Few-shot Classification for Image-based Crack Detection

Yan Gao[a], Haijiang Li[a*], Weiqi Fu[b], Guanyu Xiong[a]

[a] BIM for Smart Engineering Centre, School of Engineering, Cardiff University, Cardiff, CF24 3AA, UK

[b] Institute of Material Medica, Chinese Academy of Medical Sciences, Beijing, 10050, China

ABSTRACT

Current approaches for autonomous crack detection are mainly based on supervised learning, which relies on a large number of annotated examples for training. However, it requires a time-consuming and labor-intensive image acquisition process. This paper proposes a transductive approach based on the improved Prototypical Network (ProtoNet) for few-shot crack detection to solve this issue. Its architecture consists of a cross-domain feature encoder and a linear classifier. The feature embedding is achieved through pre-trained DNN backbones from ImageNet, and the transductive inference is based on Euclidean distance after embedding normalization. The approach is validated on a public dataset for automatic bridge crack detection, which can achieve over 94% mean accuracy for 2-way 5-shot crack identification in the test set through pre-trained GoogleNet backbone after fine-tuning. The approach is also tested using real bridge inspection images, demonstrating its capability of fast implementation for crack detection with weakly supervised information under practical scenarios.

Keywords: Few-shot classification; Crack detection; Domain adaption; Feature embedding; Prototypical network

1 Introduction

Autonomous crack detection is a real-world challenge in visual inspection for infrastructures due to various materials, surface coatings, and changing light and weather conditions [1]. Its approaches are mainly based on inductive supervised learning, which requires many labeled examples for model training [2]. This results in a time-consuming and labor-intensive image acquisition process. Although supervised transfer learning [3], [4] was expected to solve this issue, it tends to be overfitting with very limited labeled examples (i.e., under few-shot conditions). Meanwhile, there is little research about appropriate domain adaption for crack detection. Therefore, this work aims to propose an efficient framework and approach for few-shot crack detection, which is beneficial to practical inspection under complex circumstances with weakly supervised information.

This work proposes an approach based on the improved Prototypical Network (ProtoNet) for crack detection under few-shot conditions, consisting of a cross-domain feature encoder and a linear classifier. The Prototypical Network is improved with embedding normalization and a linear classifier. The feature embedding is achieved by domain adaption through pre-trained deep-learning models based on ImageNet. Different deep neural network (DNN) backbones are compared in the experiment. The binary classification is achieved based on transduction through a linear classifier using Euclidean distances between the query item and prototypes in the support set, and the performance can be further improved through fine-tuning.

The approach is validated on a public image set [5] for automatic bridge crack detection. A dedicated CNN architecture for this dataset can achieve an accuracy of 96.37% on the test set based on supervised learning [5]. The experiment demonstrates that the domain adaption from the ImageNet is an effective way for crack feature embedding, although there are no specific crack-related tasks in the ImageNet. The pre-trained GoogleNet and Swim-Transformer backbones perform as the best two embedding functions, achieving a mean accuracy of over 94% after fine-tuning for 2-way 5-shot crack identification on the test set, which is close to the supervised learning performance.

Furthermore, the proposed approach can be utilized for crack detection based on the split image patches to indicate the crack location, area, skeleton, and direction. The real bridge inspection images [1] are employed in the experiment, and the result demonstrates the excellent capability of fast implementation for crack detection with weakly supervised information under practical scenarios. The approach is promising to be developed for multi-defect detection using ensemble learning and region proposal.

2 Literature review

2.1 Crack detection

The most straightforward task in crack inspection is crack identification, i.e., determining whether a crack exists in an image. It is a binary classification task and can also be extended to crack detection, providing more information about the crack, such as location, area, skeleton, and direction [6]. A typical crack detection method is based on patches, which is implemented by splitting the image into patches or sliding a window on the image. Then the classifier is applied on each patch, followed by stitching them back [7], as shown in Fig. 1. Another method for object detection is based on a bounding box, such as YOLO and SSD. However, this method is not always the best choice for crack detection. The generated bounding box will involve many undefective sub-regions, e.g., using a large rectangle bounding box to mark an oblique crack.

