DA-RDD: Toward Domain Adaptive Road Damage Detection Across Different Countries

Chunmian Lin, Daxin Tian, *Senior Member, IEEE*, Xuting Duan, *Member, IEEE*, Jianshan Zhou, Dezong Zhao, *Senior Member, IEEE*, and Dongpu Cao

***Abstract*— Recent advances on road damage detection relies on a large amount of labeled data, whilst collecting pavement image is labor-intensive and time-consuming. Unsupervised Domain Adaptation (UDA) provides a promising solution to adapt a source domain to the target domain, however, cross-domain crack detection is still an open problem. In this paper, we propose domain adaptive road damage detection termed as DA-RDD, by incorporating image-level with instance-level feature alignment for domain-invariant representation learning in an adversarial manner. Specifically, importance weighting is introduced to evaluate the intermediate samples for image-level alignment between domains, and we aggregate RoI-wise feature with multi-scale contextual information to recover the crack details for progressive domain alignment at instance level. Additionally, a large-scale road damage dataset (based on Road Damage Dataset 2020 (RDD2020)) named as RDD2021 is constructed with 100*k* synthetic labeled distress images. Extensive experimental results on damage detection across different countries demonstrate the universality and superiority of DA-RDD, and empirical studies on RDD2021 further claim its effectiveness and advancement. To our best knowledge, it is the first time to investigate domain adaptative pavement crack detection, and we expect the contributions in this work would facilitate the development of generalized road damage detection in the future.**

***Index Terms*— Road damage detection, domain adaptation, pavement distress dataset, adversarial learning, intelligent transportation systems.**

I. Introduction

AS A crucial public infrastructure, pavement surface irregularly wears and deteriorates over time from different factors related to aging, adverse weather, frequent overload, etc., which would evolve to different degrees of pavement damage gradually. Poor road condition seriously reduces road service life, and contributes to ever-growing traffic accidents, posing a potential threat to driving safety. Therefore, governments and transportation agencies from many countries attach great importance to road condition surveys and launch several pavement maintenance projects [2]. Road damage detection is the most fundamental step in road surface assessment, and traditional pavement distress recognition mainly relies on laborious and time-consuming manual inspection, which might be prone to human visual error and subjective judgement. Recent progresses in image processing and deep learning techniques open a new era of intelligent transportation system applications [3], [4], [5], [6], e.g., automatically detecting pavement damages.

There are a variety of empirical studies on crack detection using representative image feature, e.g., Gabor Filter (GF) [7], Histogram of Oriented Gradient (HOG) [8], Local Binary Pattern (LBP) [9], [10]. Despite encouraging recognition results are reported, these methods are not robust enough to address feature diversity in the crack appearance and type in the complex environment. Deep learning technique based on convolutional neural network (CNN) has emerged impressive results on computer vision tasks, including object classification [11], detection [12], semantic segmentation [13], as well as crack detection [14], [15], [16], [17], [18]. The mainstream of CNN-based crack detection is to build the encoder-decoder architecture to perform pixel-wise segmentation on high-quality pavement image, that is acquired from laser line-scan camera or three-dimensional (3D) camera mounted on the dedicated vehicles. However, these segmentation approaches suffer from the loss of spatial resolution with high-level feature encoding, and inefficient pixel-wise classification hinders its application and deployment in the edge device. More importantly, such pavement image typically depends on high-resolution laser imaging equipment, which is too expensive for local agencies with limited budgets to perform effective road maintenance. From 2018 to 2020, the Road Damage Detection Challenge (RDDC) develops a new horizon of cost-effective road condition monitoring and damage detection. This challenge receives wide attention from researchers all over the world, and several novel methods along the pipeline of 2D object detection [12], [19] are proposed for improving automatic pavement distress identification performance [1], [20], [21], [22], [23]. In this paper, we would also investigate efficient yet effective algorithm for low-cost road damage detection.

Previous works motivate several pavement distress datasets, i.e., CrackTree200 [25] and GAPs [26], each of which contains

Manuscript received 30 January 2022; revised 6 June 2022; accepted 7 November 2022. This work was supported in part by the National Natural Science Foundation of China under Grant U20A20155, Grant 62061130221, and Grant 62173012; in part by the Natural Science Foundation of Beijing Municipality under Grant L191001; in part by the Zhuoyue Program of Beihang University (Postdoctoral Fellowship); and in part by the China Postdoctoral Science Foundation under Grant 2020M680299. The Associate Editor for this article was J. Alvarez. *(Corresponding author: Daxin Tian.)*

Chunmian Lin, Daxin Tian, Xuting Duan, and Jianshan Zhou are with the Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems and Safety Control, Beijing Advanced Innovation Center for Big Data and Brain Computing, School of Transportation Science and Engineering, Beihang University, Beijing 100191, China (e-mail: dtian@buaa.edu.cn).

Dezong Zhao is with the James Watt School of Engineering, University of Glasgow, Glasgow G12 8QQ, U.K.

Dongpu Cao is with the Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, ON N2L3G1, Canada.

Digital Object Identifier 10.1109/TITS.2022.3221067

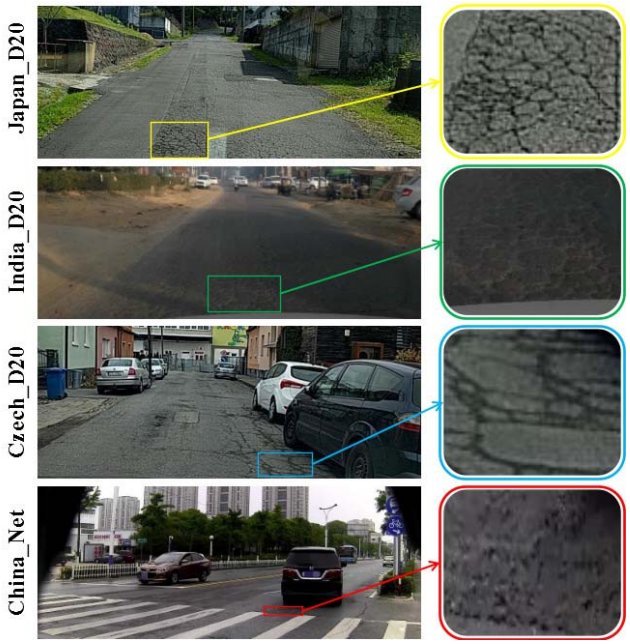


Fig. 1. Damage samples on the same caterogy from different countries. Noted that *Net* corresponds to the alligator crack (*D*20) in the case of different countries. It is challenging to perform cross-domain road damage detection due to the great variations in crack appearance and scenario layouts.

hundreds of top-view images with fine-grained pixel annotations. The road damage dataset (RDD2020) [1] is publicly available with 26620 front-view samples collected from different countries (Japan, India and Czech) using smartphone, with bounding-box annotation for various distress types. Similarly, the damage dataset recently released by Jiangsu Pro., China (CN2021) [27] contains 1366 available training with 5 crack classes in China. Nevertheless, there is a dilemma that deep learning model requires large-scale training data to achieve satisfactory performance, whereas the magnitude of pavement distress dataset is far insufficient. It is desirable to build up a large-sacle road damage dataset for model training and method evaluation. Above all, crack detector trained on labeled pavement image typically do not generalize to the new test data because of large variations in illumination, background, and image quality in real-world situation, as shown in Fig.1. This may cause a considerable domain shift [28], [29], and results in a significant performance gap.

Unsupervised Domain Adaptation (UDA) [30] is regarded as an encouraging approach to address this problem by transferring the representation from an annotated source domain to the target domain without Ground-Truth (GT) labels, which aligns different data distributions to alleviate the domain disparity effectively. It explicitly learns invariant representation via domain adversarial learning [31] or pixel-level adaptation [24], and achieves remarkable progress in image classification [32], [33], [34], detection [29], [35], [36], [37], [38] and segmentation [39], [40], [41]. Specifically, UDA in cross-domain object detection attempts to combine adversarial learning with off-the-shelf 2D detector, e.g., Faster RCNN [12], to minimize the domain discrepancy via image-instance adaptation [29], strong-weak feature alignment [35], harmonized transferability and discriminability [38], etc. In the context of pavement distress detection, UDA has not been explored yet, and it is non-trivial to investigate a generalized damage detector to adapt from a labeled source domain to the unlabeled target domain.

