applied sciences

MDPI

Article

A Review of Deep Learning Advancements in Road Analysis for Autonomous Driving

Adrian-Paul Botezatu [1], Adrian Burlacu [1,*] and Ciprian Orhei [2]

[1] Department of Automatic Control and Applied Informatics, Faculty of Automatic Control and Computer Engineering, "Gheorghe Asachi" Technical University of Iasi, 700050 Iasi, Romania; adrian-paul.botezatu@academic.tuiasi.ro

[2] Department of Communications, Politehnica University of Timișoara, 300006 Timișoara, Romania; ciprian.orhei@upt.ro

* Correspondence: adrian.burlacu@academic.tuiasi.ro

Abstract: The rapid advancement of autonomous vehicle technology has brought into focus the critical need for enhanced road safety systems, particularly in the areas of road damage detection and surface classification. This paper explores these two essential components, highlighting their importance in autonomous driving. In the domain of road damage detection, this study explores a range of deep learning methods, particularly focusing on one-stage and two-stage detectors. These methodologies, including notable ones like YOLO and SSD for one-stage detection and Faster R-CNN for two-stage detection, are critically analyzed for their efficacy in identifying various road damages under diverse conditions. The review provides insights into their comparative advantages, balancing between real-time processing and accuracy in damage localization. For road surface classification, the paper investigates the classification techniques based on both environmental conditions and material road composition. It highlights the role of different convolutional neural network architectures and innovations at the neural level in enhancing classification accuracy under varying road and weather conditions. The main finding of this work is that it offers a comprehensive overview of the current state of the art, showcasing significant strides in utilizing deep learning for road analysis in autonomous vehicle systems. The study concludes by underscoring the importance of continued research in these areas to further refine and improve the safety and efficiency of autonomous driving.

Keywords: road damage detection; road surface classification; CNN

Citation: Botezatu, A.-P.; Burlacu, A.; Orhei, C. A Review of Deep Learning Advancements in Road Analysis for Autonomous Driving. *Appl. Sci.* **2024**, *14*, 4705. https://doi.org/10.3390/app14114705

Academic Editor: Luís Picado Santos

Received: 30 April 2024
Revised: 25 May 2024
Accepted: 27 May 2024
Published: 30 May 2024

1. Introduction

Autonomous vehicle technology has been a rapidly evolving field, with many companies actively developing systems for various types of vehicles. The automotive industry has witnessed the implementation of partially automated systems over the past few years, while driverless systems are still undergoing advanced tests. The pursuit of autonomous driving technology has been a key focus of research and development efforts across car manufacturers, academia, and research institutes since 1980.

Today, a key priority for the automotive industry is ensuring driver safety and preventing accidents through the use of innovative technologies found in Advanced Driver Assistance Systems (ADAS) [1–3]. These systems utilize a range of methods to interpret the environment, equipping them with the ability to detect and avoid obstacles on the road [4–6].

In this context, the analysis of road elements has emerged as a critical component for enhancing road safety in autonomous driving systems. The accurate detection of road damage, such as cracks, potholes, and surface wear, is imperative for autonomous vehicles to navigate safely and efficiently. Similarly, the classification of road surfaces plays a vital role. By understanding the type of surface (for example, asphalt, concrete, or gravel) or what external factor is on the road (dry, wet, or snow), autonomous vehicles can optimize their driving patterns and stability controls, adapting to different conditions to maintain safety and comfort.

After road damage detection or road surface classification is performed by an autonomous vehicle, several critical decisions are made to ensure safe and efficient operation. First, the vehicle could assess the severity of the detected hazard, such as the size and depth of a pothole, or the type of road surface condition, such as wet, icy, or muddy. Based on this evaluation, the vehicle adjusts its speed to match the road conditions, reducing speed significantly in cases of severe damage or slippery surfaces to prevent loss of control. Additionally, the vehicle modifies its trajectory to avoid the hazard, making necessary steering adjustments or rerouting entirely if the damage is extensive. Furthermore, the system may activate safety protocols such as increased braking sensitivity and stability control to maintain optimal safety. These real-time decisions are essential for maintaining the integrity of the autonomous driving system and ensuring passenger safety. By integrating advanced sensor data and machine learning algorithms, the vehicle can dynamically respond to changing road conditions, enhancing safety and ride quality [2,7].

In this review, we concentrate on the following two critical aspects of the autonomous car system, specifically within the context of road safety:

- Road Damage Detection (RDD): This involves using Deep Learning (DL) methods to identify and classify road damages such as cracks, potholes, or other surface irregularities. Accurate detection of such damages is crucial for autonomous vehicles to ensure safe navigation and avoid potential dangers
- Road Surface Classification (RSC): This refers to the capability of an autonomous system to distinguish between different road types (like concrete, cobblestone, asphalt, paved/unpaved) or the external factors that can make a driver less safe on the road (like dry, wet, snow, mud).

This paper is organized into four main sections, each focusing on a specific aspect of autonomous vehicle technology and road safety. Section 2 outlines the review methodology, detailing the comprehensive approach used to gather and analyze the relevant literature from various databases, focusing on the most recent and significant studies in the field. Section 3 delves into RDD, exploring the advancements in one-stage and two-stage detectors for identifying road damage through DL techniques. This section assesses various models and approaches, highlighting their effectiveness in different scenarios. Section 4 shifts the focus to RSC, examining how different methodologies and neural network models are employed to categorize road surfaces based on material composition and environmental conditions. This section underscores the importance of accurate RSC in enhancing the safety and efficiency of autonomous driving systems. Finally, Section 5 presents the conclusions drawn from the review, summarizing the key findings and contributions of the paper while suggesting potential avenues for future research in this rapidly evolving field.

2. Review Methodology

This study's narrative review was predominantly conducted using the Web of Science (WoS) database, complemented by the Google Scholar database, centered on journal and conference articles published between 2018 and the beginning of 2024. Figure 1 presents a bibliometric analysis regarding the number of publications reviewed in this work in accordance with the publication year.

Search criteria based on various keywords influenced this bibliometric analysis. In the case of RDD, these keywords were mainly road damage detection, road crack detection, and asphalt damage. In the same idea, for RSC, the main search keywords were road surface classification, road surface detection, road surface monitoring, and road classification. Figures 2 and 3 present a keyword analysis resulting from the analyzed papers for RDD and RSC, respectively. In each network, a node represents a keyword, with the node size reflecting how frequently the keyword appears. Links between nodes indicate keyword co-occurrence; thicker links represent higher co-occurrence frequency. Colors differentiate thematic clusters, grouping together multiple related keywords.

Figure 4 presents a pie chart that provides a clear visual comparison of publications categorized into conference papers and journal articles, highlighting their status with respect to WOS indexing. It illustrates that a higher proportion of journal articles (33 out of 39) are indexed by WoS compared to conference papers, where only 28 out of 46 are WoS indexed. This discrepancy may suggest that journal articles tend to meet Wos's indexing criteria more consistently than conference papers, potentially reflecting higher standards or more rigorous peer review processes.

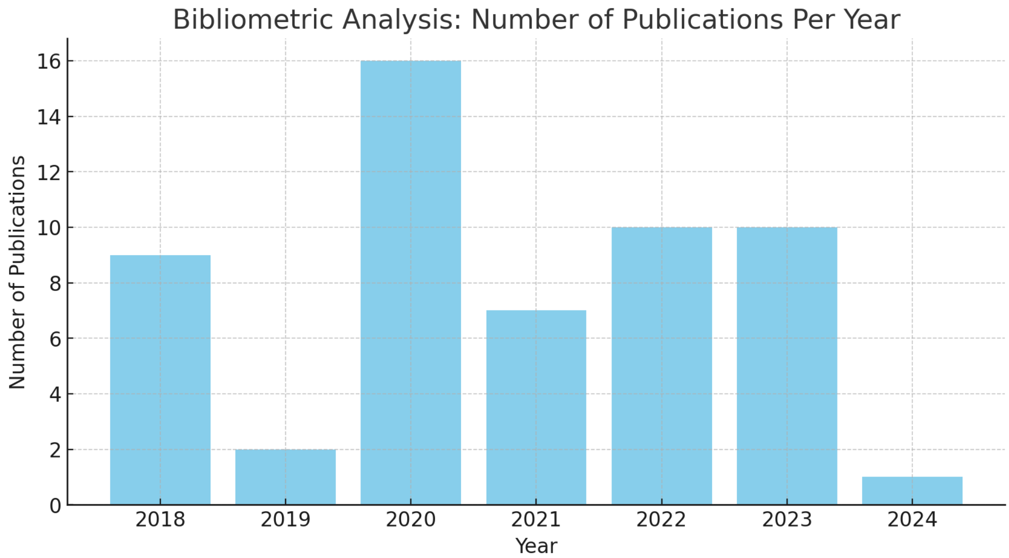

Figure 1. Bibliometric analysis: number of publications per year.

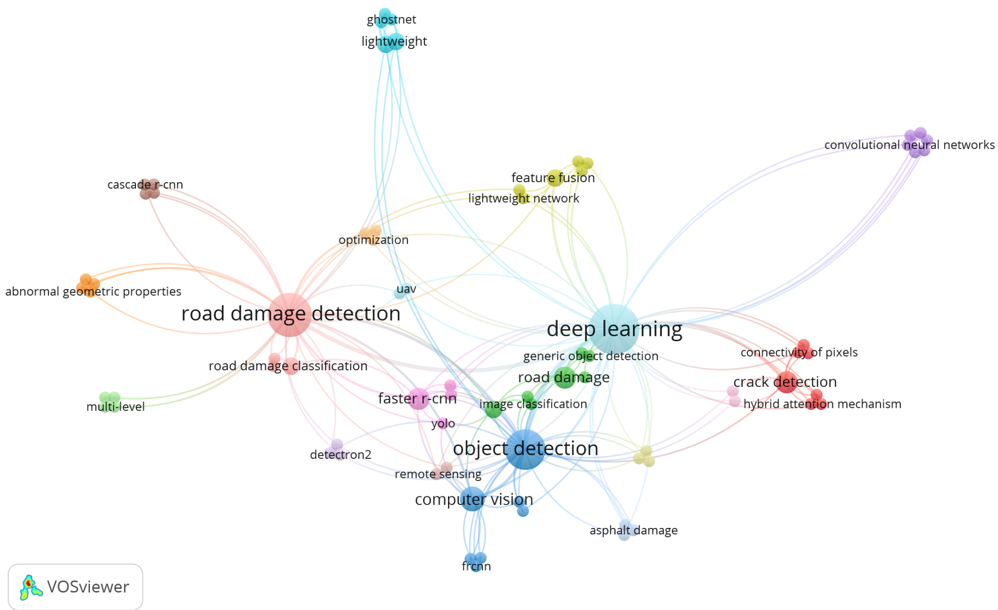

Figure 2. Keyword analysis for Road Damage Detection (RDD) papers considered.

The reasoning behind centering our investigation on DL methods, particularly Convolutional Neural Networks (CNNs), stems from the significant strides these technologies have made in the automotive area. CNNs, owing to their advanced feature extraction abilities, are remarkably effective for tasks that require detailed visual data interpretation, such as identifying various forms of road damage and categorizing road surface types. These networks could handle the diverse and unpredictable nature of road environments, including changes in texture, lighting, and weather conditions.

