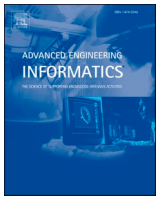
Advanced Engineering Informatics 60 (2024) 102388

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei



From global challenges to local solutions: A review of cross-country collaborations and winning strategies in road damage detection

Deeksha Arya [a,*], Hiroya Maeda [b], Yoshihide Sekimoto [a]

[a] Centre for Spatial Information Science, The University of Tokyo, Japan
[b] UrbanX Technologies, Inc., Tokyo, Japan

ARTICLE INFO

Keywords:
Automation
Big data
Deep learning
Global road damage detection
Intelligent transport
Road safety

ABSTRACT

Monitoring road conditions is crucial for safe and efficient transportation infrastructure, but developing effective models for automatic road damage detection is challenging requiring large-scale annotated datasets. Cross-country collaboration provide access to diverse datasets and insights into factors affecting road damage detection models. This paper presents a review of winning strategies of the Crowdsensing-based Road Damage Detection Challenge (CRDDC) held in 2022 as a Big Data Cup, with 90+ teams from 20+ countries proposing solutions for six countries: India, Japan, the Czech Republic, Norway, the United States, and China. The best solution achieved an F1-score of 77 % for all six countries, which is 2.7 % better than the 2nd ranked solution. This study explores the impact of factors influencing dataset and model selection by CRDDC winners. The study's insights can guide future research in making data-related choices and developing more effective road damage detection models accounting for the diverse road conditions across different countries.

1. Introduction

The transportation sector is a critical driver of economic growth and development, with roads serving as its backbone [1]. However, road safety remains a significant concern worldwide, with approximately 1.35 million road traffic deaths reported each year [2]. Automating road inspection is becoming increasingly essential to address this safety issue, apart from making the cities smarter and reducing road maintenance costs [3–5]. Automated road inspection [6] involves using advanced technologies and methods to detect and monitor road damage, such as cracks, potholes, and other defects that can compromise safety. By enabling quick and accurate detection of potential hazards, automated road inspection enhances road safety and reduces the cost and time associated with manual inspections.

Among the various technologies and high-precision methods used for automating road inspection [7–10], image-based approaches have gained prominence [11,12]. These approaches employ machine learning models to analyze images captured by cameras attached to vehicles or drones, offering a cost-effective solution [13,14]. Road images can be captured from two primary views: horizontal and top-down. Horizontal view is mostly considered for object-level detection and classification, requiring the model to prioritize the road region and minimize the influence of irrelevant areas. In contrast, top-down view also supports pixel-level detection by separating damage and background into different categories. Pixel-level detection provides detailed information about the damaged area but struggles with accurately identifying damage types. On the other hand, object-level detection addresses this limitation by utilizing deep semantic information to detect and classify various road damage instances. However, training effective deep learning-based object detection and classification models requires a substantial amount of data [15].

While generic datasets like PASCAL-VOC [16,17], KITTY [18], and MSCOCO [19] have been extensively explored [20], domain-specific datasets are required for the road sector. In recent years, various datasets have been introduced to support road damage detection, and researchers have experimented with data from different locations, emphasizing the need for location-specific datasets and models [11]. Nonetheless, another concern is the solutions proposed utilizing a given dataset. The simplest approach is to release the data and let the research community explore state-of-the-art solutions based on the data. An alternative approach is to organize data cup challenges, where researchers are invited to provide solutions to a defined problem using the provided dataset. Multiple researchers working on the same dataset can lead to diverse and optimized solutions, avoiding bias and enabling a

* Corresponding author.
E-mail address: deeksha@iis.u-tokyo.ac.jp (D. Arya).

https://doi.org/10.1016/j.aei.2024.102388
Received 26 September 2023; Received in revised form 9 January 2024; Accepted 26 January 2024
Available online 7 March 2024

comprehensive evaluation of state-of-the-art approaches. Data cup challenges differ from the general case of publicly available data, where researchers may train their own models, as it ensures fairness by fixing the train-test data split and providing standardized evaluation frameworks, with a timeline. These challenges provide insights into the contemporary methods and approaches employed by researchers worldwide, representing the current status quo in the field [16,17,21–23].

The Crowdsensing-based Road Damage Detection Challenge (CRDDC) is one such data cup that focused on automating road condition monitoring in six countries: India, Japan, the Czech Republic, Norway, the United States, and China [24]). The challenge released a dataset, Road Damage Detection Dataset, RDD2022, [25] consisting of 47,420 road images capturing different views of road surface using drones, vehicle-mounted smartphones, or high-resolution cameras ([Dataset]: [26]. Sample images are shown in Fig. 1. The dataset has been annotated with over 55,000 instances of four road damage categories: Longitudinal Cracks (D00), Transverse Cracks (D10), Alligator Cracks (D20), and Pothole (D40). The challenge attracted participation from 90 + teams representing 20 + countries, and their solutions were evaluated using five leaderboards based on their performance on unseen test images from the underlying six countries. The top-performing model [27] achieved remarkable results by employing ensemble learning techniques with YOLO [28,29] and Faster-RCNN [30] series models, achieving an impressive F1 score of approximately 77 % for the combined test data from all six countries. Table 1 presents the information of Ranks and Scores of top 11 teams in CRDDC.

Further information about the CRDDC dataset and other details for tasks, contributors, winners etc. are provided in [24] and [25]. Details of the Global Road Damage Detection Challenge (GRDDC) organized in 2020, targeting approaches for India, Japan, and Czech Republic are provided in [22]. Here it may be noted that, multi-view learning and handling multi-country data present significant challenges, such as missing or noisy information in multi-view data and variations in road damage data across different countries. Managing large-scale data in multi-country datasets also poses scalability issues. Extensive research is needed to overcome these challenges and develop innovative approaches for processing multi-view data and effectively managing multi-country datasets in road damage detection. This manuscript addresses the need and provides an in-depth analysis of the winning strategies from the CRDDC offering valuable insights into data analysis, models, and country-specific solutions for automating road condition monitoring in multiple countries. The key contributions are summarized as follows:

(i) Dataset selection: The manuscript identifies and analyzes the factors considered by the top CRDDC participants when selecting datasets for training their models. This is crucial due to the dataset's diversity in terms of location, angle, and capture method, and it can guide future research in making informed decisions regarding data selection.
(ii) Model selection: The manuscript presents a comparative analysis of the model choices made by CRDDC winners, including a diverse range of models, both state-of-the-art and previously tested ones (through GRDDC'2020). Furthermore, some teams proposed modifications to the underlying model architecture to improve performance for road damage detection. This comprehensive analysis can also guide researchers in selecting appropriate models for their work.
(iii) Data from other countries: The manuscript describes experiments conducted in the CRDDC, aiming to develop multi-national solutions using data from six countries with different views. It discusses the factors considered by participants when incorporating inter-country data to train target models. These insights assist researchers in making informed decisions about including data from other countries in their training datasets.

Overall, this manuscript provides valuable insights into the challenges and opportunities of multi-view learning and handling multi-country data in the context of road damage detection, enabling researchers to make informed decisions in their future work.

2. Related work

2.1. Datasets and data-driven solutions for RDD

RDD (Road Damage Detection) is an important area of research due

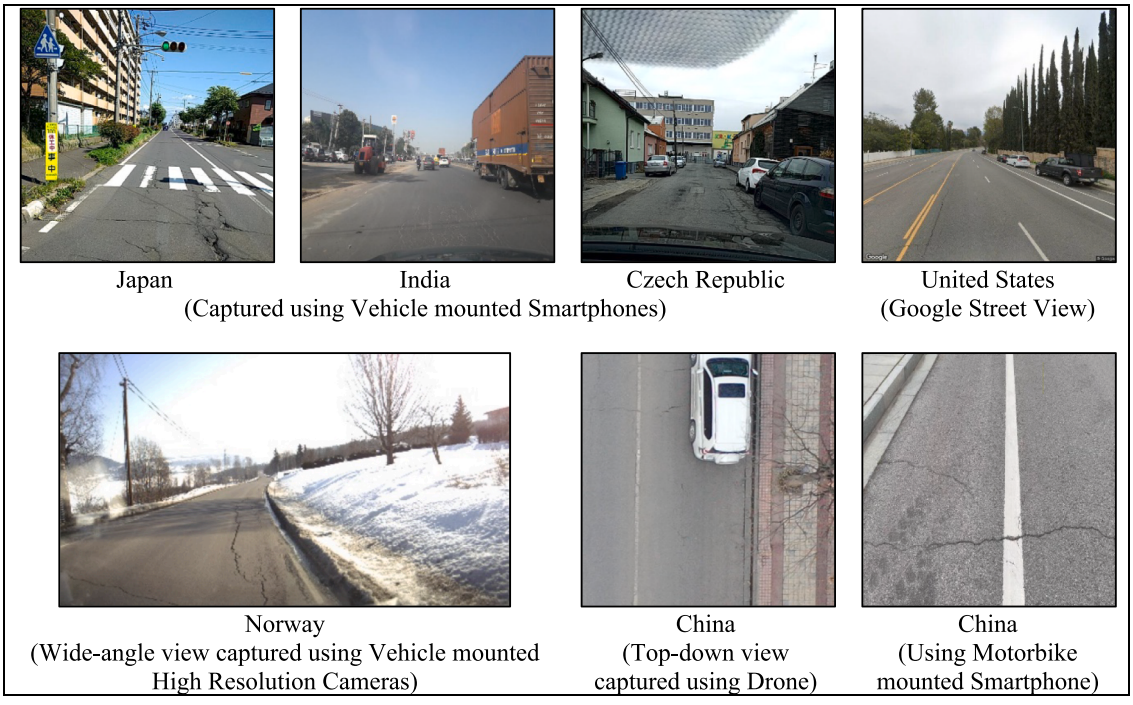


Fig. 1. Sample images for road views considered in the CRDDC data [25] from the six countries.

Table 1
Details of CRDDC winning teams (Source: [24], table II).