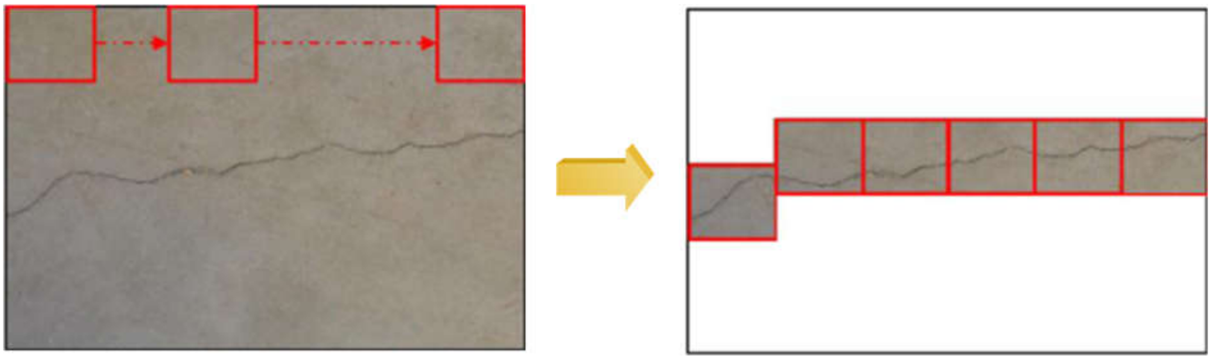

Figure 1: concrete crack detection by patch splitting and classification [7]

Many AI-based approaches have been developed for crack detection to assist routine visual inspection, including traditional machine learning (ML) and deep learning (DL). In the former, image processing for feature extraction is still required for feature extraction. Contrastingly, supervised DL can extract features automatically, such as DCNNs [5], [8]–[10] and transformers [11]. Cha et al. [12] employed supervised learning with DCNN for the first time to identify crack images without hand-crafted features. The model was trained on 40k crack and non-crack images (256 ×256) and then integrated with a sliding window technique to scan any image larger than 256 ×256 for crack detection, as shown in Fig. 1. It demonstrates the availability of the patch-based method for crack detection under practical scenarios. There are also efforts for crack detection using the bounding-box method, e.g., Xiang et al. [11] integrated a transformer module in YOLOv5 for road crack detection. However, most previous works are based on supervised learning, in which a sufficient dataset of crack and non-crack images is still required [13], [14]. Hence, a weakly-supervised approach for crack detection with only a few labeled examples becomes necessary. Although cross-domain transfer learning has the potential to solve this problem, it suffers from overfitting or difficulty in convergence under few-shot conditions. Meanwhile, some critical questions have not been solved, such as reliable source domains for crack detection through transfer learning.

2.2 Few-shot classification

Few-shot learning (FSL) was initially taken as an example of meta-learning. A meta-learner was trained through a series of works (episodic training) for unseen but related tasks with just a few examples. A few approaches have been developed for few-shot classification based on meta-learning, such as ProtoNet [15] is little research on applying meta-learning approaches for few-shot crack detection. The only related one is an attribute-based approach proposed by Xu et al. [16] for few-shot damage classification through meta-learning. However, the approach is still based on episodic training through a series of related tasks (including crack) and is not developed to the level of crack detection.

Recently, a few works [17], [18] have demonstrated that cross-domain transfer can achieve similar performance to the meta-learning approaches, such as Baseline and Baseline++ in [17], which is more efficient than episodic training. A similar work recently is an approach [19] for few-shot plant disease classification based on transfer learning across a Plant and Pest (PP) image set, which uses a transformer for feature embedding and Mahalanobis distance for evaluation. In this work, the cross-domain effect is achieved by excluding the query classes (i.e., plants and diseases) from the training set. However, as expected, this inter-class difference is insufficient because they are all plant-related images, and the approach still requires a prepared dataset of different plants and diseases. Therefore, it becomes a question to achieve few-shot crack detection using distinct cross-domain transfer learning from a popular public dataset (such as the ImageNet) to crack detection. This work proposes an approach for crack detection that does not require time-consuming data acquisition and can be quickly implemented with weakly supervised information under practical scenarios.

