jassmi, H. 2021. Performance evaluation of deep cnn-based crack detection and localization techniques for concrete structures, *Sensors* **21**: 1688–1670.

Amhaz, R. and Chambon, S. 2016. Automatic crack detection on two-dimensional pavement images: an algorithm based on minimal path selection, *IEEE Transactions on Intelligent Transportation Systems* **17**: 2718–2729.

Baohua, S. and Shijie, Z. 2015. A stereovision-based crack width detection approach for concrete surface assessment, *KSCE Journal of Civil Engineering* **20**: 215–227.

Cho, H., Yoon, H. and Jung, J. Y. 2018. Image-based crack detection using crack width transform (cwt) algorithm, *IEEE Access* **6**: 60100–60114.

Chun, C. and Ryu, S.-K. 2019. Road surface damage detection using fully convolutional neural networks and semi-supervised learning, *Sensors* **19**: 155–175.

Deeksha, A. and Hiroya, M. 2021. Deep learning-based road damage detection and classification for multiple countries, *Automation in Construction* **132**: 103–115.

Denis, S., Igor, V. and Dong, Z. 2021. Editorial of special issue mathematical optimization theory, operations research and applications, *International Journal of Artificial Intelligence* **19**: 113–115.

Eduardo, Z., Margarita, G. and Jose, E. 2018. A robust application in vessel recognition based on neural classification of acoustic fingerprint, *International Journal of Artificial Intelligence* **16**: 195–213.

Ferjani, I. and Alsaif, S. 2022. How to get best predictions for road monitoring using machine learning techniques, *PeerJ Comput. Sci.* **8**: 941–952.

Huong, N. and Long, N. 2021. Deep learning methods for classification of road defects, *International Journal of Artificial Intelligence* **19**: 178–192.

Jonathan, R., Carlos, J. H. and Luis, M. G. 2018. Ernead: Training of artificial neural networks based on a genetic algorithm and finite automata theory, *International Journal of Artificial Intelligence* **16**: 214–253.

Kramarov, S., Popov, O. and Dzhariev, I. 2023. Dynamics of information in networks structured on the basis of predictive terms, *Russian Technological Journal* pp. 17–29.

Laith, A. and Zhang, J. 2021. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions, *Journal of Big Data* **53**: 875–877.

Liu, M., Xu, X. and Hu, K. 2022. Real-time unstructured road detection based on cnn and gibbs energy function, *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology* **42**: 2471–2489.

Liu, Y., Yao, J., Lu, X. and Xie, R. 2019. A deep hierarchical feature learning architecture for crack segmentation, *Neurocomputing* **338**: 139–153.

Miguel, A. M. 2022. A component free derivation of backpropagation in deep learning, *International Journal of Artificial Intelligence* **20**: 93–104.

Mir, B. A. and Sasaki, M. 2022. Machine learning-based evaluation of the damage caused by cracks on concrete structures, *Precision Engineering* **76**: 314–327.

Rezaie, A., Achanta, R. and Godio, M. 2020. Comparison of crack segmentation using digital image correlation measurements and deep learning, *Constr. Build. Mater.* **261**: 121–132.

Sadia, F. and Khan, A. 2020. Detection and classification of road damage using r-cnn and faster r-cnn: A deep learning approach, *Cyber Security and Computer Science* pp. 730–741.

Wei, Z. and Yunfei, Z. 2022. Road defect detection from on-board cameras with scarce and cross-domain data, *Automation in Construction* **144**: 104–124.

Wu, J., Zhao, Z. and Sun, C. 2021. Learning from class imbalanced data with a model agnostic framework for machine intelligent diagnosis, *Reliability Engineering & System Safety* **216**: 107–129.

Yamashkina, E., Yamashkin, S. and Platonova, O. 2022. Development of a neural network model for spatial data analysis, *Russian Technological Journal* pp. 28–37.

Zhong, Q. and Chen, Y. X. 2019. The algorithm of concrete surface crack detection based on the genetic programming and percolation model, *IEEE Access* **7**: 54564–54573.