Rank	Team Reference	F1-Score for the five leaderboards in CRDDC				
		All 6 countries	India	Japan	Norway	United States
1	ShiYu_SeaView [27]	0.770	0.583	0.789	0.595	0.844
2	DongjunJeong [31]	0.743	0.540	0.750	0.538	0.801
3	MDPT [32]	0.741	0.516	0.735	0.504	0.817
4	SGG-RS-Group [33]	0.727	0.545	0.727	0.481	0.779
5	IRCV-URV [34]	0.726	0.518	0.735	0.498	0.779
6	IMSC [35]	0.728	0.542	0.725	0.478	0.774
7	NJUPT [36]	0.718	0.559	0.720	0.458	0.743
8	TUT [37]	0.694	0.519	0.773	0.464	0.727
9	MILA [38]	0.697	0.494	0.716	0.462	0.775
10	SIAI [39]	0.612	0.417	0.608	0.424	0.663
11	Kubapok [40]	0.603	0.421	0.594	0.383	0.649

tiem nang

to its potential to improve road safety and reduce maintenance costs. To develop effective data-driven solutions for RDD, it is necessary to have access to high-quality datasets. Several datasets have been introduced in the recent years [41–44], each with its own characteristics and challenges [12]. Some notable examples are: RDD2018 [13], RDD2019 [45], EdmCrack600 [14]; RDD2020 [11] and [46], RDD2022 [26]; CQU-BPDD [47] and CQU-BPMDD [48]. These datasets differ in various aspects, such as the study area, number of images, image size, data collection device, data acquisition method, and view captured, etc. A comparative summary of some of these datasets is presented with the RDD2022 dataset [25] proposed through CRDDC'2022 [24] in Table 2.

It may be noted that most of the datasets consider either the top-down view, focussing on the road surface [47–50], or the horizontal-direction capturing a wide view of the road, road objects and surroundings [13,22,51,52]. Some authors like Majidifard et al. [54] have considered a combination of images capturing horizontal as well as top-down view of the road. Like Majidifard et al. [54], the RDD2022 dataset includes images captured from both the directions. Additionally, RDD2022 also includes some extra-wide angled images captured using two cameras in Norway. Also, the RDD2022 captures the highest level of

Table 2
Comparative summary of existing road damage datasets with the RDD2022 dataset [25] proposed through CRDDC'2022 [24].

S. N.	Dataset Reference	Study Area	Number of images	Image Resolution	Data Collection Device	Vehicle involved	Annotation Level	Road View Captured
1	**GAPs v1** [49]	Germany	1,969	1920 × 1080	Professional Cameras	Specialized vehicles for pavement inspection	Image-level	Top-down (Only Road surface)
2	**GAPs v2** [50]	Germany	2,468	1920 × 1080	Professional Cameras			
3	**CQU-BPMDD** [48]	South-western China	38,994 (9,851 diseased images)	3692 × 2147	In-vehicle cameras			
4	**CQU-BPDD** [47]	Southern China	60,056 (16,726 diseased images)	1200 × 900	In-vehicle cameras			
5	Maeda et al. [13]	Japan	9,053	600 × 600	Smartphone	General-purpose	Boundary-box	Horizontal (Wide view)
6	Angulo et al. [51]	1) Japan, 2) Italy, 3) Mexico	18,034	600 × 600	Smartphone			
7	Roberts et al. [52]	Italy	7000	600 × 600	Smartphone			
8	**RDD2020** [53]	1) Japan, 2) India, 3) Czech Republic	26,620	600 × 600, 720 × 720	Smartphone			
9	Majidifard et al. [54]	United States	7,237	640 × 640	Google API	N/A	Boundary-box	1) Top-down, 2) Horizontal Wide view
10	**RDD2022 (proposed through CRDDC'2022)**l	1) Japan, 2) India, 3) Czech Republic, 4) Norway, 5) United States, 6) China	47,420	512 × 512, 600 × 600, 720 × 720, 3650 × 2044	Smartphones, High-resolution Cameras, Google Street View images	A combination of general-purpose and specialized vehicles	Boundary-box	1) Top-down, 2) Horizontal Wide view, 3) Extra-wide view

geographical diversity as compared to the other datasets, making it suitable for training highly robust deep learning-based models.

Apart from the datasets discussed above, several other datasets such as those including pixel-level annotations for road damage [14,55,56], are widely studied by the researchers to support segmentation-based applications. To effectively analyze the available datasets and develop data-driven solutions, researchers have used a variety of techniques, including computer vision, machine learning, and deep learning [57–60]. Deep learning techniques such as convolutional neural networks (CNNs) have proven to be particularly effective for RDD, achieving state-of-the-art performance on many datasets [15,58,59,61]. In addition to the development of datasets and data-driven solutions, researchers are also exploring ways to improve the efficiency and accuracy of RDD [15,62–65]. Approaches like data augmentation [66], addition of attention mechanisms [67–71], and optimizing network architecture [61] to suit the RDD domain [62,72], etc. are generally considered. One major approach is to use ensemble models, which combine the outputs of multiple models to achieve better performance. However, the suitability of an ensemble model depends on factors such as the availability of computational resources and the trade-off between inference time and accuracy. Further, there are multiple approaches to perform ensemble. The factors specific to RDD domain needs to be considered to effectively decide the ensemble approach and the base models to use. The factors considered by CRDDC winners in this regard are discussed in section 3.2. A brief overview of deep learning-based object detection approaches and attention mechanisms is provided in following subsections.

2.2. *Deep learning-based object detection*

The rapid progress in deep learning algorithms has revolutionized the field of object detection and visual tracking in real-world scenarios [73]. Deep learning algorithms, such as CNNs, can learn both low-level and high-level visual features. They excel at detecting and identifying target objects within an image by assigning rectangular bounding boxes that indicate the presence and certainty of objects. Two main approaches to deep learning-based object detection in computer vision are one-stage [28,29,74,75] and two-stage methods [30,76,77]. One-stage methods like YOLO [28] and SSD [78] perform detection and classification in a single pass through the network, making them faster and simpler to implement compared to two-stage methods. However, one-stage methods may struggle with detecting small objects or objects with complex shapes. On the other hand, two-stage methods like Faster R-CNN [30] and Mask R-CNN [77] utilize a separate region proposal network to generate candidate object regions, which are then classified and refined in a second pass through the network. This two-stage process enhances their accuracy in detecting objects of various sizes and shapes but comes at the cost of increased complexity and slower implementation. The choice between one-stage and two-stage methods depends on the specific requirements of the application and the desired trade-off between speed and accuracy. Further, the quality and characteristics of the underlying data is a critical aspect in training effective one-stage or two-stage models. With the prevalence of deep learning-based approaches, it becomes crucial to consider various factors when designing a context-specific solution for a particular domain like road damage detection. The subsequent section delves into the factors considered by the CRDDC winners in this direction.

2.3. *Attention mechanisms in computer vision*

Attention mechanisms in computer vision [67,79,80], are inspired by how humans selectively focus on key information in a scene. These methods dynamically assign weights to visual features, guiding the learning process and decision-making of deep learning models, akin to the human eyes adjusting focus to prioritize details while overlooking background noise [58,59]. Some of the attention mechanisms relevant to the current work are summarized as follows.

(i) Channel Attention [81,82]: These mechanisms adjust the relative importance of different feature channels within a given spatial location. For example, Squeeze-and-Excitation [83,84] dynamically amplifies channels containing prominent features while suppressing those with background noise, enhancing object-background separation.

(ii) Coordinate Attention [85]: This approach expands upon channel attention incorporating spatial cues by factoring it into separate 1D encodings along each spatial dimension. This allows the model to capture long-range dependencies in one direction while preserving precise positional information in the other, like zooming in on specific parts of the scene while maintaining contextual awareness. This empowers tasks like precise object localization and action recognition.

(iii) Spatial Attention ([86,87]: Spatial attention focuses on specific regions within a feature map, akin to human gaze scanning across a scene, and are used widely in several different forms. For instance,

a. Transformer Attentions [88], [89]: Transformer decoders utilize spatial attention to establish relationships between distant elements, even if they are not physically close. Additionally, by attending to multiple locations simultaneously, transformer attentions efficiently capture long-range dependencies within an image.

b. Swin Transformers [90]: To overcome the high computational cost of traditional transformers, Swin transformers leverage shifted window-based attention mechanisms. They process local patches within the image efficiently, then stitch them together to build a global understanding. This makes them ideal for resource-constrained scenarios like mobile devices.

These diverse attention mechanisms, strategically applied, provide models with the ability to selectively focus on crucial features, enhance localization and precision, and overcome challenges like occlusions and cluttered backgrounds. Researchers are actively exploring new architectures, such as hybrid attention mechanisms that combine different types of attention and applying these techniques to solve increasingly challenging tasks like 3D vision and scene understanding, including the field of road damage detection [69]. For instance, the CRDDC participants utilized several mechanisms such as Pyramid Squeeze Attention [91], Atrous Spatial Pyramid Pooling [92,93] etc. Pyramid Squeeze Attention refines feature maps at multiple scales within the network for better object detection of varying sizes, while Atrous Spatial Pyramid Pooling utilizes dilated convolutions to aggregate multi-scale information for more accurate semantic segmentation. The application of these mechanisms for road damage detection is described in the following section.

3. Learnings from winning strategies

Fig. 2 presents a comparative performance visualization for 11 winners of CRDDC, corresponding to Table 1 (source: [24], table II). Team ShiYu_SeaView stands out as the overall leader with a remarkable F1-Score across all nations. Teams DongjunJeong and MDPT closely follow, displaying consistent performance. Further, the teams SGG-RS-Group, IRCV-URV, and IMSC also demonstrate competitive F1-Scores, highlighting their effectiveness in diverse settings. While TUT and MILA exhibit proficiency in specific countries, SIAI and Kubapok, with comparatively lower overall F1-Scores, still make meaningful contributions to the challenge. The rankings showcase varying strengths among teams, emphasizing the adaptability and robustness of their models across different geographical contexts. Detailed analysis of the strategies adopted by the winning teams is presented in the following subsections.

