Improving Pavement and Concrete Crack Detection Through Synthetic Data Generation

Anonymous CVPR submission

Paper ID *****

Abstract

Road safety surveillance is crucial to avoid traffic jams and accidents. In crack detection, pixel-accurate predictions are necessary to measure the width – an important indicator of the severity of a crack. However, manual annotation of images to train supervised models is a hard and time-consuming task. Because of this, manual annotations tend to be inaccurate, particularly at pixel-accurate level. The learning bias introduced by this inaccuracy hinders pixel-accurate crack detection. We propose a novel tool aimed for synthetic image generation with accurate crack labels – Syncrack. This parametrizable tool also provides a method to introduce controlled noise to annotations, emulating human inaccuracy. By using this, first we do a robustness study of the impact of training with inaccurate labels. This study quantifies the detrimental effect of inaccurate annotations in the final prediction scores. Afterwards, we show the advantages of using Syncrack generated images with accurate annotations for crack detection on real road images.

1. Introduction

For structural monitoring, crack inspection plays an important role. For many constructions, such as roads [2] or concrete structures [12], the cracks' width is one of the indicators of the damage severity and future durability. Measuring the width requires a pixel-accurate crack detection, which is still a challenging task [1]. The most successful methods for crack detection are based on supervised learning, which rely on manual annotations for training [4, 10, 11]. However, image annotation is a tedious and highly time-consuming task. Furthermore, these manual annotations tend to be inaccurate. More precisely, it is usual that manual annotations are wider than the actual cracks.

Inaccurate annotations are a particular case of noisy labels, and learning in presence of noise is a highly studied topic [3]. However, the problem of inaccurate annotations for crack detection is barely discussed in the literature (beyond the use of tolerance margins).

To fill this gap in the field, the contributions of our work can be summarized as:

The Syncrack generator. We developed this open-source tool to generate parametrizable synthetic images of cracked pavement/concrete-like textures. It provides accurate annotations to alleviate the crack labeling task and parametrizable noisy annotations to study the robustness of crack detection methods.

A robustness study of the impact of inaccurate labels. We studied the detrimental impact of training under different label noise conditions, with accurate annotations for evaluation. We measure the impact on prediction using supervised and unsupervised scores.

An improved crack width detection. By training solely with Syncrack-generated images, we produced predictions competitive with those obtained by training with real-life images. These predictions exhibit an improved crack width with respect to the ones obtained using real images.

2. The Syncrack generator

Syncrack allows user customization to generate datasets with different properties. In this work, we present 3 examples of datasets generated with Syncrack (see Fig. 1). Our tool consists of 4 main modules: 1) Creating a background image, 2) Creating crack shapes, 3) Adding cracks to the background, 4) Creating noisy annotations from pixel-accurate crack masks (Fig. 2). In this work, we created Syncrack datasets with 3 difficulty levels (200 images each) varying 2 user hyperparameters: the background average smoothness (*bas*) and the crack average contrast (*cac*) (see Table 1). These values represent means for Gaussian distributions, using Syncrack's default values for standard deviation: 0.1 for smoothness and 0.03 for contrast.

To study the effect of different label noise levels in annotations used for training, we created noisy labels for the 3 mentioned datasets. We introduce 5 label noise levels varying the *np* user parameter. This noise is analyzed in Table 2;

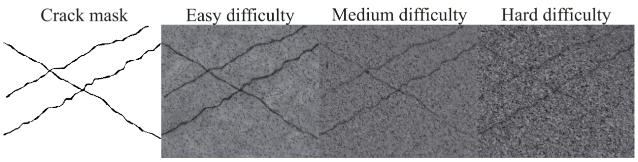

Figure 1. Example of Syncrack-generated images. The leftmost image is a synthetically generated crack shape. The rest of the images are synthetically generated backgrounds with the same inserted crack.

