

Review

A Comprehensive Review of Deep Learning-Based Crack Detection Approaches

Younes Hamishebahar ^{1,*} , Hong Guan ^{2,*} , Stephen So ² and Jun Jo ¹

¹ School of Information and Communication Technology, Gold Coast Campus, Griffith University, Southport, QLD 4222, Australia; j.jo@griffith.edu.au

² School of Engineering and Built Environment, Gold Coast Campus, Griffith University, Southport, QLD 4222, Australia; s.so@griffith.edu.au

* Correspondence: younes.hamishebahar@griffithuni.edu.au (Y.H.); h.guan@griffith.edu.au (H.G.)

Abstract: The application of deep architectures inspired by the fields of artificial intelligence and computer vision has made a significant impact on the task of crack detection. As the number of studies being published in this field is growing fast, it is important to categorize the studies at deeper levels. In this paper, a comprehensive literature review of deep learning-based crack detection studies and the contributions they have made to the field is presented. The studies are categorised according to the computer vision aspect and at deeper levels to facilitate exploring the studies that utilised similar approaches to address the crack detection problem. Moreover, the authors perform a comparison between the studies which use the same publicly available data sets, in order to find the most promising crack detection approaches. Critical future directions for research are proposed, based on these reviewed studies as well as on trends and developments in areas similar to the crack detection area.

Keywords: structural health monitoring; crack detection; deep learning; image classification; object recognition; semantic segmentation



Citation: Hamishebahar, Y.; Guan, H.; So, S.; Jo, J. A Comprehensive Review of Deep Learning-Based Crack Detection Approaches. *Appl. Sci.* **2022**, *12*, 1374. <https://doi.org/10.3390/app12031374>

Academic Editors: Visar Farhangi, Moses Karakouzian and Nikos D. Lagaros

Received: 16 December 2021

Accepted: 24 January 2022

Published: 27 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/>).

1. Introduction

Infrastructure, such as highways, road networks, bridges, and dams, are key assets of any society. Structural defects caused by ageing and environmental variations can adversely affect the reliability of structures highlighting the importance of infrastructure monitoring strategies. By and large, infrastructure management has the fundamental goal of preserving and extending the service life of long-standing structures [1]. To maintain the integrity of structures, it is important to detect the onset of any defect. Cracks are the most common defects on pavements and the surface of concrete structures and this structural degradation can potentially threaten safety and reduce service life [2].

The first and most common procedure to evaluate the health state of a structure is to perform visual inspection, which is costly and labour-intensive. Moreover, manual inspection requires highly trained experts in the field and is subject to the judgement of the specialist. These limitations have motivated work on automatic crack-detection approaches in both industry and academia [3,4]. Image-based approaches are considered to be more cost-effective because of the widespread availability of cameras and smartphones [5].

To automate the image-based crack detection procedure, computer vision techniques have been found to be effective and their application has become a research challenge in the past decades [2]. The main step of performing crack detection using computer vision techniques is to extract crack sensitive features which can be done by leveraging either Image Processing techniques (IPTs) or deep architectures. Accordingly, image-based crack detection studies are generally categorised into (i) manual and (ii) automatic feature extraction-based approaches.

Simply speaking, crack sensitive features are manually extracted via the application of traditional IPTs. Early studies concentrated on the application of methods such as edge detectors [6], morphological operations [7], and thresholding [8]. However, recent studies have taken advantage of the classification ability of classical machine learning algorithms. The extracted handcrafted features are fed to classical machine learning algorithms such as support vector machines (SVMs) [9], Random Forest [10], and neural networks (NNs) [11] to improve the accuracy of approaches in this category. Despite the wealth of research in this category, given that the main performance comes from the chosen handcrafted feature, handcrafted feature-based crack detection approaches are still sensitive to noise and different illumination conditions [4]. This problem has led to the application of deep architectures for the task of crack detection where no handcrafted features are involved.

Deep learning is a branch of machine learning in which NNs with deep layers are used to gradually perform feature extraction and the architecture itself decides which features are relevant [12]. Numerous deep learning-based crack detection studies have been proposed which are categorised into studies based on (i) image classification (IC), (ii) object recognition (OR), and (iii) semantic segmentation (SS) settings [13]. Mainly, the difference between each setting is the level at which the crack detection is performed (e.g., image-, image patch-, or pixel-level). To gain insight into the application trend of each settings in the crack detection area, Figure 1 is provided. As can be seen, the application of SS has become the most popular setting in the crack detection area in the past years. It is obvious that the number of published studies based on deep learning architectures in the crack detection area is growing fast. Thus more detailed review and categorization of these studies at deeper levels and from different points of view is necessary, to help interested scholars to more fully understand and compare the proposed approaches.

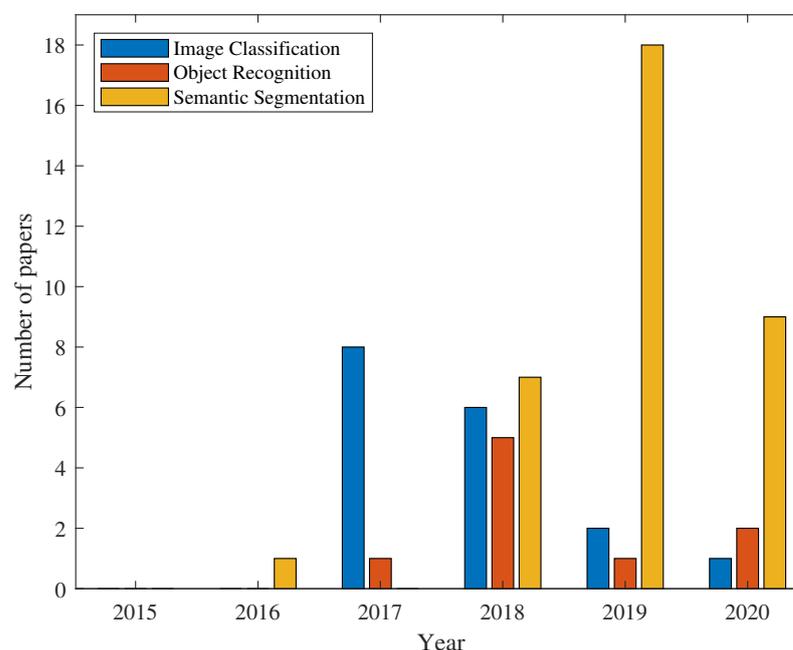


Figure 1. The number of published studies in each setting by year of publication between years 2016 and 2020.

To the best of the authors' knowledge, only three review papers have been published in this research area. Gopalakrishnan [14] provided a narrative review of deep learning-based crack detection approaches used before 2018. In addition, a comparison was performed between deep learning software frameworks, network architectures, and hyper-parameters employed by each study reviewed. Prior to 2018, there were few studies of crack detection with deep architecture. In [15], in addition to reviewing deep learning-based crack detection approaches, a comparison between three types of 3D data representation is performed.

They also reviewed traditional and deep learning-based crack detection methods that employed 3D data in detail. However, a limited number of studies have been reviewed in this study. Although in [13], a higher number of crack detection studies have been considered for review in compare to [14] and [15], these authors only categorised the published studies in the three general categories. As the number of studies in this area is increasing, it is required to further categorize the studies into sub-categories.

Therefore, in this paper, a new and more comprehensive categorization of deep learning-based crack detection studies is provided where not only are more studies covered, but categorization at deeper levels also is performed. The goal is to provide readers with different categories of deep learning-based crack detection approaches so that exploring newly proposed approaches and finding approaches similar to them can be simplified. Moreover, compared to the studies mentioned above, more recently published crack detection studies are critically reviewed and the main novelties and contributions of each study to the crack detection area are extracted and discussed. Given that the main ideas behind crack detection approaches are inspired by proposed approaches and techniques in the computer vision area, categorization of the crack detection studies based on the well-known deep architectures in the computer vision area is also considered in this paper. To give the interested readers insights into the selection of the state-of-the-art crack detection approaches among the reviewed studies, a set of criteria is selected and extracted from the studies that have performed crack detection on the same publicly available data sets. Moreover, critical directions for future work, based on the information provided in this review paper, are suggested.

The rest of this paper is organised as follows. Section 2 reviews the main contribution of the deep learning-based crack detection approaches reviewed in this paper in detail. Section 3 provides details of two publicly available data sets in the crack detection area that have been widely employed by different studies. Section 4 provides detailed information about similar studies in the SS category. In Section 5, future directions and possible research gaps in the crack detection area are identified.

2. Automatic Feature Extraction

Farrar and Worden [16] state that to develop a structural health monitoring system based on classical machine learning algorithms, four steps are required to be taken including (i) operational evaluation, (ii) data acquisition, (iii) feature extraction, and (iv) statistical modelling. Operational evaluation attempts to provide answers to questions related to the life-safety and economical justification of developing such systems. In addition, it addresses what types of damage are of concern and in case of multiple damage scenarios, which types are of most concern. Financial limitations play a strong role at the second step, which involves choosing the data acquisition hardware. Feature extraction is defined as deriving informative and non-redundant values from the raw data acquired in the previous step to facilitate the subsequent training and generalisation step. Lastly, the training and evaluation of the statistical machine learning algorithm is known as the statistical modelling step. Given the difference between data modelling using classical machine learning and deep learning algorithms, the aforementioned framework can be modified for the task of deep learning-based crack detection.

Deep architectures are called the new generation of neural networks in which multiple deep layers are used to gradually extract high-level features from raw inputs [12]. Deep learning models have been utilised in different areas such as speech and audio processing [17], language processing [18], information retrieval [19], and computer vision [20]. The most important characteristic of deep architectures can be implied from the aforementioned definition, which is the fact that the process of feature extraction is carried out by the architecture itself, without necessitating the step of engineering handcrafted features. A transition from manually extracting handcrafted features to the application of deep learning architecture for automatic feature extraction has occurred in different areas, such as medical imaging [21], and video analysis [22], as well as crack detection [13–15]. Given

the provided definition for deep architectures and the proposed framework by Farrar and Worden [16], the procedure of developing a deep learning-based crack detection system can be summarised in three steps where feature extraction and statistical modelling are performed in one step. A summarised diagram is shown in Figure 2.

The studies focusing on the application of deep architectures to perform the task of crack detection can be categorised in different ways. In [13], they are categorised based on the final output of the approach. According to this categorization, by and large, they are classified in three groups of (i) IC, (ii) OR, and (iii) SS studies. To summarise, the categorisation of image-based crack detection approaches is shown in Figure 3.

