Joint Training of Image Generator and Detector for Road Defect Detection

Kuan-Chuan Peng*
Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA
kpeng@merl.com

Abstract

*Road defect detection is important for road authorities to reduce the vehicle damage caused by road defects. Considering the practical scenarios where the defect detectors are typically deployed on edge devices with limited memory and computational resource, we aim at performing road defect detection without using ensemble-based methods or test-time augmentation (TTA). To this end, we propose to **J**ointly **T**rain the image **G**enerator and **D**etector for road defect detection (dubbed as JTGD). We design the dual discriminators for the generative model to enforce both the synthesized defect patches and overall images to look plausible. The synthesized image quality is improved by our proposed CLIP-based Fréchet Inception Distance loss. The generative model in JTGD is trained jointly with the detector to encourage the generative model to synthesize harder examples for the detector. Since harder synthesized images of better quality caused by the aforesaid design are used in the data augmentation, JTGD outperforms the state-of-the-art method in the RDD2022 road defect detection benchmark across various countries under the condition of no ensemble and TTA. JTGD only uses less than 20% of the number of parameters compared with the competing baseline, which makes it more suitable for deployment on edge devices in practice.*

1. Introduction

Potholes and road defects are menacing threats to the safety of the road users, including the drivers, bicyclists, and pedestrians. In the period spanning 2018 to 2020, these road imperfections tragically claimed the lives of more than 5000 individuals in India [2]. In addition, out of every ten drivers, one experienced vehicle damage substantial to necessitate repairs following encounters with potholes, which resulted in an average repair cost nearing $600 per incident, culminating in a staggering total of $26.5 billion in damages for the year 2021 alone [3]. The aforesaid statistics makes detecting road defects an indispensable task for the road authorities and motivates road damage detection challenges in computer vision, *e.g.*, the CRDDC 2022 challenge [5].

*Acknowledgement: This work was done in collaboration with Ramy Mounir (Univ. of South Florida; ramy@usf.edu) when he interned at MERL.

The state-of-the-art (SOTA) defect detection methods and the top performing methods of the CRDDC 2022 challenge often require ensemble from multiple models [7, 9, 12, 17, 22] or performing test-time augmentation (TTA) [9, 22]. Given the real-world scenarios where the defect detectors are commonly deployed on edge devices characterized by constrained memory and computational resources (*e.g.*, no GPU) [18], using ensemble-based method or performing TTA is impractical due to their computational overhead at the testing time, and the trained model ideally needs to be lightweight. Therefore, we challenge ourselves the following problem: *Without using any ensemble-based techniques or any form of test-time augmentation, what can practitioners do to improve the performance of road defect detection using fewer number of model parameters at the testing time*?

One common technique to improve the performance of the road defect detectors is to perform data augmentation [9, 12, 17, 21]. Nevertheless, these methods typically only use limited predefined low-level augmentation techniques such as color jitter, horizontal flip, *etc*. Some methods use generative models to synthesize realistic and diverse examples [15, 25]. However, the models trained by these methods are only trained to optimize the synthesized defect patch to be plausible, but they are not trained to make the entire image plausible as well, *e.g.*, [15, 25] use the Poisson blending to blend the synthesized defect patch into the background image post hoc. [15] even identifies that such blending method can be the cause of their suboptimal performance. To resolve the issue in these works and optimize the entire image to be plausible at the training time, we propose to use the dual discriminator architecture in the generative model such that the realism of both the synthesized defect patch and the overall image (with the synthesized defect in it) is enforced.

Another issue of using the generative model which is totally overlooked by the road defect detection literature is that the generative model is typically trained separately from the detector in advance. Due to such practice, the generative model in the prior works is unaware of the existence of the detector and its objective (*i.e.*, detect the defects). The only objective of the generative model in the prior works is to synthesize plausible images which have no guarantee to be beneficial to the detector's performance, and hopefully such implicit objective can enhance the detector's

performance. Instead of following the suboptimal practice of the prior works, we propose to train the generative model and detector **jointly** such that enhancing the detector's performance becomes part of the major objective of the generative model, which we believe is a novel and strong departure from the prior works. Specifically, we use the concept of hard example mining and propose a novel loss term for the generative model to encourage the generative model to synthesize more challenging examples that the detector fails to detect.

