Multiple data sources and domain generalization learning method for road surface defect classification

Linh Trinh, Ali Anwar, Siegfried Mercelis
Faculty of Applied Engineering, IDLab, University of Antwerp-imec, Belgium
{linh.trinh, ali.anwar, siegfried.mercelis}@uantwerpen.be

Abstract-Roads are an essential mode of transportation, and maintaining them is critical to economic growth and citizen well-being. With the continued advancement of AI, road surface inspection based on camera images has recently been extensively researched and can be performed automatically. However, because almost all of the deep learning methods for detecting road surface defects were optimized for a specific dataset, they are difficult to apply to a new, previously unseen dataset. Furthermore, there is a lack of research on training an efficient model using multiple data sources. In this paper, we propose a method for classifying road surface defects using camera images. In our method, we propose a scheme for dealing with the invariance of multiple data sources while training a model on multiple data sources. Furthermore, we present a domain generalization training algorithm for developing a generalized model that can work with new, completely unseen data sources without requiring model updates. We validate our method using an experiment with six data sources corresponding to six countries from the RDD2022 dataset. The results show that our method can efficiently classify road surface defects on previously unseen data.

I. INTRODUCTION

Roads play a crucial role in transportation. Research domains especially road engineering study the problem of road surface condition inspection for road maintenance. Periodic road maintenance is an important task, cities and municipalities allocate large sums of budget to repair the defective parts and therefore require automated road defects measurements for cost-effective road maintenance. Multiple criteria exist for evaluating visual surface quality, including the Pavement Condition Index (PCI) [1], Pavement Surface Condition Index (PSCI) [2], and Pavement Surface Evaluation Rating (PASER) [3]. These standards provide measurement instructions for pavement characteristics used in pavement assessments. Nowadays, with the growth and computational support of AI and deep learning solutions, road maintenance can be performed sufficiently accurately and automatically using data collected on the road (with the supporting of compliance of regulations [4], [5]). More specifically, a set of sensors such as a camera, lidar, radar, or GPS can be mounted on a vehicle and driven along the road to collect data, after which a developed machine learning model is used to determine the road surface conditions for assessment. With the focus on these tasks, there has been a strong emphasis on a large data-driven approach to building them. In addition, various recent studies on datasets and methods from a variety of countries have recently been

announced, providing visual data for road surface condition assessments using AI/deep learning methods, such as LTPP-FHWA in the USA [6], GAPs in Germany [7], CFD from China [8], and RDD in China, Japan, India, the Czech Republic, Norway, and United States [9], [10], [11]. This provides a strong evidence that it is an appealing topic for research on transportation.

There has been a growing interest in studying road surface inspection within computer vision community in recent years. Several recent approaches have been proposed in the field of road surface classification. The majority of the proposed methods rely on the architecture of Convolutional Neural Networks (CNN), as exemplified by [12], [13], [14], [15], [16]. These methods involve proposing the architecture through the design or modification of neural network structures. Alternative approaches utilize a pretrained backbone on ImageNet dataset, such as ResNet50, InceptionNetV3 [17], or ConvNeXt [18] to construct the classification. Another approach involves integrating Convolutional Neural Networks (CNN) with other neural network architectures, such as Transformer [19] and Long Short-Term Memory (LSTM) [20]. Alternatively, certain approaches incorporate data fusion or feature fusion techniques to improve the classification task, as demonstrated in studies such as [18], [21], [19]. In the classification of road conditions, several approaches rely on Convolutional Neural Networks (CNN) to construct classifiers, as demonstrated in the works of [22], [23], [24], [25]. The approach of constructing a classifier using a pretrained backbone network has been extensively researched, with notable examples including VGG16, MobileNetV2 [26], [27], [28], ResNet18 [29], and ResNet101 [30]. Several other efforts targeted the improvement of the inspection of road surface conditions by incorporating additional signals such as wheel-road interactions [31], accelerometers [32], orthoframes [30].

However, existing road surface defect classification methods are designed to train models on just one source of data. When data source invariance occurs, these methods may not be applicable to leveraging multiple different data sources for training purposes. Furthermore, because those methods were not trained on this dataset, they may be inefficient when applied to new, previously unseen data sets. In this paper, we take on the context in which multiple data sources are available and how to train a model efficiently on these data sources. Furthermore, we consider the issue of the model's

low accuracy on the new, unseen dataset. So we present a method for overcoming both of these challenges.

In summary, our main contributions are as follows:

- We present our method for classifying road surface defects based on camera images. Our method aims to address two issues: resolving invariance when leveraging multiple data sources for training models, and generalizing the model to work with unseen datasets without training the model on them. Our method consists of two parts: contrastive learning-based training and a new algorithm for domain generalization training.
- We conduct extensive experiments to evaluate our method using six data sources from the RDD2022 public road surface defect dataset. The experiment compares the results of our method to recent studies on classification for road surface defect classification, demonstrating that our method is efficient and wellsuited to the unseen dataset.