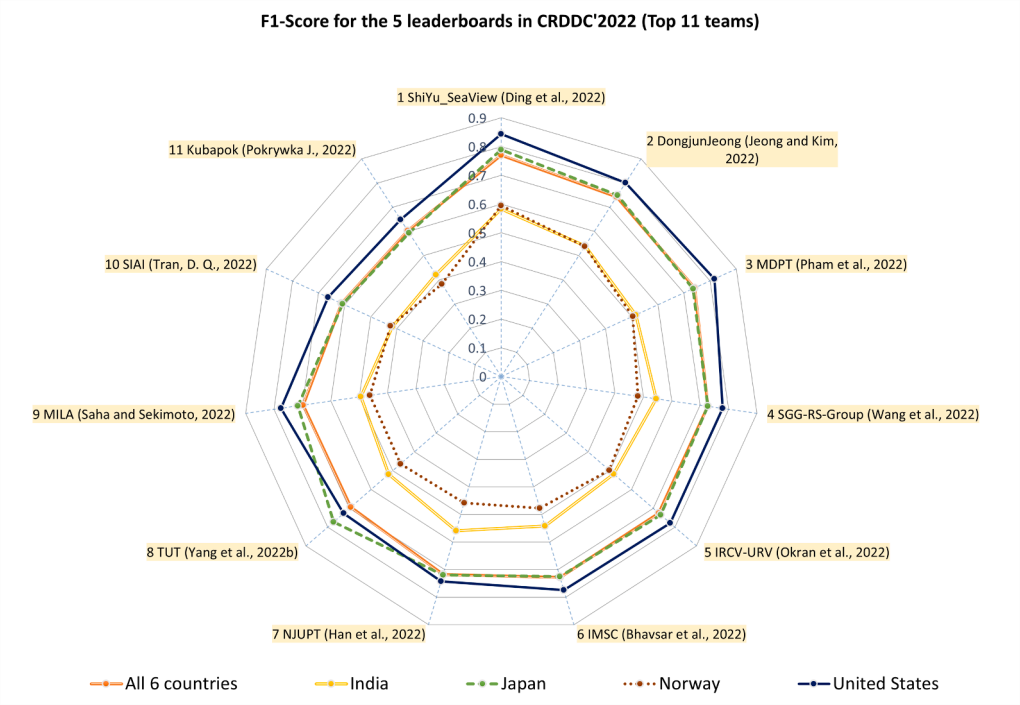


Fig. 2. Comparative performance visualization of CRDDC winning teams for the 5 leaderboards (Source: [24], table II).

3.1. *Data-related analysis*

The data is the main part of the challenge, as well as the solutions proposed. The data released through CRDDC is diverse in terms of various factors, including: the source, location, method of capturing etc. The Table 3 provides information on the CRDDC winners' choices for data and augmentation techniques to train models for four different countries: India, Japan, Norway, and the United States. It may be noted that, the number of images for China and Czech Republic were comparatively low in the RDD2022 dataset, so CRDDC did not include models targeting these two countries specifically. Further, the participants were allowed to propose models targeting one or multiple countries using any combination of datasets (comprising multi-view images from 6 countries) released through CRDDC. The main aspects considered by the participants in selecting appropriate data for model training are listed as follows.

(i) **Data Characteristics**

a. **Number of images for each country**: The CRDDC participants analyzed the number of images available for each country in the dataset to identify any data imbalance issues. They found that some countries had significantly fewer images than others, which could affect the model's performance. The teams that specifically looked at this aspect were ShiYu_SeaView, IMSC, Dongjun_Jeong, MDPT, IRCV_URV, NJUPT, and kubapok.

b. **Number of class-wise labels for each country**: The CRDDC participants also looked at the number of labels for each class of damage for each country to identify potential class imbalance issues. They found that some classes of damage had fewer labels than others, which could affect the model's ability to detect those types of damage. The teams that specifically looked at this aspect were ShiYu_SeaView, IMSC, MDPT, IRCV_URV, TUT, MILA, and kubapok. The teams proposed several approaches to address the imbalance issues. For instance, team MILA utilized weighted image selection during training.

c. **Number of labels per 100 images of each country**: The number of labels per 100 images is an aspect related to data sparsity. Some countries had a low number of labels per 100 images, indicating that there may not be enough labeled data to train an effective model for those countries. The team that specifically looked at this aspect was kubapok.

d. **Number of positive images for each country**: This aspect is related to the prevalence of damage in the images. Images containing at least one instance of road damage are referred as positive images, representing their significance in training the model to recognize specific types of road damage. The teams, MDPT and MILA, found that some countries had a higher percentage of positive images than others, which could affect the model's ability to detect damage accurately.

(ii) **View Captured and Image Properties**

a. **Scenic similarity between the view captured from different countries**: Teams like MILA and MDPT considered the scenic similarity between images captured from different countries. They found that some countries had similar scenery, while others were different. This aspect could affect the model's ability to generalize to new countries. The teams utilized this aspect to combine data from countries to propose a single model. While team MDPT proposed to have a combined model for India and Japan, team MILA chose to combine data from Japan and US. Further, team SIAI emphasized the distinction in weather conditions captured in the images from Norway compared to other countries, leading to the need of specific data enhancement.

b. **Image size (Resolution):** The CRDDC participants analyzed the image sizes and resolutions for each country to ensure that the models could handle images of different sizes and resolutions. They found that some countries (Norway, and a small portion of Japan) had images with a higher resolution than others, which could affect the model's performance and resource requirement. The teams that specifically looked at this aspect were Dongjun_Jeong, MDPT, SGG_RS_Group, IRCV_URV, NJUPT, TUT, MILA, and SIAI.

Table 3
Details of data used to train models for different countries by the top 11 teams of CRDDC (overall rank).

Team Name	Data used for India	Data used for Japan	Data used for United States	Data used for Norway	Data Augmentation/Comments
ShiYu_SeaView	RDD2022	RDD2022	RDD2022	RDD2022	HSV, image translation, image scale, horizontal flip, Mosaic, Mixup and Random Left-Right Flip
DongjunJeong	RDD2022	RDD2022	RDD2022	RDD2022	HSV color transformation, image scaling, vertical flip, horizontal flip, mosaic, Image patch (640 x 640 and 1024 x 1024 with stride 400).
MDPT	RDD2022	RDD2022	RDD2022	RDD2022	The default YOLOv7 image augmentation parameters are used with the following changes. Scale, mosaic, mixup, and paste_in are slightly reduced to avoid unwanted, unrealistic effects. Additionally, since road damages collected using cameras placed on car dashboards have some perspective, shear (0.01) and perspective (0.0001) are used.
	Validation Data: India and Japan		Validation Data: United States	Validation Data: Norway	
SGG-RS-Group	RDD2022 excluding China_Drone	RDD2022 excluding China_Drone	RDD2022 excluding China_Drone	RDD2022 excluding China_Drone	Mosaic augmentation during training, and test time augmentation. Test time augmentation improved F1-score by about 4 %.
IRCV-URV	RDD2022	RDD2022	RDD2022	Ensemble of models trained on RDD2022 and Norway data	Hue, saturation, and value for HSV, which produce images with different colors, image translation through moving the image along X or Y directions (or even both directions), image scaling, image flipping (flip the image from left to right), and mosaic. In the mosaic augmentation, multiple training images are combined into one image in specific ratios.
IMSC	RDD2022	RDD2022	RDD2022	RDD2022	Not used. Ensemble of country specific and general model used.
NJUPT	RDD2022 excluding China_Drone	RDD2022 excluding China_Drone	RDD2022 excluding China_Drone	RDD2022 excluding China_Drone	Blur, MedianBlur, ToGray, CLAHE, ImageCompression, to improve the generalization of the model and the instances of the minority class. Rotation avoided to preserve the characteristics of Longitudinal and Transverse cracks.
TUT	RDD2022	RDD2022	RDD2022	RDD2022	Only for Japan, Images with D40 class. LinearContrast, Affine, Dropout, Rot90, Fliplr, and GaussianBlur. At 1024 resolution, the F1-score improved by about 5 %.
MILA	RDD2022 data for India, only positive images	RDD2022 data for Japan and the United States, only positive images		RDD2022 data for Norway, only positive images	Combination of 3 techniques randomly selected to augment an image: CLAHE, color jitter, downscale, emboss, equalize, fancy PCA, hue saturation value, median blur, posterize, multiplicative noise, random brightness contrast, random gamma, random tone curve, RGB shift, unsharp mask, horizontal flip.
	Country specific models are ensembled with model trained using data from all 6 countries (except China_Drone).				
SIAI	RDD2022 excluding Norway	RDD2022 excluding Norway	RDD2022 excluding China Drone and Norway	RDD2022 excluding China_Drone	VFNet training: Data Augmentation via PhotometricDistortion is utilized in MMDetection YOLOv5: Augmentation by modifying hyperparameters
kubapok	RDD2022	RDD2022	RDD2022	RDD2022	Test Time Augmentation and augmentation during training.

c. **Placement position of the bounding box in the image annotations**: The CRDDC participants analyzed the placement position of the bounding box in the image annotations to ensure that the model would not miss any damage. They found that data from Norway had bounding boxes that were placed differently, which could affect the model's ability to detect damage accurately. The teams utilized this aspect to deal with the high resolution of the images and comparatively smaller aspect of damage captured in the Norway data. The teams proposed to use cropping or patch strategy focussing on lower left part of the Norway images, where most of the bounding boxes were placed. The teams that specifically considered this aspect were MDPT, IRCV_URV, IMSC, and kubapok.

d. **Camera viewing direction**: The CRDDC participants paid attention to the camera viewing direction in their analysis, considering the multi-view aspect of the datasets. They recognized that images taken from different angles could impact the model's accuracy in detecting road damage. The teams that specifically looked at this aspect were IRCV_URV, IMSC, and kubapok. Team MDPT proposed to use Shear and Perspective for data augmentation to account for the camera viewing angle. Further, the images captured from China using drone captures a completely different view of road surface as compared to other datasets. As shown in Table 2, different teams made different choices for utilizing this data to train their model despite the exclusion of China_Drone images from test set, based on available resources and the intent to improve the robustness of trained model.

e. **Shape of the images**: The CRDDC participants analyzed the shape of the images to propose models to handle both square and rectangular images. They found that some datasets had square images, while Norway had rectangular images. The team SIAI proposed two different models to handle square and rectangular images.

