



Article A Novel Road Crack Detection Technology Based on Deep Dictionary Learning and Encoding Networks

Li Fan * and Jiancheng Zou

School of Electrical and Control Engineering, North China University of Technology, Beijing 100144, China; zjc@ncut.edu.cn

* Correspondence: fanli29@126.com

Abstract: Road crack detection is an important indicator of road detection. In real life, it is very meaningful work to detect road cracks. With the rapid development of science and technology, especially computer science and technology, quite a lot of methods have been applied to crack detection. Traditional detection methods rely on manual identification, which is inefficient and prone to errors. In addition, the commonly used image processing methods are affected by many factors, such as illumination, road stains, etc., so the results are unstable. In the research on pavement crack detection, many research studies mainly focus on the recognition and classification of cracks, lacking the analysis of the specific characteristics of cracks, and the feature values of cracks cannot be measured. Starting from the deep learning method in computer science and technology, this paper proposes a road crack detection technology based on deep learning. It relies on a new deep dictionary learning and encoding network DDLCN, establishes a new activation function MeLU, and adopts a new differentiable calculation method. The technology relies on the traditional Mask-RCNN algorithm and is implemented after improvement. In the comparison of evaluation indicators, the values of recall, precision, and F1-score reflect certain superiority. Experiments show that the proposed method has good implementability and performance in road crack detection and crack feature measurement.

Keywords: crack detection; deep learning; dictionary learning; encoding networks; crack measurement

1. Introduction

Crack is a common road disease, which can provide important data for road maintenance. Under the background of the rapid development of modern high-tech, computer science technology represented by machine learning and image processing technology has developed rapidly, providing new tools for the analysis and treatment of cracks. Deep learning evolved through machine learning, where models are trained through computation. Moreover, the deep learning method can extract features and analyze data according to the characteristics of the original image, and the accuracy of the calculation is higher than that of the traditional method.

Previously, the traditional method was mainly based on the texture, edge, or contrast of the crack and other characteristics to obtain the specific position and size of the column. This method has certain effects on specific datasets. Subsequent is the use of image processing methods. At present, there are many algorithms for extracting road cracks based on image processing [1]. However, the background structure of pavement crack images is very complex, there are different types of noise, and they are subject to a lot of external interference. In order to improve the accuracy of crack extraction, to reduce these interference factors, it is necessary to preprocess the original image. Sealed et al. [2] proposed an algorithm to remove pavement markers and then detected crack pixels based on a local minimum algorithm. Zou et al. [3] proposed a geodesic shadow removal algorithm, which can eliminate the pavement shadow while preserving the cracks. At the same time, a



Citation: Fan, L.; Zou, J. A Novel Road Crack Detection Technology Based on Deep Dictionary Learning and Encoding Networks. *Appl. Sci.* 2023, *13*, 12299. https://doi.org/ 10.3390/app132212299

Academic Editors: Jiangyun Li, Tianxiang Zhang and Peixian Zhuang

Received: 29 September 2023 Revised: 28 October 2023 Accepted: 28 October 2023 Published: 14 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). crack probability map is established by using tensor voting, which enhances the connection between cracks and makes the cracks have good continuity. Zhang Hong et al. [4] proposed a noise reduction and enhancement method for cement pavement. According to the noise spectrum characteristics of the picture, the relative position of the noise in the picture is inferred and filtered by the filter, which can effectively remove the noise interference of cement pavement. Ying [5] proposed an automatic detection method of pavement cracks in digital images based on wavelet transform. This method uses the pavement crack image enhancement algorithm to correct the uneven background illumination by calculating the multiplicative factor that eliminates the background illumination change. This method can effectively extract pavement cracks in a variety of pavement images.

The method discussed in this paper is a method of crack feature extraction and analysis based on deep learning. Visual feature learning based on deep learning is a core research problem in today's computer vision tasks. Its key is to quickly learn visual feature expressions with high discrimination and generalization through deep learning network models, which is the result of the joint action of sample mining, the network model, and the supervised optimization of the loss function. Among them, the loss function plays a supervisory role in the model, which realizes the training and optimization of the network model by punishing the error samples. In the current research on computer vision tasks, the algorithmic models based on deep learning widely use deep metric learning loss functions (usually applied to vision tasks such as image retrieval [6], image classification [7], and object recognition [8]) and class-based probability loss functions (usually applied to vision tasks such as object detection and image segmentation). They impact the discrimination ability and efficiency of feature learning, in different scenarios and tasks, with different challenges. The performance of the loss function is affected by sample mining. Excellent sample mining methods can quickly and automatically select informative samples from a large number of complex samples. In addition, the deep network model also affects the performance of visual feature learning. Therefore, in order to ensure the efficiency, discrimination, and generalization of visual feature learning, it is necessary to study the design and improvement of algorithm frameworks for different scenarios and challenges from the three aspects of the loss function, sample mining, and network model, so as to meet the actual application requirements. Research on the key technologies of visual feature learning aims to explore and mine the problems and challenges existing in deep learning models and improve and optimize them, so as to obtain model sub-modules (e.g., sample mining, objective function, and network structure) that can ensure the discrimination, efficiency, and generalization of feature learning.

In recent decades, benefiting from the development of artificial intelligence technology, various research efforts have been devoted to establishing road crack classification models. Researchers are attempting to apply traditional machine learning or deep learning algorithms to object crack detection [9–11]. These artificial-intelligence-based recognition methods make precise identification of cracks possible. The key to artificial intelligence is to learn features and apply the learned features to new data. Traditional machine learning algorithms such as logistic regression [12], random forest [13], artificial neural networks [14], support vector machine (SVM), etc., have been widely used in crack detection. Artificial neural network models with appropriate thresholds have been used to separate crack pixels from the background [15]. In order to overcome the shortcomings of traditional machine learning that require prior knowledge to extract target features, deep learning techniques represented by deep convolutional neural networks [16] have been widely applied in fields such as image classification, object detection, and image segmentation and have shown potential to compete with humans [17]. Unlike traditional machine learning, deep learning models automatically learn the relationship between features and tasks. In recent years, many high-performance deep convolutional neural network models based on the large universal raw dataset ImageNet [18], such as AlexNet [19], VGGNet [20], ResNet [21], and DenseNet [22], have been developed, and their classification performance is sufficient to match humans [23]. Benefiting from these deep convolutional neural network models, many crack classification algorithms [24–28] based on deep convolutional neural networks have been proposed and successfully applied in various fields [29–36].

Visual feature learning based on deep learning is considered to be a challenging task in the era of artificial intelligence, especially when deep learning is widely used and prevalent. Rapid and accurate expression of features can be widely and successfully used in social practice, which is an important research problem at present, for example, video analysis and face recognition in the field of intelligent security, scene recognition and road crack detection in the field of traffic, product search and matching recommendation in the field of daily life, image positioning and scene recognition in the field of unmanned driving, etc. The above fields require high discrimination, efficiency, and generalization of visual feature learning, which is related to people's quality of life and social security. However, current research faces different problems in different scenarios, which is far from meeting the actual needs. Therefore, the research on the key technologies of visual feature learning based on deep learning has practical needs and social significance.

Fundamentally, the ultimate goal of deep learning is to build a machine learning architecture that consists of multiple processing layers that learn multiple levels of abstraction to represent data. The use of this method greatly improves the development level of speech recognition, visual target recognition, target detection, and other subjects. Deep learning finds complex structures in large datasets and expresses them through a backpropagation algorithm.