We organize the rest of the paper as follows. We present our literature studies on the related works in Section II. Then, in Section III, we present our methodology. Section IV discusses our experiment and results. Finally, we conclude our work in Section V.

II. RELATED WORKS

Several previous studies have focused on deep learningbased methodologies for classifying road surface defects. Li et al. [33] proposed multiple-stage networks composed of CNNs for a classifier model. This method has been verified on a private dataset containing four defect classes: alligator, block, edge, logitudinal, and transverse and reflection cracks. Nhat et al. [34] proposed a framework for determining pavement fatigue severity using decision trees, deep neural networks, and CNN. The model was tested on a private dataset containing alligator, block, and minor fatigue area cracks. More recently, Mai et al. (2024) proposed a multilabel convolutional neural network to classify asphalt distress. The model was validated using a private dataset that included four distress classes: alligator, block, longitudinal/transverse cracks, and pothole. According to [35]'s literature review, the most commonly used classifiers for transportation are VGG19, VGG16, ResNet50, DenseNet121, and generic CNN. Chen et al. [36] propose a combination of MobileNetV3-large and Convolutional Block Attention Module for crack classification. This study also used focal loss for training. The model was tested on the surface defect of the bridge road. Silva et al. [37] propose a transfer learning approach based on VGG16 to classify cracks and non-cracks on concrete roads. Dais et al. [38] propose a transfer learning approach using VGG16, MobileNet, and ResNet for non-crack and crack classification. Ali et al. [39] proposed a customized network based on CNN for crack classification of concrete roads. The model was evaluated against VGG16, VGG19, ResNet50, and InceptionV3. Liu et al. [40] proposed a deep learning method and utilized infrared thermography for crack classification in asphalt pavement. Models included MobileNet, DenseNet, Efficient-Net, and ResNet. Dong et al. [41] propose a deep learning method based on metric learning for multi-target few shot distress classification. Hsieh et al. [42] used ConvNets based on ResNet50 to classify cracks in concrete pavement using 3D images. Various works also validated their research on privately datasets, such as Gagliardi et al. [31] propose a method to evaluate pavement quality of road infrastructure using acoustic data of wheel-road interaction. This model is tested on a privately collected dataset with four categories: silence, unknown, good quality, and ruined. Rateke et al. [22] propose a CNN network architecture to classify images based on road surface condition. Hasanaath et al. [27] propose a transfer learning method that fine-tunes a classification model using a pre-trained backbone such as VGG16 or MobileNetV2. This model was tested on a private dataset collected in the Kingdom of Saudi Arabia, with four classifications: good, medium, bad, and unpaved. Chaudhary et al. [28] propose a classification model based on pre-trained MobileNetV2 with supplementary layers. The evaluation of an RTK dataset with three main categories: good, normal, and poor quality. Andrades et al. [32] propose a classification model based on the self-organizing map (SOM) algorithm. This method, which was tested on private data, consists of several sensors, including an accelerometer and a microcontroller. Riid et al. [30] propose a classification method based on ConvNets and ResNet101 to detect pavement distress using orthoframes collected by a mobile mapping system in Italy. Stricker et al. [43] use pretrained ResNet34 to improve road condition assessment through transfer learning. This model was tested on privately collected data from Germany, with six classes of road conditions. Hsieh et al. [29] developed a transfer learning-based classification model using ResNet18 to detect pavement raveling. The model was tested on a private dataset collected in the United States. Sajad et al. [23] developed a transfer learning-based classification model for pavement bleeding inspection. The method was tested using a private dataset collected in Iran. Gopalakrishnan et al. [26] developed a Deep Neural Networks classification using VGG-16 to identify pavement cracks in a dataset collected in the United States. Nguyen et al. [24] and Nhat et al. [25] created a CNN architecture to detect pavement cracks on the Vietnamese roads. Li et al. [44] created a customized CNN model to detect pavement cracks in China.

To summarize, transfer learning is a common approach of developing classification models that can be tailored to road surface defect classification. However, existing methods were developed for just one source of data, which may be inefficient when training across multiple data sources. Furthermore, these studies are only trained and evaluated on the same dataset, so they cannot be guaranteed to perform well on new data sets. In the following section, we shall explain our methods for tackled both of these challenges.