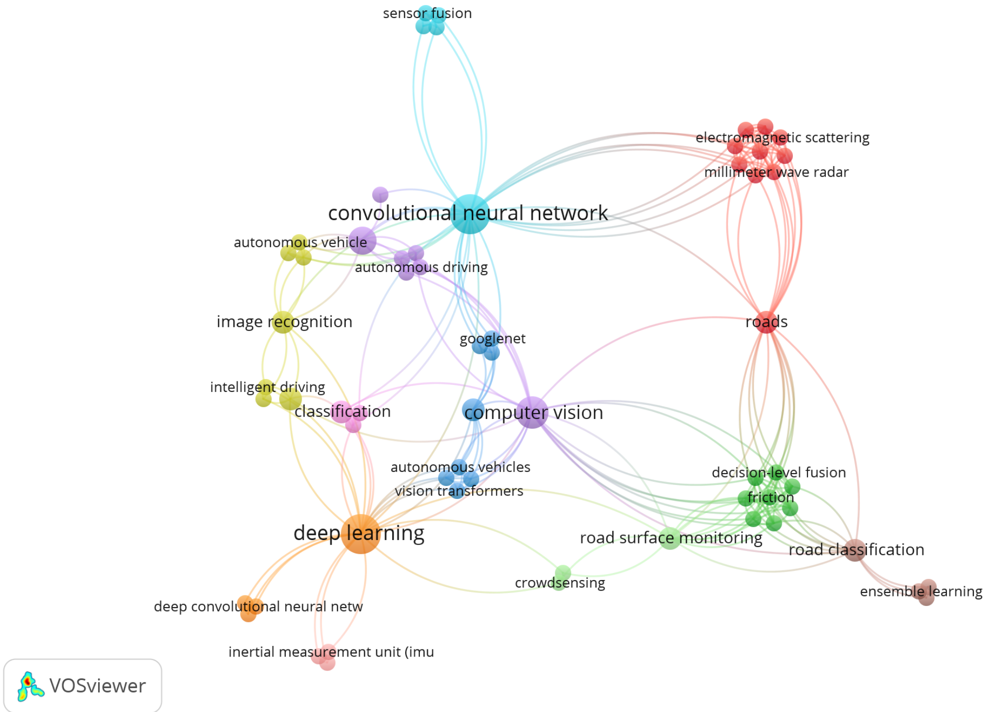

Figure 3. Keyword analysis for Road Surface Classification (RSC) papers considered.

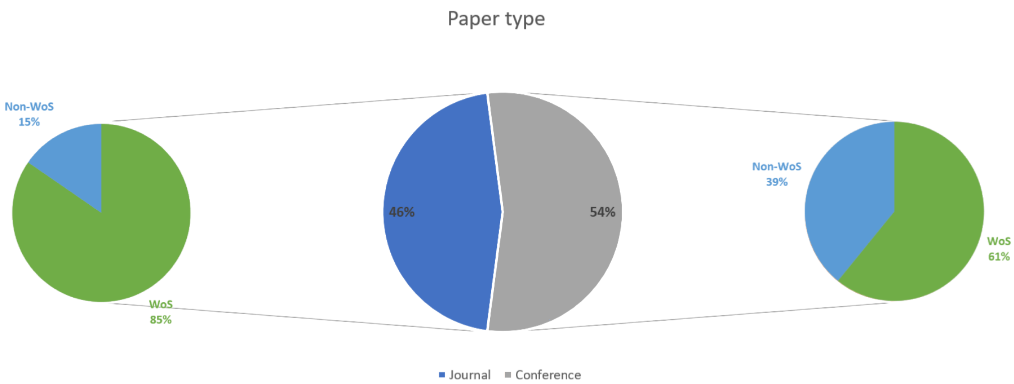

Figure 4. Distribution of Web of Science (WoS) Indexed and Non-WoS Indexed Conference Papers.

Traditional methods often struggle with such complexity and the sheer volume of data, whereas CNNs have repeatedly set new performance standards in similar image-based applications. This trend underscores their superiority and dependability compared to older techniques.

Our survey, therefore, delves into how DL, with a focus on CNNs, is reshaping road infrastructure analysis, offering insights into the current State Of The Art (SOTA) and suggesting directions for future research in this crucial field.

The articles for the first part of the survey, RDD, were categorized into two distinct groups based on the methods employed: one-stage detectors or two-stage detectors. This classification was undertaken to provide a structured and comprehensive overview of the field, reflecting the diverse approaches and technologies currently in use.

Regarding RDD, one-stage detectors, known for their speed and efficiency in real-time applications, represent a crucial segment of research, particularly appealing for scenarios demanding rapid processing. In contrast, two-stage detectors, with their heightened accuracy and detailed object assessment, are pivotal in applications where P is paramount.

The second part of the survey is dedicated to RSC, where the articles are divided into two primary categories: the nature of a road's surface material and the nature of the environmental conditions that have an impact on a surface. This bifurcation is designed to systematically address the two key aspects that define road surface characteristics.

The first category encompasses studies focusing on differentiating road types based on material composition, such as concrete, cobblestone, asphalt, and distinctions between paved and unpaved surfaces. This classification is critical for understanding the inherent properties and durability of different road types, which directly influences maintenance strategies and vehicle interaction.

The second category delves into how external factors like weather conditions—dry, wet, snowy, or muddy—affect road surfaces. This is particularly important for adaptive driving systems and safety protocols, as these conditions significantly alter driving dynamics.

By categorizing the articles in this manner, the survey aims to provide a nuanced understanding of RSC, highlighting how both material composition and environmental factors play a vital role in road maintenance, safety, and the overall driving experience. This approach ensures comprehensive coverage of the field, offering valuable insights for future research and practical applications in road infrastructure and vehicle technology. A summary of the review methodology is described in Figure 5.

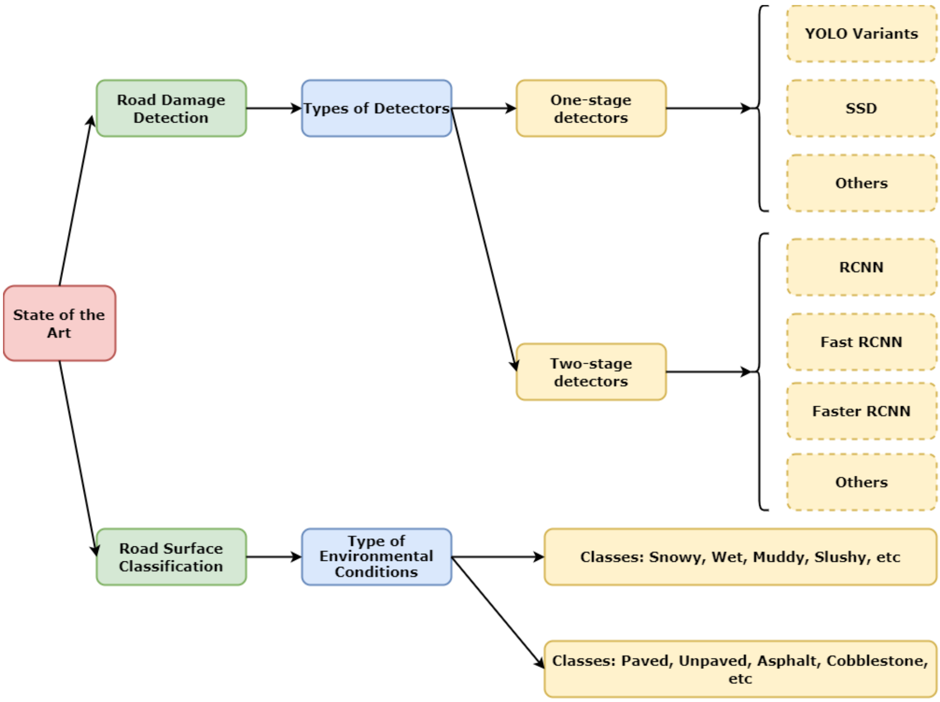

Figure 5. Summary of the review methodology.

3. Road Damage Detection

In this chapter, we delve into the methodologies that define the current SOTA in RDD. This field, crucial for maintaining road infrastructure and ensuring safety, has seen remarkable advancements thanks to the intersection of Computer Vision (CV), Machine Learning (ML), and sensor technology. Our focus will be on exploring and evaluating the diverse range of methods that have been developed and refined to address the challenges inherent in accurately detecting road damage.

The traditional approach to RDD, largely manual and visually based, has been significantly overshadowed by the emergence of automated systems powered by advanced computational techniques. Among these, DL, particularly the use of CNNs, has emerged as a leading method.

The approach to RDD aligns closely with Object Detection (OD), an established technique in CV. In this application, the primary targets for detection are various types of road damage, such as cracks and potholes. Figure 6 presents an example of this type of detection.

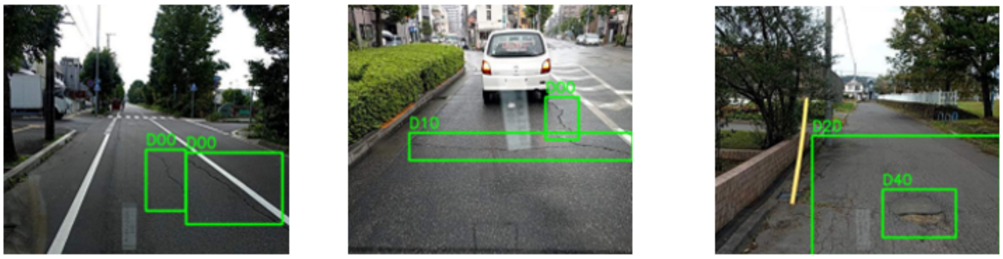

Figure 6. Example of Road Damage Detection (RDD) detection from [8].

Recognizing RDD as a facet of OD is essential, as it guides the application of specialized DL models adept at pointing and identifying objects within the visual data. Two of the most popular methods in OD are the one-stage detector and two-stage detector due to their effectiveness in balancing the critical trade-offs between speed and accuracy. Given this aspect, we analyzed the SOTA through these two methods.

To better understand the effect of each solution presented, we have centralized the most used evaluation metrics in our analyzed papers in Table 1 regarding RDD. The brief description alongside the mathematical formula serves as a quick reference to the essential statistical tools for evaluating model predictions.

Table 1. Most used performance metrics.

Metric	Description	Formula
Precision (P)	The ratio of correctly predicted positive observations to the total predicted positives.	$P = \frac{TP}{TP+FP}$
Recall (R)	The ratio of correctly predicted positive observations to the total actual positives.	$R = \frac{TP}{TP+FN}$
Average Precision (AP)	Sum of products of P at each R threshold and the change in R from the previous threshold.	$AP = \sum_n (R_n - R_{n-1}) P_n$
mAP	Mean of the average precision scores across different classes or queries, averaging the area under the precision–recall curve.	$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i$
F1 Score	Harmonic mean of precision and recall.	$F1 = 2 \times \frac{P \cdot R}{P+R}$
mAP@50	mAP calculated specifically at an Intersection over Union (IoU) threshold of 0.50.	mAP at IoU ≥ 0.50
Accuracy	Ratio of correctly predicted observations to the total observations.	$Acc = \frac{TP+TN}{\text{Total samples}}$

Note: TP: True Positives; FP: False Positives; TN: True Negatives; FN: False Negatives.

3.1. One-Stage Detectors for Road Damage Analysis

The influence of one-stage detectors in the field of RDD has been transformative, marking a significant shift towards more efficient methods. Characterized by their ability to perform OD in single-pass, one-stage detectors like YOLO (You Only Look Once) [9] or SSD (Single Shot Multibox Detector) [10] have streamlined the process of identifying road damages. An example in this sense is the work from [8] where the authors evaluate two specific models: SSD using Inception-V2 [11] and SSD using MobileNet [12] frameworks. These models were chosen for their balance of accuracy and efficiency, which is particularly important for real-time processing on devices with limited computational resources. In terms of performance, the work reveals the finding that SSD using MobileNet demonstrates superior results compared to the usage of Inception V2. This idea is sustained by computing high Recall (R) and Precision (P) in detecting various types of road damage in detecting longitudinal, lateral, and alligator cracks. The research also delves into the inference speed of the models, an essential factor for real-time applications. It is noted that the SSD MobileNet model could process data in 1.5 seconds on a smartphone, enabling the potential for real-time damage detection in moving vehicles. Another pivotal aspect is the development of a large-scale dataset of road damage images captured through a smartphone, which includes 9053 images with 15,435 instances of road damage across eight categories ([8]).

By working with the same dataset, the authors from [13] use RetinaNet [14] as the one-stage detector. The innovative core of their methodology lies in the application of RetinaNet's focal loss function, which significantly improves the model's performance on class imbalance—a common issue in this dataset where damaged areas are sparse compared to the undamaged. In the same manner, the authors from [15] utilized the YOLO network, with variations in the number of training iterations, minimum confidence threshold, and non-maximum suppression. Through the last one, they achieved the best results by obtaining an $F1$ score of 0.62.