As the most prominent deep learning algorithm in the supervised algorithm, the convolutional neural network has gradually become the most popular algorithm in the field of image recognition and a hot research tool. Scholars have applied convolutional neural networks to civil infrastructure inspection and monitoring and have made numerous achievements. The results of [37] suggest that automatic pavement crack detection is an effective intelligent health monitoring method [38]. Researchers use the research of CNN on crack detection mainly obtaining network models in three ways: one is the self-created shallow neural network; the second one is to directly use the existing network model with good performance; and the third one is based on the existing deep feature extraction network to make improvements.

In the early stage of research, scholars conduct crack detection research by self-building shallow convolutional neural network models. Zhang et al. [39] proposed the use of CNN for crack detection in 2016. Their study showed that CNN in pavement crack detection outperforms traditional machine learning techniques. However, the class labels assigned to individual pixels are still based on the local area around the pixel in context, which leads to an overestimation of the crack width. Zhang et al. [40] proposed an efficient network based on CNN in 2017. The network architecture, called CrackNet, is used to automatically detect pavement cracks on asphalt surfaces.

Popular models in recent years include the following: FCN (2018 and 2019) [41] and SegNet (2019) [42]. In the research on pixel segmentation of cracks in the past two years, a large number of scholars have targeted cracks on the basis of existing models. Seam detection has proposed its own model, such as ACNet (2020) based on CrackWeb (2021) [43] and U-HDN (2020) [44]; DeepCrack is proposed based on SegNet (2019) [45]; HCNNFP is proposed based on DeepCrack (2021) [46]; RRCE-Net is proposed based on the CE-Net (2021) [47].

In the research on pavement crack detection, most scholars mainly focus on the recognition and classification of cracks and lack the analysis of the specific characteristics of cracks. In this paper, based on the deep learning of computer science and technology, a road crack detection technique based on deep learning is proposed, which uses a new deep dictionary learning and coding network, establishes a new activation function, and uses a new differentiable programming method. The contribution of this paper is to combine deep learning with the encoding network, and on this basis, a new activation function is introduced for calculation. Of course, this is also the difference between the proposed method and other methods. Experiments show that the proposed method is operable and has good performance in road crack detection and crack characteristics measurement.

2. Methods

2.1. Deep Learning

2.1.1. Fundamental Theory

Deep learning is a computing model that expresses the relevant rules inside the sample data through continuous training and learning. Deep learning will give us a lot of help in the process of calculation, such as image parsing, sound recognition, and the use of text. Through repeated training and repeated learning, deep learning enables computers to achieve the same ability to analyze and process problems as the human brain. The effect of deep learning exceeds that of many previous technologies, and it reflects the rapid development of computer science and technology [48,49].See the next section for a detailed introduction, and see Figure 1.



Figure 1. Convolution neural network model.

2.1.2. Convolutional Neural Network

Input layer. The input layer is the structure of the nonlinear input. In image processing, the image itself is the input layer. Grayscale images or three-channel RGB images can be used as input images.

Summary layer. A parameterized convolution kernel (usually 3×3 or 5×5) is added up by multiplying it with local input features in the form of a sliding window. The main purpose of the convolution operation is to extract different features of the input GPR image, and this is achieved by the convolution operation on each convolution kernel and the input matrix. The operation of the convolution kernel mainly includes the cross-submatrix multiplication and convolution kernel matrix. The initial weights of the convolution kernel matrix were randomly generated by the system and were adjusted by using the stochastic gradient descent method in the training process. The accumulated partial values are obtained by crossmultiplication.

Pooling layer. The pooling layer is usually sandwiched between continuous convolutional layers to compress the amount of data and parameters, achieve feature dimensionality reduction, and reduce model overfitting. For GPR images, the main function of the pooling layer is to compress the image.

The active layer. There are a large number of neurons in a neural network, and the relationship between them is mathematically expressed as an activation function. The individual neurons can transfer information between each other, and it is very complex information. Pull down the layers. By adding dropout layers between active and fully connected layers, the possibility of model overfitting can be effectively reduced. The dropout layer can reduce the influence of overfitting neurons in each layer on subsequent computations by randomly disconnecting some neurons.

Connection layer. The connections of the connection layer are very comprehensive and able to map information. The construction of fully connected layers can enhance the robustness of image classification, but the fully connected layers are not sensitive to the spatial structure of the image and are not suitable for object localization and image segmentation. In addition, the parameters of the fully connected layer are large, which can easily cause overfitting.

Output layer. The output layer has different output forms for different requirements. For the classification problem, the output layer is the classifier, and the output layer is the target class. For object location and segmentation problems, the output of the neural network also includes the location of the object.

2.2. Deep Dictionary Learning and Encoding Networks

The reconstruction ability of the signal can be reconstructed by a good dictionary, so sparse representation is a key step in dictionary learning, and it has been studied since the 1990s. Dictionary learning was first proposed by S. Mallat et al. [50] in 1993. Their over-complete dictionary theory laid a solid foundation for dictionary learning. B.A. Olshausen et al. [51,52] proposed a form of dictionary learning based on sparsity, which promoted the development of sparse representation and aroused more scholars' longing for dictionary learning. Due to the fact that some dictionaries are manually designed under certain mathematical constraints and are not flexible enough to represent complex natural image structures, in recent years, researchers have shifted to learning dictionaries directly from image data and developed many dictionary learning methods [53–56]. The purpose of dictionary learning is to obtain a sparse representation of the original signal, and the resulting sparse representation has strong representation ability for the original data [57]. Dictionary learning can compress the vast majority of redundant information in existing data, thereby obtaining information with a certain utilization value. Dictionary training is the process of reconstructing raw data, that is, obtaining the corresponding dictionary matrix through training. Such dictionaries can be divided into two types. The first type is fixed dictionaries, such as the Discrete Cosine Transform (DCT) dictionary [58], Curve let dictionary [59], Wavelets dictionary [60], etc. The second type is an adaptive dictionary, which learns the features of the specified image to better achieve sparsity and is suitable for special signals.

Dictionary learning is widely used in image processing. It has been applied in many aspects in practice and can solve problems very well, such as removing noise from images, restoring images through various transformations [61], analyzing the pixel composition of images [62], distinguishing sounds [63], and classifying images according to different situations [64]. The content of this paper is crack detection to identify and analyze the crack image. We adopt a method that combines dictionary learning and deep learning, which has good representation ability and is called DDLCN [65]. DDLCN combines the advantages of both and uses a new learning encoding layer to replace the original convolution layer. The core of DDLCN is to replace the traditional convolution layer with a dictionary learning encoding layer for efficient computation.

DDLCN has better function calculation ability, which mainly stems from its ability to deal with nonlinear problems linearly. See Figure 2. It takes into account the basic vector in the coding layer and carries out a deeper layout, and the extraction of each feature is more explicit. It is layered in multiple directions, and the data representation is more optimized.



Figure 2. Framework of DDLCN.

2.2.1. Feature Extraction Layer

F is a feature extractor, and we adopt it to extract a set. The set is m-dimensional local descriptors which are defined $Y = [y_1, ..., y_l] \in \mathbb{R}^{m*l}$ from the input image *I*. The *l* represents the total number of local descriptors. We only use a single feature extractor to highlight the effectiveness of the proposed method. To further improve performance, multiple feature extractors can be used.