III. METHOD

We aim to develop a generalized model using n different available data sources D_i , i = 1, ..., n. Denote a collection of

data sources as $\mathbf{D} = \{D_1, D_2, ..., D_n\}$, and \mathbf{x}_{ij} is the j-th image of data source D_i . There are C target classes that each sample of D_i belongs to. We aim to train a generalized classification model that works well on new unseen data sources \mathbf{D}_{unseen} without updating the model on them. We construct a classification model \mathcal{F}_{θ} that consists of a feature extraction $f_e(;\theta_e)$ (θ_e is the parameter of f_e) and a fully connected layer for generating logits, where θ represents the model's learnable parameters of whole network \mathcal{F}_{θ} . Figure

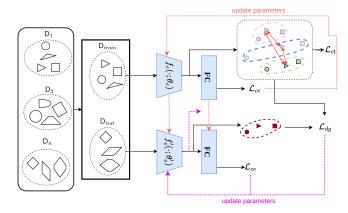


Fig. 1: The diagram illustrates our method for training domain generalization and contrastive learning on multiple data sources.

1 illustrates the workflow of our method. We divided the training data source \mathbf{D} into two parts: \mathbf{D}_{train} and \mathbf{D}_{test} . In our method, we train model \mathcal{F}_{θ} on data \mathbf{D}_{train} while optimizing generalization by prioritizing testing data \mathbf{D}_{test} . To be more specific, after each step of training a batch of \mathbf{D}_{train} , we optimize the generalization process on \mathbf{D}_{test} .

Train model on D_{train} . Training data D_{train} can be composed from multiple data sources, then we apply contrastive learning [45], [46] to enhance model learning on invariance domain via embedding feature extracted by $f_e(;\theta_e)$. Given a pair of samples $(\mathbf{x}_{ki},\mathbf{x}_{hj})$, we denote the extracted semantic embedding feature as \mathbf{p}_{ki} and \mathbf{p}_{hj} , respectively. In training, we apply contrastive loss to each positive pair $(\mathbf{x}_{ki},\mathbf{x}_{hj})$ in each training batch of B_{tr} , as shown in the equation below.

$$\mathcal{L}_{ct}(\mathbf{x}_{ki}, \mathbf{x}_{hj}) = -\log \frac{S(\mathbf{x}_{ki}, \mathbf{x}_{hj})}{\sum_{\mathbf{x}_{lb} \in B_{tr}} I(\mathbf{x}_{ki}, \mathbf{x}_{lb}) \cdot S(\mathbf{x}_{ki}, \mathbf{x}_{hj})}$$
(1)

where I denote an indicator function as below equation:

$$I(\mathbf{x}_{ki}, \mathbf{x}_{lb}) = \begin{cases} 0 & \mathbf{x}_{ki}, \mathbf{x}_{lb} \text{ is same class} \\ 1 & \text{otherwise} \end{cases}$$
 (2)

Rather than directly calculating the similarity between two original images \mathbf{x}_{ki} and \mathbf{x}_{hj} , we use the corresponding embedded feature. For more information, we use the cosine distance between two features to determine the similarity of two vectors. The following equation determines the similarity

of a pair $(\mathbf{x}_{ki}, \mathbf{x}_{hj})$:

$$S(\mathbf{x}_{ki}, \mathbf{x}_{hj}) = \exp\left(\frac{SIM(\mathbf{x}_{ki}, \mathbf{x}_{hj})}{\tau}\right)$$
(3)

where τ is a temperature hyperparameter, and:

$$SIM(\mathbf{x}_{ki}, \mathbf{x}_{hj}) \approx cosine(\mathbf{p}_{ki}, \mathbf{p}_{hj}) = \frac{\mathbf{p}_{ki}^T \times \mathbf{p}_{hj}}{||\mathbf{p}_{ki}|| \times ||\mathbf{p}_{hj}||}$$
(4)

The contrastive loss function is computed for all similar class pairs in batch training for both $(\mathbf{x}_{ki},\mathbf{x}_{hj})$ and $(\mathbf{x}_{hj},\mathbf{x}_{ki})$. Training in this fashion improves the model's ability to explore the invariance of representations, allowing embedding features from the same class to be close together while keeping those from different classes separated.

The training loss on D_{train} is described below:

$$\mathcal{L}_{tr} = \mathcal{L}_{ce} + \lambda \mathcal{L}_{ct} \tag{5}$$

where λ is a hyperparameter for controlling the weight of contrastive loss, and \mathcal{L}_{ce} is the cross-entropy loss for multilabel classification task.

Domain generalization training. We propose a strategy for training generalized models to work with unseen datasets. Our method was inspired by recent domain generalization techniques [47], [48], [49], [50]. The main generalization training of these methods is to align the distribution of features on meta tests with the distribution of features in the meta domain. Unlike most of these techniques, which deal with face recognition and then design metric learning based on triplet loss, we design a new generalization training by optimizing the alignment of feature distribution of testing data source D_{test} to training data source D_{train} , ulitizing Mahalanobis distance from a point to a distribution. For the distribution of the differential feature vector on the training data source D_{train} , we compute the covariance matrix Ξ and the mean vector μ . We compute the covariance matrix on the entire training data D_{train} for each class $c \in C$ using the following equation:

$$\Xi_c = \frac{1}{N} \sum_{i=1}^{N} (z_i - \mu_c) (z_i - \mu_c)^T$$
 (6)

where N is the number of sample pair of class c z_i is a different vector of i-th pair in this N pair, and $z_i = f_e(\mathbf{x}_{pg};\theta_e) - f_e(\mathbf{x}_{qh};\theta_e)$ where training pair $\mathbf{x}_{pg},\mathbf{x}_{qh}$ is the same class c, and μ_c is the mean vector of differential distribution, which is $\mu_c = mean(\mathbf{z})$. After that, with each differential feature vector of testing meta \mathbf{D}_{test} , we compute its distance to the distribution of \mathbf{D}_{train} using the Mahalanobis distance as shown below:

$$\mathcal{M}(\mathbf{z}_i; \Xi_c, \mu_c) = \sqrt{(\mathbf{z}_i - \mu_c)^T \Xi_c(\mathbf{z}_i - \mu_c)} \tag{7}$$

Finally, we use D_{test} to train domain generalization using a loss function as shown in the equation below:

$$\mathcal{L}_{dg}(\mathsf{B}_{te}; \mathcal{F}_{\theta'}) = \frac{1}{C} \sum_{c \in C} \frac{1}{N_c} \sum_{i=1}^{N_c} \mathcal{M}(\mathsf{z}_i; \Xi_c, \mu_c)$$
(8)

where $z_i = f_e(x_{lk}; \theta') - f_e(x_{gh}; \theta')$ with x_{lk} and x_{lk} in B_{te} and same class c; $\mathcal{M}(z_i; \Xi_c, \mu_c)$ is the Mahalanobis distance from a point z_i to distribution $\mathcal{N}(\mu_c, \Xi_c)$.

The algorithm 1 show a detailed description of our method's model training process.

Algorithm 1 Training domain generalization model

```
Input: n data source \mathbf{D} = \{D_1, D_2, ..., D_n\}
     Output: trained model \mathcal{F}_{\theta}
 1: for i = 1, 2, ..., N_{epoch} do
            for Each D_{test} \in \mathbf{D} do
                  \chi \leftarrow None
 3:
                  \mathsf{D}_{train} = \mathbf{D} \setminus \mathsf{D}_{test}
 4:
 5:
                  for Sample batch B_{te} in D_{test} do
                       Compute \Xi, \mu = \{\Xi_c\}, \{\mu_c\}, \forall c \in C \text{ on } \mathsf{D}_{train}
 6:
                      for Sample batch B_{tr} from D_{train} do
 7:
                            \mathcal{L}_{tr} \leftarrow \mathcal{L}_{ct}(\mathsf{B}_{tr}; \mathcal{F}_{\theta}) + \mathcal{L}_{ce}(\mathsf{B}_{tr}; \mathcal{F}_{\theta})
 8:
                            Update weight \mathcal{F}_{\theta'} \leftarrow \mathcal{L}_{tr}.backward()
 9.
10:
                      \begin{aligned} &\mathcal{L}_{te} \leftarrow \mathcal{L}_{ce}(\mathsf{B}_{te}; \mathcal{F}_{\theta'}) + \mathcal{L}_{dg}(\Xi, \mathsf{B}_{te}; \mathcal{F}_{\theta'}) \\ &\chi \leftarrow \chi + \gamma \nabla_{\theta'} \mathcal{L}_{te} \end{aligned}
11:
12:
                  end for
13:
                 Update weight of \mathcal{F}: \theta'' \leftarrow \theta' - \eta \times \chi
14:
15:
             end for
16: end for
17: return optimal \mathcal{F}_{\theta^*}
```

IV. EXPERIMENT

A. Settings

Dataset. In this experiment, we use RDD2022 [51], [52], [53], a public data set for road surface defect study, which data was collected from six countries: China, Japan, Czech, India, Norway, and the United States. We use six data sources for this dataset, which correspond to six countries: China MotorBike, Czech, India, Japan, Norway, and the United States. We crop the bounding boxes from the original dataset based on the annotations provided by the dataset. For consistencies in all six datasets, we retain bounding boxes belonging to four defect classes: longitudinal crack (D00), transverse crack (D10), alligator crack (D20), and a combination of rutting bump, separation cracks, and pothole (D40). We remove small bounding boxes with squared sizes less than 400 *pixel*². The prepared dataset for our experiment is summarized in the table I and are available here¹.

Method settings. Our model for feature extraction uses several backbones for $f_e(;\theta)$, including ResNet18, VGG16, MobileNet-V2, and Inception-V3. These backbones are already pretrained on ImageNet22K. We use four data sources from each training dataset (D_{train}) and the remaining data sources for D_{test} . The learning rate η is 0.001, the weight parameter γ is 0.7, and there are four defect classes (C). The image will be resized to 64x64 for classification purposes. We train the classification models in a batch size of 32

TABLE I: Statistic of preprocessed data from RDD2022 for our experiment.