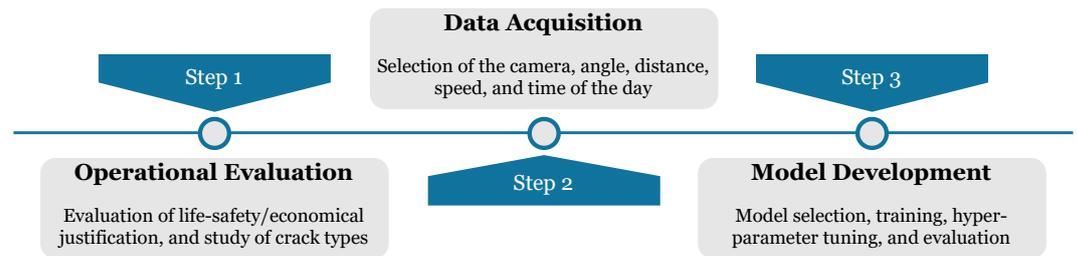


Figure 2. Developing a deep learning-based crack detection system framework.

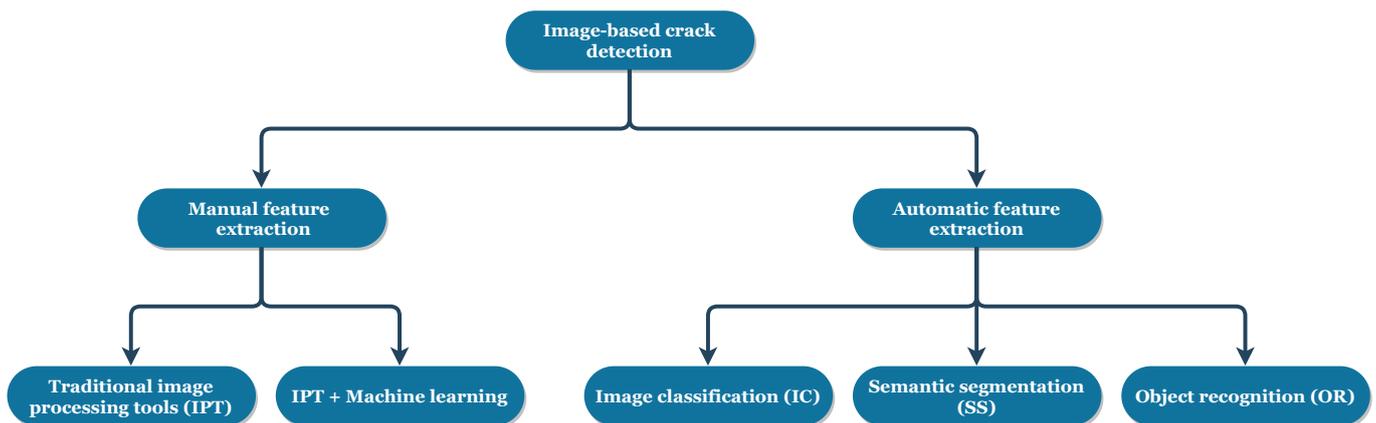


Figure 3. Categorization of image-based crack detection approaches.

2.1. Image Classification

In deep crack detection approaches based on the IC settings, the decision is limited to image/image patch-level and the trained architecture decides whether the new input is a crack contained one or not. Once the classification is made, the overall skeleton of the crack can be achieved by stacking the positive patches. A typical output of a crack detection approach based on the IC setting is shown in Figure 4. The overall architecture that is employed to perform IC consists of two parts. The first part is responsible for extracting meaningful features layer by layer from raw images, which is formed by successive convolutional and max-pooling layers. The second part consists of fully connected layers and is responsible for classifying the features extracted by the first part into certain classes [23].

To the best of the authors’ knowledge, the first application of deep learning to perform crack detection with an IC setting was performed on 2017 [24], which investigated crack detection of concrete images where the image data set includes five different classes. The considered classes are cracked, chalk letter, joint, surface, and other. After this first application, researchers started investigating different aspects of the problem. Pauly et al. [25] studied the effect of depth (the number of deep layers) on the crack detection capability of the network. They concluded that the deeper the networks are, the more information can be learnt by the architecture. In [26], four architectures with different receptive fields were utilised and compared.



Figure 4. Typical output of a deep learning-based crack detection approach based on the IC setting [23].

The main research objective, in [27,28], was a direct comparative study between crack detection approaches based on handcrafted features and deep architectures. The crack detection performance of different edge detectors, e.g., Sobel, Canny, Prewitt, Butterworth, etc., was compared with the application of deep architectures in an IC setting. Kim et al. [29] performed a comparative study between the application of the “speeded-up robust features” method based on the Hessian matrix and Haar wavelet and automatic feature extraction using convolutional neural network (CNN) for the task of crack detection.

Several studies have focused on looking at the problem from a multi-class classification point of view and considering crack type classification in addition to crack detection [26,30–32]. Wang and Hu [31] investigated crack type classification using the principal component analysis (PCA) algorithm after crack detection using CNN. ResNet [33] architecture was employed by Feng et al. [30] to detect and classify three types of defects on concrete surface images. In the case of road images, a multi-class classification approach based on CNN was proposed by Park et al. [32] resulting in classifying road areas into the crack, road marking, and intact areas. In [26], in addition to classification of crack type into five classes, inspired by AlexNet [34] and LeNet [35], four CNNs with different depths were considered and compared.

It is well-known that the training of deep architectures requires a large amount of annotated data, making their application in different areas challenging. To solve this challenge, it has been proven that the application of pre-trained networks on large-scale annotated image data sets (e.g., ImageNet), which is known as transfer learning could be effective [36]. Gopalakrishnan et al. [37] utilised the pre-trained VGG-16 [38] architecture on ImageNet data set to extract features and inputted them into different classical machine learning algorithms to perform crack detection. In [39,40], the pre-trained VGG-16 [38] and AlexNet [34] on ImageNet data set, respectively, were employed to extract features. Then, the extracted features were fed to fully connected layers and a soft-max layer to perform the classification instead of classical machine learning algorithms. Dorafshan et al. [41] performed a comparative study between the application of transfer learning concept and trained from scratch in the crack detection area with an IC setting.

Several studies with unique research objectives have been published in the IC category. Cha et al. [23] used MatConvNet [42] to perform crack detection in concrete images. They proposed the application of a sliding window technique resulting in scanning images with any resolution in the testing phase. The specific suggested architecture was employed in [43] to propose a UAV-based crack detection system. As the novelty in this study, in addition to crack detection in the images, the geo-localization of cracks was also performed. Wang et al. [44] considered a large data set (5000 3D pavement images) including diverse examples to help the architecture to learn the possible complexity and diversity of the road surface. The German asphalt pavement distress (GAPs) data set was collected and presented as a publicly available data set in [45]. LeNet [35] and ANIVOS (inspired by both

VGG-16 [38] and AlexNet [34]) architectures were employed to perform crack detection on the proposed data set.

In the crack detection area, in addition to the detection of crack images/image patches using an IC setting, the localization of cracks in the images has high importance. Although a considerable research has been conducted on the application of this setting for crack detection, the main drawback is that the overall skeleton of cracks in the images, obtained by stacking the positive patches together (Figure 4), is coarse and fails to provide a detailed localization of the cracks. A summary of the studies of deep crack detection approaches based on the IC setting is shown in Table 1.

Table 1. Summary of deep crack detection approaches based on the IC setting.

Ref	Novelty/Novelties	Core Architecture
[24]	The first application of deep learning for crack classification using an IC setting.	-
[23]	Application of a sliding window technique resulting in scanning images with any resolution at the testing phase.	MatConvNet
[25]	Investigation of the effect of the number of layers on the performance.	-
[31]	1. Investigation of the effect of patch size on the performance. 2. Crack type classification using PCA.	-
[30]	1. Crack type classification. 2. Applying active learning in the training phase.	ResNet
[37]	1. Application of the pre-trained VGG-16 on ImageNet data for feature extraction. 2. Application of classical machine learning algorithms to classify the extracted features.	VGG-16
[44]	Utilization of a larger data set with complex and diverse pavement surfaces (3D images).	-
[45]	1. Application of a shallow network (inspired by LeNet) and ANIVOS architecture (inspired by VGG and AlexNet) to perform deep crack detection. 2. Collecting and making GAPs data set publicly available.	LeNet, AlexNet, VGG-models
[27]	IC of road images into crack and non-crack images using deep learning and comparing the performance with handcrafted features-based approaches.	-
[41]	Performing a comparison between the application of AlexNet for crack detection in two settings of (i) training from scratch and (ii) transfer learning.	AlexNet
[28]	Performing a comparative study between the application of deep architectures relying on transfer learning and different edge detectors to perform crack detection.	AlexNet
[43]	In addition to crack detection, performing crack geo-localization using ultrasonic beacon system to cope with the limitations of GPS.	MatConvNet
[39]	Performing crack detection in an IC setting using CNN and relying on transfer learning.	VGG-16
[40]	1. Application of pre-trained AlexNet on ImageNet data to perform crack detection. 2. Considering a richer data set where non-crack objects are included.	AlexNet
[29]	Performing a comparison between two frameworks of crack detection using (i) fully connected CNN and (ii) based on speeded-up robust features method.	AlexNet
[32]	Focusing on crack detection in black-box images and classifying images into the crack, road marking, and intact areas.	-
[26]	1. Investigation of the effect of receptive field size on the performance. 2. Multi-class classification of different types of cracks.	Inspired by AlexNet and LeNet

2.2. Object Recognition

OR is defined as a collection of related tasks to identify objects of interest in images using bounding boxes [46]. In the computer vision area, its application has improved the accuracy of object detection and localization in the images compared with the application of the sliding window technique for the same purpose [47]. In the crack detection area, the output of this framework is the input images where the crack areas are detected using bounding boxes as can be seen in Figure 5.

In the area of computer vision, there are several families of architectures that are used to perform OR, such as region-based convolutional neural networks (R-CNN) [47], you only look once (YOLO) [48], and single shot detector (SSD) [49]. Members of the R-CNN family have been most used in the area of crack detection. To the best of the authors' knowledge, SSD and YOLO architectures are only used once to perform crack

detection as the main architecture. However, as rivals to the main architecture for the sake of comparison, both have been utilised several times. In the following sections, the studies in this subcategory will be reviewed and categorised based on the main architecture. In addition to this categorization, the novelties discussed in the studies will be reviewed.