In this paper, we incorporate unsupervised domain adaptation (UDA) with pavement distress detection, and propose a general framework termed as DA-RDD for domain adaptive road damage detection, which integrates Faster RCNN with image-level and instance-level feature alignment methods to learn domain-invariance crack representation in an adversarial manner. Specifically,an intermediate space is introduced via CycleGAN [24] to bridge the domain gaps, and feature importance is reweighted for image-level alignment across different domains. To highlight the complementary effect of multi-level features, we further aggregate the RoI-wise feature with multi-scale contextual representations to recover the damage details for local instance discriminability enhancement and progressive cross-domain alignment. Extensive cross-domain damage detection experiments are conducted on RDD2020 and CN2021, and results demonstrate the generality and adaptability of DA-RDD, which outperform the baseline and other counterparts over a substantial margin. Additionally, we build up a large-scale road damage dataset (based on Road Damage Dataset 2020 (RDD2020) [1]) namely RDD2021, comprising 100*k* front-view synthetic images with bounding-box annotations on four damage categories across three different countries. Empirical studies verify its effectiveness and advancement, with substantial performance gains on crack identification. To the best of our knowledge, it is the first time to explore the adapted distress detection across different countries, and we expect the proposed DA-RDD and RDD2021 would beneficial to the generalized road damage detection in the future. The contributions can be mainly summarized as follows:

1) DA-RDD is proposed by integrating Faster RCNN with image-level and instance-level alignment methods for domain-invariant representation learning and domain adaptive road damage detection in an adversarial manner.

2) Importance weighting method is developed to measure the intermediate feature space for image-level alignment between domains; furthermore, RoI feature is concatenated with multi-level context information to recover the damage details for hierarchical cross-domain instance alignment.

3) A large-scale road damage dataset RDD2021 is constructed with 100*k* front-view synthetic samples, annotated on four distress categories across three different countries.

4) Domain adaptation experimental results on RDD2020 and CN2021 illustrate the universality and transferability of DA-RDD for cross-domain damage detection; moreover, empirical studies on RDD2021 suggest the effectiveness and superiority of large-scale crack data, with substantial detection performance improvements.

The reminder in this paper is organized as follows: we review the related works in Section II, and describe our proposed methods in Section III. Section IV presents the experimental setup and results, and we conclude the whole paper in Section V.

II. Related Works

This section would briefly review the related works on road damage detection, pavement distress dataset, and unsupervised domain adaptation.

A. Road Damage Detection

The conventional road damage detection is mainly based on low-level image feature. For instance, Medina et al. [7] combine visual appearance and geometrical information to extract the profile of road crack, while Quintana et al. [9] utilize local binary pattern (LBP) within non-overlapping grid cell to detect and classify cracks on the road surfaces. To enhance the robustness of method, Gradient LBP (GLBP) [10] is further formulated to suppress the effect of illumination and background noise on the detection result.

With the introduction of convolutional neural network (CNN) in civil engineering, there are a variety of segmentation-based and detection-based approaches to pavement distress detection. On one hand, Zhang et al. propose CrackNet [14] and ins variant [16] for pixel-level crack identification on the 3D asphalt surface. Moreover, Fei et al. [42] stack several convolutional layers with multi-scale kernels to achieve high damage segmentation performance. Based on semantic segmentation model [43], Zou et al. [17] design an encoder-encoder architecture DeepCrack with pairwise feature fusion under different scales for accurate distress segmentation. Yang et al. [18] integrate pyramid architecture with hierarchical boosting network to aggregate low-level feature and context information for non-continuity crack segmentation. On the other hand, recent Road Damage Detection (RDD) Challenge motivates to develop efficient road damage detectors. Wang et al. [22] introduce data augmentation to balance the training sample, and adopt Faster RCNN [12] to classify different cracks. Besides, Team IMSC [44] trains a YOLOv5 detector [19] with parameter adjustment and architecture fine-tuned to detect various types of pavement distress. The tricks of ensemble learning and test time augmentation (TTA) also facilitate the model performance. Generally, detection-based methods provide a simple yet accurate crack detection scheme, and we would follow this pipeline to carry out road damage detection in this paper.

B. Pavement Distress Dataset

Regarding pavement distress dataset, it can be roughly divided into pixel-level segmentation [25], [45], [46] and instance-level box [26], [27], [47] according to the label format. CrackTree [25] comprises of 206 top-view pavement images with pixel-level annotations, each of which has a size of 800×600. Based on expensive mobile mapping system, GAP dataset [26] consists of 1969 high-definition pavement images with a size of 1920×1080, and an actual crack is annotated with a delicated box. The relevant damage classes include crack, pothole, applied patch, open joint and bleeding. RDD2020 [1] is a heterogeneous road damage dataset[1] with 26620 front-view images acquired from Japan, India and Czech using smartphone. The resolution of image in different countries varies from 600×600 to 720×960, and there are 8 distress types: linear crack (wheel-marked part marked as D00, construction joint part marked as D01, equal interval marked as D10 and construction joint part marked as D11), alligator crack (marked as D20) and others (pothole marked as D40, cross walk blur marked as D43 and white line blur marked as D44. Recently, CN2021 [27] is publicly available with 1365 labeled training and 900 test data collected from an on-vehicle dashcam in China, which focuses on marking, net, abnormalmanhole, pothole and crack identification. However, the size of these dataset is inadequate to conduct comprehensive model training and evaluation, and it is also troublesome to collect massive labeled pavement images due to the expensive manual annotation. Although Maeda et al. [47] construct a progressive growing generative adversarial network (PG-GAN) for pseudo-image generation, only pothole class (marked as D40) is augmented without the consideration of multiple classes augmentation. In this work, we would generate synthetic labeled samples from different layouts and scenes, and build up a large-scale road damage dataset.

[1] https://github.com/sekilab/RoadDamageDetector

C. Unsupervised Domain Adaptation

Unsupervised domain adaptation (UDA) provides an effective solution to tackle with the rare of labeled data and domain shift problem, which aims to transfer the knowledge from the source domain to another one. Currently, statistic matching and adversarial learning are two main UDA techniques. The former attempts to match high-order statistic of feature distributions across different domains in latent space. DAN [48] matches the implicit representations of task specific in a reproducing kernel Hilbert space, and reduces distributional disparity by the optimal multi-kernel selection. CORAL [49] aligns the second-order statistic of distribution with a linear transformation between two domains. Motivated by the two-player game in generative adversarial network (GAN) [50], adversarial learning is the prevalent UDA method that explicitly exploits domain-invariant representations via adversarial training or pixel-level adaptation, which has been widely utilized for domain adaptive object detection [29], [35], [36], [37], [38], [51], [52] and semantic segmentation [39], [40], [41]. Based on Faster RCNN model, Chen et al. [29] firstly design different adaptation methods to align the image and instance features between domains, and learns a domain classifier to reduce the discrepancy in an adversarial training manner. After that, Saito et al. [35] combine strong local alignment with weak global alignment to match multi-level features, while Guan et al. [37] propose an uncertainty-aware domain adaptation network (UaDA) to adjust the well- and poor-aligned samples via conditional adversarial learning. To consider the object of interest in local, Chen et al. [38] harmonize the transferability and discriminability for cross-domain feature alignment, and Zhu et al. [51] focus on discriminative region alignment across domains. Similarly, Cai et al. [52] regularize the domain consistency by mining oobject relations in a teacher-student scheme. Furthermore, CycleGAN [24] adopts cycle consistency loss to generate an image without paired