The feature extractor we used has its own characteristics compared with other such feature extractors, which can perform feature transformation without changing its metric boundary (SIFT) [66]. *F* is the feature extractor, $Y = [y_1, ..., y_l] \in \mathbb{R}^{m*l}$, *I* is the image, and the specific process can be described as $y_i = F(I)$, $i \in [1, ..., l]$.

2.2.2. Dictionary Learning and Encoding

Next, we proceed to the operations of dictionary learning and encoding. The number of dictionaries of the first order bits of each layer's dictionary is denoted as q, and p images are selected from them for training. r is the total number of datasets. Note that this process is random, and the number of dictionaries is denoted by $D_1 = r * q$. After that, we have the following formula:

$$\min_{V_i} \left[\frac{1}{2} ||y_i - V_i a_i||_2^2 \right] s.t._{||a_i||_1 < =\lambda}$$
(1)

Next, each encoding layer performs a subtransformation, which is the specific operation we perform for encoding. Of course, this process takes place after learning V, the body structure of the encoding, and $\gamma^1 = [\gamma_1^1, \gamma_2^1, \dots, \gamma_l^1] \in \mathbb{R}^{D_1 * l}$. In general, coding goes like this:

$$\min_{\gamma_i^1} \left[\sum_{i=1}^l \frac{1}{2} \| y_i - V \gamma_i^1 \|_2^2 + \beta \| \gamma_i^1 \Theta \zeta_i^1 \|_1 \right] s.t. 1^T \gamma_i^1 = 1$$
(2)

After multiple dictionary learning and feature encoding, we can get a deeper dictionary learning and encoding framework:

$$\min_{\gamma_i^n} \left[\frac{1}{2} \| d_i^{n-1} - D^n \gamma_i^n \|_2^2 + \beta \| \gamma_i^n \Theta \zeta_i^n \|_1 \right] s.t. 1^T \gamma_i^1 = 1$$
(3)

Through deep learning and coding, the input image can be clearly expressed, which is the operation of DDLCN for output features. In addition, what we can see is that DDLCN improves the separability and reduces the possibility of excessive error. In the equation above, ζ_i^n is one of the basis vectors adopted in the feature representation, γ_i^n is the nth

layer coding, $d_i^{n-1} \in D^{n-1}$ is at the (n-1) coding layer, and it is expressed by a vector. To remove constraints, let us do it at another level. The diagram is shown in Figure 3.



Figure 3. Multi-layer coding strategy.

As an example, let us look at a two-level framework of tasks, augmented with a transpose, which is augmented from the representation of the first level of code. This shows that we are able to obtain such a result that the feature representation can be finally realized. Here is the formula:

$$\left[\gamma_i^1(v_j), \gamma_i^1(v_j) \left[\gamma_j^2(u_1), \gamma_j^2(u_2), \dots, \gamma_j^2(u_{s_2})\right]\right]^T$$
(4)

By concatenating the codes, we can get a $D_1 \times (1 + D_2)$ -dimensional vector, which is the final encoded representation we want. Not only that, DDLCN has good performance in each task because it can perform well in iteration and can complete better feature representation [67–70].

2.3. A New Activation Function

Now that we have established our new deep learning framework, let us address some of the details. DDLCN has better depth representation ability, but in order to better transform nonlinear problems into linear problems for calculation, we need to introduce a new activation function. This activation function is improved from ReLU to avoid the vanishing gradient problem and allow more reliable parameter learning.

Why do we use a new activation function? Or why are activation functions so important? Taking the convolutional neural network (CNN) as an example, the activation function can be regarded as a special layer in the CNN model, namely the nonlinear mapping layer. After the linear transformation of the convolutional neural network, a nonlinear activation function is superimposed on the back of the convolutional neural network, and the data distribution is remapped under the action of the nonlinear activation function to increase the nonlinear expression ability of the convolutional neural network. From the perspective of imitating human neuroscience, the action process of activation functions on data in the model is to simulate the processing process of biological neurons for electrical signals. The function process of biological neurons is to set a certain threshold to activate or inhibit the received electrical signals to carry out the propagation of biological information and signals. To simulate the action process of biological neurons, the ideal activation function should be to directly output the input data into "0" and "1" through a certain threshold. However, the convolutional neural network model requires the activation function to have the properties of continuity and differentiability in the process of forward propagation and error backpropagation. Obviously, the ideal biological neuron activation function does not meet the requirements. The self-function property of the activation function determines the advantages and disadvantages of the action process. It has become an important research topic to study the properties of activation functions, analyze the correlation between the properties of activation functions and their advantages and disadvantages, and find activation functions that are efficient in time, space, and feature collection.

The activation function is the core of the deep neural network structure. Currently, common activation functions include sigmoid and tanh functions in the sigmoid system, ReLU in the ReLU system [71], LReLU functions, etc. However, there is a gradient vanishing problem during the backward transfer of functions in the sigmoid system, which greatly reduces the training speed [72]. The ReLU function can effectively alleviate the problem of gradient vanishing by training deep neural networks in a supervised manner, without relying on unsupervised layer-by-layer pre-training, significantly improving the performance of deep neural networks. Moreover, ReLU can output piecewise linear functions, meeting the characteristics of local approximation. As the number of segments increases, the approximation error of the function decreases until the required approximation error is achieved. Krizhevsky A et al. [19] tested the commonly used activation functions ReLU, sigmoid, and tanh and demonstrated that the performance of ReLU functions is superior to sigmoid system functions. The PRe LU function is an improved version of the ReLU function, which is unsaturated and can alleviate the problem of mean shift and neuronal death.

The activation function we used is an activation function called Mexican ReLU (MeLU for short) [73] that can be computed efficiently. Our ultimate goal is to be able to compute efficiently. MeLU is built from PReLU and several Mexican hat functions to achieve our goal, which is efficient computation. Its biggest feature is that it can improve the network expression ability. It has a very large number of parameters, the total is a hyperparameter, made up of many learnable parameters.

2.3.1. Mexican ReLU

MeLU is a sum of multiple Mexican hat functions and PReLU, which is a piecewise linear activation function. Its parameter is a hyperparameter. This parameter tends to be perfect, and therefore, it has good expressive ability. Meanwhile, the range of its learning parameters is infinite, which can range from 0 to infinity. Additionally, its gradient is uneven and will not saturate anywhere. It can approximate any continuous function on a compact set if and only if the number of its parameters approaches infinity. Finally, it should be pointed out that its optimization steps are more simplified because parameter changes only occur within a very small range.

MeLU is expressed by the following formula:

$$\phi_{a,\lambda}(x) = \max(\lambda - |x - a|, 0) \tag{5}$$

It can be seen that MeLU has good integration and can solve a variety of unexpected situations in training. It is called a "Mexican hat" function because it is represented graphically like a Mexican hat, and the associated derivative in the middle is decreasing, resulting in a shape that is somewhat similar to the shape of a wavelet. The relevant parameters are recursively chosen, positional determined, and trainable. MeLU is rendered by the following expression:

$$y_i = MeLU(x_i) = PEeLU(x_i) + \sum_{j=1}^{k-1} c_j \phi_{a_j,\lambda_j}(x_i)$$
(6)

2.3.2. Applications and Features

MeLU works well for transfer learning because of its features. Its expression is continuous and differently computed in every segment of the range. MeLU's approximation is reflected in its hidden layers.