In comparison with the solutions above, the authors from [16] revolve around the application of RDD by the usage of SSD network enhanced with the pre-trained networks VGG [17] and Residual Networks (ResNet) [18], which provide robust feature extraction capabilities. They employed an ensemble strategy that involves training multiple instances of the SSD model, each under different configurations and data subsets, by using higher-resolution inputs (1024 $\times$ 1024 pixels) and a random cropping technique. The ensemble combines predictions from these models using non-maximum suppression to filter overlapping detections alongside a high confidence threshold to minimize false positives. Additionally, Intersection over Union (IoU) thresholds are applied for refining predictions.

Another approach based on the dataset from [8] comes from [19] where the authors used a YOLOv4 [20] model with a Proportional–Integral–Differential (PID) optimizer. This modification aims to address the overshoot problem commonly encountered in DL, thus improving training efficiency. The PID optimizer, integrated into YOLOv4, leverages the derivative term of the PID control theory, focusing on the rate of change of the gradient during model training. This approach leads to faster convergence and reduces training oscillations, effectively improving the model's performance. The modified YOLOv4 model demonstrates superior mean average precision (mAP) compared to the original YOLOv4, with a notable increase of 5.29% (on the dataset from [8]) and 5.22% (on the dataset from [21]).

A similar approach is the work from [22], which presents an approach that involves a fusion of low-level features from the early layer of the network with high-level features from the deeper to capture both semantic information and detailed visual characteristics of road damages. Also, they propose a novel loss function that is sensitive to the direction of road crack extension, addressing the common issue of incomplete bounding box predictions. This direction-aware loss function assigns different weights to the x and y dimensions of the bounding box, depending on the class of road damage.

An approach based on densely connected convolutional layers is presented in [23], where the authors employed a CNN known as DenseNet (Densely Connected Convolutional Network) [24] for automatic pavement crack detection. In their methodology, the authors utilized DenseNet201, a variant of DenseNet with 201 layers, to process images of pavement cracks. A novel aspect of their approach was the introduction of a unique loss function that considers the connectivity of pixels. This function aimed to address the challenges posed by the output of transposed convolution layers, which often lead to sparse and disconnected detections. By focusing on the connectivity aspect, the new loss function helped to generate more coherent and accurate crack predictions. The approach is validated using two datasets: one obtained from a handheld smartphone (CFD dataset [25]) and another one acquired by them from a high-speed camera mounted on a moving car (EdmCrack1000).

A different approach is presented in [26], where the authors used EfficientDet [27], a single-stage detector that combines the benefits of EfficientNet (used for feature extraction) and Bi-directional Feature Pyramid Network (BiFPN) [28] for feature fusion. Varying the input image resolution, the batch size, and the learning rate, the paper reports $F1$-scores ranging from 52% to 56%. The authors used the dataset from [29], an upgraded version of the dataset from [8], consisting of 26,620 images collected from India, the Czech Republic, and partially from Slovakia, with four damage categories: longitudinal cracks, lateral cracks, alligator cracks, and potholes. On this dataset, the authors from [30] analyzed their dataset using SSD models based on MobileNet architectures. They trained and evaluated 16 different models across 30 scenarios using various data combinations from the three countries. This strategy enhanced the models' generalization capabilities, making them more robust and effective in different settings.

In [31], the authors present an innovative approach to RDD by employing a deep ensemble learning model. The paper explores the effectiveness of various YOLO-v4 models trained under different conditions, highlighting the impact of training iterations and image resolutions on detection performance. The ensemble approach, which joins 30 multiple-model predictions, demonstrates notable success, achieving $F1$ scores of 0.628 and 0.6358 on the dataset. In the same manner of working with an ensemble strategy, in [32], the proposed method is based on adopting ultralytics-YOLO (u-YOLO) [33] as the core model, enhanced through three ensemble learning strategies: ensemble prediction, ensemble model, and a hybrid approach between them. These strategies involve test time augmentation and averaging predictions from multiple trained models, aiming to improve robustness and accuracy. The approaches are evaluated using the dataset from [29], demonstrating an F1 score up to 0.67.

The authors from [34] present an approach based on image segmentation and YOLOV4 to identify road damages. The methodology starts with using the DeepLabV3Plus [35] model for image segmentation to isolate road areas, creating what they refer to as a 'road-interest map'. This map is then used as the basis for training their OD model. A different methodology is proposed in [36], where the authors introduced a novel model for RDD, innovatively enhancing the YOLOv5s framework. Their approach, YOLO-LRDD, incorporates significant modifications such as a new backbone network, Shuffle-ECANet, which combines ShuffleNetV2 [37] with the efficient channel attention module for increased efficiency and speed. The model also uses Weighted BiFPN instead of PANet for better feature fusion and introduces FOCAL-EIOU loss to address sample imbalance and improve anchor box quality. These improvements result in a more lightweight and faster model than the standard YOLOv5 while maintaining a comparable *Acc*.

In the same manner of constructing a lightweight model, the authors from [38] introduce E-EfficientDet, a model based on EfficientDet [27]. They integrate a feature extraction enhancement module, utilizing semi-dense connectivity in its feature pyramid, focusing on improving multi-scale feature extraction and information reuse across different network layers. By employing a compound scaling strategy, E-EfficientDet offers scalable solutions that adjust to varying hardware capabilities, ensuring efficient deployment on both high-end and mobile devices. The model outperforms existing benchmarks, like YOLOv5s and YOLOv7-tiny [39], in *Acc* and speed, demonstrating a 2.41% increase in detection P over the original EfficientDet-D0.

An interesting approach is presented in [40], where the authors proposed a multi-faceted system design, integrating roadside data collection, edge computing, and cloud-based data transfer to achieve real-time road damage visualization on client devices. The deployment on the NVIDIA Jetson NX platform (NVIDIA, Santa Clara, CA, USA), optimized with NVIDIA's TensorRT and DeepStream technologies, showcases significant advancements in processing speed and detection accuracy. Experimental outcomes demonstrate the system's capability to accurately and swiftly identify road damages, outperforming previous YOLO versions.

Based on the RDD2020 dataset from [29], the same authors examined an upgraded dataset, RDD2022 [41], with 47,420 road images from six countries: Japan, India, Czech Republic, Norway, the United States, and China. The annotations are based on four types of road damage: longitudinal cracks, transverse cracks, alligator cracks, and potholes. In [29] it is presented the evolution of the dataset. Starting from this dataset, the authors from [42] adopt an interesting approach by utilizing YOLOv7, enhanced with coordinate attention to better capture spatial features of the road damages. Their methodology is distinguished by the use of Google Street View for extensive data across the United States to collect the coordinate attention, which was integrated with the neural model. The approach yields promising results on this specific dataset, indicating the potential for generalization to the original dataset [41].

A different approach is the work from [43], where the innovation lies in their use of image tiling—a technique that breaks down high-resolution images into smaller, standardized patches suitable for YOLO processing—to effectively integrate high-resolution images within the dataset. They employed an ensemble of twelve YOLOv5 models to detect four types of road damages, with their methodology focusing on adapting the image input size and optimizing the model using various data augmentation strategies, including flips, scaling, and color adjustments. Based on ensemble learning, the authors from [44] are using multiple approaches by taking YOLOv5 as the baseline: the first one considers training on the images from all countries, the second one uses multiple models where each is trained on the images from a specific country, the third one uses new anchor boxes for training the YOLOv5, and the fourth one creates various ensemble models by combining the previous three approaches.

In [45], the authors present the model YOLOv5s-DSG, focused on addressing the challenges of detecting small-scale road damages under varying resolutions. The model innovates by optimizing both the depth and width of the YOLOv5s architecture to minimize the impact on detection performance while reducing computational overhead. A critical innovation is the incorporation of the Ghost module, which substitutes traditional convolution operations to decrease the model's parameter count, thereby rendering the model more lightweight and increasing its processing speed. Furthermore, the introduction of the Space-to-depth-Conv module is specifically designed to enhance the model's capability in detecting low-resolution and small objects, a common challenge in road disease detection scenarios.

In [46], the authors propose another variation starting from a baseline YOLOv8 [33] model, BL-YOLOv8. Central to this model is the adoption of a reconstructed neck structure, incorporating the BiFPN, a move aimed at diminishing the model's parameters and computational complexity. A notable innovation is the introduction of the LSK-attention mechanism, a dynamic large convolutional kernel attention strategy devised to broaden the model's receptive field, thereby boosting the P of OD. Additionally, the model integrates a SimSPPF module within its feature pyramid layer, a strategic enhancement designed to expedite the model's operational speed. Through these methodological advancements, L-YOLOv8 achieves a 3.3% improvement in mAP@0.5 over the baseline YOLOv8 model, alongside a remarkable reduction in both parameter volume by 29.92% and computational load by 11.45%.

Other remarkable results have been obtained by working with other datasets. For example, in [47], the authors labeled 665 Brazilian highway images with two classes, cracks and potholes. Using data augmentation, they deployed YOLO to identify these road damages and obtained a 91% defect detection rate. In the same manner of working with YOLO, the authors from [48] work with a Spanish roads dataset of 600 images with two labels, potholes, and cracks. With some image augmentation techniques, they tested three types of YOLO models with promising results. Another approach based on YOLO is the work from [49], where the authors propose CrackYOLO, a model based on YOLOv5, incorporating several innovative crack detection modules. Their tests are performed on a dataset consisting of 9801 rural roads with three main categories of cracks as labels: transverse cracks, longitudinal cracks, and reticular cracks.

In [50], the authors collected 2900 images from the Fengtai District of Beijing, labeling seven classes: longitudinal cracks, transverse cracks, alligator cracks, potholes, manhole covers, longitudinal patch, and transverse patch. They used YOLOv5s-M, a model incorporating a generalized feature pyramid network for effective cross-layer and cross-scale feature fusion and a diagonal intersection over union loss for boundary box regression.

3.2. Two-Stage Detectors for Road Damage Analysis

The evolution from R-CNN [51] to Faster R-CNN [52] has notably advanced the field of RDD, providing a nuanced approach to identifying road damages. Unlike their one-stage counterparts, the two-stage detectors generate region proposals and then classify those regions into object categories in the second stage. This hierarchical method could enhance the accuracy of detecting various road damages by meticulously examining potential areas of interest before classification, thereby offering a more precise and reliable analysis of road conditions.

Using the dataset from [8], the authors from [53] employed the Faster R-CNN framework to analyze aspect ratios and sizes of damaged areas in the dataset. The core of their methodology involves the utilization of two variants of the ResNet architecture, ResNet-101 and ResNet-152, as the backbone for feature extraction within the Faster R-CNN model. This choice is pivotal to their approach, as ResNet architectures are known for their deep residual learning framework, enabling the training of much deeper networks by addressing the vanishing gradient problem often encountered in such scenarios.

A different approach, but on the same dataset, is the method proposed in [54]. DenseNet [24] is the backbone for feature extraction within Mask R-CNN, a method that promises improvements for the model's efficiency in learning necessary features for accurate detection. This novel combination leverages the strength of DenseNet in feature extraction, characterized by its dense connectivity that ensures maximum information flow between layers. By employing DenseNet as the backbone for Mask R-CNN, the study not only enhances the model's capability in distinguishing subtle features of road damage but also improves the model's efficiency by reducing the number of parameters compared to traditional architectures.

Another approach has been developed, starting with the RDD 2020 dataset from [29]. In [55], the authors investigate the application of Detectron2's [56] implementation of Faster R-CNN across various configurations and base models. Their methodological innovation lies in evaluating Detectron2's Faster R-CNN with an X101-FPN base model. The study achieves $F1$ scores of 51.0% and 51.4% on two distinct test sets, indicating a promising direction for further refining the balance between P and R in RDD tasks.

A different approach is presented in [57] where, at the beginning, the authors classify the geographic origin of each image by employing a two-step process. Initially, a Country-Classifier, developed through the adaptation of Faster R-CNN, determined the image's country of origin by labeling the entire image with a country-specific bounding box. Subsequently, the image was directed to a specialized regional expert Faster R-CNN model, finely tuned to detect and classify road damages specific to that country.

An interesting method is presented in [58], where the authors fine-tune the Faster R-CNN architecture, employing ResNet and Feature Pyramid Networks (FPN) as backbones to enhance the model's detection capabilities. They conduct a thorough analysis to identify the most effective configurations, comparing the performance of models with different backbone networks like ResNet-50 and ResNet-101. The experiments reveal insights into the impact of various hyperparameters and training strategies on the model's ability to generalize across countries with diverse road conditions.