Ideally, the images synthesized by the generative models should be of the same distribution as the real training data for the purpose of data augmentation. Therefore, we expect that the quality of the synthesized images and that of the real training images should be as close as possible. To further improve the synthesized image quality, we propose to minimize the Fréchet inception distance (FID) [11] as part of the optimization when training the generative model. Although FID has become a standardized evaluation metric to assess image quality, surprisingly, we are unaware of any defect detection works which explicitly optimize FID during training to improve the synthesized image quality for the data augmentation purpose.

To the best of our knowledge, we are the first in road defect detection to jointly train the generative model and detector, use dual discriminators, and explicitly optimize the FID as part of the objectives when training a generative model for data augmentation. We make the following contributions:

1. We propose to **J**ointly **T**rain an image **G**enerator and **D**etector for the task of road defect detection (dubbed as JTGD). Under the condition of no ensemble-based techniques or test-time augmentation, JTGD outperforms the state-of-the-art (SOTA) defect detection method on the RDD2022 dataset [6] in most cases of country-wise and overall performance across six countries.
2. For road defect detection, we propose the hard example synthesis loss for jointly training the generative model and detector, the CLIP-based Fréchet Inception Distance loss, and the dual discriminator architecture to improve the image synthesis quality of the generative model.
3. Outperforming the SOTA baseline, JTGD only uses less than 20% of the test-time model parameters compared with the SOTA baseline, which supports that JTGD is a more suitable defect detection method to be deployed on the edge devices in the real-world applications where the memory and computational resources are limited.
4. We provide both the qualitative and quantitative ablation study to support that jointly training the generative model and detector, and enhancing the synthesized image quality using our proposed techniques can improve the defect detection performance.

2. Proposed Method – JTGD

Our proposed method JTGD is illustrated in Fig. 1, where we propose to use the dual discriminators (*i.e.*, image discriminator D_i and patch discriminator D_p) and the Fréchet inception distance loss $\mathcal{L}_{\text{FID}}$ with the CLIP [19] features to improve the quality of the images synthesized by the image generator ψ for the data augmentation purpose. Furthermore, ψ is trained jointly with the detector D such that the detection result from D serves as the feedback for ψ to encourage ψ to generate harder examples for D to detect. In Fig. 1, we only show the part of training the generative model jointly with the detector for clarity. Given a set of clean (defect-free) real images I_c^r and their masks M indicating where and what types of the defects we would like to add to I_c^r (the number in the box of M in Fig. 1 illustrates the defect type ID), the goal of the image generator ψ is to synthesize the fake images I^f with the defect types and locations specified in M such that the defects seamlessly blend into the same background of I_c^r. Different from the prior works [15, 25] which only employ one discriminator (D_p in Fig. 1) to enforce that the synthesized defect patches are indistinguishable from the defect patches of the real training images, we add an additional image discriminator D_i to enforce that the entire I^f are also indistinguishable from the entire real defected images I_d^r. Such dual discriminator design encourages not only the synthesized defect patches but also the overall synthesized defected images to be plausible. Formally, the objective function of D_i (dubbed as the image discriminator loss $\mathcal{L}_{id}$) can be written as:

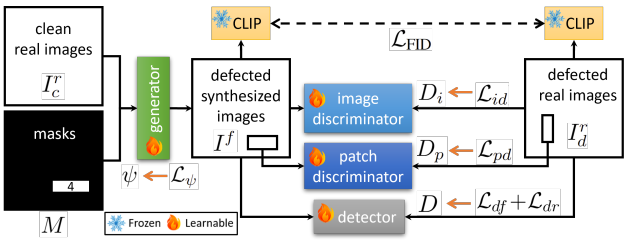

Figure 1. The overview of our method JTGD, where we propose to use the dual discriminators (D_i and D_p) and the Fréchet inception distance loss $\mathcal{L}_{\text{FID}}$ computed using the CLIP features (frozen) to improve the quality of the images synthesized by the generator ψ for data augmentation when training the defect detector D. We train ψ and D **jointly** such that ψ can learn to synthesize harder examples based on D's detection results, which makes it possible for ψ to update itself based on D's objective (defect detection). The orange arrows point to the locations where each loss is enforced on. See Sec. 2 for details.