MeLU has a fairly large number of parameters, which means that it can be represented adequately, but it also brings some problems. The most obvious problem that can arise is that support is overfitting at a certain point.

We use MeLU to maximize the superiority of computation in its support phase. We will introduce the specific calculation mode in the next section.

2.4. Improved Calculation Method

Earlier, we improved the activation function; now, we want to solve the calculation problem in the new framework. We employ a novel differentiable approach for deep dictionary learning. The method has improved discriminate power and jointly learns the depth metric and the associated depth transform. We use DeTraMe-Net [74] to ensure the flexibility and anti-interference of the framework structure, and it also has a strong learning ability.

2.4.1. Methodology Overview

What we need to do is rephrase the deep dictionary network to ensure that it has strong discriminate ability and can provide us with scale solutions. The first and most important thing is to transform it into a deep cascading pairing structure. Secondly, the convolution neural network and RNN are combined to achieve differentiable training. By decoupling metrics and double frame operators (pseudoinverse of dictionaries) into two independent variables, additional flexibility and computational power are introduced. Therefore, by integrating the RNN part into various CNN architectures, such as Plain CNN [75] and ResNets, different new DeTraMe networks can be obtained.

We want to make a deep connection between the DDL method and the combination of linear layers and RNNS and focus on improving its computational power. This is because converting dictionary learning to transform learning and q-metric learning and converting DDL to DeTraMe-Net is an effective way to realize the connection of linear layers and RNNS.

The q measure can also be viewed as a nonlinear activation function. The advantage is that nonlinear operators (soft max, max pooling, average pooling) are rarely used in existing architectures. The transformation is used for the dictionary, and the Q-measure uses the dual frame operator for the dictionary and finally uses two independent variables for learning. It will have a stronger discriminatory power and be able to learn the program as soon as possible.

2.4.2. Joint Depth Metrics and Transformation Learning

First, we complete the proximal explanation. We need to identify an equivalent and profound solution for M_D .

 M_D could be solved exactly by a proximal operator, and it is related to transform learning with a metric *Q*:

$$M_D(x) = prox_{\lambda\psi}^Q (Fx - c) \tag{7}$$

$$Q = D^{T}D + aI, F = Q^{-1}D^{T}, c = Q^{-1}d$$
(8)

Indeed, for every $D \in \mathbb{R}^{m * k}$, $a \in \mathbb{R}^k$, $x \in \mathbb{R}^m$; we have the following equation:

$$L^{R}(D, a, x) = \frac{1}{2} \Big(\|x\|^{2} - 2x^{T}Da + a^{T} \Big(D^{T}D + aI \Big) a \Big) + \lambda \psi(a) + d^{T}a = L^{R}(D, a, x) + \frac{1}{2} \Big(\|x\|^{2} - \|Fx\|_{Q}^{2} - \|c\|_{Q}^{2} \Big) + x^{T}Dc$$
(9)

Obviously, computing the proximity operator is equivalent to determining the optimal sparse representation of $x \in R^m$.

$$M_D(X) = \underset{a \in R^k}{\operatorname{argmin}} L^R(D, a, x) = prox_{\lambda\psi}^Q(Fx - c)$$
(10)

Next, we proceed to the multi-layer representation. Deep dictionary learning models are able to represent more comprehensively and concisely:

$$\hat{y} = \phi \circ A^{s+1} \circ prox_{\lambda\psi_s}^{Q^{(s)}} \circ A^{(s)} \circ prox_{\lambda\psi_{s-1}}^{Q(s-1)} \circ \dots \circ prox_{\lambda\psi_1}^{Q(1)} \circ A^{(1)}\left(x^{(0)}\right)$$
(11)

The affine operators $A^{(r)}$ mapping $z^{r-1} \in \mathbb{R}^{k_{r-1}}$ to $z^{(r)} \in \mathbb{R}^k$, $1 \le r \le s$, by a shift term $c^{(r)}$, by an analysis transform $W^{(r)}$, it has the following representation:

$$\forall r \in \{1, \dots, s\}, A^{(r)} : R^{k_{r-1}} \to R^{k_r}, z^{(r-1)} \mapsto W^{(r)} z^{(r)} - c^{(r)}$$
(12)

With $k_0 = m_1, W^{(1)} = F^{(1)}, \forall r \in \{2, \dots, s\},\$

$$W^{(r)} = F^{(r)} p^{(r-1)},$$

$$W^{(s+1)} = C p^{(s)}$$

$$\forall r \in \{1, \dots s\},$$

$$Q^{(r)} = \left(D^{(r)}\right)^{T} D^{(r)} + aI,$$

$$F^{(r)} = \left(Q^{(r)}\right)^{-1} \left(D^{(r)}\right)^{T},$$

$$c^{(r)} = \left(Q^{(r)}\right)^{-1} d^{(r)}.$$
(13)

Equations (11) and (12) show that our model is feasible and widely used. This effect can be achieved because we introduce some quantitative and qualitatively different measures for these operators. In FNN, the activation function is called the proximity operator of a convex function [76]. Next, it is time to adopt a method for efficient learning.

2.4.3. Measurement Expression and Framework Implementation

Reformulations (11)–(13) are very useful and have huge advantages. If we want to get enough benefits from the algorithmic framework developed by FNNs, then we must calculate the following equation accurately:

$$prox_{\lambda\psi}^{Q}(Z) = \underset{U \in \mathbb{R}^{k*N}}{\operatorname{argmin}} \frac{1}{2} \|U - Z\|_{F,Q}^{2} + \lambda\psi(U)$$
(14)

Among the rest, $\|\cdot\|_{F,Q} = \sqrt{tr((\cdot)Q(\cdot)^T)}$ is the norm, and it is Q-weighted Frobenius. It is apparent that Z is a matrix where the N samples associated with the training set have been stacked columnwise. A similar convention is used to construct X and Y from and (u) (x_i)

$$y_{1 \leq j \leq N}$$
 and $(y_j)_{1 \leq j \leq N}$

An elastic-net-like regularization is chosen by setting

$$\psi = \| \cdot \|_{1} + l_{[0,+\infty)}^{k*N} + \frac{\beta}{2\lambda} \| \cdot \|_{F}^{2} \text{ with } \beta \in (0,+\infty)$$
(15)

We observed that the last quadratic term has a positive influence on increasing stability and avoiding overfitting. Equation (14) is actually equivalent to solving the next problem:

$$\min_{U \in [0,+\infty)^{k*N}} \frac{1}{2} \|D(U-Z)\|_F^2 + \frac{\alpha}{2} \|U-Z\|_F^2 + \frac{\beta}{2} \|U\|_F^2 + \lambda \|U\|_1$$
(16)

Various iterative splitting methods can be used to discover the unique minimization of the optimized convex function [77,78]. Our goal is to develop an algorithmic solution for which classical NN learning techniques can be applied in a fast and convenient manner. We show next the following property.

The solution of Equation (16) is obtained as an iteration of the form:

$$U_{t+1} = \operatorname{Re}LU\left(\left(h1^{T}\right)\Theta Z + \widetilde{W}(U_{t} - Z) - b1^{T}\right)$$
(17)

Among the rest, \widetilde{W} is a symmetric k*k matrix, $h \in [0, 1]^k$, $b \in [0, +\infty)^k$, and one more point $1 = [1, ..., 1]^T \in \mathbb{R}^N$.