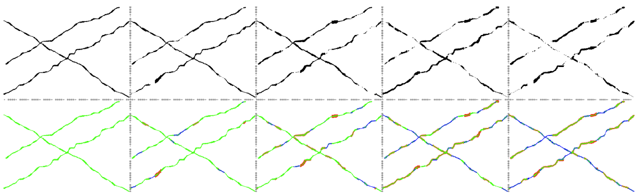

Figure 2. Examples of label noise levels 0 to 4 (see Table 2). The second row shows: Green) Crack pixels; Blue) Crack pixels mislabeled as background; Red) Background pixels mislabeled as crack.

Table 1. Hyperparameters used to create the 3 difficulty versions.

Difficulty level	Easy	Medium	Hard
Background average smoothness (*bas*)	6.0	3.0	1.5
Crack average contrast (*cac*)	0.5	0.7	0.7

Table 2. Label noise levels used for experiments.

Label noise level	0	1	2	3	4
np	0.00	0.25	0.50	0.75	1.00
DSC (%)	100.0	88.93	78.41	68.39	58.94
Pr (%)	100.0	88.44	77.18	68.53	58.29
Re (%)	100.0	89.92	80.61	69.41	60.66
H_{crack}	3.984	4.064	4.117	4.138	4.159
H^2_{crack}	7.391	7.498	7.589	7.613	7.675
K-S	0.6846	0.6299	0.5768	0.5373	0.4911

we show the average precision, recall and DSC (equivalent to F-score) per image with respect to the clean annotations. We also propose 3 additional unsupervised scores to evaluate the quality of the noisy annotations; these scores are proposed for further evaluation on real images, in which the supervised scores are biased because of the inaccuracy of the manual annotations.

The first of these unsupervised scores is the crack region entropy H_{crack}, based on the region entropy [6]. With a good segmentation, the intensity distribution within the crack-predicted region is skewed towards dark pixels; this will minimize H_{crack}. We also use a second-order crack region entropy H^2_{crack}. The second-order entropy relies on co-occurrence matrices [5] rather than pixel intensities. A good segmentation will produce a reduced amount of textures within the crack-predicted region, reducing H^2_{crack}.

From a probabilistic point of view, we assume that background and cracks are two different distributions. The Kolmogorov-Smirnov score (K-S) [9] increases with respect to how much two distributions differ. Therefore, a good segmentation should maximize this score.

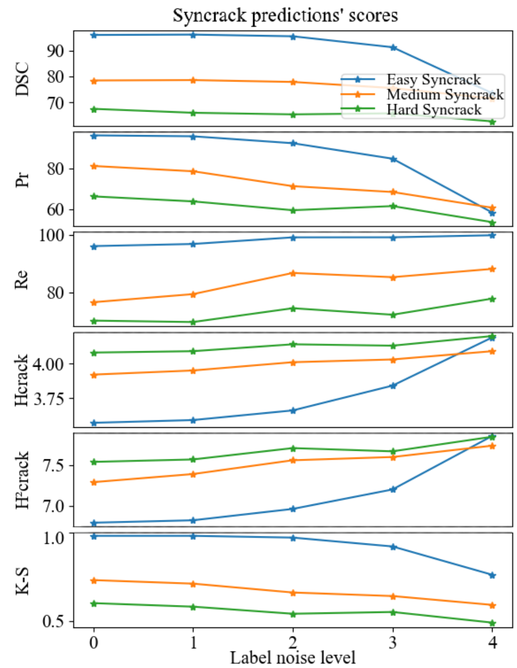

Figure 3. Prediction scores obtained in the validation splits of the Syncrack's 3 difficulty levels evaluating with accurate annotations.

3. EFFECT OF DIFFERENT NOISE CONDITIONS

To validate the unsupervised segmentation scores proposed in this paper, we analyze their behavior under controlled conditions. To do this, we use the noisy annotations obtained with the Syncrack generator. We use these inaccurate annotations to train U-VGG19 [7], a state-of-the-art network for road crack detection, and we analyze the relation between the unsupervised scores and the prediction quality of trained models. These results are plotted in Fig. 3. We see a direct relation between the supervised precision and the unsupervised scores: when the precision decreases, the K-S decrease and the entropies increase. In fact, the relation between precision and the unsupervised scores can be approximated by a liner function.