In [59], the authors propose to integrate a consistency filtering mechanism with the Cascade R-CNN model, aimed at enhancing the performance of OD systems in identifying RDD from images. Their methodology leverages unlabeled data through self-supervised learning, employing high-confidence samples with pseudo-labels for training and applying various data augmentation techniques to improve model robustness. The study emphasizes using advanced segmentation methods and model ensembling through weighted box fusion to refine the final predictions.

Based on the RDD-2022 dataset [41], in [60], the authors introduce an ensemble approach that combines the strengths of a one-stage and a two-stage detector, specifically YOLO and Faster R-CNN. As a one-stage detector, they used YOLOv5 and YOLOv7 as baseline, respectively, Faster RCNN with Swin Transformer as the backbone for the two-stage detector part.

Advancements have also been made by working with different datasets. For example, Ref. [61] presents a dataset of 323 images with a resolution of 4128 × 2322, captured under various shooting conditions, including different weather scenarios and illumination levels, for road crack detection. The evaluation is made with Faster R-CNN, which promises good results in terms of accuracy. A different approach is presented in [62], where the authors introduced CrdNet, an architecture that integrates a novel backbone network named LrNet [63] for enhanced feature extraction, a multi-scale and multi-aspect ratios anchor mechanism for improved sample generation, and an adaptive proposal assignment strategy for precise damage localization and classification. R-CNN and Faster R-CNN were compared in [64] with better results obtained for Faster R-CNN. Also, the paper introduced a novel dataset consisting of 1100 images from Bangladesh with three types of road damages: potholes, cracks and raveling.

3.3. Discussions

As analyzed previously in this section, the advancement of CNN architectures in the problem of RDD has been pivotal. This task refers to detecting and classifying road damages, including but not limited to cracks, potholes, or various surface anomalies. The precise identification of these damages is essential for the operational safety of autonomous vehicles, facilitating secure navigation and avoiding potential dangers. Table 2 provides a systematic overview of various methods applied to this task. It acts as a consolidated reference, detailing the performance metrics, datasets used and the architectural highlights of each proposed solution. The first column reveals the paper reference, the second column is the architecture that obtained the best metric value reported in the work, column 3 reveals the used metric, while column 4 underlines the metric value. The last column presents the dataset for which the architecture was trained and tested.

From the neural architectures point of view, the most used CNNs are the one-shot detectors, especially the YOLO family of models and their subsequent versions. These characteristics make them well-suited for real-time applications where rapid and reliable RDD is critical. An alternative to this is the SSD architecture, which also promises good speed and *Acc* but not better performance in detecting small objects compared to YOLO. Given that this aspect can be critical in detecting cracks or potholes, YOLO could be preferred due to its advantageous processing of the entire image at once, allowing it to recognize context and reduce false positives effectively. The Faster R-CNN family emerges as a notable alternative, particularly for its precision in detecting a wide range of damage sizes.

Table 2. Summarization of state-of-the-art road damage detection methods.

Work	Architecture	Metric	Value	Dataset
[8]	SSD-Mobile Net	P	0.99	RDD-2018
[13]	RetinaNet with VGG19 as backbone	mAP	0.83	RDD-2018
[15]	YOLOv3	F1	0.62	RDD-2018
[16]	Ensemble (16 models)	F1	0.66	RDD-2018
[19]	Yolov4 with PID optimizer	mAP	0.47	RDD-2018
[19]	Yolov3 with improvements	mAP@50	0.52	AerialCrackDataset [21]
[22]	Yolov4 with PID optimizer	mAP	0.63	RDD-2018
[23]	DenseNet	F1	0.75	EDmCrack1000
[26]	EfficientDet	F1	0.56	RDD-2020
[31]	YOLOv4 with ensemble approach	F1	0.64	RDD-2020
[32]	u-YOLO with ensemble approach	F1	0.67	RDD-2020
[34]	Ensemble model based on YOLOv3 and Faster R-CNN	F1	0.57	RDD-2020

Table 2. *Cont.*

Work	Architecture	Metric	Value	Dataset
[36]	YOLO-LRDD	F1	0.59	RDD2020 + new Chinese road damages
[38]	E-EfficientDet	F1	0.47	RDD-2020
[40]	YOLOv5	mAP	0.55	RDD-2020
[42]	YOLOv7	F1	0.66	RDD-2022
[43]	YOLOv5x with ensemble approach	F1	0.67	RDD-2022
[44]	YOLOv5x with ensemble approach	F1	0.73	RDD-2022
[45]	YOLOv5s-DSG	mAP@50	0.63	RDD-2022
[46]	BL-YOLOv8	F1	0.87	RDD-2022
[47]	YOLO	P	0.91	BrazilDNIT
[48]	YOLOv7	mAP@50	0.73	UAV Spanish roads
[49]	CrackYOLO	mAP	0.70	Rural roads dataset
[50]	YOLOv5s	F1	0.67	Beijing dataset
[53]	Faster R-CNN with ResNet-152 as backbone	F1	0.63	RDD-2018
[54]	Mask R-CNN with DenseNet as backbone	Visual results	-	RDD-2018
[55]	Faster R-CNN with X101-FPN base model	F1	0.51	RDD-2020
[57]	Faster R-CNN	F1	0.49	RDD-2020
[58]	Faster R-CNN with ResNet-50	F1	0.54	RDD-2020
[59]	Cascade R-CNN with CFM	F1	0.63	RDD-2020
[61]	Faster R-CNN	Visual results	-	Cracks dataset
[62]	CrdNet	mAP	0.91	China roads dataset
[64]	Faster R-CNN	Acc	0.98	Bangladesh roads dataset
[60]	Ensemble approach of YOLO and Faster R-CNN	F1	0.77	RDD-2022

Unlike YOLO and SSD, which excel in speed and are adept at handling smaller objects to an extent, Faster R-CNN is distinguished by its two-stage process, where it first generates region proposals and then classifies those regions into object categories. This methodical approach allows for more accurate detection, making Faster R-CNN highly effective in scenarios where the precision of damage identification outweighs the need for real-time processing. Other used architectures are DenseNet and EfficientDet, each bringing their contributions to the field. DenseNet's architecture facilitates feature reuse throughout the network, which can be advantageous for identifying the intricate features of road damage. EfficientDet, on the other hand, is designed for efficiency, scaling systematically with the dataset and task complexity.

Interestingly, the table also includes novel architectures like CrackYolo and CrdNet, tailored specifically for the task of RDD. These specialized networks potentially offer optimizations over more general architectures and reflect the ongoing innovation in the domain. A visual overview of the most used neural networks in RDD is shown in Figure 7.

Regarding the most used metrics from the RDD state-of-the-art analysis, see Tables 1 and 2; multiple comments can be made. For instance, one can observe that while SSD-Mobile Net boasts a high P, it may not be the most comprehensive metric to consider, as models with high $F1$ scores, like YOLOv7, suggest a better balance between P and R, which is vital in practical scenarios where both false positives and false negatives carry significant costs. Generalizing this idea, there is a clear predilection towards YOLO variants for achieving high $F1$ scores, indicative of a balanced P and R. This is contrasted by the deployment of Faster R-CNN models, preferred for their meticulous localization capabilities, essential for detailed damage assessments. The emergence of specialized models, such as CrackYOLO and CrdNet, tailored to specific damage types or conditions, highlights the trend towards optimizing architectures for enhanced performance on specialized datasets. Ensemble models that integrate the strengths of YOLO and Faster R-CNN architectures demonstrate superior $F1$ scores, suggesting that a collaborative approach may yield the best results in applications requiring both rapid detection and accuracy.

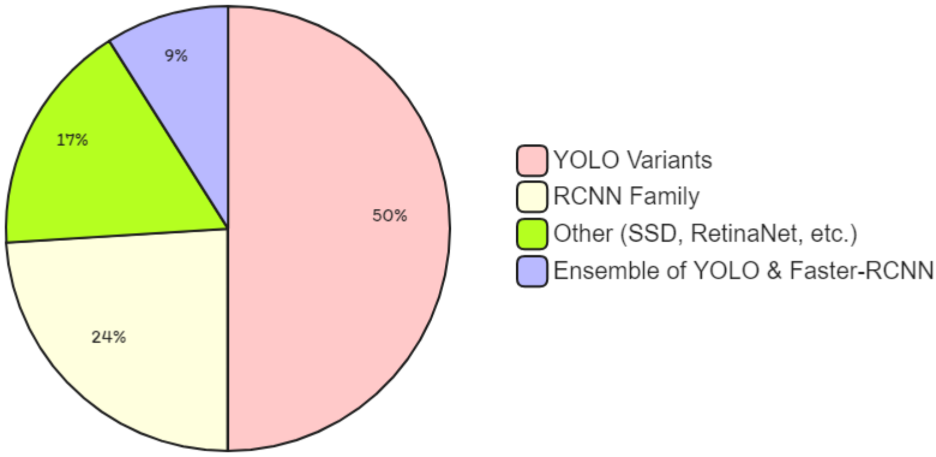

Figure 7. Distribution of the used Convolutional Neural Network (CNN) in Road Damage Detection (RDD).

The various studies on RDD also reveal insightful trends about the datasets used, highlighting their pivotal role in developing and benchmarking neural network models. Table 3 presents the datasets used in RDD, comparing them by size, number of classes, and the countries where the images were acquired. Notably, the RDD Challenge with the datasets RDD-2018, RDD-2020, and RDD-2022 emerge as popular choices, underscoring their comprehensive nature and relevance in capturing a wide array of road damage types, likely making them benchmark datasets in the field. The appearance of specialized datasets, such as the AerialCrackDataset and those focusing on specific geographic regions (e.g., Brazilian highways, rural roads, and UAV images of Spanish roads), points to the field's evolving need for diverse datasets. Figure 8 presents a pie chart reflecting the distribution of the dataset used in RDD.

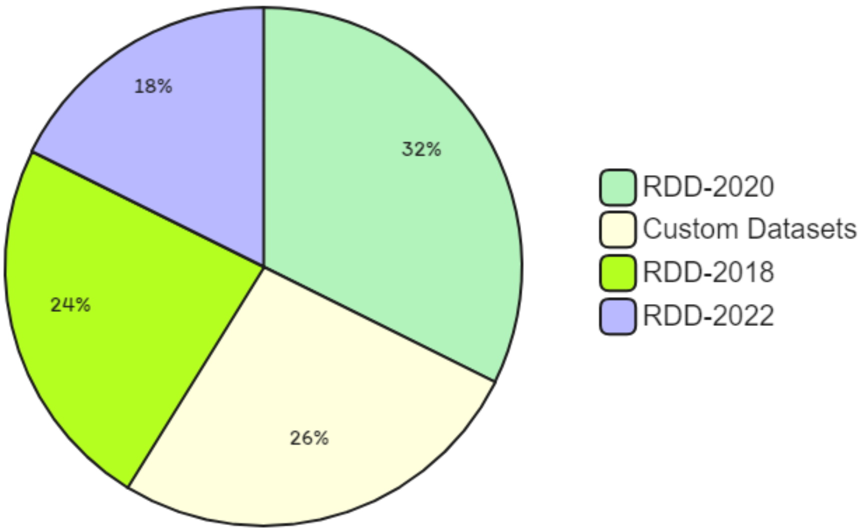

Figure 8. Distribution of the datasets used in Road Damage Detection (RDD).

Table 3. Global road damage challenge and custom datasets.