By a complete subdifferent calculus, the solution U to the problem (14) satisfies the following important optimal condition:

$$0 \in Q(U-Z) + \beta U + \lambda \partial \widetilde{\psi}(U) \tag{18}$$

where $\tilde{\psi} = \| \cdot \|_1 + l_{[0,+\infty)}^{k*N}$. For Equation (16), every $i \in \{1, \dots, k\}$, and $j \in \{1, \dots, N\}$, and the following equality holds:

$$0 \in \sum_{l=1}^{k} q_{i,l} \left(u_{lj} - z_{lj} \right) + \beta u_{i,j} + \begin{cases} (-\infty, \lambda], u_{i,j} = 0\\ \lambda, u_{i,j} > 0\\ \Phi, u_{i,j} < 0 \end{cases}$$
(19)

Let us adopt a block-coordinate approach and update the i row of U by fixing all the other ones. As Q is a positive definite matrix, $q_{i,i} > 0$, and Equation (19) implies that all of these things can be expressed and calculated.

Although W, h, and b are defined on the basis of matrix Q, for more flexibility, we will treat them as decoupled variables. Then, given independent \widetilde{W} , h, and b, they can be viewed as an RNN structure for which U_t is the hidden variable and Z is a constant input over time. By taking advantage of existing gradient backpropagation techniques for RNNs, (\widetilde{W}, h, b) can thus be directly computed in order to minimize the global loss L. This shows that q-metric learning has been recast as the training of a specific RNN.

Note that Q is a $k \times k$ symmetric matrix. In order to reduce the number of parameters and increase the ease of optimizing them, we choose a block-diagonal structure for Q. In addition, for each of the blocks, either an arbitrary or convolution structure can be adopted. Since the structure of Q is reflected by the structure of \widetilde{W} , this leads to fully connected or convolution layers where the channel outputs are linked to nonoverlapping blocks of the inputs.

We finally transformed our deep dictionary learning network approach in an alternation of linear layers and specific RNNs. This not only simplifies the implementation of the resulting DeTraMe-Net by making use of standard NN tools but also allows us to employ well-established stochastic gradient-based learning strategies.

3. Results

In this paper, an improved road crack detection technology based on Mask-RCNN is proposed by using an improved deep dictionary learning network framework DDLCN, using a new activation function MeLU, adopting a new differentiable calculation method, combining with the traditional RCNN algorithm, and then classifying and recognizing the pavement crack images. The basic feature information of the outlet surface is extracted and calculated.

3.1. Improved Algorithm and Model Constructing

3.1.1. Other Deep Learning Methods

For a more intuitive comparison, we briefly introduce the other two methods.

Improved U-Net Method

Parallel convolution module introducing cavity convolution. Set the size of the characteristic graph of the input parallel convolution module as $M \times N \times K$. The first branch road passes through 3×3 after ordinary convolution (Padding = Same); an $M \times N \times K/4$ feature map is obtained, which focuses on local information near the center point. The second branch passes 3 with a rate of 3×3 . After hole convolution (Padding = Same), an $M \times N \times K/4$ feature map is obtained, which focuses on global information. Then, stack the two outputs to get an $M \times N \times K/2$ feature map, complete the fusion of local information and global information, and help improve the accuracy of crack segmentation. After the convolution of the parallel convolution module, the ReLU activation function of the modified linear unit is added to improve the expression ability of the network. In addition, a batch normalization layer BN is added between the convolution and ReLU activation functions in the U-Net network to speed up the network training speed, enable the network to converge faster, and prevent the network from overfitting.

The attention mechanism is similar to visual attention, allowing the network to focus on the target in the image and ignore redundant information. SENet is a simple and effective channel attention mechanism with low complexity and a small amount of computation. It improves the network segmentation effect by changing the weight of the feature map channel. SENet includes three parts: squeeze, exception, and channel weighting. In order to reduce the impact of background and noise and improve the feature recovery capability of the decoder, the parallel convolution module is integrated with the SENet attention mechanism, the channel relationship of the dual scale information feature map is adjusted, the weight of the channel containing the crack feature information is increased, the weight of other channels is reduced, and the weight distribution of features such as the crack edge and shape is strengthened, so as to capture more important semantic information. It can better realize the detection of concrete cracks under a complex background [79].

Mask-RCNN Method

The Mask-RCNN algorithm is composed of a basic network, region candidate network (region proposal network (RPN)), classification network, and Mask partition network. The basic network is composed of a residual network that removes the full connection layer and uses feature pyramid networks (FPNs) to resample images, fuse the two feature maps, and output feature maps with different depths, so that the feature map has more complete image information for subsequent operations. The purpose of RPN is to use the trained parameters in the generated candidate box to provide a candidate region alignment layer of interest for the classification network (Region of Interest Align, RoIAlign). The input of RPN is the feature map after feature extraction, and many candidate boxes are generated on the feature map. RPN first realizes binary category judgment and boundary box prediction for candidate boxes and then uses the nonmaximum suppression method to filter candidate boxes, reducing invalid candidate boxes and leaving valid candidate boxes, so as to improve the efficiency of the classification network and regression tasks to be carried out later.

Compared with RoIpooling, Ro IAlign cancels quantization operation, thus reducing the impact of dislocation caused by quantization operation on algorithm accuracy. The function of the classification network is to predict the category of detected targets within the RoIAlign range and adjust the position of RoIAlign to reduce the impact on algorithm accuracy caused by the inaccurate position of RoIAlign. The principle of the mask segmentation network is to classify the target to be detected pixel by pixel based on the RoIAlign position and classification network results obtained previously [80].

3.1.2. Improved Algorithm

In order to improve the detection accuracy of the Mask-RCNN algorithm for the ground crack dataset, this paper improves the Mask-RCNN algorithm [81]. The main improvements are as follows.

Improving the RPN network. In this paper, according to the size of the pavement crack dataset, the nine target boxes with different sizes in the original RPN network are changed to 64×64 and 128×128 , which can improve the detection effect of small cracks and the detection rate of the target detection frame.

Improving the FPN network. The original Mask-RCNN algorithm uses a top-down approach to integrate high-level semantic features and low-level semantic features to improve the classification ability of FPN. In this paper, the bottom-up method is used to shorten the path of high-level semantic features to obtain low-level semantic features and improve the precise positioning ability brought by the feature pyramid architecture.

Finally, the post-processing module is added, using the target detection box located by the previously improved Mask-RCNN algorithm and using the Otsu threshold algorithm of the adaptive threshold method to further distinguish the pavement cracks and the pavement background in the detection box.

3.1.3. Model Constructing

The modified Mask-RCNN algorithm is a neural network consisting of a fully connected layer containing at least one hidden layer. In this network, each node is an artificial neuron, and the output of one artificial neuron can serve as an input to another artificial neuron, and the output of each hidden layer neuron is transformed by the activation function. A remarkable feature of MLP is the ability of the model to represent arbitrary mathematical functions when the network is large enough.

In general, the more layers of the convolution neural network, the better the data fitting effect. However, the longer the calculation time, the greater the probability of the overfitting phenomenon. For small batches of data, the number of layers does not need to be too high. In this paper, a five-layer network structure model is selected for training.