Figure 5. Typical output of a deep learning-based crack detection approach based on the OR setting [3].

2.2.1. R-CNN Family

R-CNN [47], Fast R-CNN [50], and faster R-CNN [51] are members of this family, and faster R-CNN has been the most popular in the crack detection area. As the first member of the R-CNN family, R-CNN [47] architecture has been applied in the crack detection area only once [52]. In that study, the selective search was first used to detect candidate regions of cracks. Then, a CNN pre-trained on ImageNet and Cifar-10 data sets was utilised for crack classification and bounding box regression. To the best of the authors' knowledge, Cha et al. [3] were the first to perform crack detection using an OR setting. That study is an improvement to their previous study where crack detection was carried out in an IC setting using a sliding window technique [23]. The improvements are two-fold. The first is inclusion of a richer data set and the second is employment of Faster R-CNN architecture and solve the crack detection challenge in an OR setting. After the first application, researchers considered different research objectives in this subcategory. In general, the research objectives can be categorised as (i) improving the Faster R-CNN performance via the application of extra modules and (ii) considering a richer data set that includes more complex images.

Li et al. [53] designed a multi-scale defect region proposal network (RPN) that proposes candidate bounding boxes in different layers resulting in higher detection accuracy. They also performed crack geo-localization using an extra deep architecture and a database of geo-tagged images. It is noteworthy that both the crack detection network and the geo-localization module are incorporated in one network. An improved version of faster R-CNN in the crack detection area called CrackDN was proposed in [54]. In this framework, a pre-trained CNN is used to extract deep features and along with it, a sensitivity detection network is employed leading to a faster training procedure. In [4], the novelty is incorporation of two extra modules into the Faster R-CNN architecture to suppress false detection. The first module is a simple and compact deep CNN, which is responsible for crack orientation estimation of the identified crack patches. The second module is a Bayesian integration algorithm to exploit local consistency and reduce the uncertainties of detected crack patches.

With a goal of making the data set more challenging in the OR setting, in [54–56], complex data sets were considered to check the performance of Faster R-CNN architecture in more realistic situations. Huyan et al. [54] included images with unbalanced illumination, cracks on marking, and shades. In contrast, Deng et al. [55] considered a concrete crack data set where images with handwritten scripts are included in the data set as well. They concluded that the handwritten scripts can be considered as a strong noise in concrete

images. A richer data set was collected by Ciaparrone et al. [56]. To annotate the data for the task of OR, the regions of interest for each video frame containing the damaged part of the road were extracted. The authors managed to extract 14 different damage classes and perform crack detection on the collected data set using faster R-CNN architecture.

2.2.2. SSD

The name of SSD stems from the fact that there is no need for an extra module to find candidate regions of objects in an image and it encapsulates all the computation in a single network. This makes SSD easy to train and straightforward to integrate into frameworks that need a detection component. It was first proposed in 2016 by Liu et al. [49] and outperformed faster R-CNN in terms of speed on the VOC2007 data set for the task of OR. Maeda et al. [57] employed SSD framework to perform crack detection in a comprehensive data set using an OR setting. They utilised Inception V2 [58] and MobileNet [59] as the backbone feature extractor module in the SSD framework. It is worth mentioning that the authors made the collected and annotated data set publicly available.

2.2.3. YOLO Family

The YOLO name stems from the fact the YOLO architectures need to look at the image once and perform the task of OR in one forward propagation. YOLO [48], YOLO9000 [60], and YOLOv3 [61] are members in this family. YOLOv3 is able to perform the OR task with accuracy comparable to other methods with a notably faster processing time. To the best of the authors' knowledge, as the main architecture from the YOLO family, YOLO9000 has been employed in the crack detection area only once. Mandal et al. [62] utilised YOLO9000 to perform crack detection in a diverse data set that included different types of cracks and defects.

The reason OR failed to become the most popular setting after the IC setting (Figure 1) is the undesirable performance in terms of crack localization in this setting as well. The OR setting also results in a coarse crack localization output. Moreover, in this setting, the overlapping of bounding boxes when the algorithm fails to capture the whole crack as a single object is problematic. A summary of studies of deep crack detection approaches based on object the recognition setting is shown in Table 2.

Table 2. Summary of deep crack detection approaches based on the OR setting.

Ref	Novelty/Novelties	Core Architecture
[3]	1. First application of the OR setting for the task of defect detection. 2. Considering a richer data set, including five types of defects in comparison to the previous study.	Faster R-CNN
[57]	1. Collecting and making publicly available a new data set including eight types of road damage. 2. Applying the SSD framework with two different backbone architectures for the sake of comparison.	SSD
[53]	1. Proposing a new RPN dealing with different scales and improving accuracy. 2. Considering defect localization in the map using the NetVLAD module.	Faster R-CNN
[56]	1. Considering 14 defect types in the data set. 2. Performing a comparative study between OR and SS settings.	Faster R-CNN
[62]	Application of YOLO9000 architecture for crack detection of roads in an OR setting.	YOLO v2
[52]	1. Application of R-CNN framework to perform crack detection in an OR setting. 2. Employing a pre-trained CNN on ImageNet and Cifar-10 data sets to perform crack classification and bounding box regression.	R-CNN
[54]	1. Proposing CrackDN consisting of Faster R-CNN, a sensitivity detection network, and a region proposal refinement network. 2. Considering a richer data set where different backgrounds of normal, unbalanced illuminations, with markings and shadings, are included.	Faster R-CNN
[55]	Considering complex backgrounds in the images including handwritten scripts and applying Faster R-CNN to identify cracks in the images.	Faster R-CNN
[4]	Combining Faster R-CNN with a new Bayesian integration algorithm to suppress false detection and address challenging vision problems where a simple end-to-end learning strategy might not be effective.	Faster R-CNN

2.3. Semantic Segmentation

In the area of computer vision, performing the classification at the pixel-level is known as SS [63]. SS is one of the fundamental topics in the computer vision area and has been applied in many areas, such as autonomous driving [64–66], 3D reconstruction [67–69], and medical image analysis [70–72] areas. The application of SS has the obvious benefit of locating the object in the image more precisely than the application of OR and IC settings for the same purpose [73]. The output of any SS framework proposed in the crack detection area is the input image where the crack pixels are distinguishable with a different colour from the background pixels. A typical output is shown in Figure 6.

In the area of crack detection, SS is the most popular setting (Figure 1). As the number of published crack detection studies based on the SS setting is the highest, we further categorised the deep crack segmentation approaches. Depending on whether they are involved with either an IC or OR setting, deep crack segmentation approaches can be categorised into (i) hybrid and (ii) pure approaches, each with two sub-categories, as shown in Figure 7.



Figure 6. Typical output of a deep learning-based crack detection approach based on the SS setting [74].

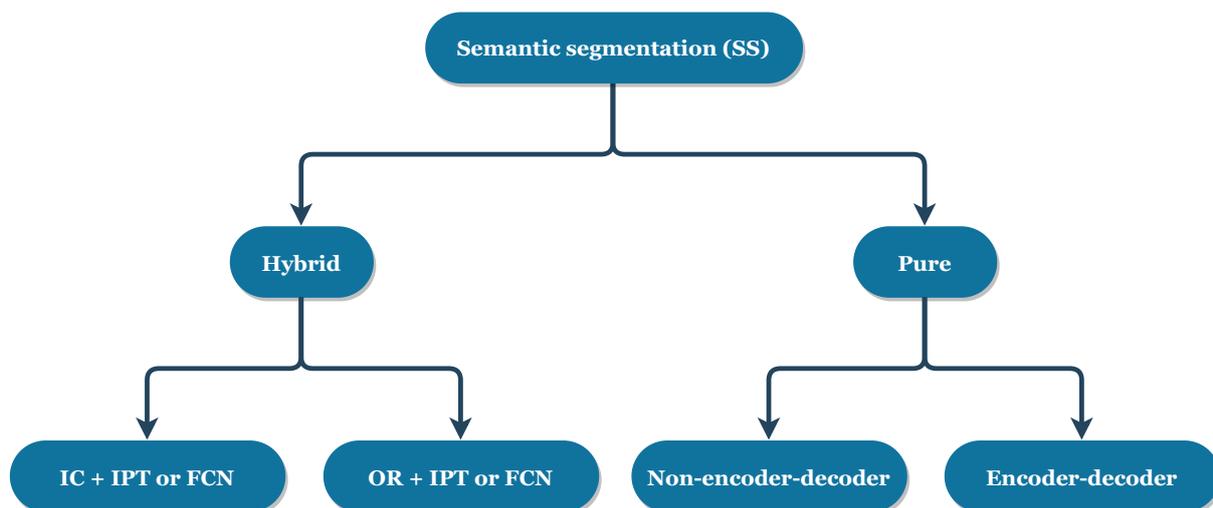


Figure 7. Categorization of deep crack detection approaches based on the SS setting.

2.3.1. Hybrid Semantic Segmentation

We use the word “hybrid” in this subcategory because the procedure of crack segmentation is involved with the application of the two aforementioned settings (i.e., either IC or OR). In other words, as the first step, either an IC setting (or an OR setting) is performed to find crack patches (or crack candidate regions) and then, a technique is employed to segment the crack pixels in the identified crack patches. The employed technique could be either an IPT or a shallow fully convolutional network (FCN). For the former case, several

IPTs, such as structured Random Forest edge detection [75], tabularity flow [76], Otsu's thresholding [77], and fast block-wise segmentation [78] have been used. For the latter case, mainly the Mask R-CNN [79] framework which is an extended version of faster R-CNN has been used. Mask R-CNN consists of an extra shallow FCN being utilised to perform SS on the bounding boxes which contain cracks [75,80,81].

Several studies have first used classification architectures to find crack patches, and then different IPTs to extract the crack map at the pixel-level. Ni et al. [77] used GoogLeNet [82] and ResNet [33] classifiers to detect crack patches. Then, Otsu's thresholding was employed to segment the identified crack patches, followed by median filtering and the Hessian matrix, which were used to eliminate the influence of illumination and to enhance the crack structures, respectively. In [78], the idea of transfer learning was employed at the crack patch identification step using a pre-trained architecture on the ImageNet data set. Then, fast blockwise segmentation and tensor voting curve detection methods were employed to create the crack mask and improve the accuracy of crack localization, respectively.