Dataset	Year	Number of Images	Classes	Countries
Global Road Damage Challenge Datasets				
RDD-2018	2018	9053	8 classes: longitudinal crack (wheel mark part), longitudinal crack (construction joint part), lateral crack (equal interval), lateral crack (construction joint part), alligator crack, pothole, white line blur, crosswalk blur	Japan
RDD-2020	2020	26,336	4 classes: longitudinal crack, lateral crack, alligator crack, potholes	Japan, India, Czech Republic
RDD-2022	2022	47,420	4 classes: longitudinal crack, lateral crack, alligator crack, potholes	Japan, USA, Norway, Czech Republic, India, China Motorbike, China Drone
Custom Datasets				
AerialCrackDataset	2017	3525	2 classes: horizontal crack, longitudinal crack	China
EDmCrack1000	2020	1000	1 class: cracks	Canada
Bangladesh roads dataset	2020	1100	3 classes: crack, pothole, raveling	Bangladesh
Cracks dataset	2021	400	1 class: cracks	not specified
China roads dataset	2021	7283	7 classes: longitudinal cracks, transverse cracks, repair, poor repair, pothole, crazing, block repair	China
BrazilDNIT	2023	665	2 classes: cracks and potholes	Brazil
UAV Spanish roads	2023	1362	2 classes: cracks and potholes	Spain
Beijing dataset	2023	2900	7 classes: longitudinal cracks, transverse cracks, alligator cracks, pothole, manhole cover, longitudinal patch, transverse patch	China
Rural roads dataset	2024	9801	3 classes: transverse cracks, longitudinal cracks, reticular cracks	China

Thus, multiple challenges appear that reflect the varied conditions encountered in real-world scenarios. These diverse datasets facilitate the tailoring of neural network architectures to specific detection tasks, enhancing their robustness and adaptability across different environmental conditions and damage types. For instance, models trained on datasets with aerial images or specific damage classes (such as the longitudinal, transverse, and reticular cracks in the rural road dataset) indicate targeted efforts to address the nuances of RDD more effectively. Moreover, the evolution of datasets from RDD-2018 to RDD-2022 and the inclusion of new, specialized datasets reflect the field's continuous pursuit of more accurate, efficient, and context-aware detection systems.

There is a significant underutilization in real-time testing facilitated through high-performance mobile computing devices. From the RDD perspective, in [40], the authors use the NVIDIA Jetson NX platform in order to illustrate the application of a mobile platform in deploying DL models for real-time detection. The gap in mobile platform usage in similar studies could be attributed to several factors. The field's access to mobile computing resources might be limited, as traditional cloud computing resources have been the norm for processing large datasets in DL applications. Also, the technical complexity and challenges associated with deploying and optimizing DL models on mobile platforms might discourage researchers. These challenges include managing the computational constraints of mobile devices, such as limited processing power and memory, which can affect the model's performance and accuracy.

4. Road Surface Classification

In this chapter, we examine the advancement of methodologies in the domain of RSC, another critical component in the maintenance of road analysis. This task, while sharing common principles with image classification and CV, requests a deeper approach. It demands not only the detection of distinctive features indicative of various road conditions and types but also an advanced level of interpretative analysis to categorize road surfaces into predefined classes accurately. This task is further complicated by the requirement to discern subtle distinctions and contextual nuances to different road surfaces. The SOTA will be analyzed in this chapter through two perspectives, as stated in Figure 5: the nature of surface material (an example of this type of classification being presented in

Figure 9–bottom and the nature of the environmental conditions (an example is shown in Figure 9–top).

Figure 9. Example of Road Surface Classification (RSC) detection from [65].

4.1. The Nature of Environmental Conditions

Environmental conditions have a significant impact on road surfaces, influencing their integrity and performance. Factors such as rain, snow, or ice could lead to either deterioration (cracking or potholing) or safe driving. Recognizing these environmental influences enables the development of more resilient road surfaces and the implementation of timely maintenance strategies, ensuring safer and more reliable roadways in order to develop autonomous driving. As we review the SOTA in this domain, we will highlight innovative approaches and methodologies that have been developed to accurately identify and classify road surface types, underscoring the significance of environmental conditions.

For example, in [66] the authors utilize a pre-trained VGG16 [17] for RSC on a dataset collected from two highway sections in South-Western Ontario, Canada. The dataset was analyzed under three granularity levels: two-class (Bare, With Snow/Ice), three-class (Bare, Partly Snow Covered, Fully Snow Covered), and five-class descriptions, to offer a nuanced understanding of winter road conditions. The five-class description provides the most detailed classification, distinguishing between various degrees of snow coverage. The model achieved the highest classification accuracy with the two-class description, suggesting that while detailed classifications provide valuable insights into road conditions, simpler classification schemes may benefit from higher accuracy and robustness in practical applications.

A different approach is presented in [67], where the authors implemented a comprehensive two-stage approach to classify RSC and road friction estimate (RFE) using front-view camera images from vehicles. In the first stage, they explored various CNN architectures, including SqueezeNet [68] and traditional CNN models with different depths, to identify region-specific features. They focused on features from drivable surface textures, the sky, and surrounding areas, achieving high classification *Acc*. In the second stage, a rule-based model segmented the ego-lane drivable surface into patches, classifying them to assess RFE. This innovative method allowed for the estimation of friction levels across different road conditions by evaluating the patchiness of the road surface. The dataset for this study was derived from publicly available sources, consisting of 3750 training images and 1550 test images, with four classes: dry, wet, slush, and snow/ice. Similarly, the authors from [69] curated a mixed dataset from public datasets (KITI [70], Oxford Robocar Dataset [71]), and images sourced from the web to train two CNN models: ResNet50 and InceptionV3 [72]. These models aimed to distinguish between six types of road surfaces: asphalt, dirt, grass, wet asphalt, cobblestone, and snow. The best result achieved was a 92% average classification accuracy, provided by ResNet50.

An innovation at the neural level is the work from [73], where the authors present a modified activation function based on the Rectified Linear Units (ReLU) [74]. This adjustment, named Gai-ReLU, aims to improve classification accuracy by addressing the issue of data leakage for negative inputs. Their CNN architecture, optimized for this task, demonstrated superior performance with a classification *Acc* of 94.89% on 10,000 images from the Oxford RobotCar dataset and KITTI, with 5 labels: dry, wet, snowy, muddy roads, and others (road surface conditions that do not specifically fall into the primary defined categories). In the same manner, in [75], the authors propose RCNet, a CNN-based model which classifies the road into five categories: curvy, dry, ice, rough, and wet. The RCNet architecture, although based on CNNs, is optimized for extracting significant features from road images under varying conditions, such as illumination and weather changes. The dataset used for training and validation is the Oxford RobotCar dataset, supplemented with YouTube video frames for the curvy road category, totaling over 20,000 images.

Another innovation and the neural level is presented in [65], where the authors use a dual-stream CNN model that integrates both visible light and thermal images. The dual-stream CNN architecture consists of two independent input streams, each with its convolutional block fed by one type of image (either visible light or thermal). The CNN model itself is adapted from VGG16 but restructured into a simpler seven-layer configuration to increase processing speed while maintaining performance. An $F1$ score of 0.94 was computed by training the model on 4244 pairs of images acquired from various roads in Montreal, Canada, during winter and divided into four distinct classes: snowy, icy, wet, and slushy.

An innovative approach is presented in [76], where the authors introduce the usage of Vision Transformers (ViTs) models [77]. Unlike traditional CNNs, ViTs apply attention mechanisms to dynamically assign weights to pixels, focusing the model on relevant image features. Another innovation of the paper consists of creating a dataset of 2498 images from online videos from the United Kingdom from which seven classes are derived from weather conditions: sunny, cloudy, foggy, rainy, wet, clear, and snowy.

A comparison between multiple CNNs is made in [78], where the authors evaluated seven SOTA models such as InceptionV3, Inception-Resnet-v2 [79], Xception [80], DenseNet169, MobileNetV2 [81], and NASNet [82], all of which were pre-trained on the ImageNet dataset. Additionally, they proposed a baseline model inspired by AlexNet [83] and VGG [17], which was significantly less complex in terms of layers and parameters. The core contribution was their experiment in simplifying the baseline model through an ablation study, successfully reducing its complexity without substantially affecting performance. The models were trained on a dataset consisting of 14,000 labeled images collected from 40 stations connected to the Ontario Road Weather Information System (RWIS) during the 2017–2018 Winter season.

Another work based on multiple comparisons is presented in [84], where the authors compared AlexNet, GoogleNet [11], and ResNet18 adapted via transfer learning to identify road surfaces based on three weather conditions (clear, light snow, and heavy snow) and three surface conditions (dry, snowy, wet/slushy). The dataset includes 30,000 images (15,000 images from road surface conditions) obtained from roadside webcams from Wyoming, USA, with the best results for the ResNet18 model, reaching 97% *Acc* for weather detection and 99% for RSC. In [85] the authors compare both classical ML techniques (Support Vector Machine—SVM, K-Nearest Neighbor—KNN, K Means Clustering—KMC) and advanced DL approaches (ResNet50 and MobileNet) to classify three road conditions: dry, wet, and icy. The best results were obtained using DL techniques by training the models on 12,545 images from various roads in China and Australia.

4.2. The Nature of Surface Material

The composition of a road's surface material plays an important role in its characteristics, durability, and suitability for various environmental conditions and traffic demands. Understanding these distinctions is crucial for a multitude of applications, ranging from infrastructure maintenance and urban planning to the calibration of autonomous vehicle navigation systems. As we review the SOTA in this domain, we will highlight innovative approaches and methodologies that have been developed to accurately identify and classify road surface types, underscoring the significance of material composition in the context of RSC.

In [86], the authors developed a deep ensemble network utilizing various sensor data obtained from the vehicle's Controller Area Network, such as data from IMU sensors, steering angle sensors, vehicle speed sensors, and acceleration sensors. The dataset comprised four distinct road types: flat road, sinusoidal road, manhole and pothole, and bump, which is training information for a deep ensemble network that incorporates Long-Short Term Memory (LSTM) units. In the same direction, but working with accelerometer data and GPS information as input information, in [87], the authors utilize CNNs, LSTM network, and reservoir computing models to analyze the data and predict road conditions like asphalt, concrete, cobblestone, and dirt roads. The best result is obtained with the custom CNN, which was designed to process the sequences of oscillations produced by the accelerometers of mobile devices, adapting the inputs (temporal multidimensional 1D samples) into a form suitable for a 2D convolutional process.

Similarly, the work from [88] lies in the application of CNN to time–frequency representations of data from accelerometers and gyroscopes. This method, termed CNN-SP, is tested against traditional CNN applications on time–domain data (CNN-1D) and shallow ML algorithms on a dataset collected from 12 different vehicles over more than 40 km of road, achieving an impressive 97.2% identification *Acc*. In a similar manner, in [89], the author uses data collected from accelerometers and gyroscopes from nine different vehicles, resulting in nine datasets. They tested 34 computational models, employing techniques like KMC, SVM, KNN, LSTM, CNN, and hybrid CNN–LSTM models. The best-performing model in the study was a CNN-based deep neural network. This model achieved a validation *Acc* of 93.17% in classifying surfaces between segments of dirt, cobblestone, or asphalt roads.

An innovative approach regarding the dataset presented in [90], where the authors use as neural input of a 2D representation of the tire–pavement interaction noise signals. These signals are acoustic noises generated by the interaction between the tire and road surface, which are captured using a microphone, and transformed into 2D images through a continuous wavelet transform. By doing so, 11,040 images were generated, considering two classes: asphalt and snowy roads. On this dataset, a simple CNN architecture was based on [91] for which an *Acc* of 91.84% was obtained on the data from the front wheel and 91.18% on the data from the rear wheel.

A different approach is presented in [92] where the authors introduce a dataset called Road Traversing Knowledge (RTK) with 77,547 road images from Brazil, classified into three classes: asphalt, paved, and unpaved. On this dataset, they trained a CNN consisting of a simple structure for which the training stage was also performed on 5093 images from the KITTI dataset and 1401 images from CaRINA [93]. An *Acc* of 94.57% was obtained on the RTK dataset with augmentation, trying to leave each surface type with a similar amount of samples. Using the same dataset, which also includes the semantic segmentation data for each label, the authors from [94] utilize the U-Net [95] architecture paired with different ResNet architectures. Using eleven semantic segmentation classes—asphalt, paved, unpaved, markings, speed-bump, cats-eye, storm-drain, patch, water-puddle, pothole, and cracks, they trained and tested multiple neural architectures, the best result being obtained with ResNet50 with weight added in the loss function.