A five-layer convolution neural network is established. The first three layers are convolution layers (Con layers), which are used to extract the characteristics of the crack image. Each convolution layer is connected with a pool layer. The last two layers are the full connection layer (Fc layer). After the full connection layer, a soft max classifier is connected to achieve image classification. In order to prevent overfitting in training, each pool layer is followed by a dropout layer to make some neurons randomly inactivated, so as to ensure good results.

A total of 2000 runway images and crack images were used for the model training, from which 20% of the data were randomly selected as the validation set to improve the generalization ability of the model. The runway pictures and crack pictures we used were collected from the school runway and the cracks in and around the school. We adjusted the images to a uniform size and then performed training and detection. The environment is Python 3.9, deep learning framework PyTorch, and Windows 10 system with NVIDIA GTX 2080Ti and 48 G memory.

The shape of the runway is more regular compared to the cracks. If the method we use can first extract the features of the runway, we will be more confident in extracting the features of the cracks. By extracting the runway, we find that the method is feasible and effective, and the following is the extraction of the crack image. The results can be seen in Figures 4 and 5.



Figure 5. Extracting crack features of road.

3.2. Evaluation Indicators and Analysis

The recall, precision, and F1-score are used to evaluate the performance of the pavement crack test set images. The calculation formula is as follows:

$$Recall = TP/(TP + FN)$$

$$Precision = TP/(TP + FP)$$
(20)

 $F1 = 2 \times Precision \times Recall/(Precision + Recall)$

Formula: *TP* (true positive) represents the pixels of the correct crack, *FP* (false positive) represents the detected pixels of the crack that are actually the pixels of the crack background, and *FN* (false negative) represents the detected pixels of the crack background that are actually the pixels of the crack.

In this paper, compared with the original Mask-RCNN algorithm, the fracture detection experimental results are shown in Figure 6, and the comparison results of the evaluation indexes recall, precision, and F1 values are shown in Table 1.



Figure 6. Extracting road crack features using different methods.

	Evaluation Indicators%		
Method	Recall	Precision	F1-score
Improved U-Net algorithm	85.62	86.93	86.79
Mask-RCNN algorithm	88.57	89.56	88.39
The method of this paper	91.42	91.93	90.98

Table 1. Comparison of the crack detection results.

From the comparison of the detection results in the example diagram of Figure 6, the improved algorithm can reduce the non-crack. The interference of the seam category improves the detection accuracy and performs better than the two example algorithms.

As can be seen from Table 1, the experimental results recall, precision, and F1-score of the algorithm are higher than the improved U-NET algorithm and the original Mask-RCNN algorithm, which proves the effectiveness and advancement of the algorithm.

Crack Feature Extraction and Quantification

After obtaining the binary images of cracks and markers, it is necessary to quantitatively evaluate the cracks by calculating the area, length, and average width of the cracks. For the calculation of pixel scale, the marker size used is 296 mm \times 227 mm green rectangular paperboard, and the pixel scale is obtained by calculating the ratio of the actual marker area to the pixel area S in the binary image of the marker [82]. The calculation formula is

Scale =
$$\frac{296 \times 227}{S}$$
 mm²/ pixel² (21)

The number of crack pixels and the number of skeleton pixels are counted, respectively, to get the area and length of crack pixels, and then they are multiplied by the scale to get the actual area and length. The average width is the ratio of area to crack length.

In order to verify the performance of the measurement method in this paper, without considering the error of the human-marked dataset, the size of the manually marked crack image is taken as the real physical size.

The test results of the system proposed in this paper in the test set are shown in Table 2. The test results show that the crack detection system in this paper has high accuracy, in which the average accuracy of the area measurement of the test sample is 0.935, the average accuracy of the length measurement is 0.916, and the average width measurement accuracy is 0.879. The proportion of sample error is shown in Figure 7.



Blue: area, red: length, gray: width Figure 7. Error proportional histogram.

Error Category	Maximum Error Ratio	Minimum Error Ratio	Average Error Ratio
Area	0.185	0.002	0.065
Length	0.187	0.016	0.073
Average width	0.215	0.004	0.115

Table 2. Results of various indicators.

4. Discussion

This article proposes a road crack detection technology based on deep learning methods in computer science and technology. It relies on a new deep dictionary learning and encoding network DDLCN, establishes a new activation function MeLU, and adopts a new differentiable computing method. This technology relies on the traditional Mask-RCNN algorithm and is improved. The method proposed in this paper is operable and realizable. However, as we can see from the results, its superiority is not so obvious. The method is somewhat complex and difficult to understand, and the calculation process is also more cumbersome. However, we believe that the application of deep learning combined with a coding network is a breakthrough, whether it is the introduction of a new activation function in this paper or the optimization of any link of deep learning after the combination of the two, and it is possible to produce different effects. Of course, this needs our further exploration and discovery. In the future, it is more practical to continue to find new coding networks according to this idea and see if they can be better integrated with deep learning methods.

5. Conclusions

Road crack is one of the main disasters of road, and the detection method of road crack is very important for road maintenance. In the research on road crack detection, deep learning in computer science and technology has been gradually applied and developed rapidly. Based on the development of computer science and technology, this paper proposes a road crack detection technology based on deep learning. It relies on a new deep dictionary learning and coding network, establishes a new activation function, and uses a new differentiable calculation method. The technology relies on the traditional Mask-RCNN algorithm and is implemented after improvement. Experiments show that the method has good operability and performance in road crack detection and crack feature measurement. This technology improves the detection accuracy of the crack detection algorithm to a certain extent and can effectively detect cracks and evaluate the specific characteristics of cracks. In the future, this technology will be optimized to reduce the amount of calculation and achieve better results.