Faster R-CNN architecture has also been used, combined with IPTs to extract the crack mask, to identify crack regions as the first step. In [83], a Bayesian fusion algorithm was employed to suppress false alarms based on the orientation of the identified crack regions. Then, a set of IPTs, known as morphological operations, were considered to extract the crack mask. Kang et al. [76] proposed to perform crack segmentation in the same manner. However, in their proposed approach, after crack region identification by Faster R-CNN, a modified tabularity flow field, an IPT, was employed to perform crack segmentation on concrete images taken from both indoor and outdoor environments.

In the case of approaches in which first, crack patches and regions are identified and then, an FCN is used to perform crack segmentation at the pixel-level, mask R-CNN has been most employed [75,80,81,84]. It is noteworthy to mention that Tan et al. [84] defined a new threshold resulting in a more desirable segmentation of irregular long-thin cracks. Wei et al. [80] employed mask R-CNN to detect concrete bugholes at the pixel-level. In [75], a comprehensive comparison was performed between the application of mask R-CNN and a framework where faster R-CNN combined with structured random forest edge detection was employed for the task of crack segmentation. They proved that mask R-CNN achieves higher accuracy in comparison to that combination for the task of crack segmentation.

Three more hybrid SS approaches were found in the literature, where after patch detection, deep architectures have been utilised to output the crack segmentation map. In [85], GoogLeNet [82] was employed for crack patch detection. Then, the detected patches were fed into a feature fusion module and a set of consecutive convolution layers to perform the crack segmentation. Zhang et al. [86] proposed a Sobel-edge adaptive sliding window technique to extract crack patches, which is computationally more efficient than the standard sliding window technique. Moreover, in the crack patch extraction step, to further reduce overall processing time and suppress the identified crack patches but keep the patches with significant local edge texture, a non-maximum image patch suppression strategy is proposed. Once the crack patches are detected, SegNet [87] architecture is employed to output the crack mask. In [88], firstly, the pre-trained AlexNet on ImageNet data set is used to classify the image patches into the crack, sealed crack, and background classes. Then, the knowledge acquired by the classification network is transferred into an FCN equipped with dilated convolution to perform crack classification at the pixel-level.

It is worth mentioning that since the crack detection is performed at the pixel-level with a SS setting, crack quantification has been considered frequently in studies in this subcategory via the application of different methods [76,77,80,81]. To name a few applied procedures in this subcategory, Otsu's method and Canny edge detector, distance transform method, and Zernike moment operator have been considered in [76,77,81], respectively, to estimate the width and length of the detected cracks. A summary of the studies of deep crack segmentation approaches based on the hybrid SS setting is shown in Table 3.

Table 3. Summary of deep crack segmentation approaches based on the hybrid SS setting.

Ref	Novelty/Novelties	Method	Core Architecture
[78]	1. Application of transfer learning to identify crack and sealed crack patches. 2. Applying fast block-wise segmentation based on linear regression to segment the identified patches. 3. Application of tensor voting curve detection to extract the detected crack curves.	IC + IP	-
[88]	1. Application of pre-trained AlexNet on ImageNet data set to classify road images into the crack, sealed crack, and background patches. 2. Application of an FCN combined with dilated convolution to segment the detected crack patches.	IC + FCN	AlexNet
[85]	Proposing a crack delineation network including a generic pre-trained CNN model (GoogLeNet) combined with a feature pyramid network to achieve feature-map fusion.	IC + FCN	GoogLeNet
[77]	1. Proposing a dual-scale CNN to perform crack patch classification and segmentation. 2. Application of Zernike moment operator for quantitative crack width estimation.	IC + IP	GoogLeNet
[80]	Application of mask R-CNN for bughole segmentation in concrete surface images and quantification of the segmented bugholes.	OR + FCN	Mask R-CNN
[84]	Application of mask R-CNN to detect cracks in pavement image data sets	OR + FCN	Mask R-CNN
[83]	Application of faster R-CNN combined with a Bayesian probability algorithm to suppress false detection and a set of IPTs to perform crack segmentation.	OR + IP	Faster R-CNN
[81]	1. Application of mask R-CNN to output crack masks. 2. Applying IPTs to quantify the detected masks.	OR + FCN	Mask R-CNN
[86]	Proposing a new hybrid crack segmentation approach based on the global non-overlapping sliding windows and Sobel-edge detector to identify crack patches combined with a deep encoder–decoder architecture (SegNet) to perform crack segmentation.	IC + FCN	SegNet
[75]	Performing a comparative study between two crack segmentation frameworks of (i) Faster R-CNN combined with structured Random Forest edge detection and (ii) Mask R-CNN.	OR + IP and OR + FCN	Faster R-CNN, Mask R-CNN
[76]	1. Performing crack segmentation using Faster R-CNN combined with a modified tabularity flow field. 2. Performing crack quantification using a modified distance transform method.	OR + IP	Faster R-CNN

2.3.2. Pure Semantic Segmentation

Pure crack segmentation is performed with no crack patch or crack candidate region identification beforehand. Therefore, no IC or OR approach is involved in this procedure. In the literature, several approaches have been considered for performing pure SS with an application for crack detection. Pure SS can be done by the typical architecture utilised for IC if the fully connected layers are replaced with convolutional layers resulting in an FCN [89,90] (aka an encoder–decoder structure). This opens a window to different FCN structures that can be used for SS in the area of computer vision. In the area of crack detection, mostly, the pure crack segmentation approaches are performed with the application of encoder–decoder structures. However, several studies have employed other techniques to perform crack segmentation at the pixel-level. Firstly, these approaches and then, the encoder–decoder approaches will now be reviewed.

Crack segmentation at the pixel-level was done with an application of the IC architecture in [91,92]. In both of these studies, once the patches are extracted from the pavement images, crack patches (i.e., positive patches) are defined based on a criterion of whether the centre pixels are crack ones or not. After training and classification of the positive patches, the crack mask at the pixel-level can be achieved by stacking the positive pixels together. The difference between these studies is that in [91], more than one pixel in the centre define whether the patch is positive or not resulting in thicker crack masks after detection. On the other hand, in [92], only the centre pixel is considered as the criterion. The same idea of crack structure prediction in [92] was employed to perform crack detection with high accuracy by Fan et al. [93]. It is worth noting that they removed the pooling layers from the main structure to avoid losing spatial information and crack quantification also was considered [93]. CrackNet II [94] and CrackNet-V [95] as improvements to the CrackNet [96], in terms of both learning capability and the required processing time, were proposed. It is worth mentioning that the CrackNet framework is based on both handcrafted features and deep learning. However, CrackNet II and CrackNet-V are deep learning-based approaches utilised for SS of 3D crack images. Both approaches do not include any pooling layers and the image width and height are invariant through the consecutive convolution layers resulting in a supervised classification at the pixel-level. The difference between CrackNet II and CrackNet-V approaches is that CrackNet-V, in addition to a set of consecutive convolution layers, consists of a pre-processing module based on the median filter and Z-score normalization. It must be added that the same research group also proposed CrackNet-R [97] as another improved version of CrackNet architecture, based on the application of recurrent neural network (RNN) for the task of crack segmentation.

In order to perform SS via the application of an encoder–decoder structure, a backbone architecture is used to extract deep features which is formed by a sequence of convolution, pooling, and activation layers. Since the width and height of the input image decrease once it passes through the backbone architecture, a decoder module, which is formed by a sequence of deconvolution (aka transposed convolution or fractionally strided convolution) layers, is employed to restore the size of the features to the original image size so that the classification can be done at the pixel-level [98]. The encoder–decoder structure has been abundantly applied to perform crack segmentation. Among various well-known architectures that have been proposed for performing SS in the computer vision area, U-Net [99], SegNet [87], and FC-DenseNet [100] have been the most considered in the crack detection area. It must be added that to achieve higher accuracy in the SS setting, it is important to let the contextual information flow in the architecture [101]. To achieve this, various approaches have been considered in the crack detection area, as follows.

U-Net is an architecture first proposed in the area of medical image analysis for biomedical image segmentation [99]. To the best of the authors' knowledge, David Jenkins et al. [102] performed the first application of U-Net architecture in the crack detection area and showed its superiority for the task of crack segmentation over approaches based on handcrafted features combined with classical machine learning algorithms. It must be added that in the crack detection area, there is a challenge of class imbalance because of having access to more background pixels compared to crack pixels. To solve this challenge, different techniques and approaches have been proposed that, in the case of some studies, could be considered as one of the contributions of the study to the field. In [103], U-Net architecture was equipped with a new loss function based on distance transform to deal with the challenge of class imbalance in the crack detection area. To further improve the performance of U-Net for crack segmentation, Konig et al. [104] combined the U-Net architecture with residual connections and an attention gating mechanism. An ablation study related to the application of U-Net architecture for crack segmentation was carried out in [105]. Three U-Net architectures with different depths and number of layers were utilised and compared to perform crack segmentation on challenging data sets of CFD and Aigle-RN. In another study [106], to cope with the class imbalance challenge in the crack detection area, the focal loss function was employed for training of the U-Net architecture to ensure

higher generalization performance. Fan et al. [107] proposed a modified version of U-Net where two modules are added to the architecture to boost the performance of crack detection. The multi-dilation module equipped with dilation convolution and hierarchical feature learning module are the added modules which are responsible for obtaining crack features of multiple context sizes and deriving multi-scale features from high- to low-level convolution layers, respectively [107]. Zhang et al. [108] asserted that in industrial pixel-level crack detection, the aforementioned challenge results in a common problem called “All Black”. “All Black” happens when the algorithm classifies all the pixels of the pavement image as the background and still achieves good accuracy. To solve this problem, they proposed an approach called CrackGAN, which is based on generative adversarial networks (GANs) [109]. In this approach, the asymmetric U-Net is employed as the backbone architecture of the generator network providing the ability to work with images of arbitrary sizes.

SegNet is an encoder–decoder architecture where the encoder is the VGG-16 [38] architecture without the fully connected layers [87]. The main feature of SegNet is that during the decoding, the max-pooling indices at the corresponding encoder layer are recalled and used to up-sample in the decoder module. This makes SegNet faster than U-Net. A pavement and bridge crack segmentation network inspired by SegNet architecture was proposed by Chen et al. [110]. To train the end-to-end deep learning model proposed in this study, the “AdaDelta” optimizer and the cross-entropy loss function were used. SegNet architecture was employed and modified to have a better performance on line like object segmentation in the area of crack detection by Zou et al. [111]. The proposed approach in this study is called “DeepCrack” and feature fusion at different scales was considered as the main novelty. Particularly, they fused the sparse features at smaller scales with continuous features at larger scales to have a better performance for the task of crack segmentation. It is stated in [112] that the proposed encoder–decoder structure in the study which is inspired by SegNet [87], FCN [89], and ZFNet [113] was the first application of deep learning to find cracks in black-box images. A pre-trained ResNet [33] on the ImageNet data set was employed as the encoder module. Also, in the decoder module, the deconvolution technique of ZFNet and SegNet, which is based on storing the location information in the max-pooling layers and boundary information in the encoder feature maps, was considered in the three deconvolutional layers.