In [96], the authors trained and tested AlexNet on 500 original images from Poland roads from which nine types of road surfaces were labeled: asphalt, concrete, trylinka (road surface made of large hexagonal concrete paver blocks), concrete paver blocks, granite paver blocks, openwork surface, gravel, sand, and grass. The study conducted two sets of experiments: one for recognizing specific types of road surfaces and another for determining the general condition of the surfaces. For general condition classification, the images were divided into three types: good ones, i.e., asphalt and concrete; medium ones, i.e., trylinka, concrete cubes, granite cubes, and openwork surfaces; and bad ones, i.e., sand, gravel, and grass. The results showed high accuracy rates, over 85% in identifying specific road types and over 91% in determining their general condition. In comparison, in [97], the authors innovatively employ a deep CNN using a ResNet50 architecture, coupled with a "one-cycle" learning rate (LR) optimization technique, to classify five pavement types: asphalt, concrete, interlocking, cobblestone, and unpaved. The dataset used consists of 49,173 images sourced from highways and local roads in Ceará State, Brazil, for which, in the testing phase, the authors claimed near-perfect accuracy, P, and R values close to 100%. The model's success lies in its use of transfer learning with pre-trained ResNet50 weights, fine-tuned for pavement surface type classification.

An innovative approach is presented in [98], where the authors integrate visual and tactile data in an early fusion method (VTF). This integration is achieved through a CNN-Transformer network, for which the visual data are obtained from road images, while tactile data come from a specially designed intelligent tire system equipped with piezoelectric sensors. The CNN-Transformer architecture effectively fuses these two modalities. The network comprises modality-specific SE-CNNs (Squeeze-and-Excitation Convolutional Neural Networks) for initial feature extraction from both visual and tactile inputs. This method reports an accuracy of 99.48% on a dataset consisting of 40,000 images, which includes four classes of road surfaces: marble road, painted road, gravel road, and asphalt road.

Combining classes that define the surface material with those that characterize the ambient circumstances resulted in the proposal of various techniques. For example, in [99], the authors used a dataset consisting of 12,440 images captured on Lithuanian roads, which were categorized into six classes, representing combinations of three road surface types (asphalt, cobblestone, and gravel) under two different conditions (wet and dry), namely: gravel wet, gravel dry, cobblestone wet, cobblestone dry, asphalt wet, and asphalt dry. Using AlexNet as a baseline network, they obtained an Acc of 88.3% on the dataset. In the same manner of creating a comprehensive dataset (Road Surface Condition Dataset—RSCD), in [100], the authors conducted a real-vehicle image acquisition on Beijing roads, resulting in approximately 1 million image samples. Using three detailed road level annotations: material (asphalt, concrete, dirt/mud, gravel), friction (dry, wet, water, fresh snow, melted snow, and ice) and unevenness (smooth, slight, or severe), they obtained 27 combined classes for which an EfficientNet-B0-based CNN was trained, with a decision-level fusion method based on the Dempster–Shafer evidence theory [101]. The paper reports an Acc of 92.05% for the CNN model and 97.50% after the fusion process.

Using the same dataset, the authors from [102] introduce Attention-ReXNet, an advanced CNN that builds upon the ReXNet [103] architecture. This model enhances spatial perception using a depth-wise separable convolution structure and differentiates feature extraction through stage-wise network construction. Another aspect of their approach is the incorporation of a feature fuse module, which fuses high-dimensional and low-dimensional features to improve global attention. Using Balance Softmax Cross-Entropy, they tested the model performance on the last five less representative classes from the RSCD dataset by the number of images (Water Asphalt Slight, Water Concrete Slight, Wet Concrete Severe, Water Asphalt Severe, and Wet Asphalt Severe—see Figure 2 from [102]) and obtained an Acc of 90.7%. In comparison, in all 27 classes, they obtained an Acc of 88.52%.

4.3. Discussions

As explored in the preceding discussion, the evolution of CNN architectures has played a crucial role in the realm of RSC. This task includes the ability of autonomous systems to differentiate between various road types as well as recognize external conditions that may compromise driver safety. Accurate classification of road surfaces is vital for enhancing the operational safety and efficiency of autonomous vehicles, enabling them to make informed navigation decisions under diverse environmental conditions. Table 4 presents a comprehensive summary of SOTA methodologies applied to this challenge. It serves as an extensive resource, elaborating on the performance metrics, datasets utilized, and the specific features of each proposed CNN solution. This comprehensive review emphasizes the importance of advanced neural network designs in improving the classification of road surfaces and the ongoing efforts to improve these models' performance and flexibility in a range of road settings. In Table 4, the first column provides the paper reference, while the second column identifies the architecture that achieved the highest metric value reported in the study. Column three specifies the metric used, and column four highlights the metric value. The final column is divided into three subcolumns, each presenting different details about the dataset: the number of images, the number of classes, their descriptions, and the countries where the images were collected.

Table 4. Summarization of state-of-the-art road surface classifications.

Work	Architecture	Metric	Value	Dataset		
				Images	**Classes**	**Countries**
[66]	VGG16	Acc	0.91	5 k	5: bare, partly snow covered (<25%, 25–50%, 50–75%, fully snow covered)	Canada
[67]	SqueezeNet	Acc	0.97	5.3 k	4: dry, wet, slush, snow/ice	Public domain
[69]	ResNet50	Acc	0.92	18.5 k	6: asphalt, dirt, grass, wet asphalt, cobblestone, snow	UK, Germany
[73]	Own CNN with Gai-ReLU	Acc	0.95	10 k	5: dry, wet, snowy, muddy roads, others	UK, Germany
[73]	RCNet	Acc	0.99	20 k	5: dry, curvy, ice, rough, wet	UK, Germany
[65]	Dual-stream CNN	F1	0.94	4.3 k	4: snowy, icy, wet, slushy	Canada
[76]	Vision Transformers	Acc	0.92	2.5 k	7: sunny, cloudy, foggy, rainy, wet, clear, snowy	UK
[78]	Baseline (AlexNet + VGG)	Acc	0.91	14 k	3: bare pavement, partial snow cover, full snow cover	Canada
[84]	ResNet18	Acc	0.97	15 k	3: clear, light snow, heavy snow	USA
[84]	ResNet18	Acc	0.99	15 k	3: dry, snowy, wet/slushy	USA
[85]	ResNet50	Acc	0.98	12.5 k	3: dry, wet, icy	China, Australia
[86]	Deep ensemble with LSTM	Acc	0.95	510 k	4: flat road, sinusoidal road, manhole and pothole, bump	South Korea
[87]	Custom CNN	Acc	0.85	9.2 k	4: asphalt, concrete, cobblestone, dirt roads	Argentina
[88]	CNN-SP	Acc	0.97	9.8 k	20: normal road segments, road anomalies, obstacles	Italy
[89]	Custom CNN	Acc	0.93	9 datasets	3: dirt, cobblestone, asphalt	Brazil
[90]	Custom CNN	Acc	0.92	11 k	2: asphalt, snow	South Korea
[92]	Custom CNN	Acc	0.95	62 k	3: asphalt, paved, unpaved	Brazil
[92]	Custom CNN	Acc	0.94	5 k	3: asphalt, paved, unpaved	Germany
[92]	Custom CNN	Acc	0.99	1.4 k	3: asphalt, paved, unpaved	Brazil
[94]	U-NET with ResNet50	Acc	0.97	70 k	11: asphalt, paved, unpaved, markings, speed-bump, cats-eye, storm-drain, patch, water-puddle, pothole, cracks	Brazil, Germany, Japan, UK, Italy
[96]	AlexNet	Acc	0.85	0.5 k	9: asphalt, concrete, trylinka, concrete paver blocks, granite paver blocks, openwork surface, gravel, sand, grass	Poland
[97]	ResNet50 with LR optimization	Acc	0.99	49 k	5: asphalt, concrete, interlocking, cobblestone, unpaved	Brazil
[98]	CNN-Transformer with VTF	Acc	0.99	4 k	4: marble, painted, gravel, asphalt roads	China
[99]	AlexNet	Acc	0.88	12.5 k	6: gravel wet, gravel dry, cobblestone wet, cobblestone dry, asphalt wet, asphalt dry	Lithuania
[100]	Custom EfficientNet-B0	Acc	0.98	1 M	27 combined classes	China
[102]	Attention-RexNet	Acc	0.89	RSCD	27 classes	China
[102]	Attention-RexNet	Acc	0.91	RSCD	5 classes (wet-asphalt-smooth, water-asphalt-smooth, dry-concrete-smooth, dry-asphalt-smooth, fresh-snow)	China

In the domain of RSC, the diversity of CNN architectures is evident, with a noticeable preference for models such as ResNet, VGG16, and custom CNN designs, alongside novel approaches like Vision Transformers and attention mechanisms. ResNet, particularly ResNet50 and ResNet18, stands out for its deep residual learning framework, enabling the training of networks that are much deeper than previously possible, thereby improving accuracy. Its use across different datasets underscores its robustness and adaptability to varying road surface conditions. VGG16, known for its simplicity and depth, offers an advantage in extracting rich feature representations, making it suitable for classifying complex road surface images. Custom CNNs tailored to specific datasets indicate a trend toward optimizing network structures to achieve higher performance, demonstrating the field's move towards specialization. Vision Transformers, a more recent development, introduces an architecture that departs from traditional convolutional approaches, focusing instead on self-attention mechanisms, which can capture global dependencies within an image, offering promising results in classifying road surfaces under diverse conditions.

The variation in RSC datasets reflects the need to capture a wide range of surface types and conditions, including seasonal changes, weather effects, and material differences, to develop models that are truly adaptable and robust across global road infrastructures. This approach ensures that the developed models can generalize well across different geographic locations, weather conditions, and road types, addressing the critical requirement for comprehensive RSC in support of autonomous driving and road maintenance.

The most utilized metric in our analyzed paper from RSC is *Acc*, a contrast to the multiple metrics seen in Table 1. The metric provides a straightforward and intuitive measure of a model's performance across these datasets' varied and often unbalanced classes. It reflects the proportion of total correct predictions made out of all predictions, offering a clear indicator of a model's effectiveness in classifying road surfaces. However, this singular focus on accuracy might not fully capture the nuances of model performance, especially in datasets with class imbalance or more subtle distinctions between classes, highlighting a potential area for further methodological development. Table 5 presents the distribution of accuracy values across all analyzed studies in RSC.

Table 5. Table of accuracies and corresponding number of papers.

Accuracy	0.85	0.88	0.89	0.91	0.92	0.93	0.94	0.95	0.97	0.98	0.99
No. of Papers	2	1	1	3	4	1	2	3	3	2	4

A comparison of how the accuracy varies if we compare the size of the dataset and the number of classes is presented as a bubble chart in Figure 10. Each analyzed paper in RSC is represented by a bubble that is larger in size, the greater the number of classes. Five papers reported an *Acc* = 0.99 ([73,84,94,97,98]), each with a number of classes ranging from 3 to 5. The biggest difference comes from the size of the dataset. The high accuracy could be influenced by the simplicity of the classification task, compared to other studies where the number of classes is higher. In comparison, Ref. [100] presents the largest classes in their bubble chart, with 27 categories and the largest dataset (approximately 1 million patches of surface condition). However, they report facing the long-tail dataset problem in their work.

Similarly, in the RDD case, there is a lack of mobile platform usage for real-time testing. Notable mentions are the works from [100], where the authors use a NVIDIA Xavier NX embedded platform to demonstrate the capability to operate in real time or [99] where Nvidia Jetson TX2 is used to leverage the computational power of the mobile platform for surface classification.