3 Methodology

3.1 Few-shot problem definition

In the FSL, the dataset D is separated into $D_{sup\,port}$ and D_{query} with input x and label y, shown in Eq. 1 and Eq. 2. For the N-way K-shot classification, $D_{support}$ comes from N categories with K samples per category, so there are $N \times K$ support examples. D_{query} contains samples from N categories with Q samples per category, and the goal is to classify the $N \times Q$ samples into N categories with weakly supervised information from $D_{support}$. K is usually from 1 to 5, and here, crack detection is a binary classification problem, so $N = 2$.

$$D_{support} = \{(x_i, y_i)\}_{i=1}^{I=N\times K} \tag{1}$$

$$D_{query} = \{x_j\}_{j=1}^{J=N\times Q} \tag{2}$$

3.2 Proposed approach

This work proposes a transductive approach based on the improved ProtoNet, which consists of feature embedding, transduction, and linear classification. Its architecture is shown in Fig. 2 using an example of 2-way 3-shot crack detection. The patches with the blue boundary belong to the support set, while the others belong to the query set.

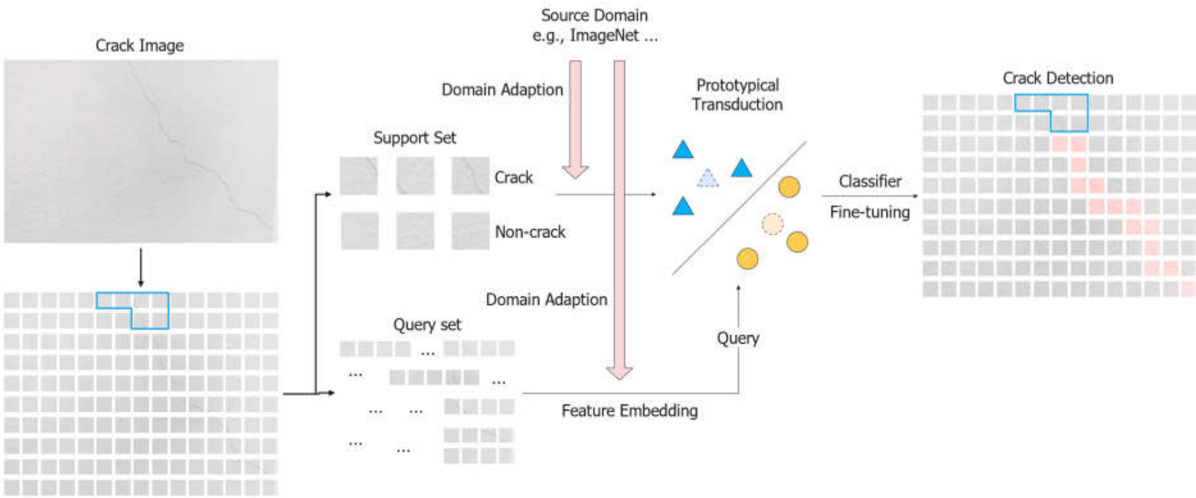

Fig. 1 Proposed approach and architecture for crack detection

The ImageNet is utilized as the source domain. The feature embedding is achieved by domain adaption using the pre-trained DL backbones based on the ImageNet rather than episodic training in the original ProtoNet. Therefore, the approach does not require a collection of defect images in advance to train a supervised model or a meta-learner. The ProtoNet is improved with embedding normalization and a linear classifier. The mean vector of support embeddings is computed as the prototype for each class.

The transductive inference is based on the prototypes and Euclidean distance (Eq. 3) through a linear classifier (Eq. 4). Here, $n = 2$ for binary classification. $d_{n\times1}$ represents Euclidean distances between a support example and each prototype. $W_{n\times n}$ and $b_{n\times1}$ are parameters and bias, respectively.

$$d = dist(v, w) = \left(\sum |v - w|^2\right)^{\frac{1}{2}} \tag{3}$$

$$y_{n\times1} = soft\,max(W_{n\times n} \cdot d_{n\times1} + b_{n\times1}) \tag{4}$$

4 Ablation study

4.1 Preparation

The approach is validated on a public dataset for automatic bridge crack detection [5], including the 2014 background and 4055 crack images (224×224). A dedicated CNN model based on supervised learning can reach 96.37% accuracy on the test set (train-test split of 80%:20%) [14]. Hence, the experiment is designed to test the performance of the proposed approach on the test set under few-shot conditions. The query accuracies are shown in a boxplot based on 5000 samplings. Furthermore, real bridge inspection images from the CODEBRIM dataset [1] are utilized to demonstrate its capability of fast implementation for practical bridge inspection.