DenseNet was first proposed for the task of IC as an improvement to the ResNet [33] architecture [114]. ResNet architecture is involved with a large number of parameters because each layer has its weights and it has been proven that many layers in the ResNet architecture contribute very little. DenseNet architecture is proposed as a solution to this challenge. The main feature of the DenseNet architecture is the application of dense blocks, which benefits the task by alleviating the vanishing gradient problem, improving feature reuse, and significantly decreasing the number of parameters. Later the same year, the FCN version of DenseNet known as FC-DenseNet was proposed by Jégou et al. [100] for the task of SS. To the best of the authors’ knowledge, the first application of FC-DenseNet for the task of crack segmentation was done by Li et al. [115]. In this study, an FCN-based approach through the fine-tuning of the DenseNet-121 architecture was considered to perform multiple damage detection in a challenging data set. In addition to the crack images, the data set includes spalling, efflorescence, and hole images captured using a smartphone in different illumination conditions. They proved that FC-DenseNet outperforms SegNet architecture for the task of crack segmentation. An approach called “DenseCrack” was proposed by Mei and Gül [116] where an encoder–decoder structure based on dense blocks is combined with a depth-first search (DFS)-based algorithm as a refinement module. This study also can be considered as an ablation study because three architectures of DenseCrack121, 169, and 201 were employed and compared. In [117], the “ConnCrack” approach, which is based on GANs, was proposed. DenseNet121 architecture was employed as the generator of conditional Wasserstein GAN (cWGAN) in this study. Another approach based on the application of FC-DenseNet was proposed by the same

research group in [74]. In this study, to further improve the accuracy of the SS, a new loss function based on the connectivity of the pixels in the crack areas is defined. The application of the new loss function benefits the crack segmentation performance both by dealing with the class imbalance challenge and taking into consideration the connectivity of crack pixels. They made a comprehensive comparative study between the proposed approach and different published state-of-the-art crack segmentation approaches on three different data sets.

Several other studies have been conducted in the crack detection area with an application of encoder–decoder structures. However, the proposed approaches are not inspired by well-known architectures in the computer vision area. In these studies, the technique that is employed to let the contextual information flow between the encoder and the decoder module is notable. In Yang et al. [118], the up-sampling part is the core novelty of the proposed architecture. The up-sampling part combines global information and local information by adding specific convolutional and deconvolutional layers resulting in the ability of the proposed architecture to deal with multi-scale and multi-level images. In [119], the ideas of transfer learning and dilated convolution before the decoder module are combined to perform crack detection at the pixel-level. Dilated convolution (aka atrous convolution) was originally designed to aggregate contextual information at multi-scales for SS [120]. Another approach called “DeepCrack” was proposed by Liu et al. [121]. In this study, to let the contextual information flow, side-output layers are inserted after convolutional layers. After deep supervision at each side-output layer, the outputs are concatenated to form a final output layer that acquires multi-scale and multi-level features. Then, the fused prediction is refined by conditional random field (CRF) and Guided Filtering (GF) modules to improve the accuracy of SS. The same technique of considering a side network including 1×1 convolution layers applied at each level of information and associated deconvolution layers was employed by Yang et al. [2]. However, they noted that the side-outputs of low-level layers are messy which stems from lacking context information. To solve this issue, the authors utilised a feature pyramid module between the backbone architecture and the side network resulting in combining multi-scale context information into low-level feature maps. They comprehensively compared the performance of their proposed approach with state-of-the-art deep learning-based crack segmentation approaches over five different data sets. On the other hand, without considering any technique to deal with the fusing of feature maps at different scales, simple encoder–decoder structures were considered in [73,122]. However, in the former study, a comparison between the application of three different pre-trained deep CNNs (i.e., VGG-16 [38], InceptionV3 [123], and ResNet [33]) as the backbone architectures in the encoder module was performed. The main contribution of the latter study, to solve the problem of class imbalance, was development of an image generation algorithm using the Brownian motion process and Gaussian kernel to generate simulated crack images. A summary of the studies of deep crack segmentation approaches based on the pure SS setting is presented in Table 4.

Table 4. Summary of deep crack segmentation approaches based on the pure SS setting.

Ref	Novelty/Novelties	Method
[91]	The first application of deep learning for the task of crack segmentation, where using ConvNet, the feature extraction is done on raw images.	Centre crack pixels in the patches
[94]	Proposing an improved version of CrackNet called CrackNet II resulting in higher performance in terms of accuracy and speed.	Consecutive conv layers with an invariant spatial size
[118]	1. Application of an encoder–decoder structure to perform crack segmentation without employing the sliding windows technique. 2. Extracting the geometric characteristic of cracks via the application of morphological operations.	Encoder–decoder (feature fusion)
[102]	The first application of U-Net architecture in the crack detection area to cope with several limitations of applying CNN for the task of crack detection.	Encoder–decoder (U-Net)
[103]	Application of U-Net architecture equipped with a new proposed loss function based on distance transform to perform the task of crack segmentation.	Encoder–decoder (U-Net)
[92]	1. Predicting the crack structure using CNN. 2. Proposing a strategy to deal with the class imbalance challenge.	Centre crack pixels in the patches
[97]	Proposing an improved version of CrackNet called CrackNet-R based on RNN, including a new recurrent unit.	RNN
[122]	Investigation of the performance of FCN architectures in the crack detection area for the task of crack segmentation.	Encoder–decoder
[104]	Proposing an encoder–decoder structure based on U-Net architecture combined with attention gating and residual connections to improve the performance.	Encoder–decoder (U-Net)
[105]	Investigating the effect of the depth of the U-Net architecture on the crack segmentation performance.	Encoder–decoder (U-Net)
[111]	Designing a new end-to-end trainable neural network based on SegNet architecture for robust crack detection.	Encoder–decoder (SegNet)
[115]	Proposing a deeper and more comprehensive FCN architecture to detect four concrete damages where it requires no sliding window technique.	Encoder–decoder (FC-DenseNet)
[112]	Performing a successful application of deep learning methods for detecting road cracks in black-box images.	Encoder–decoder (ResNet + SegNet, FCN, ZFNet)
[73]	1. Proving the superiority of the SS over OR setting. 2. Proposing a method based on Gaussian kernel and Brownian motion process to generate arbitrary simulated crack images.	Encoder–decoder
[119]	Application of an FCN architecture based on dilated convolution to perform SS of crack images.	Encoder–decoder (dilated convolution)
[121]	1. Proposing an FCN combined with condition random field and guided filtering methods. 2. Designing a new loss function to deal with the class imbalance challenge. 3. Making the employed data set publicly available.	Encoder–decoder (feature fusion)
[106]	Application of U-Net architecture for crack segmentation where Adam algorithm and focal loss function are used as the optimizer and evaluation function, respectively.	Encoder–decoder (U-Net)
[2]	Proposing a feature pyramid and hierarchical boosting network to achieve more robust feature representation and deal with the class imbalance challenge.	Encoder–decoder (feature fusion)
[74]	1. Proposing a novel deep architecture based on dense blocks. 2. Proposing a novel loss function based on the connectivity of crack pixels in the crack areas.	Encoder–decoder (FC-DenseNet)
[108]	Proposing a CrackGAN framework based on the application of GAN architecture where the proposed approach is capable of working with partially annotated ground truths.	Encoder–decoder (U-Net as generator)
[117]	Proposing a crack segmentation approach based on conditional Wasserstein GAN combined with connectivity map to refine the results.	Encoder–decoder (FC-DenseNet as generator)
[116]	1. Application of skip connections in the deep architecture for feature fusion. 2. Application of a depth-first search algorithm for post-processing and improving the accuracy.	Encoder–decoder (FC-DenseNet)
[110]	Application of SegNet architecture to perform the crack segmentation where the “Adadelta” optimizer and cross-entropy loss function are used	Encoder–decoder (SegNet)
[95]	1. Proposing an improved version of CrackNet called CrackNet-V resulting in higher performance in terms of accuracy and speed. 2. Proposing a new activation function to improve the accuracy of crack segmentation for shallow cracks.	Consecutive conv layers with an invariant spatial size
[107]	Proposing U-hierarchical dilated network, a modified encoder–decoder architecture based on U-Net, where two modules of multi-dilation and hierarchical feature learning are added to improve the performance of crack segmentation.	Encoder–decoder (U-Net)

3. Public Data Sets

Based on the comprehensive literature review provided in Section 2 and the drawbacks related to different approaches in terms of localization of cracks in the images, we choose the SS setting as the most reliable setting in the crack detection area. It is worth mentioning that a higher number of studies, which have published in the crack detection area focusing on crack detection at the pixel-level, also highlights the popularity of this setting in the crack detection area. Therefore, for the rest of this paper, we will focus on exploring the details related to this setting. In terms of publicly available data sets, we realised that among the 35 studies reviewed in the SS category, the CFD [10] and Aigle-RN [124] data sets were most frequently used, in 11 and 5 studies, respectively.

3.1. CFD Data Set

The CFD data set was collected and made publicly available by Shi et al. [10]. It is stated by the authors that the data set generally reflects the condition of the urban road surface in Beijing, China. The images are labelled manually at the pixel-level. CFD contains 118 RGB road images of size 480×320 pixels where noise such as shadows, oil spots, and water stains and different illumination conditions can be found in the images. The images are captured by an iPhone5 with a focus of 4mm, an aperture of $f/2.4$, and an exposure time of $1/134$ s. It is worth mentioning that the width of the images ranges from 1 to 3 mm.

3.2. Aigle-RN Data Set

Aigle-RN is the name of the system that was used to acquire the images. The Aigle-RN data set contains 38 grey-level road images of size 991×462 pixels. This data set is acquired at traffic speed for periodical monitoring of the pavement surface condition in France [124]. It is worth mentioning that the Aigle-RN images are pre-processed to lessen the effect of non-uniform illumination condition.