$$\mathcal{L}_{id} = -\mathbb{E}_{r_i \sim R_i}[\log D_i(r_i)] - \mathbb{E}_{f_i \sim F_i}[\log(1 - D_i(f_i))], \quad (1)$$

where R_i is the data distribution of I_d^r, and F_i is the data distribution of I^f. The objective function of D_p (dubbed as the patch discriminator loss $\mathcal{L}_{pd}$) can be written as:

$$\mathcal{L}_{pd} = -\mathbb{E}_{r_p \sim R_p}[\log D_p(r_p)] - \mathbb{E}_{f_p \sim F_p}[\log(1 - D_p(f_p))], \quad (2)$$

where R_p and F_p denote the distributions of real defect patches in I_d^r and synthesized defect patches in I^f, respectively.

To further improve the synthesized image quality, we propose to add the Fréchet inception distance loss $\mathcal{L}_{\text{FID}}$ as an additional objective of ψ. The goal of $\mathcal{L}_{\text{FID}}$ is to encourage the

statistics of I^f and I^r_d to be the same. Inspired by the recent finding [13] that the FID evaluated using the ImageNet [20] features does not necessarily reflect the image quality, we modify the existing ImageNet-based FID loss implementation [16] such that the FID is evaluated using the CLIP [19] features, as suggested by [13]. To be clear, we are the first in road defect detection which utilizes the foundation model (FM) features (*e.g.*, CLIP) to improve data augmentation quality, and hence the final performance. Most of the prior works which use the concept similar to FID did not even consider leveraging the large FMs nowadays but still use the ImageNet based features which are proven to fail to reflect image quality [13], which clearly separates `JTGD` from existing works.

In addition to the aforesaid design, the most distinguishing characteristic of `JTGD` is that ψ is **jointly** trained with the detector D such that ψ can adjust the synthesized image quality according to the feedback given by D. This design improves over the traditional data augmentation pipeline where the image generator and detector are trained **separately** (*i.e.*, an image generator is trained first, and a detector is trained later using the images synthesized by the already trained and frozen image generator). One clear drawback of the traditional data augmentation pipeline is that the image generator is totally unaware of the objective of the task of interest (*e.g.*, road defect detection) because the objective of the image generator in the traditional data augmentation pipeline typically only encourages the image generator to synthesize plausible images (which is not directly related to our task of interest), instead of encouraging the image generator to *synthesize the images which will be beneficial to the objective of the task of interest*. To the best of our knowledge, we are the first in road defect detection to propose jointly training the generative model and detector to optimize the detector's performance, which is a strong departure from prior works in road defect detection.

To realize the aforementioned idea and let ψ and D be trained jointly, we feed I^f and I^r_d as the input of D and supervise D with the detection losses $\mathcal{L}_{df}$ and $\mathcal{L}_{dr}$, respectively, where $\mathcal{L}_{df}$ and $\mathcal{L}_{dr}$ are the same detection loss as that used in the Faster Swin [9] except that they correspond to different inputs (I^f and I^r_d). Since we use ψ to synthesize I^f on the fly during training, we know exactly where the synthesized defects are placed in I^f, which provides the supervision signal for $\mathcal{L}_{df}$. The supervision signal for $\mathcal{L}_{dr}$ is directly from the ground truth of the training data. Given $\mathcal{L}_{df}$ and $\mathcal{L}_{dr}$, we propose the following hard example synthesis loss $\mathcal{L}_h$ as an additional loss term to supervise ψ:

$$\mathcal{L}_h = -\mathcal{L}_{df} - \mathcal{L}_{dr}. \tag{3}$$

$\mathcal{L}_h$ encourages ψ to synthesize harder examples where D fails to detect the defects, which is similar to the spirit of hard example mining. Once the synthesized hard examples are added to the training set and we retrain D, the performance of the retrained D is expected to improve because its ability to detect harder examples is enhanced.