4.2 1-shot and 5-shot classification

The experiment starts with 2-way 1-shot and 2-way 5-shot classifications. The image size is 224×224. The ResNet18 is utilized for feature embedding. Its parameters are pre-trained on ImageNet. Meanwhile, the hardcoded mean $\mu = [0.485,0.456,0.406]$ and the standard deviation $= [0.229,0.224,0.225]$, derived statistically from the ImageNet, are adopted for image transformation. The results without fine-tuning are shown in Fig. 3. As can be seen, the mean accuracy of 2-way 1-shot classification is around 80%, and 2-way 5-shot can reach over 90% accuracy with an interquartile range (IQR) of 4%. The experiment demonstrates that the approach based on cross-domain transfer learning from ImageNet is available for few-shot crack identification. As 2-way 5-shot classification performs much better than 2-way 1-shot, it is adopted for further ablation studies.

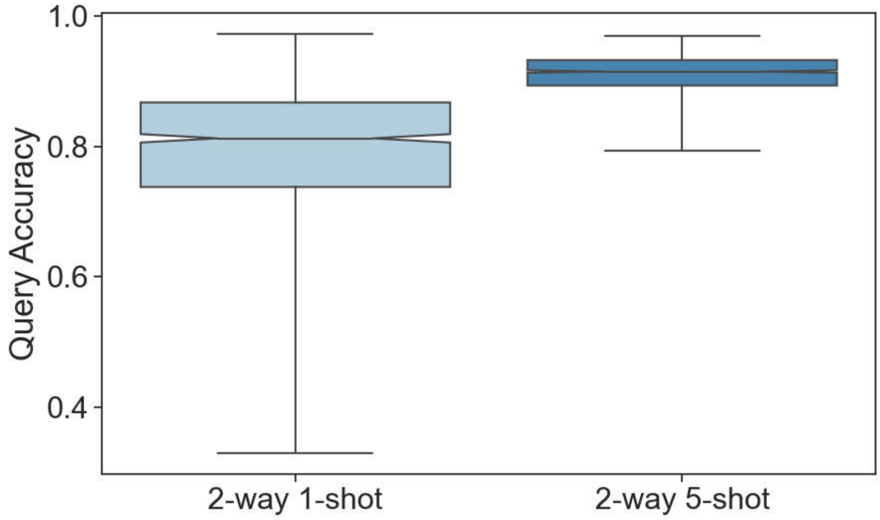

Fig. 2 2-way 1-shot and 2-way 5-shot classification performance

4.3 Different deep-learning backbones

The experiment uses different pre-trained DNN backbones based on ImageNet for feature embedding. Due to CUDA memory limitation, the image size is reduced to 84×84. The results are shown in Fig.4. As can be seen, different embedding functions have a significant impact on performance. The pre-trained GoogleNet and Swim-Transformer are the best two backbones for 2-way 5-shot crack identification in this dataset, achieving a mean accuracy of over 93% on the test set.