Author Contributions: Conceptualization, L.F. and J.Z.; methodology, L.F.; software, L.F.; validation, L.F. and J.Z.; formal analysis, J.Z.; resources, J.Z.; data curation, L.F.; writing—original draft preparation, L.F.; writing—review and editing, J.Z.; supervision, L.F.; project administration, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: Thanks a lot for the support of the project (project number: 110051360023XN278-13) "Collaborative Innovation Research on the Planning and Design System of Slow Moving Environment for Capital Human Settlements Based on the Concept of Intelligent Adaptation to Disabilities and Elderly" for this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. Syst. Man Cybern. 2007, 9, 62–66. [CrossRef]
- 2. Kamaliardakani, M.; Sun, L.; Ardakani, M.K. Sealed-crack detection algorithm using heuristic thresholding approach. *J. Comput. Civ. Eng.* **2016**, *30*, 04014110. [CrossRef]
- Zou, Q.; Cao, Y.; Li, Q.; Mao, Q.; Wang, S. CrackTree: Automatic crack detection from pavement images. *Pattern Recognit. Lett.* 2012, 33, 227–238. [CrossRef]
- 4. Koch, C.; Georgieva, K.; Kasireddy, V.; Akinci, B.; Fieguth, P. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Adv. Eng. Inform.* **2015**, *29*, 196–210. [CrossRef]
- Liu, F.; Xu, G.; Yang, Y.; Niu, X.; Pan, Y. Novel approach to pavement cracking automatic detection based on segment extending. In Proceedings of the 2008 International Symposium on Knowledge Acquisition and Modeling, Wuhan, China, 21–22 December 2008; pp. 610–614.
- 6. Li, Y.; Li, H.; Wang, H. Pixel-wise crack detection using deep local pattern predictor for robot application. *Sensors* **2018**, *18*, 3042. [CrossRef]
- 7. Lee, D.; Kim, J.; Lee, D. Robust concrete crack detection using deep learning-based semantic segmentation. *Int. J. Aeronaut. Space Sci.* 2019, 20, 287–299. [CrossRef]
- 8. Song, W.; Jia, G.; Zhu, H.; Jia, D.; Gao, L. Automated pavement crack damage detection using deep multiscale convolutional features. *J. Adv. Transp.* **2020**, 2020, 6412562. [CrossRef]
- 9. Dorafshan, S.; Thomas, R.J.; Maguire, M. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. *Constr. Build. Mater.* **2018**, *186*, 1031–1045. [CrossRef]
- 10. Mei, Q.; Gül, M.; Azim, M.R. Densely connected deep neural network considering connectivity of pixels for automatic crack detection. *Autom. Constr.* 2020, 110, 103018. [CrossRef]
- 11. Yang, X.C.; Li, H.; Yu, Y.T.; Luo, X.C.; Huang, T.; Yang, X. Automatic Pixel-Level Crack Detection and Measurement Using Fully Convolutional Network. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 1090–1109. [CrossRef]
- 12. Wright, R.E.; Rosenberg, S. Knowledge of text coherence and expository writing: A developmental study. *J. Educ. Psychol.* **1993**, *85*, 152–158. [CrossRef]
- 13. Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]
- 14. Jain, A.K.; Jianchang, M.; Mohiuddin, K.M. Artificial neural networks: A tutorial. Computer 1996, 29, 31–44. [CrossRef]
- 15. Cheng, H.; Wang, J.; Hu, Y.; Glazier, C.; Shi, X.; Chen, X. Novel approach to pavement cracking detection based on neural network. *Transp. Res. Rec. J. Transp. Res. Board* 2001, 1764, 119–127. [CrossRef]
- 16. Lecun, Y.; Boser, B.; Denker, J.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
- 17. Szegedy, C.; Wei, L.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
- Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F.F. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. J. Commun. ACM 2017, 60, 84–90. [CrossRef]
- 20. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
- 21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- Huang, G.; Liu, Z.; Laurens, V.D.M.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
- Yang, Q.; Shi, W.; Chen, J.; Lin, W. Deep convolution neural network-based transfer learning method for civil infrastructure crack detection. *Autom. Constr.* 2020, 116, 103199. [CrossRef]
- 24. Feng, C.; Zhang, H.; Wang, S.; Li, Y.; Wang, H.; Yan, F. Structural Damage Detection using Deep Convolutional Neural Network and Transfer Learning. *KSCE J. Civ. Eng.* 2019, 23, 4493. [CrossRef]
- Flah, M.; Suleiman, A.R.; Nehdi, M.L. Classification and quantification of cracks in concrete structures using deep learning image-based techniques. *Cem. Concr. Compos.* 2020, 114, 103781. [CrossRef]
- 26. Huang, X.; Liu, Z.; Zhang, X.; Kang, J.; Zhang, M.; Guo, Y. Surface damage detection for steel wire ropes using deep learning and computer vision techniques. *Measurement* **2020**, *161*, 107843. [CrossRef]
- Khani, M.M.; Vahidnia, S.; Ghasemzadeh, L.; Ozturk, Y.E.; Yuvalaklioglu, M.; Akin, S.; Ure, N.K. Deep-learning-based crack detection with applications for the structural health monitoring of gas turbines. *Struct. Health Monit.* 2020, 19, 1440–1452. [CrossRef]
- Zhou, S.; Song, W. Deep learning-based roadway crack classification using laser-scanned range images: A comparative study on hyperparameter selection. *Autom. Constr.* 2020, 114, 103171. [CrossRef]
- 29. Wang, S.; Zhang, P.; Zhou, S.; Wei, D.; Ding, F.; Li, F. A computer vision based machine learning approach for fatigue crack initiation sites recognition. *Comput. Mater. Sci.* 2020, 171, 109259. [CrossRef]

- Bastidas-Rodriguez, M.X.; Polania, L.; Gruson, A.; Prieto-Ortiz, F. Deep Learning for fractographic classification in metallic materials. *Eng. Fail. Anal.* 2020, 113, 104532. [CrossRef]
- Elapolu, M.S.R.; Shishir, M.I.R.; Tabarraei, A. A novel approach for studying crack propagation in polycrystalline graphene using machine learning algorithms. *Comput. Mater. Sci.* 2022, 201, 110878. [CrossRef]
- Perera, R.; Guzzetti, D.; Agrawal, V. Graph neural networks for simulating crack coalescence and propagation in brittle materials. Comput. Methods Appl. Mech. Eng. 2022, 395, 115021. [CrossRef]
- 33. An, Q.; Chen, X.; Du, X.; Yang, J.; Wu, S.; Ban, Y. Semantic Recognition and Location of Cracks by Fusing Cracks Segmentation and Deep Learning. *Complexity* 2021, 2021, 3159968. [CrossRef]
- 34. Han, X. A Novel Search Strategy-Based Deep Learning for City Bridge Cracks Detection in Urban Planning. *Autom. Control Comput. Sci.* 2022, 56, 428–437.
- Paramanandham, N.; Koppad, D.; Anbalagan, S. Vision Based Crack Detection in Concrete Structures Using Cutting-Edge Deep Learning Techniques. *Trait. Du Signal* 2022, 39, 485–492. [CrossRef]
- Alipour, M.; Harris, D.K. Increasing the robustness of material-specific deep learning models for crack detection across different materials. *Eng. Struct.* 2020, 206, 110157. [CrossRef]
- Spencer, B.F.; Hoskere, V.; Narazaki, Y. Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. Engineering 2019, 5, 199–222. [CrossRef]
- Wang KC, P.; Zhang, A.; Li, J.Q.; Fei, Y.; Chen, C.; Li, B. Deep Learning for Asphalt Pavement Cracking Recognition Using Convolutional Neural Network. *Airfield Highw. Pavements* 2017, 2017, 166–177.
- Zhang, L.; Yang, F.; Zhang, Y.D.; Zhu, Y.J. Road crack detection using deep convolutional neural network. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3708–3712.
- Zhang, A.; Wang K.C.P.; Li, B.; Yang, E.; Dai, X.; Peng, Y.; Fei, Y.; Liu, Y.; Li, J.Q.; Chen, C. Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces Using a Deep-Learning Network. *Comput.-Aided Civ. Infrastruct. Eng.* 2017, 32, 805–819.
 [CrossRef]
- 41. Li, S.; Zhao, X.; Zhou, G. Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. *Comput.-Aided Civ. Infrastruct. Eng.* **2019**, *34*, 616–634. [CrossRef]
- 42. Zhang, X.; Rajan, D.; Story, B. Concrete crack detection using context-aware deep semantic segmentation network. *Comput.-Aided Civ. Infrastruct. Eng.* 2019, 34, 951–971. [CrossRef]
- 43. Ghosh, S.; Singh, S.; Maity, A.; Maity, H.K. CrackWeb: A modified U-Net based segmentation architecture for crack detection. *IOP Conf. Ser. Mater. Sci. Eng.* 2021, 1080, 012002. [CrossRef]
- 44. Fan, Z.; Li, C.; Chen, Y.; Wei, J.; Loprencipe, G.; Chen, X.; Di Mascio, P. Automatic Crack Detection on Road Pavements Using Encoder-Decoder Architecture. *Materials* **2020**, *13*, 2960. [CrossRef]
- 45. Liu, Y.; Yao, J.; Lu, X.; Xie, R.; Li, L. DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing* **2019**, *338*, 139–153. [CrossRef]
- 46. Zhu, Q.; Dinh, T.H.; Phung, M.D.; Ha, Q.P. Hierarchical Convolutional Neural Network With Feature Preservation and Autotuned Thresholding for Crack Detection. *IEEE Access* 2021, 9, 60201–60214. [CrossRef]
- 47. Li, G.; Li, X.; Zhou, J.; Liu, D.; Ren, W. Pixel-level bridge crack detection using a deep fusion about recurrent residual convolution and context encoder network. *Measurement* **2021**, *176*, 109171. [CrossRef]
- 48. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444. [CrossRef]
- 49. Chen, Z. Research on Hand Gesture Recognition Based on Depth Convolution Neural Network; Shaanxi Normal University: Xi'an, China, 2016; pp. 5–12.
- 50. Mallat, S.; Zhang, Z. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **1993**, *41*, 3397–3415. [CrossRef]
- Olshausen, B.A. Learning sparse overcomplete representations of time-varying natural images. Proc. IEEE Int. Conf. Image Process. 2003, 41–44.
- 52. Sallee, P.; Olshausen, B.A. Image denoising using learned overcomplete representations. In Proceedings of the 2003 International Conference on Image Processing (Cat. No. 03CH37429), Barcelona, Spain, 14–17 September 2003; p. III-381. [CrossRef]
- 53. Tariyal, S.; Majumdar, A.; Singh, R.; Vatsa, M. Deep dictionary learning. IEEE Access 2016, 4, 10096–10109. [CrossRef]
- 54. Garcia-Cardona, C.; Wohlberg, B. Convolutional dictionary learning: A comparative review and new algorithms. *IEEE Trans. Comput. Imaging* **2018**, *4*, 366–381. [CrossRef]
- 55. Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **2006**, *54*, 4311–4322. [CrossRef]
- Wu, X.; Zhang, L.; Buades, A.; Li, X. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. J. Electron. Imaging 2011, 20, 023016. [CrossRef]
- Lodetti, S.; Ritzmann, D.; Davis, P.; Wright, P.; Brom, H.v.D.; Marais, Z.; Have, B.T. Wavelet-Based Sparse Representation of Waveforms for Type-Testing of Static Electricity Meters. *IEEE Trans. Instrum. Meas.* 2022, *71*, 9001010-1–9001010-10. [CrossRef]
- 58. Yamada, H. A pioneering study on discrete cosine transform. Commun. Stat.-Theory Methods 2022, 51, 5364–5368. [CrossRef]
- Niu, L.; Wu, M. Sparse representation and reconstruction of gravity data based on redundancy dictionary from Curvelet transform. Wutan Huatan Jisuan Jishu 2018, 40, 631–636.