(iii) **Data Utilization**

a. **Utilizing the data from other countries**: Table 3 shows most teams used the same data and augmentation techniques for all four countries. However, some teams proposed to use different training data for different target countries, such as team MILA and SIAI. Team MDPT utilized the complete data RDD2022 for training all models but proposed to vary the validation data based on the target country. Here it may be noted that utilizing the same data for all countries does not always imply the teams trained a single model for all four countries. In many cases, the teams trained multiple models by varying underlying data. For example: training two or more models for each target: one

using the country-specific data, and second, using the data from multiple countries (RDD2022 or RDD2022 except China_Drone), and finally using the ensemble of the trained models to have better performance. Further, it is worth noting that even though the China_Drone data was not included in the test data of any of the leaderboards, some teams found it helpful to improve the generalization of their models.

b. **Data Augmentation**: Image data may be enhanced [94] with several approaches to represent the domain better [95]. The augmentation techniques used by different CRDDC teams (Table 3) included HSV color transformation, image scaling, flipping, mosaic, mixup, paste-in, shear, perspective, blur, median blur, grayscale, CLAHE, image compression, contrast adjustment, dropout, rotation, Gaussian blur, fancy PCA, RGB shift, and unsharp mask. A brief analysis is presented as follows.

i. The teams considered damage distribution for each country to decide augmentation. For example, team TUT used data augmentation only for Japan, augmenting only the images with D40 category (potholes).

ii. Further, the teams also considered the domain characteristics to decide the augmentation. For example, team NJUPT pointed out that rotation blurs out the characteristics of longitudinal and transverse cracks, so should not be used for RDD domain with crack classification as a task.

iii. Furthermore, some teams used test time augmentation (TTA) to improve their model's performance. Using TTA was mostly in coherence with the winning solution from GRDDC'2020, however, the CRDDC participants also

Table 4
Details of models trained for different countries by the top 11 teams of CRDDC (overall rank).

Team Name	Model used for India	Model used for Japan	Model used for United States	Model used for Norway	Details
ShiYu_SeaView	Ensemble of 8 models			Ensemble of 9 models	YOLOv5 + YOLOv7 + Faster RCNN with SWIN Transformer used to train various models. YOLO was used for higher recall rate and Faster-RCNN for higher precision.
DongjunJeong	Ensemble of 12 models				YOLOv5x used as base model. Trained eight P5 models (166 MB) and four P6 models (269 MB). Utilized pretrained weights from GRDDC'2020 winner's model.
MDPT	Ensemble of 6 models (3 models using validation data from India, and 3 models using data from Japan)		Ensemble of 3 models (Validation data: United States).	Ensemble of 3 models (Validation data: Norway).	Trained models using YOLOv7, and its two variants by adding Coordinate attention block to head/backbone of YOLOv7. Training data was included from all countries, and validation data was specific to the target country.
SGG-RS-Group	Ensemble of 6 models			Ensemble of 4 models	YOLOv5 and its 5 variants by adding Squeeze and Excitation, and Coordinate attention blocks to YOLOv5 head/backbone.
IRCV-URV	Ensemble of 4 models (EM1)		Ensemble of 4 models (EM2)	Ensemble of 3 models (EM3)	7 models trained by varying YOLOv7 configuration, underlying data, and image size used for training. Three types of ensemble models created and tested: EM1, EM2, and EM3.
IMSC	Ensemble of 8 models (M1, M2, and 6 country-specific models)				M1: YOLOv5 Trained using RDD2022, M2: YOLOv5 trained using anchor boxes customized for RDD2022. M_C: YOLOv5-based Country specific model (6 models for 6 countries)
NJUPT	Ensemble of 6 models	Ensemble of 5 models	Ensemble of 4 models	Ensemble of 4 models	6 models trained by varying: underlying data (all countries or individual), model (YOLOv5 or YOLOv7), use of transformer attention, and image size (input for training).
TUT	Single model for all countries				YPLNet (YOLOv5 + Pyramid Squeeze Attention + Large Field Contextual Integration).
MILA	Ensemble of 3 YOLOv7 models (varying underlying data and image size).	Ensemble of YOLOv7 based model trained using (Japan + US) data with overall model		Ensemble of YOLOv7 model trained using (Norway) data with overall model	YOLOv7-based: • Country specific models trained using intra-model ensemble. • Overall model trained using RDD2022 except China_drone. Inter model ensemble used for final model.Experiment with input image size.
SIAI	Numerous image-size models trained with Yolov5x and ResNeXt101 backbone			Ensemble of VFNet model trained on RDD2022 and Norway data	Three types of models trained: • VFNet trained on data from all countries. • VFNet trained on only Norwegian dataset. • Numerous image size models with yolov5x, and ResNeXt101 backbone for countries with square images
kubapok	Ensemble of models trained with different data augmentation hyperparameter settings.				• Two architectures used: YOLOv5l6 (76.8M params, 111.4B FLOPS) and YOLOv5x6 (140.7M params, 209.8B FLOPS). • Training a model jointly on all countries' data found superior to fine-tuning to a specific country in RDD2022.

pointed that even though TTA helped improving the score, it also increased the inference time three-fold, since the model is now required to parse each image three times.

In summary, the teams' choices in data augmentation strategies showcase a sophisticated understanding of the specific challenges posed by diverse geographical contexts, domain intricacies, and the trade-off between model performance and computational efficiency. These considerations collectively contribute to a robust and context-aware approach to data augmentation in the realm of deep learning for damage detection, classification, and localization.

Collectively, these aspects highlight the importance of analyzing the underlying dataset to identify potential issues such as data imbalance, data sparsity, and domain shift. By considering these aspects, the teams were able to develop effective solutions to improve the model's performance in detecting road damage in different countries.

3.2. *Model-related choices*

Teams in CRDDC'2022 considered several factors for developing object detection models to detect road damages. These factors include choosing between one-stage and two-stage models, selecting the base model, customizing the underlying architecture, incorporating attention mechanisms, considering input image shape/dimension, choosing the image size for training, deciding whether to use ensemble or non-ensemble methods, determining the ensemble basis and number of models for ensemble, and balancing speed, accuracy, and resource requirements. CRDDC winning teams used different approaches addressing these factors. Table 4 summarizes the models used by different teams. Corresponding analysis concerning aforementioned factors is provided as follows.

(i) **Architecture Selection**

a. **One stage vs two stage detectors**: One-stage detectors like YOLO are faster and simpler but may have lower accuracy than two-stage detectors like Faster-RCNN. Even though the challenge aimed at achieving higher accuracy, only the winning team Shiyu_SeaView utilized a combination of one-stage (for higher recall rate) and two-stage detectors (for higher precision). Rest all the teams mostly utilized YOLO-based one-stage detectors, combined with several approaches to improve the accuracy. Team SIAI proposed to use VFNet for Norway and YOLO for other countries.

b. **Base model**: Teams used different base models, with some using the latest available best-performing model (YOLOv7 [29]) while others used the one that performed better in the previous case (YOLOv5 [96] for GRDDC). For instance,

i. Team ShiYu_SeaView adopted YOLOv5 and YOLOv7 as one-stage detector baselines, based on the success of ensemble models using YOLOv4 and YOLOv5 in GRDDC'2020 [22,97].

ii. Team MDPT utilized YOLOv7 due to its superior speed and accuracy compared to other object detectors. Explored various hyperparameters and custom modules to train models for road damage detection and classification.

iii. Team Dongjun_Jeong chose YOLOv5 to leverage well-pre-trained weights [97] from the previous challenge GRDDC'2020 [22] and conducted experiments within limited GPU resources and time constraints.

iv. Team IRCV-URV chose the one-stage detector YOLOv7, because of its superior speed compared to other networks in both the one-stage and two-stage detection families.

v. Team SGG-RS-Group selected YOLOv5 in the x scale (maximum depth and width). Trained six models with different variants (YOLOv5, YOLOv5_SE, YOLOv5_CA, YOLOv5_BB, YOLOv5_HD, YOLOv5_BBHD) on the RDD2022 training set from all countries.

vi. Team IMSC tested both YOLOv5 and YOLOv7, and finally chose YOLOv5 based on the test performance.

vii. Team NJUPT utilized an ensemble of YOLOv5 and YOLOv7, to get advantage of both.

viii. Team TUT chose YOLOv5s as a baseline model, to keep it lightweight and achieve fast inference speed.

ix. Team MILA selected YOLOv7 as it was proven to surpass other object detectors in terms of speed and accuracy, while also reducing parameters and computation by approximately 40 % compared to the current state-of-the-art real-time object detectors.

x. Team SIAI proposed to use different base models for square and rectangular images, showing that VFNet performed better for high-resolution rectangular images from Norway, whereas YOLOv5-based models suited the square shaped low-resolution images from other countries.

xi. Team kubapok chose YOLOv5l6 (76.8 M params, 111.4B FLOPS) and YOLOv5x6 (140.7 M params, 209.8B FLOPS) to optimize the requirement of GPU resources.

(ii) **Architecture Customization**

a. **Customizing network layers or using additional modules**: Some teams proposed changes to network layers or the use of additional modules for better performance suiting the RDD domain or other constraints, while others proposed to use the available models without any change in the underlying architecture. The major changes suggested by the winning teams to underlying architecture are summarized in the Table 5.

b. **Incorporating Attention Mechanisms**: Attention mechanisms can help models focus on important features. Some teams proposed the use of attention mechanism considering the suitability for RDD domain.

i. MDPT: The team recognized that the channel attention mechanism used in YOLOv7 does not capture important spatial information in road damage detection. This is because channel attention compresses the entire spatial space of a channel into a single value. To overcome this limitation, the team experimented with Coordinate

Table 5
Adjustments to model network architecture suggested by CRDDC winners.