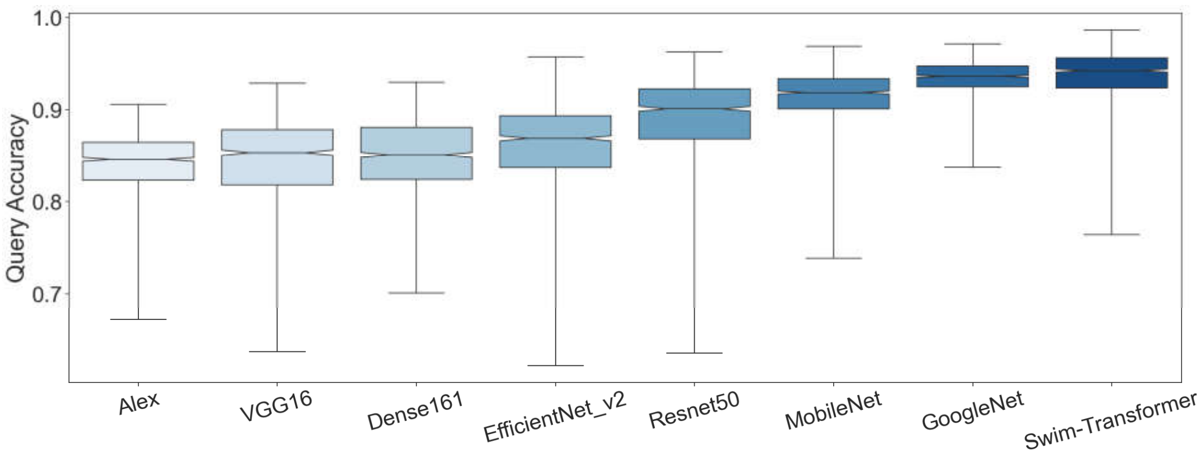

Fig. 3 Comparison of different pre-trained DL backbones from the ImageNet

4.4 Linear Classifier and Fine-tuning

Fine-tuning is taken on Eq. 4 using feature embeddings and prototypes through the pre-trained GoogleNet backbone. The RMSProp is adopted as the optimizer (learning rate 0.01). The mean query accuracy and 95% confidence interval are shown in Fig. 4.

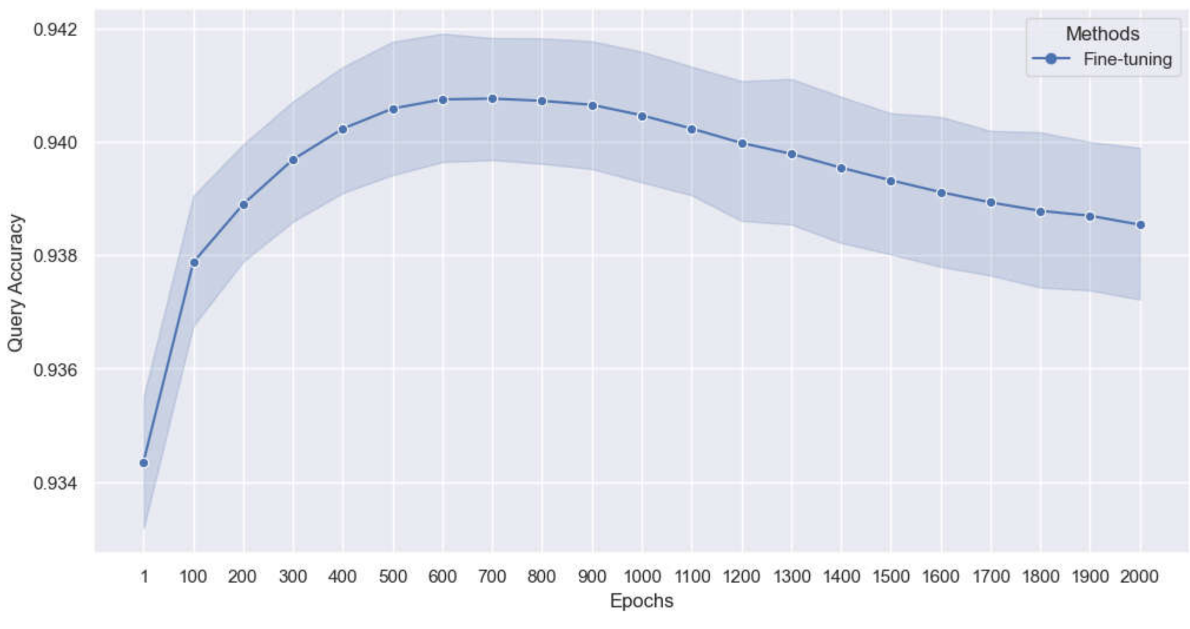

Fig. 4 Fine-tuning to improve classification performance

As can be seen, fine-tuning can enhance the mean accuracy by 0.6%, and the increment has statistical significance. Early stopping should be taken near the epoch number where the accuracy peaks, i.e., 600 epochs, which can be utilized as the empirical criteria for fine-tuning in this dataset. After the peak, the accuracy will decrease due to overfitting.

4.5 Practical crack detection

The approach is also validated using the real bridge inspection images from the CODEBRIM dataset [1]. The image is first split into patches, as shown in Fig. 5. The patches marked by blue boundary are selected as the support set, while the others are taken as the query set. The crack location and skeleton can be detected with only three labeled images for each class, as shown in the top right of Fig. 5. Moreover, the derived prototypes and classifier from the support set can be applied to a new image for crack detection without labeling, as shown in the bottom right of Fig. 5. The experiment demonstrates the approach capability of fast implementation for crack detection under practical scenarios with only a few labeled examples.