When increasing the noise, we see a tendency of the recall to actually improve. On the other hand, the precision decreases. We see an overall decrease of the DSC because the decrease in precision is greater than the increase in recall. Therefore, the noisy annotations promote the existence of false positives. These false positives are mainly caused

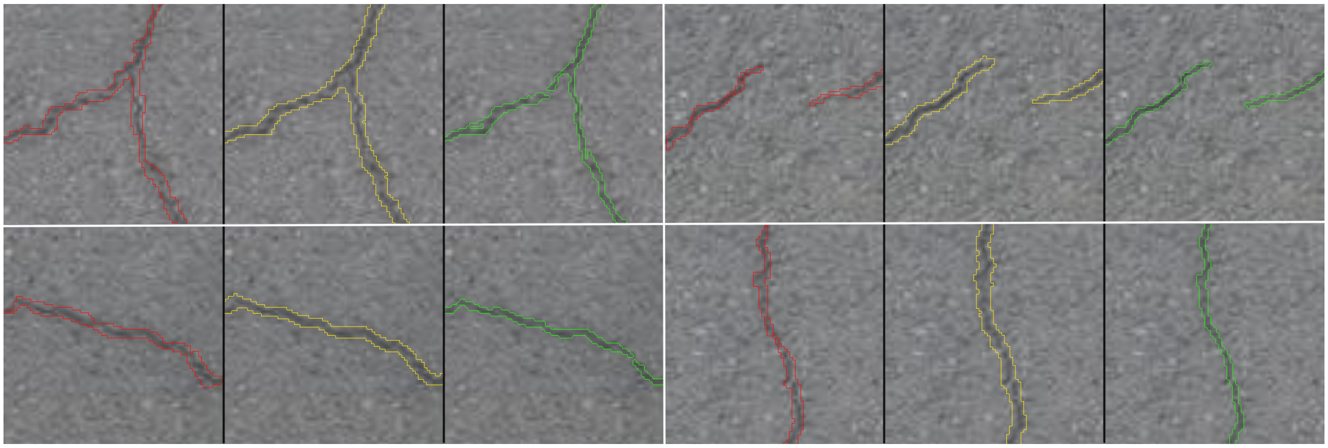

Figure 4. Examples of crack segmentation borders. The red line is the border of the manual annotation, the yellow line is the border of the prediction obtained by training on CFD, and the green line is the border of the prediction obtained by training on Syncrack hard.

by excessively wide predictions.

4. Improving real-life crack detection with Syncrack

To validate the performance of models trained with Syncrack on real-life data, we use the CrackForest Dataset (CFD) [8]. The image size is 480×320 and the cracks have a width around 3 pixels, similarly to the default Syncrack generator. As suggested by [10], we removed some images with clear severe annotation errors; we kept 108 images.

Fig. 5 shows the prediction scores of the models trained on CFD and on the 3 difficulty versions of Syncrack. We trained on the training split of each respective dataset and validated on the CFD validation split. The models trained on Syncrack were trained using accurate annotations.

The models trained with Syncrack datasets exhibit a higher precision than the model trained on real images. However, overall, the recall and DSC of the models trained with synthetic data are lower than the one of the model trained with CFD. As the Syncrack difficulty increases, the precision decreases a bit but the recall increases. The model trained with the hard Syncrack has the better trade-off between precision and recall, obtaining a DSC of 58.26% in contrast with the 69.32% obtained by training with CFD.

This DSC difference is caused only by a decreased recall. Considering that manual annotations tend to be wider than the actual cracks, a more precise segmentation will lead indeed to a lower recall. In Fig. 5b, we observe that the models trained with Syncrack have better unsupervised scores than the one trained on CFD. As we increase the Syncrack difficulty level, the entropies in the predicted crack regions increase and the K-S decreases. Even for the hard difficulty, these scores are better than for CFD.