Formally, the final objective function of ψ (dubbed as $\mathcal{L}_\psi$) can be written as: $\mathcal{L}_\psi = \mathcal{L}_g + w_h\mathcal{L}_h + w_{\text{FID}}\mathcal{L}_{\text{FID}}$, where w_h and w_{FID} are the weights of $\mathcal{L}_h$ and $\mathcal{L}_{\text{FID}}$, respectively, and

$$\mathcal{L}_g = \mathbb{E}_{f_i \sim F_i}[\log(1 - D_i(f_i))] + \mathbb{E}_{f_p \sim F_p}[\log(1 - D_p(f_p))]. \tag{4}$$

After training ψ, we use it to synthesize defected images to augment the training set for D. D is then retrained with the union of the original real training set and the set of the synthesized training images using the same detection loss as that used in the Faster Swin [9] such that the retrained D can perform well on not only the original training images but also more challenging synthesized images, which we expect to enhance D's generalization ability on the unknown testing set.

3. Experimental setup

We experiment on the RDD2022 road damage detection dataset [6], one of the largest public datasets for road defect detection, associated with the CRDDC 2022 challenge [5]. Following the evaluation protocol of the CRDDC 2022 challenge [5], we present the statistics of the RDD2022 dataset and the details of the evaluation protocol in the Apendix.

We use the InternImage-T [23] as the backbone of the detector because it is one of the SOTA lightweight backbones across general object detection benchmarks such as PASCAL VOC [10] and COCO [14]. We adapt the CycleGAN [26] as the backbone of the generative model such that it can take both a defect-free image and a mask indicating where to add synthesized defects as input and implement dual discriminators in `JTGD` using the default discriminator architecture in CycleGAN. We use the road images in the RDD2022 training set which do not have any defect annotation as the defect-free images to train the generator. We perform drivable area detection as the preprocessing step to ensure that the synthesized defects are placed within the road areas (details are in the Appendix). We train the modified CycleGAN (generator) from scratch with dual discriminators and $\mathcal{L}_{\text{FID}}$ until convergence. Once the generator is trained, we use it to synthesize road images with defects for data augmentation. We pre-train the detector using the checkpoint released by InternImage [1] and finetune it with the RDD2022 training data and the synthesized data from the generator until convergence.

For all the experiments, we follow the default experimental settings and parameters of the InternImage [23] and CycleGAN [26] unless otherwise specified. We put all the other implementation details in the Appendix.

4. Experimental result

We first conduct an ablation study showing the efficacy of each component in `JTGD` and summarize our experimental results in Table 2, where we refer to each row by the experiment ID $E_1 \sim E_5$. Except the column of parameter number (reported in million parameters), all the other numbers are reported in

experiment ID	method	# parameters (M)↓	India	Japan	Norway	USA	mean (left 4 countries)	overall (6 countries)
E_1	Faster Swin	253	39.82	60.57	34.67	**66.71**	50.44	59.20
E_5	JTGD	**49**	**39.86**	**61.88**	**47.28**	66.24	**53.82**	**63.13**

Table 1. JTGD outperforms the top-performing method Faster Swin [9] in average F1 scores across 4 listed countries and overall across 6 (including Czech Republic and China) countries in the CRDDC 2022 challenge [5], while using <20% of its parameters. All F1 values (%) are from the RDD2022 test set, evaluated by the challenge server.

experiment ID	method	ψ	$\mathcal{L}_{\text{FID}}$	$\mathcal{L}_h$	# parameters (M)↓	overall (6 countries)
E_1	Faster Swin	N/A			253	59.20 (+0)
E_2	JTGD	✗	✗	✗	**49**	61.47 (+2.27)
E_3		✓	✗	✗	**49**	62.52 (+3.32)
E_4		✓	✓	✗	**49**	62.94 (+3.74)
E_5		✓	✓	✓	**49**	**63.13 (+3.93)**

Table 2. The ablation study comparing JTGD with the top-performing method Faster Swin [9] in the CRDDC 2022 challenge [5]. The numbers (except the parm. column) are the F1 values (%) along with the gain over the Faster Swin on the RDD2022 testing set evaluated by the challenge server. Despite using less than 20% of Faster Swin's parameters, JTGD outperforms it on the average 6-country performance. Notations: ψ: the image generative model; $\mathcal{L}_{\text{FID}}$: the Fréchet Inception Distance loss; $\mathcal{L}_h$: the hard example synthesis loss.

the F1 values (%) on the RDD2022 testing set evaluated by the CRDDC 2022 challenge server. E_1 shows the performance of the SOTA method Faster Swin [9], but E_5 shows the final performance of our proposed JTGD. $E_2 \sim E_5$ show that: (1) JTGD outperforms the SOTA method; (2) each component in JTGD contributes to the final performance; (3) using all of our proposed components reaches the best overall F1 value.