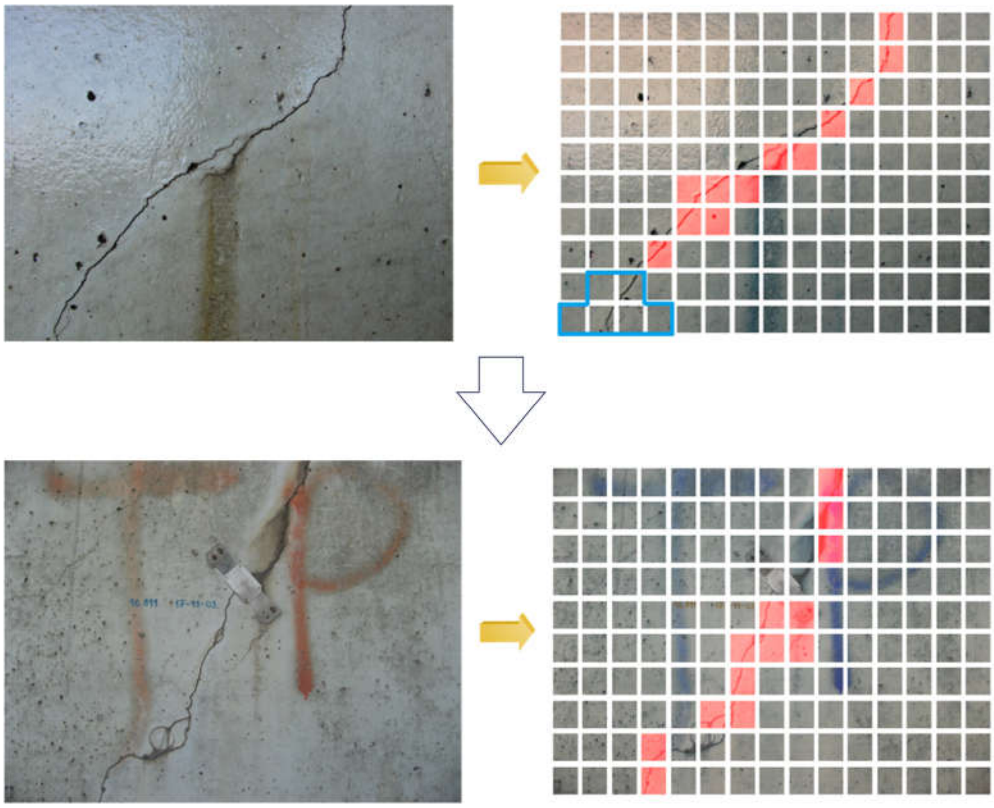

Fig. 5 Few-shot crack detection in real bridge inspection images

5 Conclusion

This work proposed an approach based on the improved ProtoNet for crack detection with only a few annotated examples. It integrates cross-domain transfer learning, transductive inference, and linear classification. The feature embedding is achieved via the pre-trained DL backbones based on ImageNet. The approach is explored on a public dataset for automatic bridge crack detection, achieving a mean accuracy of over 94% for 2-way 5-shot crack identification through the GoogleNet backbone after fine-tuning. Furthermore, the approach is also validated using real bridge inspection images, demonstrating its fast implementation capability for crack detection under practical scenarios with only a few annotated examples. The approach is also promising for the detection of other defects with weakly supervised information. However, it also has a few limitations currently. For example, the current development is only based on binary classification (i.e., defect and background), and the performance is sensitive to noise such as stains and marks. Therefore, the next step of the work is to develop the approach to the level of multi-defect detection and enhance its robustness.

6 Declaration

The authors declared that they have no conflicts of interest in this work.

7 Acknowledgment

This work was supported by the Cardiff University – China Scholarship Council (CSC) joint program.

Reference

[1] M. Mundt, S. Majumder, S. Murali, P. Panetsos, and V. Ramesh, 'Meta-learning convolutional neural architectures for multi-target concrete defect classification with the concrete defect bridge image dataset', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 11188–11197, 2019, doi: 10.1109/CVPR.2019.01145.