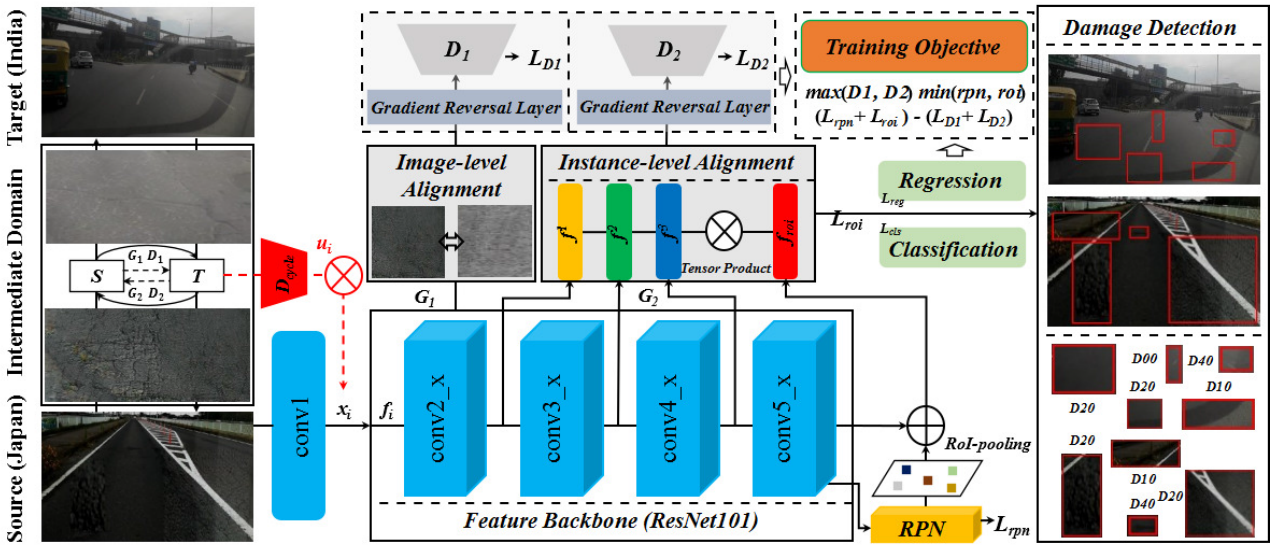


Fig. 2. Overview of the DA-RDD. Given the images from source domain (Japan) and the target domian (India), CycleGAN [24] generates the intermediate sample to reduce the domain gap, and the prediction scrore from D_{cycle} is utilized for importance estimation and feature reweight, aligning the domain distribution in image-level. Through the feature backbone (ResNet101 [11]), multi-scale contextual information from $conv2\_x$ (f^1), $conv3\_x$ (f^2) and $conv4\_x$ (f^3), is aggregated with the RoI-pooling feature for local object discriminability enhancement and progressive instance-level domain alignment. Notably, the discriminators D_1 and D_2 inversely update the gradient via GRL in an adversarial manner. Finally, the fused instance features predict the damage detection result, and the training objective for DA-RDD comprises of adversarial losses (L_{D_1} and L_{D_2}) and detection losses (L_{rpn} and L_{roi}).

examples, and achieve the pixel-level adaptation from the source to the target domains. However, cross-domain pavement distress detection is still an open problem, and we would introduce UDA methods for domain adaptive road damage detection in this paper.

III. DA-RDD: Domain Adaptive Road Damage Detection

This section would describe the technical details of DA-RDD, as shown in Fig.2. Based on Faster RCNN, we reweight feature importance from the intermediate space for image-level feature alignment between domains, and aggregate RoI feature with multi-scale contextual information to align the instance details across different domains.

The two-stage Faster RCNN [12] mainly consists of three key components: the shared feature backbone, region proposal network (RPN) and RoI classifier. Given a pavement input image, convolutional feature maps are extracted by the feature backbone extracts, and RPN generates numerous candidate proposals from these feature maps to indicate the potential object region. Simultaneously, RoI classifier predicts the object category label and bounding-box coordinates from the RoI-pooling feature. As mathematically formulated as Eq.1, the overall loss function L_{det} comprises of RPN loss L_{rpn} and RoI classifier loss L_{roi}, each of which contains classification loss L_{cls} and regression loss L_{reg} to measure the detection accuracy. More technical details can refer to the paper [12].

$$L_{det} = L_{rpn} + L_{roi} \tag{1}$$

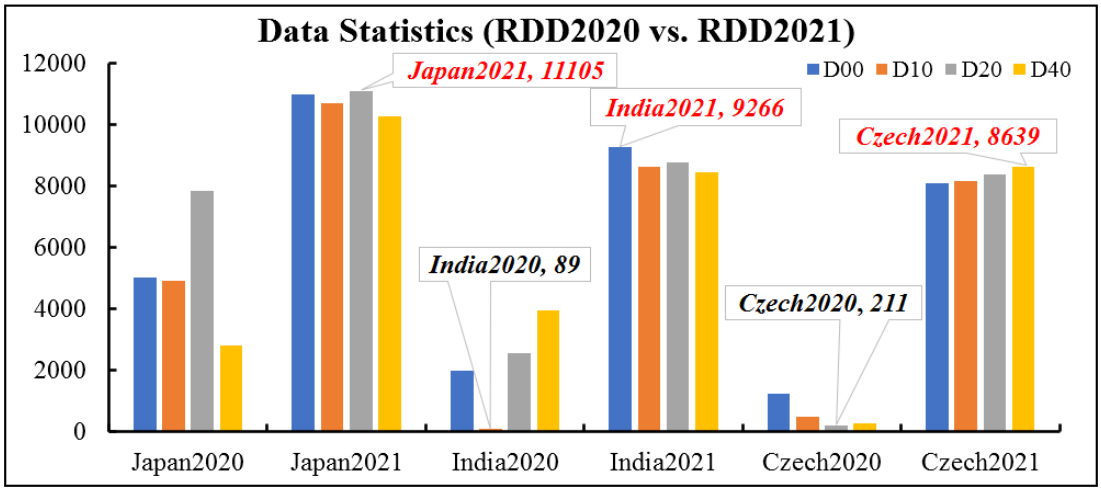
A. Problem Formulation and H-Divergence Theory

In the real-world situation, object detector generally suffers from domain shift problem where distributional disparity may result in performance degradation. Formally, we denote a labeled source domain as $\mathcal{S} = \left\{(x_i^s, b_i^s, c_i^s)_{i=1}^{N_s}\right\}$, where $b_i^s \in \mathcal{R}^{n\times 4}$ denotes the bounding-box coordinates, $c_i^s \in \mathcal{R}^{n\times 1}$ is the class label and N_s indicates the number of sample x_i^s, and define a unlabeled target domain as $\mathcal{T} = \left\{(x_j^t)_{j=1}^{N_t}\right\}$ of N_t samples. The $\mathcal{H}$-divergence theory [53] is proposed to measure the divergence between two datasets with different distributions. Given the feature vector x and a set of domain classifiers $\mathcal{H}(h(x)) = \left\{h(x_{(i,j)}^d), (d = s, t; i = 0, \ldots, N_s, j = 0, \ldots, N_t)\right\}$, the $\mathcal{H}$-divergence calculates the discrepancy between two domains [29] as follows Eq.2:

$$d\left(\mathcal{H}\left(h\left(x\right)\right)\right) = 2\left(1 - \min_{h\in\mathcal{H}}\left(\epsilon_S\left(h\left(x^S\right)\right) + \epsilon_T\left(h\left(x^T\right)\right)\right)\right) \tag{2}$$

where $\epsilon_S\left(h\left(x^S\right)\right) = \epsilon_{x_i^s}\left(h\left(x_i^s\right)\right), (i = 0, \ldots, N_s)$ and $\epsilon_T\left(h\left(x^T\right)\right) = \epsilon_{x_j^t}\left(h\left(x_j^t\right)\right), (j = 0, \ldots, N_t)$ denote the prediction error for each sample on the source domain and the target domain, respectively; a domain classifier $h(x)$ predicts the source sample x_i^s to 0 and the target data x_j^t to 1. The formulation suggests the domain disparity $d\left(\mathcal{H}\left(h\left(x\right)\right)\right)$ is inversely proportional to the domain prediction error [29]. Thus, a well-performance domain classifier is difficult to distinguish the sample between domains. In the context of deep neural network, the training objective is to minimize the domain distance for feature alignment between two domains, and we reformulate the Eq.2 as follows Eq.3:

$$\min_f d\left(\mathcal{H}\left(h\left(x\right)\right)\right) \Leftrightarrow \max_f \min_{h\in\mathcal{H}}\left(\epsilon_S\left(h\left(x^S\right)\right) + \epsilon_T\left(h\left(x^T\right)\right)\right) \tag{3}$$



	Data Statistics (RDD2020 vs. RDD2021)						CN2021
	Japan2020	**Japan2021**	India2020	**India2021**	Czech2020	**Czech2021**	China
Total Img.	10506	**40000**	7706	**30000**	2829	**30000**	1365
D00	5006	**10978**	1994	**9266**	1242	**8091**	1015 (*Crack*)
D10	4911	**10697**	89	**8631**	480	**8164**	243 (*Net*)
D20	7846	**11105**	2556	**8778**	211	**8373**	274 (*Pothole*)
D40	2798	**10283**	3947	**8440**	263	**8639**	928 (*Marking*)
Total Ins.	20561	**43063**	8586	**35115**	2196	**33267**	1532