Data source		SUM				
Data source	D00	D10	D20	D40	SUM	
China MotorBike	1,419	738	486	164	2,807	
Czech	764	300	129	154	1,347	
India	1,109	60	1,758	1,530	4,457	
Japan	2,797	2,240	4,714	1,390	11,141	
Norway	2,518	1,006	322	256	4,102	
United States	3,907	2,342	739	116	7,104	
	30,958					

for both B_{te} and B_{tr} in $N_{epoch}=20$ epochs using Adam optimizer and a learning rate of $1e^{-4}$. For our classification models, we set λ to 1.0 and temperature parameter τ to 0.05. All experiments were performed on an NVIDIA Tesla V100-SXM3-32GB GPU device with 16GB of RAM, with the PyTorch library.

Evaluation setting. We use three standard metrics to benchmark classification models: precision (P), recall (R), and F1-score (F). We compare our method to several recent transfer learning methods for classifying road surface defects. The compared methods are Hsieh *et al.* [29], Hsieh *et al.* [42], Silva *et al.* [37], Ali *et al.* [39], Hasanaath *et al.* [27], Chen *et al.* [36]. These methods utilized a backbone pretrained on ImageNet for transfer learning. To validate the model's performance, we divided the data set into training and testing purposes. In each experiment, we use one data source as an unseen test set and the remaining five data sources for training. The performance of all compared models will be reported on unseen test set.

B. Results

The table II presents detailed results of our method and recent previous methods testing on six unseen data set of RDD2022. The results show that all methods perform less accurately on an unseen dataset. It demonstrates that the models are difficult to work well without training on an unseen dataset. It is reasonable because the unseen test set contains a road defect that differs significantly from the training set due to differences in road architecture across countries. The results also show that the model with MobileNet and ResNet backbones is more stable on an unseen dataset than the VGG backbone. Compared to other methods, our method has significantly higher accuracy in all unseen test sets. Furthermore, all of the performance metrics in all of the tested backbones of our method outperform previous methods by a large margin. This result shows that our method is more generalizable than the previous method for classifying road surface defects.

Ablation study. The table III shows the ablation study results for each metric learned in our method. We analyze the impact of training on multiple invariance data sources using contrastive learning separately from training domain generalization. Our method training with only traditional classification training loss. \mathcal{L}_{ce} produces similar results to the previous method as shown in Table II. Our method improves model performance significantly by incorporating

¹https://tinyurl.com/mr45jfhn

TABLE II: Comparison results between our method and several recent methods with diverse backbones on unseen testing data from RDD2022.

			Unseen testing data source																
Method Backbone		China MotorBike		Czech		India			Japan			Norway			United States				
		P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Hsieh et al. [29]	ResNet18	0.43	0.46	0.43	0.41	0.44	0.38	0.39	0.42	0.27	0.41	0.4	0.38	0.36	0.4	0.35	0.37	0.43	0.36
Hsieh et al. [42]	ResNet50	0.44	0.46	0.45	0.45	0.48	0.42	0.38	0.4	0.28	0.38	0.37	0.35	0.33	0.36	0.31	0.39	0.41	0.35
Silva et al. [37]	VGG16	0.4	0.44	0.42	0.38	0.4	0.35	0.39	0.41	0.28	0.38	0.36	0.35	0.32	0.36	0.31	0.36	0.4	0.33
Ali et al. [39]	VGG19	0.42	0.45	0.41	0.4	0.42	0.37	0.39	0.42	0.3	0.4	0.39	0.37	0.33	0.38	0.33	0.38	0.43	0.34
Hasanaath et al. [27]	MobileNet-V2	0.43	0.44	0.42	0.38	0.39	0.35	0.39	0.4	0.26	0.39	0.38	0.36	0.36	0.38	0.34	0.38	0.41	0.34
Chen et al. [36]	MobileNet-V3	0.43	0.45	0.43	0.38	0.41	0.35	0.35	0.39	0.28	0.39	0.4	0.37	0.36	0.38	0.34	0.39	0.44	0.36
Ali et al. [39]	Inception-V3	0.41	0.43	0.41	0.4	0.42	0.38	0.36	0.41	0.3	0.4	0.4	0.37	0.35	0.38	0.35	0.39	0.44	0.37
	ResNet-18	0.5	0.53	0.5	0.48	0.5	0.46	0.45	0.5	0.35	0.47	0.47	0.44	0.45	0.48	0.44	0.43	0.5	0.43
Ours.	VGG16	0.48	0.5	0.48	0.45	0.44	0.43	0.44	0.47	0.33	0.44	0.43	0.41	0.4	0.43	0.38	0.41	0.46	0.4
Ours.	MobileNet-V2	0.51	0.53	0.49	0.43	0.44	0.42	0.44	0.49	0.34	0.45	0.44	0.42	0.44	0.47	0.42	0.45	0.49	0.44
	Inception-V3	0.49	0.5	0.47	0.46	0.45	0.43	0.42	0.47	0.32	0.46	0.45	0.43	0.43	0.47	0.42	0.44	0.48	0.43

TABLE III: The ablation study yields results with average F1-score of all unseen test sets on various training losses in our method.