4. Results and Discussion

As it has been discussed so far, the main idea of the proposed approaches in the crack segmentation area is inspired by the proposed approaches in the computer vision area. The main difference between these two areas is that in the computer vision area, there are several benchmark data sets with a specified number of training, cross-validation, and testing data. Hence, the proposed approaches by different researchers can be compared directly. On the other hand, in the crack detection area, even though the publicly available data sets of CFD [10] and Aigle-RN [124]) are used by different studies, the results cannot be compared directly. The reason is that the number of images in these data sets is small, and researchers have been required to employ different data augmentation techniques resulting in a different number of training, cross-validation, and testing examples in each study. Therefore, in this study, to compare the results and enable readers to find the state-of-the-art crack segmentation approaches, in addition to accuracy, the employed data augmentation techniques, and the number of training, cross-validation, and testing examples are extracted. The details are sorted by the year of publication and shown in Table 5.

Table 5. Implementation details related to the state-of-the-art approaches in the crack detection area.

Ref.	Comparison with	Data Set	#Training-Validation-Testing Data	Data Augmentation	#Training Patches	Precision %	Recall %	F1-Score %	Time/ Image (s)
[102]	None	CFD	80-20-18	Patch extraction with a stride of 1	160,000	92.46	82.82	87.38	≈3
[103]	None	CFD	78-0-40	Rotation, width and height shifting, horizontal flipping, and zooming	2415	92.12	95.7	93.88	-
		Aigle-RN	25-0-13		773	92.02	93.21	92.61	-
[92]	None	CFD	72-0-46	Patch extraction	741,932	91.19	94.81	92.44	-
		Aigle-RN	24-0-14		156,832	91.78	88.12	89.54	-
[104]	U-Net [102], CNN [92]	CFD	71-0-46	Patch extraction	142,000	96.37	93.55	94.94	-
		Aigle-RN	24-0-14		84,000	86.9	93.04	89.86	-
[105]	CNN [92], VGG-16	CFD	100-0-18	-	-	97.31	94.28	95.75	-
		Aigle-RN	0-0-38	-	-	93.51	82.9	87.33	-
[2]	HED [125], RCF [126], FCN [89]	Crack500	250-50-200	Patch extraction	1896	AIU = 0.489 ¹	ODS = 0.604 ²	OIS = 0.635 ³	0.197
		CFD	0-0-118	-	-	AIU = 0.173	ODS = 0.683	OIS = 0.705	0.133
		AEL ⁴	0-0-58	-	-	AIU = 0.079	ODS = 0.492	OIS = 0.507	0.259
[86]	Original-Segnet [87]	CFD	70-24-24	Patch extraction	-	82.03	82.83	82.34	2.4
		TRIMMED ⁵	31-11-11	Patch extraction	-	79.99	85.38	82.52	5.8
[116]	CrackNet-V [95]	CFD	70-0-48	Horizontal and vertical flipping, patch extraction with a stride of 16	62,790	92.02	91.13	91.58	1.66
[74]	ResNet152-FCN [112], VGG19-FCN [118], CrackNet-V [95], U-Net [106], and FPHBN [2]	CFD	70-0-48	Horizontal and vertical flipping, patch extraction with a stride of 16	62,790	91	93.22	91.99	≈3.5
[117]	ResNet152-FCN [112], VGG119-FCN [118], U-Net [106], CrackNet-V [95]	CFD	70-0-48	Patch extraction	-	96.79	87.75	91.96	≈1.5
[107]	HED [125], RCF [126], FCN [89], FPHBN [2], U-Net [99]	CFD	70-0-48	-	-	94.50	93.60	93.90	-
		Aigle-RN	24-0-14	-	-	92.10	93.10	92.40	-

¹ Average intersection over union. ² The best F1-score on the data set for a fixed scale. ³ The aggregate F1-score on the data set for the best scale in each image. ⁴ Three data sets of Aigle-RN, ESAR, and LCMS combined together [2]. ⁵ Two data sets of Aigle-RN and ESAR combined together [86].

As can be seen in Table 5, although all of the studies employed the same data sets, there is a significant difference between the number of training patches considered in each study. Given the detailed information provided in Table 5 and considering the “Comparison with” column as significant criteria, both in terms of considering a higher number of studies and more recent published studies, the authors state that the proposed approaches in [74,117] are the most promising approaches between the studies reviewed in this study.

5. Conclusions

Deep architectures, just as they have done in other areas, have improved the performance of crack detection approaches significantly. In this study, 61 deep learning-based crack detection approaches were comprehensively reviewed. More precisely, the crack detection approaches that use deep layers to perform the feature extraction step in an automated manner were reviewed and the main contributions of each study were extracted and discussed in detail. Moreover, additional details of several studies that employed the same public data sets have been provided to give the readers insights into the selection of the most promising crack segmentation approaches.

Based on the number of studies in each category, the SS setting was chosen to focus on, as the most popular setting in the crack detection area in recent years. In addition to providing higher accuracy in terms of detection, the crack localization is done at the pixel-level thereby providing higher localization accuracy. By and large, the crack segmentation approaches proposed in the crack detection area are inspired by the proposed approaches in the SS field of the computer vision area. Even though promising approaches have been proposed so far, notable future directions can be proposed in the crack segmentation area.

Identification of different types of cracks and their associated cause or causes is a vital phase in the selection of rehabilitation treatment. There has been a very strong concentration on crack detection using a SS setting. However, no study performed crack type classification with a SS setting. It is worth mentioning that the crack type classification has been considered in both IC and OR settings and promising performances have been obtained. Moreover, in addition to the considered crack types studied so far, there are other crack types, such as meandering and crescent-shaped cracks, which have not been considered. If these two types of cracks are detected and classified correctly, useful information about the base and asphalt layers can be achieved.

According to the details provided in Section 4, it is imperative to develop a universal public crack data set in the area of crack detection. Although CFD and Aigle-RN data sets have been widely considered in the literature, they suffer from consisting of a limited number of images. This has resulted in state-of-the-art approaches not being compared fairly and properly. The universal public data set should be acquired in different illumination conditions (i.e., different times of a day) so that the effectiveness of the proposed approaches by researchers in different conditions can be validated. Additionally, different kinds of noise (both similar and non-similar to cracks), such as zebra crossings, oil spots, sealed cracks, leaves, and other objects on the road, etc., should be considered. More importantly, the universal public data set needs to be released with a specified number of training, cross-validation, and testing data. This will promote a trend of validation of the proposed crack segmentation approaches on the same data set so that all of the proposed approaches can be compared directly.

Even though the main idea of crack segmentation approaches is drawn from the SS context in the computer vision area, surprisingly, there have been a few attempts to perform the task of crack segmentation in a weakly supervised or unsupervised manner. On the other hand, promising SS approaches with weakly supervised and unsupervised settings can be found in the computer vision area. It is well-known that deep architectures are data-hungry, and in a supervised mode, they require lots of annotated data to be trained with. In the case of the SS setting, data annotation at the pixel-level is considered a labour-intensive and time-consuming task. Therefore, performing crack segmentation task in

a non-supervised manner will limit human intervention and make the automatic crack detection area one step closer to full automation.

To the best of the authors' knowledge, and based on the information provided in this paper, the most promising crack segmentation performances have been acquired from the application of FCN networks based on dense blocks. Although they have shown promising results in the crack detection area, it is worth mentioning that since their proposal in the computer vision area, other different SS approaches with better performances have been proposed. Their application in the crack detection area and confirmation of their higher efficiency for the task of crack segmentation have not yet been considered.

Given the fact that one of the main categorizations of image processing-based crack detection approaches is dedicated to methods based on edge-detection, the study of "deep edge detection" (aka "deep boundary detection") approaches is proposed. It is well-known that a better understanding of boundaries and using them as prior or constraint leads to a better performance in the task of SS. In particular, the study of texture and boundary information in separate streams has resulted in higher SS accuracy in terms of thin objects in the computer vision area. On the other hand, segmentation of thin cracks still is a challenging problem in the crack detection area. The application of proposed ideas in the deep edge detection field can improve crack segmentation in the crack detection area.

Different approaches have been proposed in the crack detection area to deal with the class imbalance challenge. Not only do crack detection and the medical image analysis area share similar characteristics, the same trend of development can also be found in their history. Albeit the abnormality segmentation approaches in the medical image analysis area have taken a further step toward the application of deep anomaly segmentation techniques. The application of these techniques has resulted in solving two challenges in this area. First, having access to a large amount of annotated data is costly and requires expert knowledge. Secondly, more normal data are available in this area. As the same challenges can be found in the crack detection area, a successful crack segmentation approach using deep anomaly segmentation techniques could be beneficial in advancing this area in the future.

As part of a future plan, the authors will be considering looking at the crack detection task from a different point of view. Based on the literature, different algorithms equipped with different loss functions have been proposed in the crack detection area. However, still, due to the complexity of the crack shapes, different performances are being reported. The authors will consider performing a study on the complexity of the crack shapes in depth and the parameter correlations with the performances of the different proposed architectures and loss functions in the area.

Author Contributions: Conceptualization, Y.H. and J.J.; methodology, Y.H. and J.J.; validation, Y.H. and J.J.; formal analysis, Y.H. and J.J.; investigation, Y.H. and J.J.; resources, Y.H. and J.J.; data curation, Y.H. and J.J.; writing—original draft preparation, Y.H.; writing—review and editing, Y.H., H.G., S.S. and J.J.; visualization, Y.H.; supervision, H.G., S.S. and J.J.; project administration, Y.H. and J.J. 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: Not applicable.