Fig. 3. Data statistics information among different pavement distress dataset. In RDD2020 and CN2021, severe class imbalanced problem is observed, i.e., $India2020\_D10$ with 89 samples, $Czech2020\_D20$ with 211 and $CN2021\_Net$ with 263 images, suggesting the difficultis to receive satisfactory crack detection performance. On the contrary, the synthetic RDD2021 provides large-scale road damage examples for each category across different countries.

where f defines a neural network, and here x denotes the feature map from the source or the target set after activation. The min-max two-player game can be jointly optimized in an adversarial learning mode, and be implemented by a gradient reversal layer (GRL) [53], which updates the gradient in an opposite direction to maximize the domain discriminator loss $\epsilon(.)$. By this way, domain discrepancy is addressed by feature distribution alignment between two different domains. In this paper, we attempt to learn a domain adaptive road damage detector DA-RDD trained on the labeled source domain $\mathcal{S}$ to generalize to a new unseen target domain $\mathcal{T}$.

B. Image-Level Feature Alignment

Domain adaptation is regarded as a promising approach to cross-domain road damage detection. However, due to large variations from pavement distress appearance, type, and multi-source noise (e.g., background and layout) in the real world, transferring the knowledge directly from the source to target domain is still challenging to receive satisfactory performance gains. To this end, we append a domain classifier D_1 into the Faster RCNN framework via GRL, and introduce an interpolated domain $\mathcal{I}$ to bridge the significant gap between the source domain $\mathcal{S}$ and the target domain $\mathcal{T}$. Furthermore, importance reweighting is also utilized for highlighting the discriminative representation and aligning the feature distribution across domains.

The intermediate domain $\mathcal{I}$ comprises a set of interpolated damage images generated by a CycleGAN [24], which is a pixel-wise image translation architecture for learning dual domain mappings via a cycle consistency loss. As shown in Fig.4, it bridges the distributional gap between the source and the target data, and the pipeline of cross-domain detection convert from 'source→target' into 'source→intermediate→target', thus reducing the difficulty of domain adaptation significantly. Notably, the synthetic sample in intermediate domain has the similarity with the source data in damage content but diverging in visual appearances; simultaneously the interpolated example is similar to the target set in pixel-level distribution while varying in distress instance details. However, it is noted that GAN easily suffers from such model collapse problem in trivial or negative distress sample generation, leading to false feature alignment across domains. Thus, we further propose an importance weighting approach to measure the significance of interpolated feature, that is, desirable damage sample would receive more attention than that of irrelevant one.

To be specific, the importance is based on the similarity between domains, which implies the interpolated feature would be up-weighted when it closed to the target set, and vice versa. As demonstrated in Fig.1, we append a domain classifier and GRL to the Faster RCNN framework to learn adopt the predicted score from the target domain discriminator D_{cycle}, to measure the uncertainty of the generated damage sample

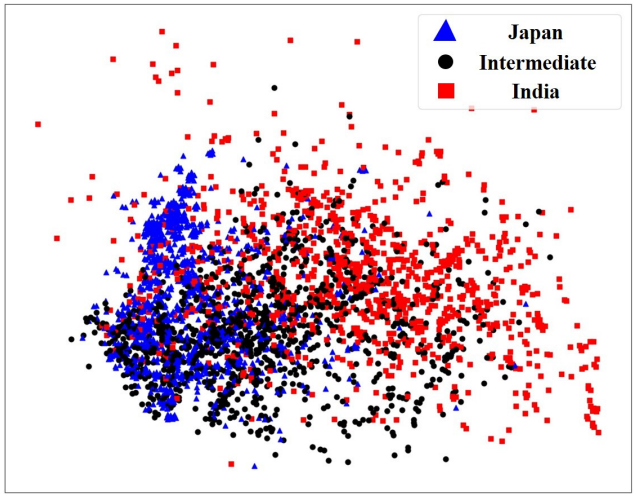


Fig. 4. Visualization on feature distribution between domians via t-distributed stochastic neighbor embedding (t-SNE [54]). Taking the adaptation from Japan to India as example, we randomly sample 1000 images from Japan2020 and India2020, and feed them into the CycleGAN [24] for the interpolated feature generation. It is clearly observed that the intermediate space bridges the data distributional gap between two domains, thus reducing the difficulty of domain adaptatiion.

as Eq.4:

$$u_{\mathcal{I}_i} = \frac{\mathcal{P}_{\mathcal{T}}(\mathcal{I}_i)}{\mathcal{P}_{\mathcal{S}}(\mathcal{I}_i) + \mathcal{P}_{\mathcal{T}}(\mathcal{I}_i)} \tag{4}$$

where $\mathcal{I}_i$ defines the synthetic target image, $\mathcal{P}_{\mathcal{S}}(\mathcal{I}_i)$ and $\mathcal{P}_{\mathcal{T}}(\mathcal{I}_i)$ are the predicted probability of $\mathcal{I}_i$ belonging to the source or target damage domain. Due to high-uncertainty feature hard to distinguish by the cycle domain classifier, we assume the intermediate space is close to the target domain in this case, and attach great weight to emphasize the crack feature importance. Given an input x_i, the reweighted representation f_i before feeding into the feature backbone is presented as Eq.5:

$$f_i = x_i \times (1 + u_i) \tag{5}$$

Therefore, the input of domain discriminator D_1 is $G_1(f_i)$, and we denote the classifier result $D_1(G_1(f_i))$ as D_1 for simplicity. The adversarial loss of domain discriminator D_1 can be formulated as Eq.6:

$$\begin{aligned} L_{D_1} &= \mathbb{E}\left(log\left(D_1\left(G_1\left(f_i^{\mathcal{S}}\right)\right)\right)\right) \\ &\quad + \mathbb{E}\left(1 - log\left(D_1\left(G_1\left(f_j^{\mathcal{T}}\right)\right)\right)\right) \\ &= -\frac{1}{N_s}\sum_{i=0}^{N_s} D_1^{\mathcal{S}} log\left(D_1^{\mathcal{S}}\right) - \frac{1}{N_t}\sum_{j=0}^{N_t} D_1^{\mathcal{T}} log\left(D_1^{\mathcal{T}}\right) \end{aligned} \tag{6}$$

C. Instance-Level Feature Alignment

Previous works on domain adaptive detection [29], [35], [36], [51], [52] mostly follow two-stage detection pipeline, and utilizes RoI-pooling feature for bounding-box refinement. Nevertheless, the instance feature after pooling operation may be ambiguous because of the increasing receptive fields, and information loss would also pose a great challenge in feature alignment across different domains, which have not been investigated in the prior studies yet. And these issues would be aggravated in the context of pavement crack identification due to the subtle damage feature. In this paper, it is assumed that contextual feature at multiple levels provides different characteristics, i.e., low-level appearance or shape, for depicting the specific distress, which can be complementary with abstracted RoI-wise representation. Hence, we propose to aggregate RoI-pooling with multi-scale contextual representations to recover the discriminative damage details and perform the cross-domain instance alignment hierarchically.

Formally, given multi-level context feature vectors ($f_c^v, v = 1, 2, 3\ldots$), and the RoI-wise crack feature as $f_{roi}^{(i,r)}$ with respect to the r-th region for the i-th image input, We firstly concatenate the context representation at different scales into a single vector f_c, and further aggregate it with the RoI $f_{roi}^{(i,r)}$ in an element-wise product. Here, we adopt context information at three resolutions from the shared feature backbone, and the fused feature $f_{fus}^{(i,r)}$ can be described as Eq.7:

$$f_{fus}^{(i,r)} = \left[f_c^1, f_c^2, f_c^3 \qquad f_{roi}^{(i,r)}\right. \tag{7}$$

where $f_{fus}^{(i,r)}$ denotes the fused feature vector, and $\bigotimes$ is the tensor product operation. Such a non-linear fusion strategy interacts the RoI-wise feature with global context information effectively, and refines the instance details to facilitate the domain alignment at different levels. In the implementation, we adopt the randomized methods [55], [56] as an unbiased estimator of tensor product for computational efficiency, and the feature fusion process can be transformed as Eq.8:

$$f_{fus}^{(i,r)} = \frac{1}{\sqrt{d}}\left(\boldsymbol{R}_1 f_c\right) \bigodot \left(\boldsymbol{R}_2 f_{roi}^{(i,r)}\right) \tag{8}$$

where $\bigodot$ denotes the Hadamard product, $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are random symmetric matrices with univariance, and in this work, they are sampled from the uniform distribution following the previous studies [38], [56]. d ($d \ll d_c \times d_{fus}$) indicates the estimated dimension of feature that is far less than the dimension product of the concatenated feature d_c and fused feature d_{fus}.