The fact that JTGD only use <20% of the number of parameters at the testing time compared with the Faster Swin shows that the performance improvement of JTGD over the Faster Swin definitely does not come from using more parameters. Instead, the design choices of JTGD, including using the InternImage-T [23] backbone and the generative model, and the introduction of $\mathcal{L}_{\text{FID}}$, and jointly training the generative model and detector with $\mathcal{L}_h$ all contribute to the final performance of JTGD. Comparing E_2 and E_1, we show that the usage of the InternImage-T backbone in JTGD already improves the performance over the Faster Swin, even under fewer number of parameters. The ablation study of JTGD presented in $E_2 \sim E_5$ shows that the usage of the generative model ψ, training ψ with the proposed Fréchet Inception Distance loss $\mathcal{L}_{\text{FID}}$, and jointly training ψ and the detector with $\mathcal{L}_h$ further enhance the performance of JTGD in the overall F1 value. E_5 shows that using all the aforesaid components reaches the best overall F1 value across the 6 countries. In the Appendix, we also show the qualitative ablation study of using $\mathcal{L}_{\text{FID}}$.

Since the CRDDC 2022 challenge server also evaluates the country-wise performance of the four countries (India, Japan, Norway, and USA), we compare the country-wise performance of JTGD versus that of the Faster Swin and report the result in Table 1, where JTGD outperforms the Faster Swin in 3 out of 4 country-wise performance and the mean of these 4 country-wise performance. The fact that E_5 is better than E_1 in most cases shows that JTGD outperforms the SOTA method in country-wise and overall performance, regardless of whether we include the Czech Republic and China testing data for evaluation or not. Please notice that this evaluation is done on the held-out testing set (which is not publicly accessible) on the CRDDC 2022 challenge server, so unfortunately it is impossible for us to diagnose the failure cases from the testing data to understand why JTGD does not outperform the Faster Swin in the USA testing data. However, we argue that in the USA testing data, JTGD still performs competitively to the SOTA method Faster Swin (0.47 F1 score difference) while using only <20% of the number of parameters. Furthermore, JTGD outperforms the Faster Swin by much larger margins 1.31, 12.61, 3.38, 3.93 F1 scores in Japan, Norway, 4-country average, and 6-country average.

In addition, E_5 versus E_4 in Table 1 shows the efficacy of jointly training ψ and D using our proposed $\mathcal{L}_h$, which supports that the data augmentation pipeline (*i.e.*, ψ) used for object detection is better informed by the information from the detection objective instead of just being trained separately from the detector (*i.e.*, D) in advance, as is typically done in the prior works of road defect detection. Although we only use a simple hard example mining loss $\mathcal{L}_h$ for jointly training ψ and D, the joint training strategy has the potential to incorporate more advanced design to train ψ from the feedback given by D (*e.g.*, explicitly utilizing the confidence and accuracy of each detection output by D to update the defect class frequency of ψ's image synthesis strategy), which is part of our planned future work.

5. Conclusion

We propose JTGD, a novel method for road defect detection that jointly trains a generative model and detector with dual discriminators and a CLIP-based Fréchet inception distance loss to enhance synthesized image quality and detector performance. Under the condition of not using the ensemble-based techniques or test-time augmentation to increase the test-time computational overhead, JTGD outperforms the state-of-the-art methods on the RDD2022 benchmark across most country-wise and overall performance across six countries, while using less than 20% of test-time parameters compared to baselines—making it well-suited for resource-constrained edge devices. Our ablations show that joint training the generative model and detector, improved image synthesis, and data augmentation boost the defect detection performance. Moreover, the dual discriminator design and CLIP-based Fréchet inception distance loss in JTGD can serve as general enhancements for detection tasks beyond road defects.