Team	Modification to network architecture and context-specific inputs
ShiYu_SeaView	Suggested using SWIN-transformer to enable better extraction of semantic information capturing long and narrow morphologies of cracks. Additionally, proposed to use ROI (Region of Interest) pooling to better adapt to special shape of road damage.
MDPT	Proposed to use additional Coordinate attention blocks in the head/backbone of YOLOv7 models. Also, utilized label smoothing considering that longitudinal and transverse crack may be mislabeled due to different perspectives or slight camera rotation.
SGG-RS-Group	Proposed five new variants of YOLOv5 by adding Squeeze and Excitation (SE) to focus on channel relationship, and Coordinate attention blocks to embed positional information to channel.
IRCV_URV	Modified YOLOv7 by replacing a block with the improved Atrous Spatial Pyramid Pooling network to detect and identify the small and big cracks in multi-scale.
IMSC	Proposed to use a new set of anchor boxes customized for RDD2022 data, for YOLOv5 model.
NJUPT	Proposed to use transformer attention with YOLOv5.
TUT	Proposed a new network called YPLNet (Yolov5s + Pyramid Squeeze Attention (PSA) + Large-field Contextual Feature Integration (LCFI)) by adding PSA and LCFI modules to YOLOv5 for multi-scale contextual feature extraction and fusion.

Table 6
Details of image size used by CRDDC winners for model training.

Team	Image Sizes	Details
ShiYu_SeaView	640 × 640 1280 × 1280 1600 × 1600 3200 × 3200	• One set of models for India, Japan, and US. • Different model training for Norway targeting larger image sizes. • To improve the effectiveness of detecting small damage in Norwegian dataset, trained the YOLOv5x and YOLOv5m on image size of 1600 and 3200.
Dongjun_Jeong	640 × 640	Adjusting input size individually for each country suggested, but trained with 640 × 640 due to time and GPU constraints
MDPT	640 × 640	Reduced training image size to 640 × 640 for all countries to accommodate GPU limitations
SGG-RS-Group	640 × 640	Scaled image sizes of all countries to 640 × 640
IRCV_URV	640 × 640 1280 × 1280	• Trained model on Norwegian data with image size 1280 × 1280. • Trained multiple models using RDD2022 dataset with varying image sizes (640 × 640, 1280 × 1280, multiscale)
IMSC	640 × 640	• 640 × 640 for YOLO variants • Size automatically adjusted by the transformer for DINO (self-supervised vision transformer)
NJUPT	640 × 640 448 × 448	Experimented with image sizes of 640 × 640 and 448 × 448
TUT	1280, 1024	• Recognized that the D40 class occupies a smaller pixel area in the images. • Scaled image resolution to 1280 or 1024 during training and validation to improve performance.
MILA	448 × 448 640 × 640 704 × 704 1280 × 1280	• Image sizes of 448 and 640 for overall training, India, Japan, the United States, and Norway with YOLOv7-X. • Image size of 704 for India and 1280 for Norway with YOLOv7-W6. • Emphasized the importance of selecting the appropriate image size for training to improve detection accuracy, especially when using country-specific data.
SIAI	1333 × 800	• Considering that images from Norway have a higher resolution and rectangular dimensions, the team trained it separately scaling the input image to 1333 × 800 pixels. • Images from other countries scaled to square size.
kubapok	1280p	Used 1280p image resolution for training and inference

Attention, a technique that introduces spatial information into channel attention. Given that damages can appear at various locations with diverse appearances due to camera perspectives, incorporating spatial information becomes crucial. By leveraging Coordinate Attention, the team successfully improved the performance of the YOLOv7 model on the RDD2022 dataset.

ii. SGG_RS_Group: The team proposed the use of Squeeze and Excitation blocks to enhance channel relationships in the model. Additionally, Coordinate Attention blocks are employed to incorporate positional information into channel attention. This approach helps improve the overall performance of the model.

iii. NJUPT: Integrated transformer's attention mechanism with YOLOv5 to improve model's ability to extract features.

iv. TUT: Added the Pyramid Squeeze Attention module to the YOLOv5 backbone network to obtain multi-scale spatial information to enrich the feature space, and to build long-term dependencies between different channels to obtain more refined features.

(iii) **Data Considerations**

a. **Input Image shape/dimension**: The input image shape or dimension can affect model performance. RDD2022 data contains both square and rectangular shaped images. The team SIAI considered this aspect and proposed to use VFNet for rectangular images and YOLOv5 for square images based on their results.

b. **Selection of image size used for training**: The selection of appropriate image sizes for training played a crucial role in improving detection accuracy, as observed by multiple teams. Various approaches were proposed, including using different sizes for different countries, scaling image sizes uniformly, and experimenting with different resolutions. The teams recognized the significance of selecting suitable image sizes, especially when dealing with country-specific data and specific challenges such as detecting small damages. Factors like GPU limitations and time constraints influenced the choice of fixed image sizes. Overall, the teams emphasized the importance of image size selection in optimizing performance and addressing the unique requirements of road damage detection. The details are provided in Table 6.

(iv) **Model Strategies**

a. **Ensemble vs non-ensemble**: Ensemble learning involves combining multiple models to improve performance. Most of the teams utilized ensemble modelling to increase the accuracy. For instance, team ShiYu_SeaView observed an improvement of 7–10 % accuracy using ensemble learning. Some teams, like MILA, also experimented with models without using ensemble learning, targeting optimized accuracy, time, and resource requirements. Team TUT proposed a single model for all countries without using ensemble learning.

b. **Ensemble basis**: Teams that used ensemble methods mostly based them either on network architecture or underlying data. Table 7 summarizes the aspects considered by different teams.

c. **Number of models for Ensemble**: The analysis of ensemble models used by different teams highlights the significance of carefully considering the number of models in an ensemble and striking a balance between accuracy and inference time. Table 4 summarizes the diverse approaches taken by the teams in this regard. Different teams have experimented with varying numbers of base models in their ensembles, yielding different results for different countries, with some opting for a larger number (8 to 12 models), while others chose a moderate number (3 to 6 models). Some teams ensembled different number of models for different countries, while some kept the number same for all countries. Team MILA and kubapok emphasized the importance of diversity and independence of base models for effective ensembling. However, it is crucial to note that increasing the number of models may lead to longer inference times without guaranteed accuracy improvement.

Table 7
Ensemble basis used by CRDDC winners for proposing best-performing models.

Team	Ensemble Basis
ShiYu_SeaView	Ensemble of different models (YOLOv5, YOLOv7, Faster RCNN) trained using various architectures to optimize for recall rate and precision.
DongjunJeong	Ensemble of 12 YOLOv5-based models (eight P5 and four P6) trained by varying underlying data (adding image patch of Norway data with image size 640 and 1024 to original images) and the use of pre-trained weights.
MDPT	Ensembled three types of models for each target: (i) YOLOv7, (ii) adding 3 coordinate attention layers in the head of YOLOv7, and (iii) 3 additional coordinate attention layers in the backbone of YOLOv7.
SGG-RS-Group	Ensembled YOLOv5 with its variants trained by adding Squeeze and Excitation (SE) and Coordinate attention blocks to its head/backbone.
IRCV-URV	Trained 7 models with different configurations of YOLOv7, varying underlying data, and image size. Three types of ensemble models (EM1, EM2, EM3) tested, with different number/combination of models ensembled, and found to have different performance for different targets (details in table 3).
IMSC	Ensemble based on both underlying data (country-specific or overall) and anchor boxes (YOLOv5 trained using default anchor boxes and boxes customized for RDD2022).
NJUPT	Ensemble of models with variations in underlying data (individual or all countries), model architecture (YOLOv5 or YOLOv7), transformer attention, and image size.
MILA	Tried both Intra-model and Inter-model ensemble. Intra-model ensemble: Weights of models obtained at different iterations of the YOLOv7 model were considered. 2) Inter-model ensemble: Weights of models obtained by training YOLOv7 models on different training data were considered.
SIAI	Ensemble of models trained using data from individual countries with model trained using data from all countries.
kubapok	Ensemble of models trained with different data augmentation hyperparameter settings.

The decision to use an ensemble model should be based on computational resources, task requirements, and the trade-off between inference time and accuracy.

(v) **Performance Balance: Speed, Accuracy or Resource Requirement (Computational Cost):** Although CRDDC used only F1 score as the evaluation metric, teams tried to strike a balance between speed, accuracy, and computational cost. It is interesting to analyse how the participants dealt with limitation of resources while maintaining performance in top solutions.

a. Team ShiYu_SeaView used Faster RCNN instead of cascade RCNN due to limitation of GPU resources.
b. Team Dongjun_Jeong emphasized the importance of hyper-parameter optimization and expressed eagerness to test with sufficient GPU resources for further improvements. The team also reported slow inference speed (1 FPS) for their proposed solution. Conversion of the trained PyTorch model to TensorRT format was suggested to improve the speed by around three times.
c. Team MDPT expressed interest in exploring various YOLOv7 configurations (YOLOv7-X, YOLOv7-W6, etc.) known for improved accuracy on the MS COCO dataset. However, limited computational resources prevented them from conducting experiments with these models.
d. Team SGG_RS_Group scaled images to size 640 × 640, citing that scaling to 1280 × 1280 will increase the training time and GPU requirements.
e. Team TUT proposed to use a lightweight baseline model, YOLOv5s, avoiding larger models like YOLOv5m, to prioritize deployment ease and detection speed in the actual road damage evaluation. They enhanced accuracy by incorporating PSA and LCFI modules to extract road damage features without significant computational overhead.
f. Team MILA analysed accuracy and inference time for both ensemble and non-ensemble models to suit resource availability and performance requirements. Also, the team used caching of images for faster training.
g. Team kubapok proposed an efficient approach that enables the generation of three models with different training settings while keeping the GPU time equivalent to three separate training runs. This efficiency is achieved by initializing the weights of two models based on a pre-trained model, rather than using random initialization.