As depicted in Fig.1, we also append a domain discriminator D_2 into the framework, to guide the fused feature aggregation via GRL and align the instance distribution between domains progressively. The adversarial loss of D_2 can be defined as Eq.9:

$$\begin{aligned} L_{D_2} &= -\frac{1}{N_s}\sum_{i=0\ (i,r)}^{N_s} log\left(D_2\left(f_{fus}^{(i,r)}\right)_{\mathcal{S}}\right) \\ &\quad -\frac{1}{N_t}\sum_{i=0\ \ i,r}^{N_t} log(1 - D_2\left(f_{fus}^{(i,r)}\right)_{\mathcal{T}}) \end{aligned} \tag{9}$$

D. Loss Function

The loss funtion of DA-RDD includes detection loss L_{det} and domain adaptation loss L_{da}. As mentioned above, the detection loss L_{det} comprises L_{cls} and L_{reg} for both RPN and RoI classifier branches, where L_{cls} utilizes binary cross-entropy function and L_{reg} adopts the smooth-$L1$ loss.

TABLE I

RESULTS ON DOMAIN ADAPTATION FROM JAPAN TO INDIA (**JAIN-DA**). PERFORMANCE IS EVALUATED BY THE MEAN AVERAGE PRECISION (mAP) AT THE THRESHOLD OF 0.5 ACROSS 4 DAMAGE CATEGORIES ($D00$, $D10$, $D20$ AND $D40$). NOTED THAT *Oracle* REFERS TO THE MODEL IS TRAINED AND TESTED ON THE SAME LABELED OR SOURCE DOMAIN

Japan2020→India2020 (JaIn-DA)					
Methods	$D00$	$D10$	$D20$	$D40$	mAP
F-RCNN	0.2015	0.0528	0.2319	0.3093	0.1989
DA-Faster	0.2506	0.2185	0.3181	0.3300	0.2793
SWDA	0.3394	0.2801	0.2612	0.4236	0.3261
P-Faster	0.4971	0.4794	0.3984	0.4676	0.4606
UaDA	0.5464	0.4631	0.4730	0.5036	0.4965
DA-RDD	**0.6556**	*0.5873*	*0.5199*	**0.6204**	**0.5958**
Oracle	*0.6038*	**0.6054**	**0.5325**	*0.6127*	*0.5886*

Meanwhile, the domain adaptation loss consists of L_{D_1} and L_{D_2}. The whole loss L can be formulated as Eq.10: where λ denotes the balanced parameter.

$$L = \max_{D_1, D_2} \min_{G_1, G_2} L_{rpn} + L_{roi} - \lambda \ L_{D_1} + L_{D_2} \tag{10}$$

IV. EXPERIMENTS

This section would evaluate the adaptation performance of DA-RDD on different pavement distress datasets. Also, ablation studies are performed to explore the contribution of each component. Besides, we build up RDD2021 with 100k composite labeled images, to investigate large-scale road damage detection performance.

A. Implementation Details

All experiments are implemented on the Ubuntu1804 LTS with PyTorch framework, using a single NVIDIA TITAN RTX GPUs. For simplicity, we evaluate the model performance on 4 distress classes ($D00$, $D10$, $D20$ and $D40$), and adopt mean average precision (mAP) under the threshold of 0.5, unless specific stated. In this work, the domain adaptation performance of DA-RDD across different countries, and the detection performance various detectors (i.e., Faster RCNN and YOLOv5) on the large-scale road damage dataset are extensively investigated, which would be described in details.

1) Domain Adaptation Task: The DA-RDD model is based on Faster RCNN [8] framework with a pretrained ResNet101 [7] on ImageNet as the feature backbone, and we follow the parameter settings in previous works [26, 34, 35, 47]. It is highlighted that we select the output of $conv2\_x$, $conv3\_x$, and $conv4\_x$ layers in ResNet101 backbone as three-level contextual features f_c^1, f_c^2, and f_c^3, respectively. The model is optimized by stochastic gradient descent (SGD), with 0.001 initial learning rate, 0.9 momentum and 0.0005 weight decay parameters.

During model training, data augmentation technique, e.g., image flipping, rotation, translation, is adopted to enhance the model robustness. Due to the unlabeled test data, we only consider the training samples on RDD2020 and CN2021, and split the RDD2020 into Japan2020, India2020, and Czech2020, respectively, as elaborated in Fig.3. For domain adaptation experiments, Japan2020 is only utilized as the source data, which is roughly divided into 9506 training and 1000 validation samples with the proportion of 9: 1. Generally, cross-domain road damage detection tasks are illustrated as follows:

Fig. 5. Samples on RDD2021. We crop the damage patch from RDD2020, and generate the synthetic data via CycleGAN [24], which is further merged into an undamaged image using Poisson Blending (PB).

Japan2020→India2020 (JaIn-DA) We adopt Japan2020 as the source domain and India2020 as the target domain. Specifically, the DA-RDD is trained on 9506 training data and evaluated on 1000 val images with 100 epochs, and we test it on India2020 with 7706 labeled samples to explore the adaptation performance of model across different countries.

Japan2020→Czech2020 (JaCz-DA) Japan2020 and Czech2020 are utilized as the source set and the target set, respectively. Similarly, we train and validate the model on Japan2020 with 80 epochs, and further evaluate the detection accuracy on 4 damage categories of DA-RDD when adapting to Czech2020.

Japan2020→CN2021 (JaCn-DA) The generalization of DA-RDD is also explored between Japan2020 and CN2021. Notably, we make some changes in category name from Japan2020 to CN2021: $D00\&D10 \rightarrow Crack$, $D20 \rightarrow Net$, $D40 \rightarrow Pothole$, and $D43\&D44 \rightarrow Marking$, respectively; others are ignored. The model is trained and verified on Japan2020 data with 80 epochs, and we assess the domain adaptation of DA-RDD from Japan to China.

2) Large-Scale Damage Detection Task: Based on RDD2020, a large-scale road damage dataset named RDD2021 is constructed with 100k labeled pavement distress images generated by CycleGAN. Then, we further adopt the prevalent two-stage Faster RCNN and one-stage YOLOv5 to evaluate the large-scale damage detection performance. Noted

TABLE II

RESULTS ON DOMAIN ADAPTATION FROM JAPAN TO CZECH (**JACZ-DA**). PERFORMACE IS EVALUATED BY THE MEAN AVERAGE PRECISION (mAP) AT THE THRESHOLD OF 0.5 ACROSS 4 DAMAGE CATEGORIES ($D00$, $D10$, $D20$ AND $D40$). NOTED THAT *Oracle* REFERS TO THE MODEL IS TRAINED AND TESTED ON THE SAME LABELED OR SOURCE DOMAIN

Japan2020→Czech2020 (JaCz-DA)					
Methods	$D00$	$D10$	$D20$	$D40$	mAP
F-RCNN	0.2066	0.0621	0.1985	0.1305	0.1494
DA-Faster	0.3237	0.2890	0.2862	0.2771	0.294
SWDA	0.2875	0.3062	0.3117	0.2946	0.3000
P-Faster	0.3428	0.3713	0.3074	0.2590	0.3201
UaDA	0.4419	0.4268	0.3533	0.4097	0.4079
DA-RDD	*0.5747*	*0.5846*	**0.5050**	*0.5496*	*0.5535*
Oracle	**0.5827**	**0.6173**	*0.5048*	**0.5710**	**0.5690**

TABLE III

RESULTS ON DOMAIN ADAPTATION FROM JAPAN TO CHINA (**JACN-DA**). PERFORMACE IS EVALUATED BY THE MEAN AVERAGE PRECISION (mAP) AT THE THRESHOLD OF 0.5 ACROSS 4 DAMAGE CATEGORIES ($Crack$, Net, $Pothole$, AND $Marking$). NOTED THAT *Oracle* REFERS TO THE MODEL IS TRAINED AND TESTED ON THE SAME LABELED OR SOURCE DOMAIN