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

References

1. Cagle, R.F. *Infrastructure Asset Management: An Emerging Direction*; AACE International Transactions: Morgantown, WV, USA, 2003; pp. PM21–PM26.
2. 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.* **2020**, *21*, 1525–1535. [[CrossRef](#)]

3. Cha, Y.J.; Choi, W.; Suh, G.; Mahmoudkhani, S.; Büyüköztürk, O. Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detecting Multiple Damage Types. *Comput.-Aided Civ. Infrastruct. Eng.* **2017**, *33*, 731–747. [[CrossRef](#)]
4. Fang, F.; Li, L.; Gu, Y.; Zhu, H.; Lim, J.H. A novel hybrid approach for crack detection. *Pattern Recognit.* **2020**, *107*, 107474. [[CrossRef](#)]
5. Jahanshahi, M.R.; Kelly, J.S.; Masri, S.F.; Sukhatme, G.S. A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures. *Struct. Infrastruct. Eng.* **2009**, *5*, 455–486. [[CrossRef](#)]
6. Abdel-Qader, I.; Abudayyeh, O.; Kelly, M.E. Analysis of Edge-Detection Techniques for Crack Identification in Bridges. *J. Comput. Civ. Eng.* **2003**, *17*, 255–263. [[CrossRef](#)]
7. Cord, A.; Chambon, S. Automatic Road Defect Detection by Textural Pattern Recognition Based on AdaBoost. *Comput.-Aided Civ. Infrastruct. Eng.* **2011**, *27*, 244–259. [[CrossRef](#)]
8. Oliveira, H.; Correia, P.L. Automatic road crack segmentation using entropy and image dynamic thresholding. In Proceedings of the 2009 17th European Signal Processing Conference, Glasgow, UK, 24–28 August 2009; pp. 622–626.
9. Prasanna, P.; Dana, K.J.; Gucunski, N.; Basily, B.B.; La, H.M.; Lim, R.S.; Parvardeh, H. Automated Crack Detection on Concrete Bridges. *IEEE Trans. Autom. Sci. Eng.* **2016**, *13*, 591–599. [[CrossRef](#)]
10. Shi, Y.; Cui, L.; Qi, Z.; Meng, F.; Chen, Z. Automatic Road Crack Detection Using Random Structured Forests. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 3434–3445. [[CrossRef](#)]
11. Chen, J.H.; Su, M.C.; Cao, R.; Hsu, S.C.; Lu, J.C. A self organizing map optimization based image recognition and processing model for bridge crack inspection. *Autom. Constr.* **2017**, *73*, 58–66. [[CrossRef](#)]
12. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; Volume 1.
13. Hsieh, Y.A.; Tsai, Y.J. Machine Learning for Crack Detection: Review and Model Performance Comparison. *J. Comput. Civ. Eng.* **2020**, *34*, 04020038. [[CrossRef](#)]
14. Gopalakrishnan, K. Deep Learning in Data-Driven Pavement Image Analysis and Automated Distress Detection: A Review. *Data* **2018**, *3*, 28. [[CrossRef](#)]
15. Cao, W.; Liu, Q.; He, Z. Review of Pavement Defect Detection Methods. *IEEE Access* **2020**, *8*, 14531–14544. [[CrossRef](#)]
16. Farrar, C.R.; Worden, K. *Structural Health Monitoring: A Machine Learning Perspective*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
17. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97. [[CrossRef](#)]
18. Ramabhadran, B.; Khudanpur, S.; Arisoy, E. Proceedings of the NAACL-HLT 2012 Workshop: Will We Ever Really Replace the N-gram Model? On the Future of Language Modeling for HLT. *IEEE Comput. Intell. Mag.* **2012**, *13*, 55–75.
19. Hinton, G.; Salakhutdinov, R. Discovering Binary Codes for Documents by Learning Deep Generative Models. *Top. Cogn. Sci.* **2010**, *3*, 74–91. [[CrossRef](#)]
20. Guo, Y.; Liu, Y.; Oerlemans, A.; Lao, S.; Wu, S.; Lew, M.S. Deep learning for visual understanding: A review. *Neurocomputing* **2013**, *187*, 27–48. [[CrossRef](#)]
21. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)]
22. Sabokrou, M.; Fayyaz, M.; Fathy, M.; Moayed, Z.; Klette, R. Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. *Comput. Vis. Image Underst.* **2018**, *172*, 88–97. [[CrossRef](#)]
23. Cha, Y.J.; Choi, W.; Büyüköztürk, O. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. *Comput.-Aided Civ. Infrastruct. Eng.* **2017**, *32*, 361–378. [[CrossRef](#)]
24. Yokoyama, S.; Matsumoto, T. Development of an Automatic Detector of Cracks in Concrete Using Machine Learning. *Procedia Eng.* **2017**, *171*, 1250–1255. [[CrossRef](#)]
25. Pauly, L.; Hogg, D.; Fuentes, R.; Peel, H. Deeper networks for pavement crack detection. In Proceedings of the 34th International Symposium in Automation and Robotics in Construction, Taipei, Taiwan, 28 June–1 July 2017; pp. 479–485.
26. Li, B.; Wang, K.C.P.; Zhang, A.; Yang, E.; Wang, G. Automatic classification of pavement crack using deep convolutional neural network. *Int. J. Pavement Eng.* **2020**, *21*, 457–463. [[CrossRef](#)]
27. Nhat-Duc, H.; Nguyen, Q.L.; Tran, V.D. Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network. *Autom. Constr.* **2018**, *94*, 203–213. [[CrossRef](#)]
28. 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](#)]
29. Kim, H.; Ahn, E.; Shin, M.; Sim, S.H. Crack and Noncrack Classification from Concrete Surface Images Using Machine Learning. *Struct. Health Monit.* **2019**, *18*, 725–738. [[CrossRef](#)]
30. Feng, C.; Liu, M.Y.; Kao, C.C.; Lee, T.Y. Deep Active Learning for Civil Infrastructure Defect Detection and Classification. *Comput. Civ. Eng.* **2017**, *2017*, 298–306. [[CrossRef](#)]
31. Wang, X.; Hu, Z. Grid-based pavement crack analysis using deep learning. In Proceedings of the 2017 4th International Conference on Transportation Information and Safety (ICTIS), Banff, AB, Canada, 8–10 August 2017. [[CrossRef](#)]
32. Park, S.; Bang, S.; Kim, H.; Kim, H. Patch-Based Crack Detection in Black Box Images Using Convolutional Neural Networks. *J. Comput. Civ. Eng.* **2019**, *33*, 04019017. [[CrossRef](#)]

33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
34. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates Inc.: New York, NY, USA, 2012; pp. 1097–1105.
35. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
36. Shin, H.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [[CrossRef](#)]
37. Gopalakrishnan, K.; Khaitan, S.K.; Choudhary, A.; Agrawal, A. Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. *Constr. Build. Mater.* **2017**, *157*, 322–330. [[CrossRef](#)]
38. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556.
39. da Silva, W.R.L.; de Lucena, D.S. Concrete Cracks Detection Based on Deep Learning Image Classification. *Proceedings* **2018**, *2*, 489. [[CrossRef](#)]
40. Kim, B.; Cho, S. Automated Vision-Based Detection of Cracks on Concrete Surfaces Using a Deep Learning Technique. *Sensors* **2018**, *18*, 3452. [[CrossRef](#)] [[PubMed](#)]
41. Dorafshan, S.; Thomas, R.J.; Coopmans, C.; Maguire, M. Deep Learning Neural Networks for sUAS-Assisted Structural Inspections: Feasibility and Application. In Proceedings of the 2018 International Conference on Unmanned Aircraft Systems (ICUAS), Dallas, TX, USA, 12–15 June 2018. [[CrossRef](#)]
42. Vedaldi, A.; Lenc, K. MatConvNet. In Proceedings of the 23rd ACM international conference on Multimedia, Brisbane, Australia, 26–30 October 2015. [[CrossRef](#)]
43. Kang, D.; Cha, Y.J. Autonomous UAVs for Structural Health Monitoring Using Deep Learning and an Ultrasonic Beacon System with Geo-Tagging. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 885–902. [[CrossRef](#)]
44. Wang, K.; Zhang, A.; Li, J.Q.; Fei, Y.; Chen, C.; Li, B. Deep learning for asphalt pavement cracking recognition using convolutional neural network. In Proceedings of the 2017 International Conference on Highway Pavements and Airfield Technology, Philadelphia, PA, USA, 27–30 August 2017.
45. Eisenbach, M.; Stricker, R.; Seichter, D.; Amende, K.; Debes, K.; Sesselmann, M.; Ebersbach, D.; Stoeckert, U.; Gross, H. How to get pavement distress detection ready for deep learning? A systematic approach. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 2039–2047. [[CrossRef](#)]
46. Szeliski, R. *Computer Vision: Algorithms and Applications*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2010.
47. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. *arXiv* **2013**, arXiv:1311.2524.
48. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2016**, arXiv:1506.02640.
49. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. *arXiv* **2016**, arXiv:1512.02325.
50. Girshick, R. Fast R-CNN. *arXiv* **2015**, arXiv:1504.08083.
51. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2015**, arXiv:1506.01497.
52. Kim, I.H.; Jeon, H.; Baek, S.C.; Hong, W.H.; Jung, H.J. Application of Crack Identification Techniques for an Aging Concrete Bridge Inspection Using an Unmanned Aerial Vehicle. *Sensors* **2018**, *18*, 1881. [[CrossRef](#)]
53. Li, R.; Yuan, Y.; Zhang, W.; Yuan, Y. Unified Vision-Based Methodology for Simultaneous Concrete Defect Detection and Geolocalization. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 527–544. [[CrossRef](#)]
54. Huyan, J.; Li, W.; Tighe, S.; Zhai, J.; Xu, Z.; Chen, Y. Detection of sealed and unsealed cracks with complex backgrounds using deep convolutional neural network. *Autom. Constr.* **2019**, *107*, 102946. [[CrossRef](#)]
55. Deng, J.; Lu, Y.; Lee, V.C.S. Concrete crack detection with handwriting script interferences using faster region-based convolutional neural network. *Comput.-Aided Civ. Infrastruct. Eng.* **2020**, *35*, 373–388. [[CrossRef](#)]
56. Ciaparrone, G.; Serra, A.; Covito, V.; Finelli, P.; Scarpato, C.A.; Tagliaferri, R. A Deep Learning Approach for Road Damage Classification. In *Lecture Notes in Electrical Engineering*; Springer: Singapore, 2018; pp. 655–661. [[CrossRef](#)]
57. Maeda, H.; Sekimoto, Y.; Seto, T.; Kashiya, T.; Omata, H. Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 1127–1141. [[CrossRef](#)]
58. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.
59. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
60. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. *arXiv* **2016**, arXiv:1612.08242.
61. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.