With this, we observe that the decrease in recall is caused

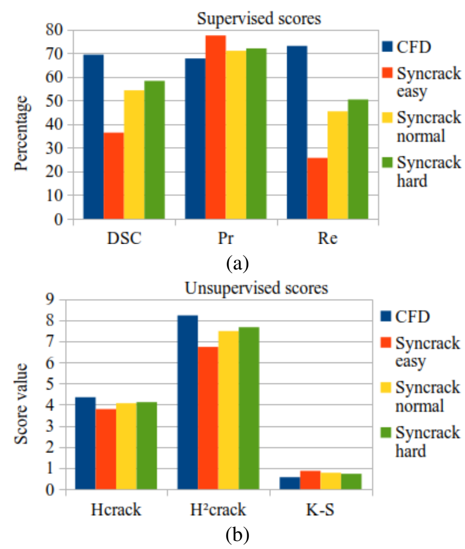

Figure 5. Scores obtained in the CFD validation set by using models trained on different datasets.

mainly by missing some pixels in the excessively wide annotations. A qualitative analysis of the predictions confirmed this. The models trained with the medium and hard difficulty do a good job by not missing cracks, and the predicted width looks more close to the actual crack that both the annotation and the prediction of CFD (see Fig. 4).

5. Conclusion

Syncrack generated images showed their promising potential for supervised crack detection without requiring labeled real-life images. The models trained solely with our synthetically generated data are transferable to real images; furthermore, they are more precise in terms of crack width.

References

[1] Sharmad Bhat, Saish Naik, Mandar Gaonkar, Pradnya Sawant, Shailendra Aswale, and Pratiksha Shetgaonkar. A Survey On Road Crack Detection Techniques. In *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, pages 1–6, Feb. 2020. 1

[2] E. Coquelle, J.-L. Gautier, and P. Dokládal. Automatic assessment of a road surface condition. In *7th Symposium on Pavement Surface Characteristics, Surf*, Norfolk, Virginia, 2012. 1

[3] Benoît Frénay and Michel Verleysen. Classification in the presence of label noise: a survey. *IEEE transactions on neural networks and learning systems*, 25(5):845–869, 2013. 1

[4] Ziping Gao, Bo Peng, Tianrui Li, and Cong Gou. Generative Adversarial Networks for Road Crack Image Segmentation. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, July 2019. ISSN: 2161-4407. 1

[5] Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6):610–621, 1973. 2

[6] Nikhil R. Pal and Dinabandhu Bhandari. Image thresholding: Some new techniques. *Signal Processing*, 33(2):139–158, Aug. 1993. 2

[7] Rodrigo Rill-García, Eva Dokládalová, and Petr Dokládal. Pixel-accurate road crack detection in presence of inaccurate annotations. *Neurocomputing*, 2022. 2

[8] Yong Shi, Limeng Cui, Zhiquan Qi, Fan Meng, and Zhensong Chen. Automatic Road Crack Detection Using Random Structured Forests. *IEEE Transactions on Intelligent Transportation Systems*, 17(12):3434–3445, Dec. 2016. Conference Name: IEEE Transactions on Intelligent Transportation Systems. 3

[9] Nikolai V Smirnov. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou*, 2(2):3–14, 1939. 2

[10] Mengyuan Sun, Runhua Guo, Jinhui Zhu, and Wenhui Fan. Roadway Crack Segmentation Based on an Encoder-decoder Deep Network with Multi-scale Convolutional Blocks. In *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0869–0874, Jan. 2020. 1, 3

[11] Fan Yang, Lei Zhang, Sijia Yu, Danil Prokhorov, Xue Mei, and Haibin Ling. Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection. *IEEE Transactions on Intelligent Transportation Systems*, 21(4):1525–1535, Apr. 2020. Conference Name: IEEE Transactions on Intelligent Transportation Systems. 1

[12] Xincong Yang, Heng Li, Yantao Yu, Xiaochun Luo, Ting Huang, and Xu Yang. Automatic Pixel-Level Crack Detection and Measurement Using Fully Convolutional Network. *Computer-Aided Civil and Infrastructure Engineering*, 33(12):1090–1109, 2018. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mice.12412. 1