The analysis recommends considering dataset characteristics and country-specific requirements when proposing new models in the road damage detection domain. Exploring ensemble methods for improved accuracy should consider the trade-off with inference time. Further, balancing speed, accuracy, and resource requirements in model and ensemble selection is crucial for optimization. By taking these recommendations into account, the researchers can optimize their models for road damage detection tasks.

3.3. *Country-specific inputs*

Teams in the CRDDC'2022 competition considered various country-specific factors when developing object detection models for road damage detection. Differences in performance were observed between countries, with certain models performing better for specific countries. Specific observations for countries include the need to address challenges like loose gravel in Indian roads. The loose gravel creates a misleading impression of alligator or longitudinal cracks in the images. The movement of vehicles may further contribute to the formation of gravel patterns resembling cracks. Likewise, the tiles along the roadside in the Czech Republic create a deceptive appearance of alligator cracks in the road images. Example images for both the cases are provided in Fig. 3. Further, Norway presented unique challenges with high-resolution and wide-angled images. The teams addressed these challenges with strategies such as image patching, customized anchor boxes, attention modules, and ensemble models with different image sizes and architectures.

Additionally, identifying the data imbalance in the individual country's dataset, some teams proposed using data from multiple countries to develop country-specific models. While some teams employed an ensemble of country-specific models along with a model trained on data from all six countries, others opted to explicitly combine the data or create a single model for multiple countries. For example, Team MILA combined data from Japan and the United States, leveraging the high-quality data from both countries. This approach, coupled with additional D40 augmentation, addressed class imbalance (particularly in the United States), improved data generalization and resulted in a strong model for both countries. However, the team found it difficult to improve the F1 score on Indian data by including images from other countries, due to several factors, including poor image quality due to environmental factors, low representation of D10 instances (~1.04 % of the total data), and a lack of similarity in scene and background compared to other countries.

In contrast, Team MDPT proposed to ensemble the best models from Japan and India, identifying the similar sets of images for the two countries. Additionally, the team also observed that the United States and Japan achieved higher accuracy due to the larger number of positive images, with the United States benefiting from the presence of easily detectable longitudinal cracks. These recommendations and insights can help optimize the development of effective road damage detection models.

(a) (b)

Fig. 3. Country-specific patterns resembling cracks resulting in false predictions by the trained deep learning-based models (a) Loose gravel in India (b) Roadside tiles in Czech Republic.

Table 8
Comparison of winning solutions of the three Road Damage Detection challenges (2018, 2020, 2022).

Factors/ Challenge	RDDC'2018 [45]	GRDDC'2020 [22]	CRDDC'2022 [24]
Countries considered	Japan	India, Japan, Czech Republic	India, Japan, Czech Republic, Norway, United States, China
Damage Categories	8	4	4
#Teams	59	121	90+
Winning Team	CMBC Challengers [98]	IMSC [97]	ShiYu_SeaView [27]
Best F1-score achieved	0.659 (for Japan)	0.670 (for 3 countries combined, average for test1 and test2)	0.780 (for Japan),0.769 (for 6 countries combined)
Best performing model	Ensemble of Faster-RCNN and SSD	YOLOv5-based ensemble model	Ensemble of Faster RCNN-series and YOLO-series models

4. Discussion and impact of the study

The data cup challenges have served as crucial milestones in advancing the Road Damage Detection field by offering a comprehensive status quo, contributing to a deeper understanding of the domain over the years. Brief comparison of winning solutions and models of the three road damage detection data cup challenges organized in 2018 [45], 2020 [22], and 2022 [24], is presented in Table 8. The impact of current study lies in providing insights into dataset analysis, context-based effective solutions, ensemble methods, customization, and optimization techniques for road damage detection, based on the winning strategies in CRDDC'2022. These factors are summarized as follows:

(i) **Dataset Analysis**: The study emphasizes the importance of analyzing the underlying dataset in road damage detection. By identifying potential issues such as data imbalance, data sparsity, and domain shift, researchers can gain insights into the characteristics and challenges of the data. This analysis helps in understanding the limitations and requirements of the dataset, leading to more effective model development.

(ii) **Utilizing multi-country data**: The analysis of CRDDC winners' solutions shows that incorporating data from multiple countries improves model generalization and performance. Ensembling country-specific models with an overall model enhances the F1 score, leveraging the rich and diverse training data.

(iii) **Context-based Solutions**: The teams in the study were able to develop effective solutions by considering dataset characteristics and country-specific requirements. This highlights the need to tailor models and techniques to the specific context of road damage detection in different countries. By addressing these specific requirements, the models can achieve better performance and accuracy.

(iv) **Ensemble Methods**: The analysis recommends exploring ensemble methods to improve accuracy. However, it also highlights the trade-off with inference time. Researchers need to carefully consider the balance between accuracy and computational efficiency when using ensemble models. This consideration ensures that the models are both accurate and practical for real-world implementation.

(v) **Customization and Optimization**: Customizing model architecture, incorporating attention mechanisms, and selecting appropriate input image size and shape are highlighted as important factors in enhancing performance. Balancing speed, accuracy, and resource requirements is crucial when selecting models and ensembles. Optimization of these factors ensures that the developed models are efficient, accurate, and suitable for real-time road damage detection applications.

(vi) **Recommendations for Researchers**: By considering the dataset analysis, country-specific requirements, ensemble methods, customization, and optimization techniques discussed in the study, researchers can optimize their models for road damage detection tasks. Further, unifying the data acquisition process, improving the quality of image captured and following standard procedures for preparing the datasets including view captured and annotations etc. can help in mitigating the impact of diverse national road environments on detection accuracy.

Overall, the study's insights on dataset selection, ensemble modeling, and addressing specific challenges contribute to the development of more accurate and efficient models. The findings serve as a valuable resource, guiding future research endeavours in similar domains.

5. Conclusion

The manuscript offers a comprehensive analysis of solutions proposed by Crowdsensing-based Road Damage Detection Challenge (CRDDC) winners, focusing on approaches tailored to road damage detection using multi-view images. The investigation has emphasized the impact of data analysis, model selection, and the incorporation of inter-country data. One significant contribution is the demonstration of how data from different countries can improve the performance of road damage detection models. This finding suggests the potential for future collaborations using techniques like Federated Learning, where countries can share model parameters instead of raw data, thereby enhancing accuracy while respecting privacy.

The manuscript serves as a valuable resource for researchers in the

road damage detection (RDD) domain or similar fields, providing multidimensional insights and highlighting avenues for further research and development. Further the scope of the study is linked with CRDDC, which included the application of state-of-the-art deep learning approaches in RDD, achieving a maximum F1 score of ~ 77 % for detecting and classifying road cracks and potholes in 6 countries: India, Japan, Czech Republic, Norway, United States, and China. The study's scope can be extended to include data from other countries, different damage categories, and alternative annotation methods. Furthermore, the research can be expanded to cover other infrastructures such as buildings and bridges. By considering the insights and recommendations from this study, researchers in the RDD domain can optimize their approaches and contribute to advancements in road damage detection technology, ultimately fostering safer transportation systems globally.

CRediT authorship contribution statement

Deeksha Arya: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. **Hiroya Maeda:** Funding acquisition, Data curation. **Yoshihide Sekimoto:** Resources, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Details for CRDDC (dataset, participants, ranking etc.) which forms the basis for current manuscript maybe accessed at https://crddc2022.sekilab.global/.

Acknowledgments

We thank the participants and our co-organizers of the challenge for their contributions. We also thank the sponsor UrbanX Technologies for providing the requisite funds. Additionally, the research is partially sponsored by the Japan Society of Civil Engineers.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used AI-assisted technologies to rephrase some portions of the text to enhance the readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