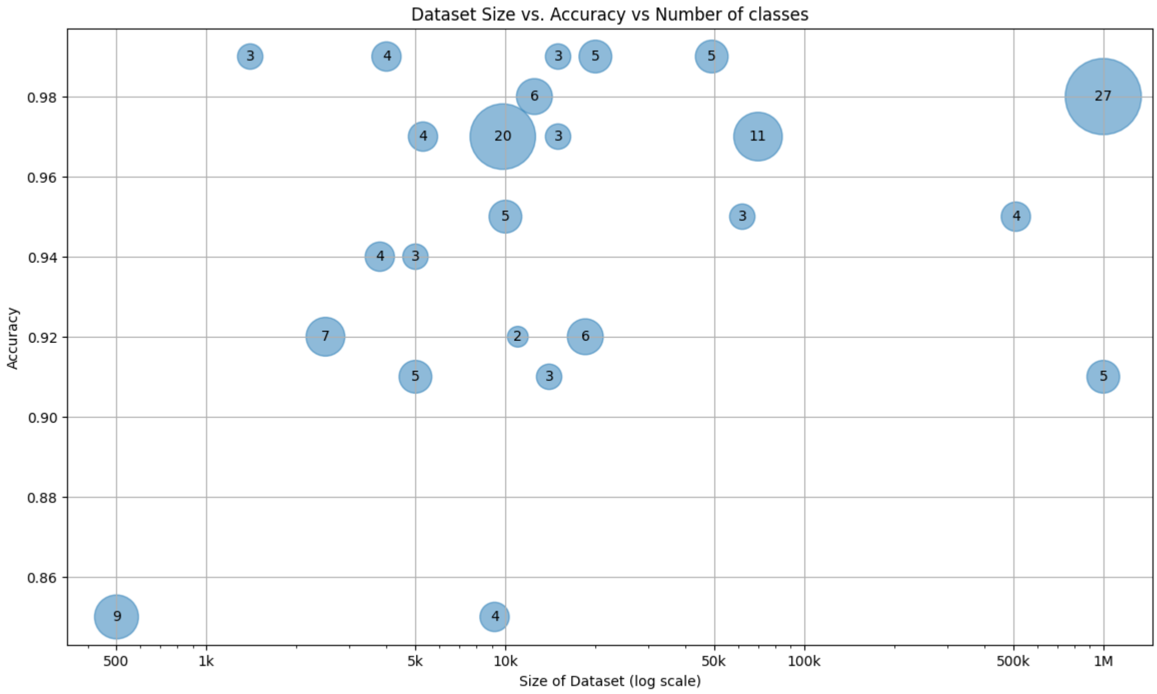

Figure 10. Accuracy vs. dataset vs. classes.

5. Conclusions

This review paper has offered a comprehensive examination of the state-of-the-art developments in autonomous vehicle technologies, with a special focus on RDD and RSC. These critical components significantly enhance autonomous driving systems' safety and operational efficiency. In the domain of RDD, our exploration into various DL methodologies, including both one-stage and two-stage detectors, has revealed their capabilities and limitations in identifying road damages such as cracks, potholes, and surface irregularities. While one-stage detectors, exemplified by YOLO and SSD, excel in providing real-time processing and rapid inspection capabilities, they might not always deliver the most precise damage localization. On the other hand, two-stage detectors like Faster R-CNN are distinguished by their exceptional accuracy in object localization, although at the cost of higher computational demands. Our study focused on detecting damages such as potholes and cracks, but there is significant potential for future research to expand this scope. Future investigations could analyze additional types of road irregularities, such as humps, ruts, or speed bumps. These features, often introduced intentionally for traffic control or formed due to extended usage and weathering, present unique challenges for detection and classification. Moreover, detecting obstacles that may appear suddenly on the road while the vehicle is in motion, such as debris or fallen branches, could further enhance the robustness of road damage detection.

Similarly, the investigation into RSC has delved into how different DL models, particularly various CNN architectures, are used to classify road surfaces based on material and environmental conditions. This section highlighted neural innovations like modified activation functions and dual-stream CNN models, which have shown promising results in achieving higher accuracy under diverse road conditions affected by environmental elements such as weather. However, while classification provides a general categorization of road surfaces, a more nuanced approach could be considered for future research. One potential direction is to explore road surface analysis as a semantic segmentation problem. Unlike classification, semantic segmentation involves labeling each pixel of an image with a class, which could allow for a more detailed and precise understanding of the road surface.

RSC emphasizes addressing a wide spectrum of road conditions, as evidenced by the diversity of datasets utilized in the area. These range from small-scale collections like Canadian Winter Roads to large-scale compilations like the RSC dataset with one million image samples. This diversity contrasts RDD, dominated by fewer, more specialized datasets.

The insights derived from this paper underscore the pivotal role of DL in pushing the boundaries of autonomous vehicle technologies, presenting viable solutions for navigating the complexities of road environments. There is a compelling case for developing integrated methods that combine RDD and RSC as we look towards the future. Such hybrid approaches promise to streamline the detection and classification processes and offer a more holistic understanding of road conditions, enhancing decision-making processes for autonomous systems. An idea for this is to incorporate weather forecast data into the RDD process. By fusing real-time meteorological information with road condition data, future models could more accurately predict road damages and surface conditions. Future research should also explore the utilization of mobile platforms. This direction is motivated by the noticeable gap in current literature regarding deploying these technologies on mobile devices, which could significantly expand their applicability and real-time functionality. The continuous evolution of datasets, alongside advancements in neural network architectures, will undoubtedly play a crucial role in realizing these future directions, demanding sustained innovation and exploration in this expanding field.

Author Contributions: Conceptualization, A.-P.B., A.B. and C.O.; methodology, A.-P.B., A.B. and C.O.; validation, A.-P.B., A.B. and C.O.; formal analysis, A.-P.B., A.B. and C.O.; investigation, A.-P.B.; resources, A.-P.B., A.B. and C.O.; writing—original draft preparation, A.-P.B.; writing—review and editing, A.-P.B., A.B. and C.O.; visualization, A.-P.B.; supervision, A.B. and C.O.; project administration, A.B. and C.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

Abbreviations

The following abbreviations are used in this manuscript:

Acc	Accuracy
ADAS	Advanced Driver Assistance Systems
BiFPN	Bi-directional Feature Pyramid Network
CNN	Convolutional Neural Networks
CV	Computer Vision
DenseNet	Densely Connected Convolutional Network
DL	Deep Learning
F1	F-measure
FN	False Negative
FP	False Positive
FPN	Feature Pyramid Networks
IoU	Intersection over Union
KMC	K Means Clustering
KNN	K-Nearest Neighbor
LSTM	Long-Short Term Memory
LR	Learning rate
ML	Machine Learning
mAP	mean average Precision
OD	Object Detection
P	Precision
PID	Proportional-Integral-Differential
R	Recall

RDD	Road Damage Detection
ReLU	Rectified Linear Units
ResNet	Residual Networks
RF	Road Friction Estimate
RSC	Road Surface Classification
RSCD	Road Surface Condition Dataset
RTK	Road Traversing Knowledge
RWIS	Road Weather Information System
SE-CNNs	Squeeze-and-Excitation Convolutional Neural Networks
SOTA	State of the Art
SSD	Single Shot Multibox Detector
SVM	Support Vector Machine
TP	True Positive
ViTs	Vision Transformers
VTF	Visual and Tactile data in an early fusion method
WoS	Web of Science
YOLO	You Only Look Once

References

1. Shaout, A.; Colella, D.; Awad, S. Advanced Driver Assistance Systems—Past, present and future. In Proceedings of the 2011 Seventh International Computer Engineering Conference (ICENCO'2011), Cairo, Egypt, 27–28 December 2011; pp. 72–82. [CrossRef]
2. Ziebinski, A.; Cupek, R.; Erdogan, H.; Waechter, S. A survey of ADAS technologies for the future perspective of sensor fusion. In Proceedings of the International Conference on Computational Collective Intelligence, Halkidiki, Greece, 28–30 September 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 135–146.
3. Weinland, D.; Ronfard, R.; Boyer, E. A survey of vision-based methods for action representation, segmentation and recognition. *Comput. Vis. Image Underst.* **2011**, *115*, 224–241. [CrossRef]
4. Sirbu, C.L.; Tomoiu, C.; Fancsali-Boldizsar, S.; Orhei, C. Real-time line matching based speed bump detection algorithm. In Proceedings of the 2021 IEEE 27th International Symposium for Design and Technology in Electronic Packaging (SIITME), Timisoara, Romania, 27–30 October 2021; pp. 246–249. [CrossRef]
5. Bălănescu, A.G.; Sîrbu, C.L.; Orhei, C. Intersection detection based on mono-camera sensor. In Proceedings of the 2022 45th International Conference on Telecommunications and Signal Processing (TSP), Virtual, 13–15 July 2022; pp. 216–220. [CrossRef]
6. Stănescu, M.; Sîrbu, C.L.; Orhei, C. Mapping the environment at range: Implications for camera calibration. In Proceedings of the 2022 IEEE 28th International Symposium for Design and Technology in Electronic Packaging (SIITME), Bucharest, Romania, 26–29 October 2022; pp. 40–43. [CrossRef]
7. Sun, S.; Petropulu, A.P.; Poor, H.V. MIMO radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges. *IEEE Signal Process. Mag.* **2020**, *37*, 98–117. [CrossRef]
8. Maeda, H.; Sekimoto, Y.; Seto, T.; Kashiyama, 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]
9. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, VA, USA, 27–30 June 2016; pp. 779–788.
10. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Part I 14; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
11. 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 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
12. 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.
13. Ale, L.; Zhang, N.; Li, L. Road damage detection using RetinaNet. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 5197–5200.
14. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
15. Alfarrarjeh, A.; Trivedi, D.; Kim, S.H.; Shahabi, C. A deep learning approach for road damage detection from smartphone images. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 5201–5204.
16. Wang, Y.J.; Ding, M.; Kan, S.; Zhang, S.; Lu, C. Deep proposal and detection networks for road damage detection and classification. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 5224–5227.

17. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
18. 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, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
19. Guo, L.; Li, R.; Jiang, B. A road surface damage detection method using yolov4 with pid optimizer. *Int. J. Innov. Comput. Inform. Control.* **2021**, *17*, 1763–1774.
20. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
21. Wang, B. *AerialCrackDataset: Towards Object Detection with Dataset*; Key Laboratory of Optoelectronic Imaging Technology and System, Ministry of Education, School of Optoelectronics, Beijing Institute of Technology: Beijing, China, 2017.
22. Wang, Q.; Mao, J.; Zhai, X.; Gui, J.; Shen, W.; Liu, Y. Improvements of YoloV3 for road damage detection. *J. Phys. Conf. Ser.* **2021**, *1903*, 012008. [CrossRef]
23. 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]
24. 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, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
25. 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]
26. Naddaf-Sh, S.; Naddaf-Sh, M.M.; Kashani, A.R.; Zargarzadeh, H. An efficient and scalable deep learning approach for road damage detection. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Virtual, 10–13 December 2020; pp. 5602–5608.
27. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 10781–10790.
28. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning. PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
29. Arya, D.; Maeda, H.; Ghosh, S.K.; Toshniwal, D.; Mraz, A.; Kashiyama, T.; Sekimoto, Y. Transfer learning-based road damage detection for multiple countries. *arXiv* **2020**, arXiv:2008.13101.
30. Arya, D.; Maeda, H.; Ghosh, S.K.; Toshniwal, D.; Mraz, A.; Kashiyama, T.; Sekimoto, Y. Deep learning-based road damage detection and classification for multiple countries. *Autom. Constr.* **2021**, *132*, 103935. [CrossRef]
31. Doshi, K.; Yilmaz, Y. Road damage detection using deep ensemble learning. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Virtual, 10–13 December 2020; pp. 5540–5544.
32. Hegde, V.; Trivedi, D.; Alfarrarjeh, A.; Deepak, A.; Kim, S.H.; Shahabi, C. Yet another deep learning approach for road damage detection using ensemble learning. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Virtual, 10–13 December 2020; pp. 5553–5558.
33. Ultralytics. YOLOv5. 2020. Available online: https://docs.ultralytics.com/yolov5/ (accessed on 26 May 2024).
34. Liu, Y.; Zhang, X.; Zhang, B.; Chen, Z. Deep network for road damage detection. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Virtual, 10–13 December 2020; pp. 5572–5576.
35. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
36. Wan, F.; Sun, C.; He, H.; Lei, G.; Xu, L.; Xiao, T. YOLO-LRDD: A lightweight method for road damage detection based on improved YOLOv5s. *EURASIP J. Adv. Signal Process.* **2022**, *2022*, 98. [CrossRef]
37. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.
38. Luo, H.; Li, C.; Wu, M.; Cai, L. An Enhanced Lightweight Network for Road Damage Detection Based on Deep Learning. *Electronics* **2023**, *12*, 2583. [CrossRef]
39. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, USA, 18–22 June 2023; pp. 7464–7475.
40. Li, M.; Wang, H.; Peng, Y.; Pei, X.; Wang, T.; Hou, T. Edge-to-client real-time road damage detection system based on Yolov5. In Proceedings of the 2021 China Automation Congress (CAC), Beijing, China, 22–24 October 2021; pp. 1221–1226.
41. Arya, D.; Maeda, H.; Ghosh, S.K.; Toshniwal, D.; Sekimoto, Y. Rdd2022: A multi-national image dataset for automatic road damage detection. *arXiv* **2022**, arXiv:2209.08538.
42. Pham, V.; Nguyen, D.; Donan, C. Road Damage Detection and Classification with YOLOv7. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17–20 December 2022; pp. 6416–6423.
43. Jeong, D.; Kim, J. Road Damage Detection using YOLO with Image Tiling about Multi-source Images. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17–20 December 2022; pp. 6401–6406.
44. Bhavsar, M.; Alfarrarjeh, A.; Baranwal, U.; Kim, S.H. Country-specific Ensemble Learning: A Deep Learning Approach for Road Damage Detection. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17–20 December 2022; pp. 6387–6394.