[2] Y. J. Cha, W. Choi, and O. Büyüköztürk, 'Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks', *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017, doi: 10.1111/mice.12263.

[3] Q. Yang, W. Shi, J. Chen, and W. Lin, 'Deep convolution neural network-based transfer learning method for civil infrastructure crack detection', *Autom Constr*, vol. 116, no. May, p. 103199, 2020, doi: 10.1016/j.autcon.2020.103199.

[4] D. Dais, İ. E. Bal, E. Smyrou, and V. Sarhosis, 'Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning', *Autom Constr*, vol. 125, no. January, 2021, doi: 10.1016/j.autcon.2021.103606.

[5] H. Xu, X. Su, Y. Wang, H. Cai, K. Cui, and X. Chen, 'Automatic bridge crack detection using a convolutional neural network', *Applied Sciences (Switzerland)*, vol. 9, no. 14, 2019, doi: 10.3390/app9142867.

[6] J. König, M. Jenkins, M. Mannion, P. Barrie, and G. Morison, 'What's Cracking? A Review and Analysis of Deep Learning Methods for Structural Crack Segmentation, Detection and Quantification', pp. 1–18, 2022.

[7] Y. J. Cha, W. Choi, and O. Büyüköztürk, 'Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks', *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017, doi: 10.1111/mice.12263.

[8] H. Fu, D. Meng, W. Li, and Y. Wang, 'Bridge crack semantic segmentation based on improved deeplabv3+', *J Mar Sci Eng*, vol. 9, no. 6, 2021, doi: 10.3390/jmse9060671.

[9] Y. Mohammed, N. Uddin, C. Tan, and Z. Shi, 'Crack Detection using Faster R-CNN and Point Feature Matching', vol. 10, no. 3, 2020, doi: 10.19080/CERJ.2020.10.555790.

[10] M. Nie and C. Wang, 'Pavement Crack Detection based on yolo v3', *Proceedings - 2019 2nd International Conference on Safety Produce Informatization, IICSPI 2019*, pp. 327–330, 2019, doi: 10.1109/IICSPI48186.2019.9095956.

[11] X. Xiang, Z. Wang, and Y. Qiao, 'An Improved YOLOv5 Crack Detection Method Combined with Transformer', *IEEE Sens J*, vol. 22, no. 14, pp. 14328–14335, 2022, doi: 10.1109/JSEN.2022.3181003.

[12] Y. J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, and O. Büyüköztürk, 'Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detecting Multiple Damage Types', *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 9, pp. 731–747, 2018, doi: 10.1111/mice.12334.

[13] Y. J. Cha, W. Choi, and O. Büyüköztürk, 'Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks', *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017, doi: 10.1111/mice.12263.

[14] H. Xu, X. Su, Y. Wang, H. Cai, K. Cui, and X. Chen, 'Automatic bridge crack detection using a convolutional neural network', *Applied Sciences (Switzerland)*, vol. 9, no. 14, 2019, doi: 10.3390/app9142867.

[15] J. Snell, K. Swersky, and R. Zemel, 'Prototypical networks for few-shot learning', in *Advances in Neural Information Processing Systems*, 2017, pp. 4078–4088.

[16] Y. Xu, Y. Bao, Y. Zhang, and H. Li, 'Attribute-based structural damage identification by few-shot meta learning with inter-class knowledge transfer', *Struct Health Monit*, vol. 20, no. 4, pp. 1494–1517, 2021, doi: 10.1177/1475921720921135.

[17] W. Y. Chen, Y. C. F. Wang, Y. C. Liu, Z. Kira, and J. Bin Huang, 'A closer look at few-shot classification', *7th International Conference on Learning Representations, ICLR 2019*, no. 2018, pp. 1–17, 2019.

[18] G. S. Dhillon, P. Chaudhari, A. Ravichandran, and S. Soatto, 'A Baseline for Few-Shot Image Classification', pp. 1–20, 2019, [Online]. Available: http://arxiv.org/abs/1909.02729

[19] S. V. Nuthalapati and A. Tunga, 'Multi-Domain Few-Shot Learning and Dataset for Agricultural Applications', *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2021-Octob, pp. 1399–1408, 2021, doi: 10.1109/ICCVW54120.2021.00161.