Japan2020→CN2021 (JaCn-DA)					
Methods	$Crack$	Net	$Pothole$	$Marking$	mAP
F-RCNN	0.3602	0.0624	0.1256	0.3415	0.2224
DA-Faster	0.4456	0.3016	0.3451	0.5210	0.4033
SWDA	0.4949	0.4465	0.3575	0.5132	0.4530
P-Faster	0.5330	0.4129	0.4203	*0.5408*	0.4768
UaDA	0.5922	0.4594	0.4751	0.5060	0.5082
DA-RDD	**0.6786**	*0.5118*	*0.5304*	0.5397	*0.5651*
Oracle	*0.6744*	**0.5326**	**0.5655**	0.5879	**0.5901**

TABLE IV

ABLATION STUDY ON EACH COMPONENT OF DA-RDD ADAPTED FROM JAPAN TO INDIA (**JAIN-DA**). PERFORMACE IS EVALUATED BY THE MEAN AVERAGE PRECISION (mAP) AT THE THRESHOLD OF 0.5 ACROSS 4 DAMAGE CATEGORIES ($D00$, $D10$, $D20$ AND $D40$). NOTED THAT W.INTER-DO AND W.IMWEIGHT DENOTE INTERMEDIATE DOMAIN AND IMPORTANCE WEIGHT, $f_c^* \bigotimes$ AND $f_c^* \bigoplus$ DEFINE CONTEXT INFORMATION AT DIFFERENT LEVELS AND ROI FEATURE INTERACTION

Ablation Study (on **JaIn-DA**)					
Methods	$D00$	$D10$	$D20$	$D40$	mAP
F-RCNN	0.2015	0.0528	0.2319	0.3093	0.1989
w.Inter-Do	0.4877	0.3031	0.3395	0.4680	0.3996
w.ImWeight	0.5649	0.4284	0.3702	0.5117	0.4688
w.$f_c^1 \bigotimes$	0.6061	0.5242	0.4530	0.5629	0.5366
w.$f_c^{(1,2)} \bigotimes$	0.6274	*0.5532*	0.4986	0.5877	0.5667
w.$f_c^{(1,2,3)} \bigoplus$	*0.6303*	0.5515	*0.5040*	*0.6101*	*0.5740*
DA-RDD	**0.6556**	**0.5873**	**0.5199**	**0.6204**	**0.5958**

that the official RDD2020 evaluation server has been closed, and currently, it is unable to make a reason comparison with counterparts under the standard setting on the leaderboard. For simplicity, we randomly select 1000, 700 and 300 samples from Japan2020, India2020 and Czech2020, respectively, which are only utilized for model evaluation.

RDD2021 As mentioned above, CycleGAN has the powerful capability of image-to-image translation, and we individually utilize it for road damage image generation. Notably, we only consider 4 main classes ($D00$, $D10$, $D20$ and $D40$) for Japan2020, India2020 and Czech2020, respectively.

Taking class $D00$ on Japan2020 and India2020 as example, we extract the crack box from each image and resize to the fixed 128×128 patches, marked as $Japan2020\_D00$. Similar process is also conducted on another three categories for each country. Subsequently, a pair of patches with the same class, e.g., $Japan2020\_D00$ and $India2020\_D00$, is fed into CycleGAN for damage sample generation. We further reshape the synthetic patch back into the real size according to the known coordinate provided by Ground-Truth box (GT box), and merge it into the corresponding original image via Poisson Blending (PB) [57]. Consequently, the old distress would be covered by a new synthesis, and a realistic crack sample could be created for augmenting the data diversity. It is highlighted that the key insight of PB algorithm is to keep the image gradient when inserting an image to another one, resulting in a high-quality synthetic image [47].

As tabulated in Fig.3, RDD2021 totally contains 100k labeled synthetic images across different countries, with 40k for Japan2021, 30k for India2021, and 30k for Czech2021, respectively. We observe the extremely imbalanced problem in RDD2020, such as only 89 samples in $India2020\_D10$ and 211 $Czech2020\_D20$ images. Therefore, we generate more instances for these rare damages, and class-imbalanced problem has been alleviated significantly on the RDD2021. The generated samples on RDD2021 can be seen in Fig.5.

Road Damage Detectors We adopt two common object detectors to evaluate the large-scale road damage detection performance. Faster RCNN is a two-stage detection method widely utilized for object detection task, and we train it on RDD2020 and RDD2021 with 80 epochs under the default settings [12], respectively. As an efficient one-stage detector, YOLOv5 presents on par with two-stage detection accuracy at real-time inference speed. More technical details can refer to [19], and likewise, it is trained with 100 epochs on two datasets to evaluate the damage detection performance.

B. Domain Adaptation Performance

We compare the proposed DA-RDD to the four domain adaptative detectors, e.g., DA-Faster RCNN (DA-Faster) [29], Strong-weak Distribution Alignment (SWDA) [35], Progressive Domain Adaptation Faster RCNN (P-Faster) [36], as well as Uncertainty-aware Domain Adaptation (UaDA) [37], to evaluate the cross-domain road damage detection performance. All methods are trained on the source Japan2020,

TABLE V

LARGE-SCALE ROAD DAMAGE DETECTION RESULTS ACHIEVED BY YOLOV5 AND FASTER RCNN ON RDD2020 AND RDD2021. PERFORMACE IS EVALUATED BY THE MEAN AVERAGE PRECISION (mAP) AT THE THRESHOLD OF 0.5 ACROSS 4 DAMAGE CATEGORIES ($D00$, $D10$, $D20$ AND $D40$). NOTED THAT THE RED BOLD FONT DENOTES THE mAP GAINS ON EACH CLASS BETWEEN RDD2020 AND RDD2021

		YOLOv5				Faster RCNN			
		$D00$	$D10$	$D20$	$D40$	$D00$	$D10$	D20	$D40$
RDD2020	Japan	0.5629	0.4267	0.5887	0.6803	0.6786	0.5188	0.8634	0.7323
	India	0.2667	0.0149	0.4322	0.3391	0.6956	0.7054	0.8325	0.7245
	Czech	0.1550	0.0962	0.2368	0.1312	0.6319	0.6682	0.8533	0.7097
	mAP	0.3282	0.1793	0.4192	0.2877	0.6687	0.6308	0.8497	0.7222
RDD2021	Japan	0.9321	0.7390	0.9365	0.9451	0.7431	0.6425	0.8848	0.8365
	India	0.5045	0.2036	0.6829	0.5428	0.8128	0.8650	0.8348	0.8282
	Czech	0.3960	0.1561	0.5064	0.3371	0.7517	0.6916	0.9644	0.8302
	mAP	0.6109	0.3662	0.7086	0.6083	0.7692	0.7330	0.8947	0.8316
mAP improvements		**+0.2827**	**+0.1869**	**+0.2894**	**+0.3206**	**+0.1005**	**+0.1022**	**+0.045**	**+0.1094**

Fig. 6. Domain adaptation results for **JaIn-DA**, **JaCz-DA** and **JaCn-DA**. For comparison, we list the Ground-Truth box (GT box), and the detection results achieved by F-RCNN (Fater RCNN without adaptation), DA-RDD and Oracle methods in each column, respectively. Clearly, DA-RDD presents significant adaptation performance on par with or closed to the Oracle, even in such difficult environment with varying illuminations and layouts. Best viewed in color. (Noted that differences in image size between countries, i.e., 600 × 600 in RDD2020 and 1920 × 1080 for CN2021, lead to the predicted box with various thickness).

and we report the domain adaptative road damage detection performance on the target domain. Notably, the F-RCNN indicates the trained model directly transfers to the target domain without adaptation methods, while the *Oracle* implies the model is trained and tested on the same labeled or source domain, e.g., trained on Japan2020 and tested on Japan2020.

Fig. 7. Visualizations on RDD2020 and RDD2021, respectively. For each row, we present the Ground-Truth box (GT box), the predicted boxes on RDD2020 and RDD2021 achieved by YOLOv5 [19] and Faster RCNN [12]. Obviously, the model suffers from low-confidence predictions or false negatives due to the scarce of sample on RDD2020, while RDD2021 provides a large amount of damage types and appearances, facilitating the more accurate and robust detection performance. Best viewed in color.