Backbone	\mathcal{L}_{ce}	$\mathcal{L}_{ce}, \mathcal{L}_{ct}$	$\mathcal{L}_{ce}, \mathcal{L}_{dg}$	$\mathcal{L}_{ce}, \mathcal{L}_{ct}, \mathcal{L}_{dg}$
ResNet-18	0.36	0.39	0.4	0.43
VGG16	0.34	0.36	0.38	0.4
MobileNet-V2	0.35	0.38	0.39	0.42
Inception-V3	0.36	0.37	0.38	0.41

contrastive learning via \mathcal{L}_{ct} (i.e., $\mathcal{L}_{ce}, \mathcal{L}_{ct}$) or domain generalization loss via \mathcal{L}_{dg} (i.e., $\mathcal{L}_{ce}, \mathcal{L}_{dg}$) separately. When testing on an unseen test set, domain generalization training achieves slightly higher accuracy than training on multiple invariance data sources. Finally, our method, which incorporates both contrastive learning and domain generalization (i.e., $\mathcal{L}_{ce}, \mathcal{L}_{ct}, \mathcal{L}_{dg}$), achieves the highest average F1-score on all unseen test sets. This result shows that our method effectively addresses both the invariance data source and generalization challenges in road surface defect classification.

V. CONCLUSIONS

In this study, we present a method for classifying road surface defects using camera images. There are two main parts to our method. Firstly, we propose an improvement to model training based on contrastive learning to address the invariance issue when developing a model using multiple datasets. Secondly, we propose an algorithm for training a generalized model that can be used with new unseen datasets without training the model on them. In the experiment, we simulate multiple data sources, including six data sets from six countries from the RDD2022 data set for road surface defects. The experiment results show that when compared to existing methods, our method is more efficient and accurate on the unsceen dataset. This indicates that our proposed method is well suited for generalizing models for road surface defect classification.

ACKNOWLEDGMENT

This work was realized in imec.ICON Hybrid AI for predictive road maintenance (HAIROAD) project, with the financial support of Flanders Innovation & Entrepreneurship (VLAIO, project no. HBC.2023.0170).

REFERENCES

- A. D5340-12, "Standard test method for airport pavement condition index surveys," 2012.
- [2] J. Mccarthy, L. Fitzgerald, J. Mclaughlin, B. Mulry, D. O'brien, and K. Dowling, "Rural flexible roads manual—pavement surface condition index," *Department of Transport, Toursim and Sports: Dublin, Ireland*, vol. 4, 2014.
- [3] D. Walker, "Asphalt-paser manual: Pavement surface evaluation and rating," 1987.
- [4] L. Trinh, P. Pham, H. Trinh, N. Bach, D. Nguyen, G. Nguyen, and H. Nguyen, "Pp4av: A benchmarking dataset for privacy-preserving autonomous driving," in *Proceedings of the IEEE/CVF Winter Con*ference on Applications of Computer Vision, 2023, pp. 1206–1215.
- [5] L. Trinh, B. Ha, and A. T. Tran, "Fisheyepp4av: A privacy-preserving method for autonomous vehicles on fisheye camera images," in 2023 RIVF International Conference on Computing and Communication Technologies (RIVF), 2023, pp. 424–429.
- [6] G. E. Elkins, B. Ostrom, et al., "Long-term pavement performance information management system user guide," United States. Federal Highway Administration. Office of Infrastructure ..., Tech. Rep., 2021.
- [7] M. Eisenbach, R. Stricker, D. Seichter, K. Amende, K. Debes, M. Sesselmann, D. Ebersbach, U. Stoeckert, and H.-M. Gross, "How to get pavement distress detection ready for deep learning? a systematic approach," in 2017 International Joint Conference on Neural Networks (IJCNN), 2017, pp. 2039–2047.
- [8] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, "Automatic road crack detection using random structured forests," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3434–3445, 2016
- [9] H. Maeda, T. Kashiyama, Y. Sekimoto, T. Seto, and H. Omata, "Generative adversarial network for road damage detection," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, pp. 47 60, 2020.
- [10] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, "Rdd2020: An annotated image dataset for automatic road damage detection using deep learning," *Data in Brief*, vol. 36, p. 107133, 2021.
- [11] D. M. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, Y. S. I. I. of Technology Roorkee, India, T. U. of Tokyo, Japan., U. Technologies, Inc., and Tokyo, "Rdd2022: A multi-national image dataset for automatic road damage detection," *ArXiv*, vol. abs/2209.08538, 2022.
- [12] V. Pereira, S. Tamura, S. Hayamizu, and H. Fukai, "Classification of paved and unpaved road image using convolutional neural network for road condition inspection system," in 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA), 2018, pp. 165–169.
- [13] D. K. Dewangan and S. P. Sahu, "Renet: road classification convolutional neural networks for intelligent vehicle system," *Intelligent Service Robotics*, vol. 14, no. 2, pp. 199–214, 2021.
- [14] L. Cheng, X. Zhang, and J. Shen, "Road surface condition classification using deep learning," *Journal of Visual Communication and Image Representation*, vol. 64, p. 102638, 2019.
- [15] J. Balcerek, A. Konieczka, K. Piniarski, and P. Pawłowski, "Classification of road surfaces using convolutional neural network," in 2020 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2020, pp. 98–103.