References

[1] Internimage pre-trained model. https://tinyurl.com/2p825mt6. 3

[2] News from the Federal. https://tinyurl.com/r39bdj8a,. 1

[3] AAA Newsroom. https://tinyurl.com/ys9mn8b4,. 1

[4] PyTorch. https://pytorch.org/. 1

[5] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, H. Omata, T. Kashiyama, and Y. Sekimoto. Crowdsensing-based road damage detection challenge (CRDDC-2022). *arXiv preprint arXiv:2211.11362*, 2022. 1, 3, 4

[6] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto. RDD2022: A multi-national image dataset for automatic road damage detection. *arXiv preprint arXiv:2209.08538*, 2022. 2, 3, 1

[7] M. Bhavsar, A. Alfarrarjeh, U. Baranwal, and S. H. Kim. Country-specific ensemble learning: A deep learning approach for road damage detection. In *IEEE International Conference on Big Data (Big Data)*, 2022. 1

[8] T. Chakraborty, U. R. K S, S. M. Naik, M. Panja, and B. Manvitha. Ten years of generative adversarial nets (GANs): A survey of the state-of-the-art. *Machine Learning: Science and Technology*, 5(1), 2024. 1

[9] W. Ding, X. Zhao, B. Zhu, Y. Du, G. Zhu, T. Yu, L. Li, and J. Wang. An ensemble of one-stage and two-stage detectors approach for road damage detection. In *IEEE International Conference on Big Data (Big Data)*, 2022. 1, 3, 4

[10] Van Gool L. Williams C. K. I. Winn J. Everingham, M. and A. Zisserman. The PASCAL visual object classes (VOC) challenge. *IJCV*, 2010. 3

[11] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *NIPS*, 2017. 2

[12] D. Jeong and J. Kim. Road damage detection using yolo with image tiling about multi-source images. In *IEEE International Conference on Big Data (Big Data)*, 2022. 1

[13] T. Kynkäänniemi, T. Karras, M. Aittala, T. Aila, and J. Lehtinen. The role of imagenet classes in Fréchet inception distance. In *ICLR*, 2023. 3

[14] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár. Microsoft COCO: Common objects in context. In *ECCV*, 2014. 3

[15] H. Maeda, T. Kashiyama, Y. Sekimoto, T. Seto, and H. Omata. Generative adversarial network for road damage detection. In *Computer-Aided Civil and Infrastructure Engineering*, 2020. 1, 2

[16] A. Mathiasen and F. Hvilshøj. Backpropagating through fréchet inception distance. *arXiv preprint arXiv:2009.14075*, 2021. 3

[17] A. M. Okran, M. Abdel-Nasser, H. A. Rashwan, and D. Puig. Effective deep learning-based ensemble model for road crack detection. In *IEEE International Conference on Big Data (Big Data)*, 2022. 1

[18] K.-C. Peng. Iterative self knowledge distillation — from pothole classification to fine-grained and covid recognition. In *ICASSP*, 2022. 1

[19] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning transferable visual models from natural language supervision. In *ICML*, 2021. 2, 3

[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. ImageNet large scale visual recognition challenge. *IJCV*, 115(3):211–252, 2015. 3

[21] P. K. Saha and Y. Sekimoto. Road damage detection for multiple countries. In *IEEE International Conference on Big Data (Big Data)*, 2022. 1

[22] S. Wang, Y. Tang, X. Liao, J. He, H. Feng, H. Jiao, X. Su, and Q. Yuan. An ensemble learning approach with multi-depth attention mechanism for road damage detection. In *IEEE International Conference on Big Data (Big Data)*, 2022. 1

[23] W. Wang, J. Dai, Z. Chen, Z. Huang, Z. Li, X. Zhu, X. Hu, T. Lu, L. Lu, H. Li, X. Wang, and Y. Qiao. InternImage: Exploring large-scale vision foundation models with deformable convolutions. In *CVPR*, 2023. 3, 4

[24] D. Wu, M. Liao, W. Zhang, X. Wang, X. Bai, W. Cheng, and W. Liu. YOLOP: You only look once for panoptic driving perception. *Machine Intelligence Research*, 2022. 1, 2

[25] J. Zhong, J. Huyan, W. Zhang, H. Cheng, J. Zhang, Z. Tong, X. Jiang, and B. Huang. A deeper generative adversarial network for grooved cement concrete pavement crack detection. In *Engineering Applications of Artificial Intelligence*, 2023. 1, 2

[26] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017. 3, 1