JaIn-DA Results As tabulated in Table I, our method shows the compelling domain adaptation performance with the mAP of 0.5958, which surpasses all comparisons by a substantial margin. DA-RDD provides at least doubled detection accuracy gains over the baseline F-RCNN, and more importantly, it fairly outperforms the *Oracle* on the $D00$ (0.6556 vs. 0.6038) and $D40$ (0.6204 vs. 0.6127) categories, as well as the mAP metric (0.5958 vs. 0.5886). The remarkable adaptation performance gains indicate the effectiveness and supriority of the proposed method.

JaCz-DA Results Table II lists the adaption results between Japan and Czech. It is observed that DA-RDD also achieves

better cross-domain performance than other adaptation methods, e.g., 0.5535 (DA-RDD) vs. 0.4079 (UaDA). Moreover, it is inferior to the $Oracle$ method by 0.0155 mAP, despite a slight improvement is reported on the $D20$ class (0.5050 vs. 0.5048). Generally, DA-RDD exhibits the considerable robustness to a new challenging environment, and generalizes well to the unseen scenario with the marginal performance degradation.

JaCn-DA Results It is much difficult to detect the road damage because of the severe class-imbalanced problem in CN2021. As demonstrated in Table III, F-RCNN reports poor mAP performance particularly in Net (0.0624) and $Pothole$ (0.1256), due to the scarce of damage sample. DA-RDD significantly facilitates the baseline in all evaluation metric, and still offers the outstanding boosts over other adaptation methods. Furthermore, the performance of DA-RDD approaches to the $Oracle$ with only 0.025 mAP apart (0.5651 vs. 0.5901), which further suggests the unversality and generalization.

Finally, domain adaptation results across different countries are visualized in Fig.6. For simplicity, we refer to the ground-truth box (GT box), and list the predictions achieved by F-RCNN (Faster RCNN without adaptation), DA-RDD and the Oracle methods. Intuitively, DA-RDD shows the significant cross-domain road damage detection performance on par with or closed to the Oracle, particularly in the challenging scenarios with different illuminations and layouts.

C. Ablation Study

Further, ablation studies are conducted to investigate the effect of each component on domain adaptation. For simplicity, all ablation experiments are performed on the adaptation from Japan to India (JaIn-DA). We adopt F-RCNN [12] as the baseline, and introduce image-level and instance-level feature alignment methods to measure the performance gain, respectively.

1) Image-Level Feature Alignment: We incorporate the baseline F-RCNN with an intermediate domain (Inter-Do) and importance weighting (ImWeight) strategy. As shown in Table IV, Inter-Do substantially improves the baseline by 0.2007 mAP (0.1989 $\rightarrow$ 0.3996), especially 0.2503 AP gains in $D10$, demonstrating the superiority of interpolated feature space. It bridges the domain gap between the source and the target features, and alleviates the difficulty of adaptation task. Besides, ImWeight receives a considerable mAP boost (0.3996 $\rightarrow$ 0.4688), which suggests the discriminative feature is beneficial to cross-domain feature alignment.

2) Instance-Level Feature Alignment: Based on image-level feature alignment, we further consider contextual feature (f_c^*) at different scales and RoI feature interaction ($\bigotimes$ or $\bigoplus$) method. As listed in Table IV, concatenating low-level context information, e.g., f_c^1 and $f_c^{(1,2)}$, with RoI feature can significantly facilitate the adaptation performance, with 0.0678 (0.4688 $\rightarrow$ 0.5366) and 0.0301 (0.5366 $\rightarrow$ 0.5667) mAP improvements respectively. It is reasonable that more object details, i.e., color and shape, can be exploited in the shallow layer to recover the damage instance. More importantly, results in the last columns claim that tensor product is more effective for RoI feature interaction when aggregating with contextual representations at three resolutions (0.5958 vs. 0.5740). In general, ablation studies verify the effectiveness and advancement of proposed methods in domain adaptation task.

D. Large-Scale Damage Detection Performance

Additionally, large-scale road damage detection using YOLOv5 [19] and Faster RCNN [12] detectors are performed on RDD2020 and RDD2021, and remarkable detection performance gains on RDD2021 can be shown in the Table V.

In each row, we observe considerable performance increments in both two methods between different datasets. Specifically, YOLOv5 receives 0.1869 $\sim$ 0.3206 mAP boosts from 4 distress classes, meanwhile Faster RCNN reports 0.1 mAP gains on average on RDD2021. For each column, Faster RCNN always presents much better performance in all damage categories on both RDD20220 and RDD2021, such as 0.3282 mAP (YOLOv5) vs. 0.6687 mAP (Faster RCNN) on RDD2020_D00, 0.3662 mAP (YOLOv5) vs. 0.7330 mAP (Faster RCNN) on RDD2021_D10. More visualization results can refer to the Fig.7. Generally, the substantial detection performance improvements on various crack types for different methods demonstrate the effectiveness and superiority of large-scale road data, and we expect RDD2021 would facilitate the empirical studies on road damage detection in the future.

V. Conclusion

In this paper, we firstly incorporate unsupervised domain adaptation (UDA) with crack identification, and propose DA-RDD for cross-domain road damage detection. To be specific, image-level feature alignment across different domains is developed with an intermediate space and feature importance weighting, and instance-level feature alignment further aggregates RoI feature with multi-scale contextual information to recover the instance details for progressive data distribution alignment between domains. Extensive domain adaptation experiments are conducted on four countries from RDD2020 and CN2021. DA-RDD reports compelling and advanced adaptation performance on par with or closed to the full supervision $Oracle$ method, and provides substantial performance gains over the baseline without adaptation and other counterpart methods by a significant margin, indicating its effectiveness and superiority. Additionally, we further develop a large-scale road damage detection dataset (based on Road Damage Dataset 2020 (RDD2020) [1]) termed as RDD2021, that comprises 100k synthetic images with bounding-box annotations on 4 main crack categories. Empirical studies demonstrate the massive pavement distress data can significantly facilitate the road damage detection performance.

In the future, we attempt to develop more effective and efficient domain adaptation in the context of road damage detection, and further explore feature alignment method for universal cross-domain detection. Furthermore, more pavement images would be generated to augment the rare crack type, i.e., $D01$, $D11$, and we expect the large-scale road damage dataset is beneficial to the development of related

works. Last but not least, more efforts on extensive domain adaptative tasks would be also investigated based on the proposed intermediate domain and feature adversarial learning strategies.

REFERENCES

[1] D. Arya et al., "Deep learning-based road damage detection and classification for multiple countries," *Autom. Construct.*, vol. 132, Dec. 2021, Art. no. 103935.

[2] National Highway Traffic Safety Administration. (Jul. 2008). *National Motor Vehicle Crash Causation Survey: Report to Congress*. [Online]. Available: https://trid.trb.org/view/880434

[3] M. Hnewa and H. Radha, "Object detection under rainy conditions for autonomous vehicles: A review of state-of-the-art and emerging techniques," *IEEE Signal Process. Mag.*, vol. 38, no. 1, pp. 53–67, Jan. 2020.

[4] Z. Liu et al., "Robust target recognition and tracking of self-driving cars with radar and camera information fusion under severe weather conditions," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6640–6653, Jul. 2022.

[5] A. Bell, T. Mantecon, C. Diaz, C. R. Del-Blanco, F. Jaureguizar, and N. Garcia, "A novel system for nighttime vehicle detection based on foveal classifiers with real-time performance," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5421–5433, Jun. 2022.

[6] Y. Cai et al., "YOLOv4–5D: An effective and efficient object detector for autonomous driving," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 4503613.

[7] R. Medina, J. Llamas, E. Zalama, and J. Gomez-Garcia-Bermejo, "Enhanced automatic detection of road surface cracks by combining 2D/3D image processing techniques," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 778–782.

[8] R. Kapela et al., "Asphalt surfaced pavement cracks detection based on histograms of oriented gradients," in *Proc. 22nd Int. Conf. Mixed Design Integr. Circuits Syst. (MIXDES)*, Jun. 2015, pp. 579–584.

[9] M. Quintana, J. Torres, and J. M. Menéndez, "A simplified computer vision system for road surface inspection and maintenance," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 3, pp. 608–619, Mar. 2016.