- [16] M.-H. Kim, J. Park, and S. Choi, "Road type identification ahead of the tire using d-cnn and reflected ultrasonic signals," *International journal of automotive technology*, vol. 22, pp. 47–54, 2021.
- [17] M. Nolte, N. Kister, and M. Maurer, "Assessment of deep convolutional neural networks for road surface classification," in 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018, pp. 381–386.
- [18] H. Zhang, Z. Li, W. Wang, L. Hu, J. Xu, M. Yuan, Z. Wang, Y. Ren, and Y. Ye, "Multi-supervised bidirectional fusion network for road-surface condition recognition," *PeerJ Computer Science*, vol. 9, p. e1446, 2023.
- [19] R. Shi, S. Yang, Y. Chen, R. Wang, M. Zhang, J. Lu, and Y. Cao, "Cnn-transformer for visual-tactile fusion applied in road recognition of autonomous vehicles," *Pattern Recognition Letters*, vol. 166, pp. 200–208, 2023.
- [20] R. Chhabra, S. Singh, and R. Kaur, "Deep-learning-based road surface classification for intelligent vehicles," in *Manufacturing Technologies* and *Production Systems*, 2024, pp. 163–170.
- [21] C. Tian, D. Jin, B. Leng, and L. Xiong, "Reliable identification of road surface condition considering shadow interference," in 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), 2021, pp. 251–257.
- [22] T. Rateke, K. A. Justen, and A. von Wangenheim, "Road surface classification with images captured from low-cost cameras – road traversing knowledge (rtk) dataset," Revista de Informática Teórica e Aplicada (RITA), 2019.
- [23] F. M. N. Sajad Ranjbar and H. Zakeri, "An image-based system for asphalt pavement bleeding inspection," *International Journal of Pavement Engineering*, vol. 23, no. 12, pp. 4080–4096, 2022.
- [24] N. T. H. Nguyen, T. H. Le, S. Perry, and T. T. Nguyen, "Pavement crack detection using convolutional neural network," in *Proceedings of* the 9th International Symposium on Information and Communication Technology, ser. SoICT '18, New York, NY, USA, 2018, p. 251–256.
- [25] H. Nhat-Duc, Q.-L. Nguyen, and V.-D. Tran, "Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network," *Automation in Construction*, vol. 94, pp. 203–213, 2018.
- [26] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and A. Agrawal, "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection," Construction and Building Materials, vol. 157, pp. 322–330, 2017.
- [27] A. A. Hasanaath, A. Moinuddeen, N. Mohammad, M. A. Khan, and A. A. Hussain, "Continuous and realtime road condition assessment using deep learning," in 2022 International Conference on Connected Systems and Intelligence (CSI), 2022, pp. 1–7.
- [28] A. Chaudhary and P. Dr., "Road surface quality detection using light weight neural network for visually impaired pedestrian," *Evergreen*, vol. 10, no. 2, pp. 706–714, 06 2023.
- [29] Y.-A. Hsieh and Y. J. Tsai, "Automated asphalt pavement raveling detection and classification using convolutional neural network and macrotexture analysis," *Transportation Research Record*, vol. 2675, no. 9, pp. 984–994, 2021.
- [30] A. Riid, R. Lõuk, R. Pihlak, A. Tepljakov, and K. Vassiljeva, "Pavement distress detection with deep learning using the orthoframes acquired by a mobile mapping system," *Applied Sciences*, vol. 9, no. 22, 2019.
- [31] A. Gagliardi, V. Staderini, and S. Saponara, "An embedded system for acoustic data processing and ai-based real-time classification for road surface analysis," *IEEE Access*, vol. 10, pp. 63 073–63 084, 2022.
- [32] I. S. Andrades, J. J. Castillo Aguilar, J. M. V. García, J. A. C. Carrillo, and M. S. Lozano, "Low-cost road-surface classification system based on self-organizing maps," *Sensors*, vol. 20, no. 21, 2020.
- [33] L. Baoxian, C. P. W. Kelvin, Z. Allen, Y. Enhui, and W. Guolong, "Automatic classification of pavement crack using deep convolutional neural network," *International Journal of Pavement Engineering*, vol. 21, no. 4, pp. 457–463, 2020.
- [34] V.-D. T. Nhat-Duc Hoang and X.-L. Tran, "Image processing-based classification of pavement fatigue severity using extremely randomized trees, deep neural network, and convolutional neural network," *International Journal of Pavement Engineering*, vol. 24, no. 1, p. 2201902, 2023
- [35] E. Eslami and H.-B. Yun, "Comparison of deep convolutional neural network classifiers and the effect of scale encoding for automated pavement assessment," *Journal of Traffic and Transportation Engi*neering (English Edition), vol. 10, no. 2, pp. 258–275, 2023.