[1] World Bank. Transport, 2020. Retrieved from. https://www.worldbank.org/en/topic/transport.
[2] World Health Organization. Road traffic injuries, 2021. Retrieved from https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.
[3] M. Gohar, M. Muzammal, A.U. Rahman, SMART TSS: Defining transportation system behavior using big data analytics in smart cities, Sustain. Cities Soc. 41 (2018) 114–119, https://doi.org/10.1016/j.scs.2018.05.008.
[4] X. Chen, S. Yongchareon, M. Knoche, A review on computer vision and machine learning techniques for automated road surface defect and distress detection, Journal of Smart Cities and Society Preprint (2023) 1–17.
[5] S. Halder, K. Afsari, Robots in inspection and monitoring of buildings and infrastructure: A systematic review, Appl. Sci. 13 (4) (2023) 2304.
[6] R. Fan, S. Guo, L. Wang, M. Junaid Bocus, Computer-aided Road inspection: Systems and algorithms, in: Recent Advances in Computer Vision Applications Using Parallel Processing, Springer International Publishing, Cham, 2023, pp. 13–39.
[7] H. Wu, L. Yao, Z. Xu, Y. Li, X. Ao, Q. Chen, B. Meng, Road pothole extraction and safety evaluation by integration of point cloud and images derived from mobile mapping sensors, Adv. Eng. Inf. 42 (2019) 100936.
[8] J. Dong, W. Meng, Y. Liu, J. Ti, A framework of pavement management system based on IoT and big data, Adv. Eng. Inf. 47 (2021) 101226.
[9] P. Mattes R. Richter J. Döllner, Detecting Road Damages in Mobile Mapping Point Clouds using Competitive Reconstruction Networks, in: Proceedings of the 26th AGILE Conference on Geographic Information Science, 2023. AGILE: GIScience Series, 4, 7, 2023. Doi: 10.5194/agile-giss-4-7-2023.
[10] C. Liu, Y. Du, G. Yue, Y. Li, D. Wu, F. Li, Advances in automatic identification of road subsurface distress using ground penetrating radar: State of the art and future trends, Autom. Constr. 158 (2024) 105185.
[11] D. Arya, H. Maeda, S.K. Ghosh, D. Toshniwal, A. Mraz, T. Kashiyama, Y. Sekimoto, Deep learning-based road damage detection and classification for multiple countries, Autom. Constr. 132 (2021) 103935, https://doi.org/10.1016/j.autcon.2021.103935.
[12] J. Guo, P. Liu, B. Xiao, L. Deng, Q. Wang, Surface defect detection of civil structures using images: Review from data perspective, Autom. Constr. 158 (2024) 105186.
[13] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, H. Omata, Road damage detection and classification using deep neural networks with smartphone images, Comput. Aided Civ. Inf. Eng. 33 (12) (2018) 1127–1141.
[14] Q. Mei, M. Gül, A cost effective solution for pavement crack inspection using cameras and deep neural networks, Constr. Build. Mater. 256 (2020) 119397.
[15] H. Gong, J. Tešić, J. Tao, X. Luo, F. Wang, Automated Pavement Crack Detection with Deep Learning Methods: What Are the Main Factors and How to Improve the Performance? Transp. Res. Rec. (2023), 03611981231161358.
[16] M. Everingham, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338.
[17] M. Everingham, S.A. Eslami, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes challenge: A retrospective, Int. J. Comput. Vis. 111 (1) (2015) 98–136.
[18] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, Vision meets robotics: The kitti dataset, The International Journal of Robotics Research 32 (11) (2013) 1231–1237, https://doi.org/10.1177/0278364913491297.
[19] T.Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, Microsoft coco: Common objects in context, in: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, Springer International Publishing, 2014, pp. 740–755, https://doi.org/10.1007/978-3-319-10602-1_48.
[20] K. Tong, Y. Wu, Rethinking PASCAL-VOC and MS-COCO dataset for small object detection, J. Vis. Commun. Image Represent. 93 (2023) 103830, https://doi.org/10.1016/j.jvcir.2023.103830.
[21] O. Vinyals, A. Toshev, S. Bengio, D. Erhan, Show and tell: Lessons learned from the 2015 mscoco image captioning challenge, IEEE Trans. Pattern Anal. Mach. Intell. 39 (4) (2016) 652–663, https://doi.org/10.1109/TPAMI.2016.2587640.
[22] D. Arya, H. Maeda, S.K. Ghosh, D. Toshniwal, H. Omata, T. Kashiyama, Y. Sekimoto. Global Road Damage Detection: State-of-the-art Solutions, IEEE International Conference on Big Data (Big Data, Atlanta, GA, USA, 2020, pp. 5533–5539, https://doi.org/10.1109/BigData50022.2020.9377790.
[23] A. Behzadian, T.W. Muturi, T. Zhang, H. Kim, A. Mullins, Y. Lu, et al., The 1st Data Science for Pavements Challenge, 2022. arXiv preprint arXiv:2206.04874.
[24] D. Arya, H. Maeda, S.K. Ghosh, D. Toshniwal, H. Omata, T. Kashiyama, Y. Sekimoto, Crowdsensing-based Road Damage Detection Challenge (CRDDC-2022), IEEE International Conference on Big Data (Big Data), Osaka, Japan (2022) 6378–6386, https://doi.org/10.1109/BigData55660.2022.10021040.
[25] D. Arya, H. Maeda, S.K. Ghosh, D. Toshniwal, Y. Sekimoto, RDD2022: A multi-national image dataset for automatic Road Damage Detection, 2022b. arXiv preprint arXiv:2209.08538.
[26] D. Arya, H. Maeda, Y. Sekimoto, H. Omata, S.K. Ghosh, D. Toshniwal, et al., RDD2022-The multi-national Road Damage Dataset released through CRDDC'2022, Figshare (2022), https://doi.org/10.6084/m9.figshare.21431547.v1.
[27] W. Ding, X. Zhao, B. Zhu, Y. Du, G. Zhu, T. Yu, et al., An Ensemble of One-Stage and Two-Stage Detectors Approach for Road Damage Detection, in: 2022 IEEE International Conference on Big Data (Big Data), IEEE, 2022, pp. 6395-6400. 10.1109/BigData55660.2022.10021000.
[28] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788.
[29] C.Y. Wang, A. Bochkovskiy, H.Y.M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in: In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7464–7475.
[30] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, Adv. Neural Inf. Proces. Syst. 28 (2015), https://doi.org/10.5555/2969239.2969250.
[31] D. Jeong, J. Kim, Road Damage Detection using YOLO with Image Tiling about Multi-source Images, in: 2022 IEEE International Conference on Big Data (Big Data), IEEE, 2022, pp. 6401–6406, https://doi.org/10.1109/BigData55660.2022.10020282.
[32] V. Pham, D. Nguyen, C. Donan, Road Damages Detection and Classification with YOLOv7, in: 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 2022, pp. 6416–6423, https://doi.org/10.1109/BigData55660.2022.10020856.
[33] S. Wang, Y. Tang, X. Liao, J. He, H. Feng, H. Jiao, Q. Yuan, An Ensemble Learning Approach with Multi-Depth Attention Mechanism for Road Damage Detection, in:

2022 IEEE International Conference on Big Data (Big Data), IEEE, 2022, pp. 6439–6444.
[34] A.M. Okran, M. Abdel-Nasser, H.A. Rashwan, D. Puig, Effective Deep Learning-Based Ensemble Model for Road Crack Detection, in: 2022 IEEE International Conference on Big Data (Big Data), IEEE, 2022, pp. 6407–6415, https://doi.org/10.1109/BigData55660.2022.10020790.
[35] M. Bhavsar, A. Alfarrarjeh, U. Baranwal, S.H. Kim, Country-specific Ensemble Learning: A Deep Learning Approach for Road Damage Detection, in: 2022 IEEE International Conference on Big Data (Big Data), IEEE, 2022, pp. 6387–6394, https://doi.org/10.1109/BigData55660.2022.10020799.
[36] S. Han, X. Haowen, Y. Jiajing, Y. Huan, J. Guoping, Z. Yingjiang, Team NJUPT Submission for Crowdsensing-based Road Damage Detection Challenge (CRDDC'2022), 2022, https://github.com/KentHan19980609/T22_034_CRDDC_2022_SourceCode (Last accessed – 06/06/2023).
[37] L. Yang, H. He, T. Liu, Road Damage Detection and Classification Based on Multi-Scale Contextual Features, in: 2022 IEEE International Conference on Big Data (Big Data), IEEE, 2022, pp. 6445–6453.
[38] P.K. Saha, Y. Sekimoto, Road Damage Detection for Multiple Countries, in: 2022 IEEE International Conference on Big Data (Big Data), IEEE, 2022, pp. 6431–6438.
[39] D.Q. Tran, Team SIAI Submission for Crowdsensing-based Road Damage Detection Challenge (CRDDC'2022), 2022, https://github.com/daitranskku/CRDDC_2022_Code (last accessed – 06/06/2023).
[40] J. Pokrywka, in: December). Efficient GPU Training of a Diversified Model Ensemble for the Crowdsensing-Based Road Damage Detection Challenge (CRDDC2022), IEEE, 2022, pp. 6424–6430, https://doi.org/10.1109/BigData55660.2022.10020877.
[41] G. Yang, K. Liu, J. Zhang, B. Zhao, Z. Zhao, X. Chen, B.M. Chen, Datasets and processing methods for boosting visual inspection of civil infrastructure: A comprehensive review and algorithm comparison for crack classification, segmentation, and detection, Constr. Build. Mater. 356 (2022) 129226, https://doi.org/10.1016/j.conbuildmat.2022.129226.
[42] C. Lin, D. Tian, X. Duan, J. Zhou, D. Zhao, D. Cao, DA-RDD: Toward Domain Adaptive Road Damage Detection Across Different Countries, IEEE Trans. Intell. Transp. Syst. (2022), https://doi.org/10.1109/TITS.2022.3221067.
[43] E. Ranyal, A. Sadhu, K. Jain, Road condition monitoring using smart sensing and artificial intelligence: A review, Sensors 22 (8) (2022) 3044.
[44] M. Ren, X. Zhang, X. Chen, B. Zhou, Z. Feng, YOLOv5s-M: A deep learning network model for road pavement damage detection from urban street-view imagery, Int. J. Appl. Earth Obs. Geoinf. 120 (2023) 103335, https://doi.org/10.1016/j.jag.2023.103335.
[45] H. Maeda, T. Kashiyama, Y. Sekimoto, T. Seto, H. Omata, Generative adversarial network for road damage detection, Comput. Aided Civ. Inf. Eng. (2020).
[46] D. Arya, H. Maeda, S.K. Ghosh, D. Toshniwal, H. Omata, T. Kashiyama, T. Seto, A. Mraz, Y. Sekimoto, RDD2020: An Image Dataset for Smartphone-based Road Damage Detection and Classification, Mendeley Data (2021), https://doi.org/10.17632/5ty2wb6gvg.2.
[47] W. Tang, S. Huang, Q. Zhao, R. Li, L. Huangfu, An iteratively optimized patch label inference network for automatic pavement distress detection, IEEE Trans. Intell. Transp. Syst. 23 (7) (2021) 8652–8661, https://doi.org/10.1109/TITS.2021.3084809.
[48] H. Liu, C. Yang, A. Li, S. Huang, X. Feng, Z. Ruan, Y. Ge, Deep Domain Adaptation for Pavement Crack Detection, IEEE Trans. Intell. Transp. Syst. (2022), https://doi.org/10.1109/TITS.2022.3225212.
[49] M. Eisenbach, R. Stricker, D. Seichter, K. Amende, K. Debes, M. Sesselmann, D. Ebersbach, U. Stoeckert, H.M. Gross, How to get pavement distress detection ready for deep learning? A systematic approach, Proceedings of the International Joint Conference on Neural Networks, 2017-May, 2017, 2039–2047. Doi: 10.1109/IJCNN.2017.7966101.
[50] R. Stricker, M. Eisenbach, M. Sesselmann, K. Debes, H.M. Gross, Improving Visual Road Condition Assessment by Extensive Experiments on the Extended GAPs Dataset, Proceedings of the International Joint Conference on Neural Networks, 2019-July, 2019. Doi: 10.1109/IJCNN.2019.8852257.
[51] A. Angulo, J.A. Vega-Fernández, L.M. Aguilar-Lobo, S. Natraj, G. Ochoa-Ruiz, Road Damage Detection Acquisition System Based on Deep Neural Networks for Physical Asset Management, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019, https://doi.org/10.1007/978-3-030-33749-0_1.
[52] R. Roberts, G. Giancontieri, L. Inzerillo, G. Di Mino, Towards low-cost pavement condition health monitoring and analysis using deep learning, Applied Sciences (Switzerland) 10 (1) (2020), https://doi.org/10.3390/APP10010319.
[53] D. Arya, H. Maeda, S.K. Ghosh, D. Toshniwal, Y. Sekimoto, RDD2020: An annotated image dataset for automatic road damage detection using deep learning, Data in Brief 36 (2021) 107133, https://doi.org/10.1016/j.dib.2021.107133.
[54] H. Majidifard, P. Jin, Y. Adu-Gyamfi, W.G. Buttlar, Pavement Image Datasets: A New Benchmark Dataset to Classify and Densify Pavement Distresses, Transp. Res. Rec. 2674 (2) (2020) 328–339, https://doi.org/10.1177/0361198120907283.
[55] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei, H. Ling, Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection, IEEE Trans. Intell. Transp. Syst. 21 (4) (2020) 1525–1535, https://doi.org/10.1109/TITS.2019.2910595.
[56] B.T. Passos, M. Cassaniga, A.M.R. Fernandes, K.B. Medeiros, E. Comunello, Cracks and Potholes in Road Images, Mendeley Data V4 (2020), https://doi.org/10.17632/t576ydh9v8.4.
[57] M.T. Cao, Q.V. Tran, N.M. Nguyen, K.T. Chang, Survey on performance of deep learning models for detecting road damages using multiple dashcam image resources, Adv. Eng. Inf. 46 (2020) 101182.
[58] G. Guo, Z. Zhang, Road damage detection algorithm for improved YOLOv5, Sci. Rep. 12 (1) (2022) 1–12, https://doi.org/10.1038/s41598-022-19674-8.
[59] M.H. Guo, T.X. Xu, J.J. Liu, Z.N. Liu, P.T. Jiang, T.J. Mu, S.M. Hu, Attention mechanisms in computer vision: A survey, Computational Visual Media 8 (3) (2022) 331–368.
[60] M.M. Hasan, S. Sakib, K. Deb, Road Damage Detection and Classification Using Deep Neural Network, in: 2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), IEEE, 2022, pp. 1–6.
[61] Z.S. Hernanda, R.W. Sudibyo, CNN-Based Hyperparameter Optimization Approach for Road Pothole and Crack Detection Systems, in: In 2022 IEEE World AI IoT Congress (AIIoT), IEEE, 2022, pp. 538–543, https://doi.org/10.1109/AIIoT54504.2022.9817316.
[62] X. Bi, S. Zhang, Y. Zhang, L. Hu, W. Zhang, W. Niu, G. Wang, October). CASA-Net: A Context-Aware Correlation Convolutional Network for Scale-Adaptive Crack Detection, in: In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022, pp. 67–76, https://doi.org/10.1145/3511808.3557252.
[63] Z. Lin, H. Wang, S. Li, Pavement anomaly detection based on transformer and self-supervised learning, Autom. Constr. 143 (2022) 104544, https://doi.org/10.1016/j.autcon.2022.104544.
[64] X. Xiang, Z. Wang, Y. Qiao, An improved YOLOv5 crack detection method combined with transformer, IEEE Sens. J. 22 (14) (2022) 14328–14335, https://doi.org/10.1109/JSEN.2022.3181003.
[65] D. Deepa, A. Sivasangari, ESSR-GAN: Enhanced super and semi supervised remora resolution based generative adversarial learning framework model for smartphone based road damage detection, Multimed. Tools Appl. (2023) 1–31.
[66] G. Zhang Z. Du P. Wu X. Zhang W. Wang Z. Wang, A crack detection network based on deformable convolution and test time augmentation, in: International Conference on Computer, Artificial Intelligence, and Control Engineering (CAICE 2022), SPIE, Vol. 12288, 2022a, pp. 46-51.
[67] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, et al., Attention is all you need, Adv. Neural Inform. Process. Syst. 30 (2017).
[68] H. Zhang, Z. Wu, Y. Qiu, X. Zhai, Z. Wang, P. Xu, N. Jiang, A New Road Damage Detection Baseline with Attention Learning, Appl. Sci. 12 (15) (2022) 7594, https://doi.org/10.3390/app12157594.
[69] A.M. Roy, J. Bhaduri, DenseSPH-YOLOv5: An automated damage detection model based on DenseNet and Swin-Transformer prediction head-enabled YOLOv5 with attention mechanism, Adv. Eng. Inf. 56 (2023) 102007.
[70] G. Yu, X. Zhou, An Improved YOLOv5 Crack Detection Method Combined with a Bottleneck Transformer, Mathematics 11 (10) (2023) 2377.
[71] C. Zhang, G. Li, Z. Zhang, R. Shao, M. Li, D. Han, M. Zhou, AAL-Net: A Lightweight Detection Method for Road Surface Defects Based on Attention and Data Augmentation, Appl. Sci. 13 (3) (2023) 1435.
[72] Y. Zhu, S. Zhang, C. Ruan, CCN: Pavement Crack Detection with Context Contrasted Net, in: Neural Information Processing: 29th International Conference, ICONIP 2022, Virtual Event, November 22–26, 2022, Proceedings, Part III, Springer International Publishing, Cham, 2023, pp. 85–96.
[73] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep learning for generic object detection: A survey, Int. J. Comput. Vis. 128 (2020) 261–318.
[74] A. Bochkovskiy, C.Y. Wang, H.Y. Liao, YOLOv4: optimal speed and accuracy of object detection, 2020. arXiv preprint arXiv:2004.10934.
[75] T. Diwan, G. Anirudh, J.V. Tembhurne, Object detection using YOLO: Challenges, architectural successors, datasets and applications, Multimed. Tools Appl. 82 (6) (2023) 9243–9275, https://doi.org/10.1007/s11042-022-13644-y.
[76] R. Girshick, Fast r-cnn, in: In Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448, https://doi.org/10.1109/ICCV.2015.169.
[77] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: In Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969.
[78] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, in: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, Springer International Publishing, 2016, pp. 21–37.
[79] C.H. Song, H.J. Han, Y. Avrithis, All the attention you need: Global-local, spatial-channel attention for image retrieval, in: In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 2754–2763.
[80] J. Wu, X. Liu, J. Dong, Strategies for inserting attention in computer vision, Multimed. Tools Appl. (2023) 1–18.
[81] A.A. Bastidas, H. Tang, Channel attention networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.
[82] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-Net: Efficient channel attention for deep convolutional neural networks, in: In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11534–11542.
[83] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[84] X. Jin, Y. Xie, X.S. Wei, B.R. Zhao, Z.M. Chen, X. Tan, Delving deep into spatial pooling for squeeze-and-excitation networks, Pattern Recogn. 121 (2022) 108159.
[85] Q. Hou, D. Zhou, J. Feng, Coordinate attention for efficient mobile network design, in: In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 13713–13722.
[86] X. Chu, Z. Tian, Y. Wang, B. Zhang, H. Ren, X. Wei, C. Shen, Twins: Revisiting the design of spatial attention in vision transformers, Adv. Neural Inf. Proces. Syst. 34 (2021) 9355–9366.

[87] X. Zhu, D. Cheng, Z. Zhang, S. Lin, J. Dai, An empirical study of spatial attention mechanisms in deep networks, in: In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6688–6697.
[88] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2020. arXiv preprint arXiv:2010.11929.
[89] M. Raghu, T. Unterthiner, S. Kornblith, C. Zhang, A. Dosovitskiy, Do vision transformers see like convolutional neural networks? Adv. Neural Inf. Proces. Syst. 34 (2021) 12116–12128.
[90] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022.
[91] H. Zhang, K. Zu, J. Lu, Y. Zou, D. Meng, EPSANet: An efficient pyramid squeeze attention block on convolutional neural network, in: In Proceedings of the Asian Conference on Computer Vision, 2022, pp. 1161–1177.
[92] L.C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, 2017a. arXiv preprint arXiv:1706.05587.
[93] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2017) 834–848.
[94] D.C. Lepcha, B. Goyal, A. Dogra, K.P. Sharma, D.N. Gupta, A Deep Journey into Image Enhancement: A Survey of Current and Emerging Trends, Information Fusion (2022), https://doi.org/10.1016/j.inffus.2022.12.012.
[95] F. Kluger, C. Reinders, K. Raetz, P. Schelske, B. Wandt, H. Ackermann, B. Rosenhahn, Region-Based Cycle-Consistent Data Augmentation for Object Detection, in: 2018 IEEE International Conference on Big Data (Big Data), IEEE, 2018, pp. 5205–5211.
[96] "YOLOv5," 2020. [Online]. Available: https://github.com/ultralytics/yolov5.
[97] V. Hegde, D. Trivedi, A. Alfarrarjeh, A. Deepak, S.H. Kim, C. Shahabi, Yet another deep learning approach for road damage detection using ensemble learning, in: 2020 IEEE International Conference on Big Data (Big Data), IEEE, 2020, pp. 5553–5558, https://doi.org/10.1109/BigData50022.2020.9377833.
[98] Y.J. Wang, M. Ding, S. Kan, S. Zhang, C. Lu, Deep Proposal and Detection Networks for Road Damage Detection and Classification2018 IEEE International Conference on Big Data (Big Data), IEEE, 2018, pp. 5224–5227.