[10] X. Liu, F. Xue, and L. Teng, "Surface defect detection based on gradient LBP," in *Proc. IEEE 3rd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2018, pp. 133–137.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[14] A. Zhang et al., "Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network," *J. Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 10, pp. 805–819, Oct. 2017.

[15] A. Zhang et al., "Automated pixel-level pavement crack detection on 3D asphalt surfaces with a recurrent neural network," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 34, no. 3, pp. 213–229, Mar. 2019.

[16] A. Zhang et al., "Deep learning–based fully automated pavement crack detection on 3D asphalt surfaces with an improved CrackNet," *J. Comput. Civil Eng.*, vol. 32, no. 5, pp. 04018041.1–04018041.14, Sep. 2018.

[17] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, "DeepCrack: Learning hierarchical convolutional features for crack detection," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1498–1512, Mar. 2018.

[18] F. Yang, L. Zhang, S. Yu, D. V. Prokhorov, X. Mei, and H. Ling, "Feature pyramid and hierarchical boosting network for pavement crack detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1525–1535, Apr. 2020.

[19] Ultraltyics. (2020). *YOLOv5*. [Online]. Available: https://github.com.ultralytics/yolov5

[20] A. Alfarrarjeh, D. Trivedi, S. H. Kim, and C. Shahabi, "A deep learning approach for road damage detection from smartphone images," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 5201–5204.

[21] F. Kluger et al., "Region-based cycle-consistent data augmentation for object detection," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 5205–5211.

[22] W. Wang, B. Wu, S. Yang, and Z. Wang, "Road damage detection and classification with faster R-CNN," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 5220–5223.

[23] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road damage detection and classification using deep neural networks with smartphone images," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 12, pp. 1127–1141, Jun. 2018.

[24] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.

[25] Q. Zou, Y. Cao, Q. Li, Q. Mao, and S. Wang, "CrackTree: Automatic crack detection from pavement images," *Pattern Recognit. Lett.*, vol. 33, no. 3, pp. 227–238, Feb. 2012.

[26] M. Eisenbach et al., "How to get pavement distress detection ready for deep learning? A systematic approach," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2039–2047.

[27] Jiangsu Province. (2021). *The Big Data Development and Application Competition of Jiangsu Province (Hualu Cup)*. [Online]. Available: https://www.marsbigdata.com/competition/details?id=21084070893568/

[28] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 1521–1528.

[29] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster R-CNN for object detection in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3339–3348.

[30] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2009.

[31] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 1180–1189.

[32] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3723–3732.

[33] C. Chen et al., "Progressive feature alignment for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 627–636.

[34] G. Kang, L. Jiang, Y. Yang, and A. G. Hauptmann, "Contrastive adaptation network for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4893–4902.

[35] K. Saito, Y. Ushiku, T. Harada, and K. Saenko, "Strong-weak distribution alignment for adaptive object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6956–6965.

[36] H.-K. Hsu et al., "Progressive domain adaptation for object detection," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 749–757.

[37] D. Guan, J. Huang, A. Xiao, S. Lu, and Y. Cao, "Uncertainty-aware unsupervised domain adaptation in object detection," *IEEE Trans. Multimedia*, vol. 24, pp. 2502–2514, 2021.

[38] C. Chen, Z. Zheng, X. Ding, Y. Huang, and Q. Dou, "Harmonizing transferability and discriminability for adapting object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8870–8878.

[39] J. Hoffman et al., "CyCADA: Cycle-consistent adversarial domain adaptation," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2018, pp. 1994–2003.

[40] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 7472–7481.

[41] X. Zhu, H. Zhou, C. Yang, J. Shi, and D. Lin, "Penalizing top performers: Conservative loss for semantic segmentation adaptation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 568–583.

[42] Y. Fei et al., "Pixel-level cracking detection on 3D asphalt pavement images through deep-learning-based CrackNet-V," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 273–284, Jan. 2020.

[43] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2016.

[44] D. Arya et al., "Global road damage detection: State-of-the-art solutions," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2020, pp. 5533–5539.

[45] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, "Automatic road crack detection using random structured forests," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3234–3445, Dec. 2016.

[46] TachibanaYoshino. (2019). *Road-Crack-Segmentation-Keras*. [Online]. Available: http://github.com/TachibanaYoshino/Road-Crack-Segmentation–Keras/
[47] H. Maeda, T. Kashiyama, Y. Sekimoto, T. Seto, and H. Omata, "Generative adversarial network for road damage detection," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 36, no. 1, pp. 47–60, 2020.
[48] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 97–105.
[49] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 443–450.
[50] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Conf. Workshop Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 2672–2680.
[51] X. Zhu, J. Pang, C. Yang, J. Shi, and D. Lin, "Adapting object detectors via selective cross-domain alignment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 687–696.
[52] Q. Cai, Y. Pan, C.-W. Ngo, X. Tian, L. Duan, and T. Yao, "Exploring object relation in mean teacher for cross-domain detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11457–11466.
[53] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Mach. Learn.*, vol. 79, nos. 1–2, pp. 151–175, May 2010.
[54] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[55] P. Kar and H. Karnick, "Random feature maps for dot product kernels," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2012, pp. 583–591.
[56] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Proc. Conf. Workshop Neural Inf. Process. Syst. (NIPS)*, 2018, pp. 1640–1650.
[57] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 313–318, Jul. 2003.

Chunmian Lin is currently pursuing the Ph.D. degree with the School of Transportation Science and Engineering, Beihang University, Beijing, China. His current research interests include autonomous driving, image processing, computer vision, artificial intelligence, and deep learning, particularly their applications in intelligent transportation systems.

Daxin Tian (Senior Member, IEEE) is currently a Professor with the School of Transportation Science and Engineering, Beihang University, Beijing, China. His current research interests include mobile computing, intelligent transportation systems, vehicular ad hoc networks, and swarm intelligence. He is an IEEE Intelligent Transportation Systems Society Member and an IEEE Vehicular Technology Society Member.

Xuting Duan (Member, IEEE) received the Ph.D. degree in traffic information engineering and control from Beihang University, Beijing, China. He is currently an Assistant Professor with the School of Transportation Science and Engineering, Beihang University. His current research interests include vehicular ad hoc networks, cooperative vehicle infrastructure systems, and the Internet of Vehicles.

Jianshan Zhou received the B.Sc. and M.Sc. degrees in traffic information engineering and control from Beihang University, Beijing, China, in 2013 and 2016, respectively, where he is currently pursuing the Ph.D. degree with the School of Transportation Science and Engineering. His current research interests include wireless communications, artificial intelligent systems, and intelligent transportation systems.

Dezong Zhao (Senior Member, IEEE) received the B.Eng. and M.S. degrees in control science and engineering from Shandong University, Jinan, China, in 2003 and 2006, respectively, and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2010. He is currently a Senior Lecturer in autonomous systems with the School of Engineering, University of Glasgow, U.K. His research interests include connected and autonomous vehicles, machine learning, and control engineering. His work has been recognized by being awarded an EPSRC Innovation Fellowship and a Royal Society-Newton Advanced Fellowship in 2018 and 2020, respectively.

Dongpu Cao received the Ph.D. degree from Concordia University, Canada, in 2008. He is currently the Canada Research Chair in driver cognition and automated driving, and also an Associate Professor and the Director of the Waterloo Cognitive Autonomous Driving (CogDrive) Laboratory, University of Waterloo, Canada. He has contributed more than 200 articles and three books. His current research interests include driver cognition, automated driving, and cognitive autonomous driving. He received the SAE Arch T. Colwell Merit Award in 2012, the IEEE VTS 2020 Best Vehicular Electronics Paper Award, and three Best Paper Awards from the ASME and IEEE Conferences. He serves on the SAE Vehicle Dynamics Standards Committee and acts as the Co-Chair of the IEEE ITSS Technical Committee on Cooperative Driving. He serves as the Deputy Editor-in-Chief for *IET Intelligent Transport Systems Journal*, and an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE/ASME TRANSACTIONS ON MECHATRONICS, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, IEEE/CAA JOURNAL OF AUTOMATICA SINICA, IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, and *ASME Journal of Dynamic Systems, Measurement and Control*. He was a Guest Editor for *Vehicle System Dynamics*, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, and IEEE INTERNET OF THINGS JOURNAL. He is an IEEE VTS Distinguished Lecturer.