62. Mandal, V.; Uong, L.; Adu-Gyamfi, Y. Automated Road Crack Detection Using Deep Convolutional Neural Networks. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018. [[CrossRef](#)]
63. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *arXiv* **2016**, arXiv:1412.7062.
64. Pohlen, T.; Hermans, A.; Mathias, M.; Leibe, B. Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 22–25 July 2017.
65. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? The KITTI vision benchmark suite. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 16–21 June 2012; pp. 3354–3361.
66. Zhang, Z.; Fidler, S.; Urtasun, R. Instance Segmentation for Autonomous Driving With Deep Densely Connected MRFs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
67. Lee, D.C.; Hebert, M.; Kanade, T. Geometric reasoning for single image structure recovery. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 2136–2143. [[CrossRef](#)]
68. Kundu, A.; Li, Y.; Dellaert, F.; Li, F.; Reh, J.M. Joint Semantic Segmentation and 3D Reconstruction from Monocular Video. In *Computer Vision—ECCV 2014*; Springer International Publishing: Cham, Switzerland, 2014; pp. 703–718. [[CrossRef](#)]
69. Pham, Q.; Hua, B.; Nguyen, T.; Yeung, S. Real-Time Progressive 3D Semantic Segmentation for Indoor Scenes. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 7–11 January 2019; pp. 1089–1098. [[CrossRef](#)]
70. Milletari, F.; Navab, N.; Ahmadi, S. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571. [[CrossRef](#)]
71. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Stoyanov, D., Taylor, Z., Carneiro, G., Syeda-Mahmood, T., Martel, A., Maier-Hein, L., Tavares, J.M.R., Bradley, A., Papa, J.P., Belagiannis, V., et al., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 3–11.
72. Alom, M.Z.; Hasan, M.; Yakopcic, C.; Taha, T.M.; Asari, V.K. Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. *arXiv* **2018**, arXiv:1802.06955.
73. 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](#)]
74. 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](#)]
75. Kalfarisi, R.; Wu, Z.Y.; Soh, K. Crack Detection and Segmentation Using Deep Learning with 3D Reality Mesh Model for Quantitative Assessment and Integrated Visualization. *J. Comput. Civ. Eng.* **2020**, *34*, 04020010. [[CrossRef](#)]
76. Kang, D.; Benipal, S.S.; Gopal, D.L.; Cha, Y.J. Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning. *Autom. Constr.* **2020**, *118*, 103291. [[CrossRef](#)]
77. Ni, F.; Zhang, J.; Chen, Z. Zernike-moment measurement of thin-crack width in images enabled by dual-scale deep learning. *Comput.-Aided Civ. Infrastruct. Eng.* **2019**, *34*, 367–384. [[CrossRef](#)]
78. Zhang, K.; Cheng, H.D.; Zhang, B. Unified Approach to Pavement Crack and Sealed Crack Detection Using Preclassification Based on Transfer Learning. *J. Comput. Civ. Eng.* **2018**, *32*, 04018001. [[CrossRef](#)]
79. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
80. Wei, F.; Yao, G.; Yang, Y.; Sun, Y. Instance-level recognition and quantification for concrete surface bughole based on deep learning. *Autom. Constr.* **2019**, *107*, 102920. [[CrossRef](#)]
81. Kim, B.; Cho, S. Image-based concrete crack assessment using mask and region-based convolutional neural network. *Struct. Control Health Monit.* **2019**, *26*, e2381. [[CrossRef](#)]
82. Szegedy, C.; Liu, W.; 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. [[CrossRef](#)]
83. Fang, F.; Li, L.; Rice, M.; Lim, J. Towards Real-Time Crack Detection Using a Deep Neural Network With a Bayesian Fusion Algorithm. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 2976–2980. [[CrossRef](#)]
84. Tan, C.; Uddin, N.; Mohammed, Y.M. Deep Learning-Based Crack Detection Using Mask R-CNN Technique. In Proceedings of the 9th International Conference on Structural Health Monitoring of Intelligent Infrastructure, St. Louis, MI, USA, 4–7 August 2019.
85. Ni, F.; Zhang, J.; Chen, Z. Pixel-level crack delineation in images with convolutional feature fusion. *Struct. Control Health Monit.* **2019**, *26*, e2286. [[CrossRef](#)]
86. 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](#)]
87. Badrinarayanan, V.; Handa, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling. *arXiv* **2015**, arXiv:1505.07293.

88. Zhang, K.; Cheng, H.; Gai, S. Efficient Dense-Dilation Network for Pavement Cracks Detection with Large Input Image Size. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 884–889. [\[CrossRef\]](#)
89. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
90. Takikawa, T.; Acuna, D.; Jampani, V.; Fidler, S. Gated-SCNN: Gated Shape CNNs for Semantic Segmentation. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
91. Zhang, L.; Yang, F.; Daniel Zhang, Y.; 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. [\[CrossRef\]](#)
92. Fan, Z.; Wu, Y.; Lu, J.; Li, W. Automatic Pavement Crack Detection Based on Structured Prediction with the Convolutional Neural Network. *arXiv* **2018**, arXiv:1802.02208.
93. Fan, Z.; Li, C.; Chen, Y.; Mascio, P.D.; Chen, X.; Zhu, G.; Loprencipe, G. Ensemble of Deep Convolutional Neural Networks for Automatic Pavement Crack Detection and Measurement. *Coatings* **2020**, *10*, 152. [\[CrossRef\]](#)
94. Zhang, A.; Wang, K.C.P.; Fei, Y.; Liu, Y.; Tao, S.; Chen, C.; Li, J.Q.; Li, B. Deep Learning-Based Fully Automated Pavement Crack Detection on 3D Asphalt Surfaces with an Improved CrackNet. *J. Comput. Civ. Eng.* **2018**, *32*, 04018041. [\[CrossRef\]](#)
95. Fei, Y.; Wang, K.C.P.; Zhang, A.; Chen, C.; Li, J.Q.; Liu, Y.; Yang, G.; Li, B. Pixel-Level Cracking Detection on 3D Asphalt Pavement Images Through Deep-Learning-Based CrackNet-V. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 273–284. [\[CrossRef\]](#)
96. 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. [\[CrossRef\]](#)
97. Zhang, A.; Wang, K.C.P.; Fei, Y.; Liu, Y.; Chen, C.; Yang, G.; Li, J.Q.; Yang, E.; Qiu, S. Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces with a Recurrent Neural Network. *Comput.-Aided Civ. Infrastruct. Eng.* **2019**, *34*, 213–229. [\[CrossRef\]](#)
98. Noh, H.; Hong, S.; Han, B. Learning Deconvolution Network for Semantic Segmentation. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
99. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Lecture Notes in Computer Science*; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241. [\[CrossRef\]](#)
100. Jegou, S.; Drozdal, M.; Vazquez, D.; Romero, A.; Bengio, Y. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Honolulu, HI, USA, 22–25 July 2017.
101. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.
102. David Jenkins, M.; Carr, T.A.; Iglesias, M.I.; Buggy, T.; Morison, G. A Deep Convolutional Neural Network for Semantic Pixel-Wise Segmentation of Road and Pavement Surface Cracks. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 2120–2124. [\[CrossRef\]](#)
103. Cheng, J.; Xiong, W.; Chen, W.; Gu, Y.; Li, Y. Pixel-level Crack Detection using U-Net. In Proceedings of the TENCON 2018—2018 IEEE Region 10 Conference, Jeju, Korea, 28–31 October 2018. [\[CrossRef\]](#)
104. Konig, J.; Jenkins, M.D.; Barrie, P.; Mannion, M.; Morison, G. A Convolutional Neural Network for Pavement Surface Crack Segmentation Using Residual Connections and Attention Gating. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019. [\[CrossRef\]](#)
105. Escalona, U.; Arce, F.; Zamora, E.; Sossa Azuela, J.H. Fully convolutional networks for automatic pavement crack segmentation. *Comput. Sist.* **2019**, *23*, 451–460. [\[CrossRef\]](#)
106. Liu, Z.; Cao, Y.; Wang, Y.; Wang, W. Computer vision-based concrete crack detection using U-net fully convolutional networks. *Autom. Constr.* **2019**, *104*, 129–139. [\[CrossRef\]](#)
107. 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. [\[CrossRef\]](#)
108. Zhang, K.; Zhang, Y.; Cheng, H. CrackGAN: Pavement Crack Detection Using Partially Accurate Ground Truths Based on Generative Adversarial Learning. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 1306–1319. [\[CrossRef\]](#)
109. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. *arXiv* **2014**, arXiv:1406.2661.
110. Chen, T.; Cai, Z.; Zhao, X.; Chen, C.; Liang, X.; Zou, T.; Wang, P. Pavement crack detection and recognition using the architecture of segNet. *J. Ind. Inf. Integr.* **2020**, *18*, 100144. [\[CrossRef\]](#)
111. Zou, Q.; Zhang, Z.; Li, Q.; Qi, X.; Wang, Q.; Wang, S. DeepCrack: Learning Hierarchical Convolutional Features for Crack Detection. *IEEE Trans. Image Process.* **2019**, *28*, 1498–1512. [\[CrossRef\]](#) [\[PubMed\]](#)
112. Bang, S.; Park, S.; Kim, H.; Kim, H. Encoder–decoder network for pixel-level road crack detection in black-box images. *Comput.-Aided Civ. Infrastruct. Eng.* **2019**, *34*, 713–727. [\[CrossRef\]](#)
113. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In *Computer Vision—ECCV 2014*; Springer International Publishing: Cham, Switzerland, 2014; pp. 818–833. [\[CrossRef\]](#)

114. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 22–25 July 2017.
115. 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](#)]
116. Mei, Q.; Gül, M. Multi-level feature fusion in densely connected deep-learning architecture and depth-first search for crack segmentation on images collected with smartphones. *Struct. Health Monit.* **2020**, *19*, 1726–1744. [[CrossRef](#)]
117. Mei, Q.; Gül, M. A cost effective solution for pavement crack inspection using cameras and deep neural networks. *Constr. Build. Mater.* **2020**, *256*, 119397. [[CrossRef](#)]
118. Yang, X.; Li, H.; Yu, Y.; Luo, X.; 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](#)]
119. Zhang, J.; Lu, C.; Wang, J.; Wang, L.; Yue, X.G. Concrete Cracks Detection Based on FCN with Dilated Convolution. *Appl. Sci.* **2019**, *9*, 2686. [[CrossRef](#)]
120. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. *arXiv* **2016**, arXiv:1511.07122.
121. 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](#)]
122. Dung, C.V.; Anh, L.D. Autonomous concrete crack detection using deep fully convolutional neural network. *Autom. Constr.* **2019**, *99*, 52–58. [[CrossRef](#)]
123. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
124. Amhaz, R.; Chambon, S.; Idier, J.; Baltazart, V. Automatic Crack Detection on Two-Dimensional Pavement Images: An Algorithm Based on Minimal Path Selection. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 2718–2729. [[CrossRef](#)]
125. Xie, S.; Tu, Z. Holistically-Nested Edge Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
126. Liu, Y.; Cheng, M.M.; Hu, X.; Wang, K.; Bai, X. Richer Convolutional Features for Edge Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 22–25 July 2017.