45. Xiang, W.; Wang, H.; Xu, Y.; Zhao, Y.; Zhang, L.; Duan, Y. Road disease detection algorithm based on YOLOv5s-DSG. *J. Real-Time Image Process.* **2023**, *20*, 56. [CrossRef]
46. Wang, X.; Gao, H.; Jia, Z.; Li, Z. BL-YOLOv8: An Improved Road Defect Detection Model Based on YOLOv8. *Sensors* **2023**, *23*, 8361. [CrossRef] [PubMed]
47. Benallal, M.A.; Tayeb, M.S. An image-based convolutional neural network system for road defects detection. *IAES Int. J. Artif. Intell.* **2023**, *12*, 577. [CrossRef]
48. Silva, L.A.; Leithardt, V.R.Q.; Batista, V.F.L.; González, G.V.; Santana, J.F.D.P. Automated Road Damage Detection using UAV Images and Deep Learning Techniques. *IEEE Access* **2023**, *11*, 62918–62931. [CrossRef]
49. Li, Y.; Sun, S.; Song, W.; Zhang, J.; Teng, Q. CrackYOLO: Rural Pavement Distress Detection Model with Complex Scenarios. *Electronics* **2024**, *13*, 312. [CrossRef]
50. Ren, M.; Zhang, X.; Chen, X.; Zhou, B.; Feng, Z. YOLOv5s-M: A deep learning network model for road pavement damage detection from urban street-view imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *120*, 103335. [CrossRef]
51. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
52. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1137–1149. [CrossRef] [PubMed]
53. Wang, W.; Wu, B.; Yang, S.; Wang, Z. Road damage detection and classification with faster R-CNN. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 5220–5223.
54. Chen, Q.; Gan, X.; Huang, W.; Feng, J.; Shim, H. Road damage detection and classification using mask R-CNN with DenseNet backbone. *Comput. Mater. Contin.* **2020**, *65*, 2201–2215. [CrossRef]
55. Pham, V.; Pham, C.; Dang, T. Road damage detection and classification with detectron2 and faster r-cnn. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Virtual, 10–13 December 2020; pp. 5592–5601.
56. Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.Y.; Girshick, R. Detectron2. 2019. Available online: https://github.com/facebookresearch/detectron2 (accessed on 26 May 2024).
57. Kortmann, F.; Talits, K.; Fassmeyer, P.; Warnecke, A.; Meier, N.; Heger, J.; Drews, P.; Funk, B. Detecting various road damage types in global countries utilizing faster r-cnn. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Virtual, 10–13 December 2020; pp. 5563–5571.
58. Vishwakarma, R.; Vennelakanti, R. CNN model & tuning for global road damage detection. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Virtual, 10–13 December 2020; pp. 5609–5615.
59. Pei, Z.; Lin, R.; Zhang, X.; Shen, H.; Tang, J.; Yang, Y. CFM: A consistency filtering mechanism for road damage detection. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Virtual, 10–13 December 2020; pp. 5584–5591.
60. Ding, W.; Zhao, X.; Zhu, B.; Du, Y.; Zhu, G.; Yu, T.; Li, L.; Wang, J. An Ensemble of One-Stage and Two-Stage Detectors Approach for Road Damage Detection. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17–20 December 2022; pp. 6395–6400.
61. Hacıefendioğlu, K.; Başağa, H.B. Concrete road crack detection using deep learning-based faster R-CNN method. *Iran. J. Sci. Technol. Trans. Civ. Eng.* **2022**, *45*, 1621–1633. [CrossRef]
62. Xu, H.; Chen, B.; Qin, J. A CNN-based length-aware cascade road damage detection approach. *Sensors* **2021**, *21*, 689. [CrossRef] [PubMed]
63. Sun, Z.; Han, Y.; Hua, Z.; Ruan, N.; Jia, W. Improving the efficiency and robustness of deepfakes detection through precise geometric features. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 3609–3618.
64. Arman, M.S.; Hasan, M.M.; Sadia, F.; Shakir, A.K.; Sarker, K.; Himu, F.A. Detection and classification of road damage using R-CNN and faster R-CNN: A deep learning approach. In Proceedings of the Cyber Security and Computer Science: Second EAI International Conference, ICONCS 2020, Dhaka, Bangladesh, 15–16 February 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 730–741.
65. Zhang, C.; Nateghinia, E.; Miranda-Moreno, L.F.; Sun, L. Winter road surface condition classification using convolutional neural network (CNN): Visible light and thermal image fusion. *Can. J. Civ. Eng.* **2022**, *49*, 569–578. [CrossRef]
66. Pan, G.; Fu, L.; Yu, R.; Muresan, M. Winter Road Surface Condition Recognition Using a Pretrained Deep Convolutional Network. 2018. Available online: http://xxx.lanl.gov/abs/1812.06858 (accessed on 26 May 2024).
67. Roychowdhury, S.; Zhao, M.; Wallin, A.; Ohlsson, N.; Jonasson, M. Machine learning models for road surface and friction estimation using front-camera images. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8.
68. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.
69. Nolte, M.; Kister, N.; Maurer, M. Assessment of deep convolutional neural networks for road surface classification. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 381–386.

70. Geiger, A.; Lenz, P.; Stiller, C.; Urtasun, R. Vision meets Robotics: The KITTI Dataset. *Int. J. Robot. Res. (IJRR)* **2013**, *32*, 1231–1237. [CrossRef]
71. Maddern, W.; Pascoe, G.; Linegar, C.; Newman, P. 1 year, 1000 km: The oxford robotcar dataset. *Int. J. Robot. Res.* **2017**, *36*, 3–15. [CrossRef]
72. 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
73. Cheng, L.; Zhang, X.; Shen, J. Road surface condition classification using deep learning. *J. Vis. Commun. Image Represent.* **2019**, *64*, 102638. [CrossRef]
74. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
75. Dewangan, D.K.; Sahu, S.P. RCNet: Road classification convolutional neural networks for intelligent vehicle system. *Intell. Serv. Robot.* **2021**, *14*, 199–214. [CrossRef]
76. Samo, M.; Mafeni Mase, J.M.; Figueredo, G. Deep learning with attention mechanisms for road weather detection. *Sensors* **2023**, *23*, 798. [CrossRef]
77. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
78. Carrillo, J.; Crowley, M.; Pan, G.; Fu, L. Design of efficient deep learning models for determining road surface condition from roadside camera images and weather data. *arXiv* **2020**, arXiv:2009.10282.
79. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–6 February 2017; Volume 31.
80. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
81. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
82. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8697–8710.
83. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
84. Khan, M.N.; Ahmed, M.M. Weather and surface condition detection based on road-side webcams: Application of pre-trained convolutional neural network. *Int. J. Transp. Sci. Technol.* **2022**, *11*, 468–483. [CrossRef]
85. Liang, H.; Zhang, H.; Sun, Z. A Comparative Study of Vision-based Road Surface Classification Methods for Dataset from Different Cities. In Proceedings of the 2022 IEEE 5th International Conference on Industrial Cyber-Physical Systems (ICPS), Coventry, UK, 24–26 May 2022; pp. 1–6.
86. Park, J.; Min, K.; Kim, H.; Lee, W.; Cho, G.; Huh, K. Road surface classification using a deep ensemble network with sensor feature selection. *Sensors* **2018**, *18*, 4342. [CrossRef]
87. Varona, B.; Monteserin, A.; Teyseyre, A. A deep learning approach to automatic road surface monitoring and pothole detection. *Pers. Ubiquitous Comput.* **2020**, *24*, 519–534. [CrossRef]
88. Baldini, G.; Giuliani, R.; Geib, F. On the application of time frequency convolutional neural networks to road anomalies' identification with accelerometers and gyroscopes. *Sensors* **2020**, *20*, 6425. [CrossRef]
89. Menegazzo, J.; von Wangenheim, A. Road surface type classification based on inertial sensors and machine learning: A comparison between classical and deep machine learning approaches for multi-contextual real-world scenarios. *Computing* **2021**, *103*, 2143–2170. [CrossRef]
90. Yoo, J.; Lee, C.H.; Jea, H.M.; Lee, S.K.; Yoon, Y.; Lee, J.; Yum, K.; Hwang, S.U. Classification of Road Surfaces Based on CNN Architecture and Tire Acoustical Signals. *Appl. Sci.* **2022**, *12*, 9521. [CrossRef]
91. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
92. Rateke, T.; Justen, K.A.; Von Wangenheim, A. Road surface classification with images captured from low-cost camera-road traversing knowledge (rtk) dataset. *Rev. Inform. Teór. Apl.* **2019**, *26*, 50–64. [CrossRef]
93. Shinzato, P.Y.; dos Santos, T.C.; Rosero, L.A.; Ridel, D.A.; Massera, C.M.; Alencar, F.; Batista, M.P.; Hata, A.Y.; Osório, F.S.; Wolf, D.F. CaRINA dataset: An emerging-country urban scenario benchmark for road detection systems. In Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, 1–4 November 2016; pp. 41–46.
94. Rateke, T.; Von Wangenheim, A. Road surface detection and differentiation considering surface damages. *Auton. Robot.* **2021**, *45*, 299–312. [CrossRef]
95. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

96. Balcerek, J.; Konieczka, A.; Piniarski, K.; Pawłowski, P. Classification of road surfaces using convolutional neural network. In Proceedings of the 2020 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), Poznan, Poland, 23–25 September 2020; pp. 98–103.
97. Espíndola, A.C.; Nobre Júnior, E.F.; Silva Júnior, E.T.d. Pavement Surface Type Classification Based on Deep Learning to the Automatic Pavement Evaluation System. In Proceedings of the Joint XLII Ibero-Latin-American Congress on Computational Methods in Engineering and III Pan-American Congress on Computational Mechanics, ABMEC-IACM, Rio de Janeiro, Brazil, 9–12 November 2021.
98. Shi, R.; Yang, S.; Chen, Y.; Wang, R.; Zhang, M.; Lu, J.; Cao, Y. CNN-Transformer for visual-tactile fusion applied in road recognition of autonomous vehicles. *Pattern Recognit. Lett.* **2023**, *166*, 200–208. [CrossRef]
99. Šabanovič, E.; Žuraulis, V.; Prentkovskis, O.; Skrickij, V. Identification of road-surface type using deep neural networks for friction coefficient estimation. *Sensors* **2020**, *20*, 612. [CrossRef] [PubMed]
100. Zhao, T.; He, J.; Lv, J.; Min, D.; Wei, Y. A Comprehensive Implementation of Road Surface Classification for Vehicle Driving Assistance: Dataset, Models, and Deployment. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 8361–8370. [CrossRef]
101. Shafer, G. Dempster-shafer theory. *Encycl. Artif. Intell.* **1992**, *1*, 330–331.
102. Guo, Y.H.; Zhu, J.R.; Yang, C.C.; Yang, B. An Attention-ReXNet Network for Long Tail Road Scene Classification. In Proceedings of the 2023 IEEE 13th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Qinhuangdao, China, 11–14 July 2023; pp. 1104–1109.
103. Han, D.; Yun, S.; Heo, B.; Yoo, Y. Rexnet: Diminishing representational bottleneck on convolutional neural network. *arXiv* **2020**, arXiv:2007.00992.