- [36] L. Chen, H. Yao, J. Fu, and C. Tai Ng, "The classification and localization of crack using lightweight convolutional neural network with cbam," *Engineering Structures*, vol. 275, p. 115291, 2023.
- [37] W. R. L. d. Silva and D. S. d. Lucena, "Concrete cracks detection based on deep learning image classification," in *Proceedings*, vol. 2, no. 8, 2018, p. 489.
- [38] D. Dais, I. E. Bal, E. Smyrou, and V. Sarhosis, "Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning," *Automation in Construction*, vol. 125, p. 103606, 2021.
- [39] L. Ali, F. Alnajjar, H. A. Jassmi, M. Gocho, W. Khan, and M. A. Serhani, "Performance evaluation of deep cnn-based crack detection and localization techniques for concrete structures," *Sensors*, vol. 21, no. 5, p. 1688, 2021.
- [40] F. Liu, J. Liu, and L. Wang, "Deep learning and infrared thermography for asphalt pavement crack severity classification," *Automation in Construction*, vol. 140, p. 104383, 2022.
- [41] H. Dong, K. Song, Q. Wang, Y. Yan, and P. Jiang, "Deep metric learning-based for multi-target few-shot pavement distress classification," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1801–1810, 2022.
- [42] Y. Hsieh, Z. Yang, and Y. James Tsai, "Convolutional neural network for automated classification of jointed plain concrete pavement conditions," *Comput.-Aided Civ. Infrastruct. Eng.*, vol. 36, no. 11, p. 1382–1397, 2021.
- [43] R. Stricker, M. Eisenbach, M. Sesselmann, K. Debes, and H.-M. Gross, "Improving visual road condition assessment by extensive experiments on the extended gaps dataset," in 2019 International Joint Conference on Neural Networks (IJCNN), 2019, pp. 1–8.
- [44] B. Li, K. C. P. Wang, E. Y. Allen Zhang, and G. Wang, "Automatic classification of pavement crack using deep convolutional neural network," *International Journal of Pavement Engineering*, vol. 21, no. 4, pp. 457–463, 2020.
- [45] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple frame-work for contrastive learning of visual representations," in *Proceedings of the 37th International Conference on Machine Learning*, ser. ICMI 220, 2020.
- [46] L. Trinh, B. Ha, and A. T. Tran, "Vqc-covid-net: Vector quantization contrastive learning for covid-19 image base classification," in 2022 9th NAFOSTED Conference on Information and Computer Science (NICS), 2022, pp. 247–251.
- [47] K. Guo and B. C. Lovell, "Domain-aware triplet loss in domain generalization," *Computer Vision and Image Understanding*, vol. 243, p. 103979, 2024.
- [48] M. Faraki, X. Yu, Y.-H. Tsai, Y. Suh, and M. Chandraker, "Cross-domain similarity learning for face recognition in unseen domains," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15292–15301.
- [49] J. Guo, X. Zhu, C. Zhao, D. Cao, Z. Lei, and S. Z. Li, "Learning meta face recognition in unseen domains," in *Proceedings of the IEEE/CVF* Conference on Computer Vision and Pattern Recognition, 2020, pp. 6163–6172.
- [50] Y. Zhao, Z. Zhong, F. Yang, Z. Luo, Y. Lin, S. Li, and N. Sebe, "Learning to generalize unseen domains via memory-based multisource meta-learning for person re-identification," in *Proceedings of* the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 6277–6286.
- [51] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, H. Omata, T. Kashiyama, and Y. Sekimoto, "Crowdsensing-based road damage detection challenge (crddc'2022)," in 2022 IEEE International Conference on Big Data (Big Data). IEEE, 2022, pp. 6378–6386.
- [52] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, "Rdd2020: An annotated image dataset for automatic road damage detection using deep learning," *Data in brief*, vol. 36, p. 107133, 2021.
- [53] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, H. Omata, T. Kashiyama, and Y. Sekimoto, "Global road damage detection: Stateof-the-art solutions," in 2020 IEEE International Conference on Big Data (Big Data). IEEE, 2020, pp. 5533–5539.