- MF, R.; Mohammed, E.; Mohammed, A. Medical Image Denoising based on Log-Gabor Wavelet Dictionary and K-SVD Algorithm. Int. J. Comput. Appl. 2016, 141, 27–32.
- Yan, C.; Li, L.; Zhang, C.; Liu, B.; Zhang, Y.; Dai, Q. Cross-modality bridging and knowledge transferring for image understanding. *IEEE Trans. Multimedia* 2019, 21, 2675–2685. [CrossRef]
- 62. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
- Jiang, Z.; Lin, Z.; Davis, L.S. Learning a discriminative dictionary for sparse coding via label consistent K-SVD. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011; pp. 1697–1704.
- 64. Fei-Fei, L.; Fergus, R.; Perona, P. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. *Comput. Vis. Image Understand.* **2007**, *106*, 59–70. [CrossRef]
- 65. Tang, H.; Liu, H.; Xiao, W.; Sebe, N. When Dictionary Learning Meets Deep Learning: Deep Dictionary Learning and Coding Network for Image Recognition With Limited Data. *IEEE Trans. Neural Networks Learn. Syst.* **2021**, *32*, 2129–2141. [CrossRef]
- 66. Lowe, D.G. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 2004, 60, 91–110. [CrossRef]
- 67. Cortes, C.; Vapnik, V. Support-vector networks. Mach. Learn. 1995, 20, 273–297. [CrossRef]
- Wang, J.; Yang, J.; Yu, K.; Lv, F.; Huang, T.; Gong, Y. Locality-constrained linear coding for image classification. In Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–18 June 2010; pp. 3360–3367.
- 69. Song, J.; Xie, X.; Shi, G.; Dong, W. Multi-layer discriminative dictionary learning with locality constraint for image classification. *Pattern Recognit.* **2019**, *91*, 135–146. [CrossRef]
- 70. Chang, C.-C.; Lin, C.-J. LIBSVM: A library for support vec-tor machines. ACM Trans. Intell. Syst. Technol. 2011, 2, 1–27. [CrossRef]
- 71. Nair, V.; Hinton, G.E. Rectified linear units improve restricted Boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning(ICML), Haifa, Israel, 21–24 June 2010; pp. 807–814.
- 72. Hochreiter, S. The vanishing gradient problem during learning re-current neural nets and problem solutions. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **1998**, *6*, 107–116. [CrossRef]
- 73. Maguolo, G.; Nanni, L.; Ghidoni, S. Ensemble of convolutional neural networks trained with different activation functions. *Expert Syst. Appl.* **2021**, *166*, 114048. [CrossRef]
- Tang, W.; Chouzenoux, E.; Pesquet, J.-C.; Krim, H. Deep transform and metric learning network: Wedding deep dictionary learning and neural network. *Neurocomputing* 2022, 509, 244–256. [CrossRef]
- 75. Springenberg, J.T.; Dosovitskiy, A.; Brox, T.; Riedmiller, M. Striving for simplicity: The all convolutional net. *arXiv* 2014, arXiv:1412.6806.
- 76. Combettes, P.L.; Pesquet, J.C. Deep neural network structures solving variational inequalities. *Set-Valued Var. Anal.* 2020, 28, 491–518. [CrossRef]
- 77. Boyd, S.; Vandenberghe, L. Convex Optimization; Cambridge University Press: London, UK, 2004.
- Komodakis, N.; Pesquet, J.C. Playing with duality: An overview of recent primal dual approaches for solving large-scale optimization problems. *IEEE Signal Process. Mag.* 2014, 32, 31–35. [CrossRef]
- Li, H.; Wu, Z.; Nie, J.; Peng, B.; Gui, Z.-C. Automatic detection algorithm of airport pavement cracks based on depth image. J. Transp. Eng. 2020, 20, 250–260.
- Yang, F.; Zhang, L.; Yu, S.; Prokhorov, D.; Mei, X.; Ling, H. Feature pyramid and hierarchical boosting network for pavement crack detection. *IEEE Trans. Intell. Transp. Syst.* 2019, 21, 1525–1535. [CrossRef]
- 81. You, J. Pavement crack detection based on improved Mask RCNN Television Technology. Dian Shi Ji Shu 2022, 46, 7–9.
- 82. Zhang, S.; He, Y.; Zhou, X. Road surface crack detection method based on deep learning. *Sci. Technol. Eng.* 2021, 21, 6380–6385.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.