Efficient Pavement Distress Classification via Deep Patch Soft Selective Learning and Knowledge Distillation

Shizheng Zhang*,[1] Wenhao Tang*,[2] Jing Wang,[2] and Sheng Huang[2]
[1] *College of Software Engineering, Zhengzhou University of Light Industry, China*
[2] *School of Big Data and Software Engineering, Chongqing University, China*
Email: huangsheng@cqu.edu.cn

Pavement distress classification is a vital step for automatic pavement inspection and maintenance. Recently, patch-based approaches have achieved promising performances and thus extensive attention in this field. However, these methods simply assume that all patches contribute equally to the distress classification, leading to weakly discriminating abilities of models. Moreover, their tedious processes also lead to a low efficiency in inference. In this letter, we present a novel patch-based pavement distress classification approach named Deep Patch Soft Selective Learning (DPS^{2}L), which addresses these issues. Similar to other patch-based approaches, DPS^{2}L partitions the pavement images into patches and aggregates the patch features to accomplish the task. To address the first issue, we introduce a succinct Soft Patch Feature Selection Network (SPFSN) to assess the importance of each patch to the distress classification with a score based on its feature. These scores will be considered as patch-wise weights for feature aggregation. In such a manner, the most discriminative patches are selected in a soft way, and thereby benefit the final classification. To address the inference efficiency issue, knowledge distillation is leveraged to transfer the classification knowledge from DPS^{2}L to the image-based approaches, such as EfficientNet-B3. This distilled model enables incorporating both the advantages of patch-based approaches in classification performance and the advantages of image-based approaches in inference efficiency. Extensive experiments on a large-scale pavement image dataset named CQU-BPDD demonstrate the superiority of our methods over baselines regardless of performance or efficiency.

Introduction: Pavement distresses jeopardize road serviceability and pose a serious threat to the public transportation system [1]. Detecting and recognizing pavement distresses, referred to as Pavement Distress Classification (PDC), is a core step in the pavement management system for determining cost-effective maintenance and rehabilitation strategies [2]. Automating this step will prominently reduce the costs of labor and finance for pavement maintenance [3].

As a common and economical sensor, camera has been widely used for inspecting pavement conditions. Many researchers have devoted themselves to investigating PDC from the perspective of computer vision [4–7]. The traditional methods often leverage image processing, hand-crafted features, and conventional classifiers to classify pavement distresses [8, 9] which relies heavily on expert knowledge and lacks universality [10]. Inspired by the remarkable advancement of deep learning, many deep learning-based PDC methods have been proposed, which consider PDC as a common image classification or object detection task, and often directly apply advanced deep learning techniques, such as Deep Convolutional Neural Network (DCNN) and faster R-CNN, as attempted solutions [11, 12]. However, these methods neglect specific characteristics inherent to this problem, such as high-resolution and low distressed area ratio [13], which leads to unsatisfactory performances.

To better incorporate these characteristics of pavement images, patch-based PDC methods have been developed recently [13, 14]. This sort of method partitions the pavement images into patches first, and then leverages CNN models to infer the patches labels for pavement distress classification. Since only image labels are available during model training, the patch label inference is often conducted by a complicated iterative optimization or weakly supervised learning process. For example, Tang et al [13] present a patch-based pavement distress detector, which elaborates an Expectation-Maximization Inspired Patch Label Distillation (EMIPLD) strategy to produce pseudo-labels for optimizing the model.

* These authors contributed equally to this work.

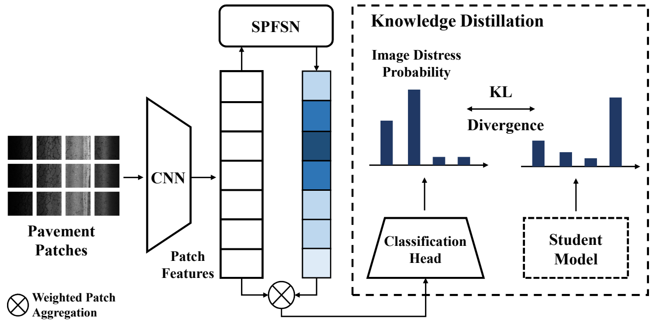

Fig 1 *The overview of the Deep Patch Soft Selective Learning (DPS^{2}L) model. As same as other patch-based approaches, DPS^{2}L partitions the pavement images into patches and extracts their features with CNN first. Then, we elaborate a Soft Patch Feature Selection Network (SPFSN) to adaptively score each patch based on its feature for highlighting the discriminative patches while suppressing the negative influences from overabundant non-distressed patches in feature aggregation. Finally, the scores are considered as weights to aggregate the patch features to a final one for accomplishing the pavement distress classification. Moreover, the knowledge distillation is employed in DPS^{2}L for speeding up the classification via transferring the classification knowledge learned by a patch-based approach, such as DPS^{2}L, to a more succinct image-based approach, such as EfficientNet-B3.*

Huang et al [14] propose a novel patch-based pavement distress recognition approach, which optimizes patch label inference networks via introducing a sparsity constraint to the patch label predictions while aggregating the patch label predictions with a comprehensive decision network. The main drawback of the current patch-based pavement distress classification methods is they consider each patch contributes equally to the final classification. However, the low distressed area ratio of pavement image implies that most patches contain no pavement distresses, which can easily dilute the features of the distressed patches, and thereby reduce the discriminating powers of models. Moreover, the patch-based approaches involve more processes compared with image-based methods in the inference phase, which reduces the efficiency in practice.

In this letter, we present a novel patch-based pavement distress classification approach named Deep Patch Soft Selective Learning (DPS^{2}L) to address these aforementioned issues. DPS^{2}L is a succinct end-to-end weakly supervised learning framework as shown in Figure 1. Like other patch-based methods, it partitions pavement images into patches and employs CNN to extract the feature of each patch individually at first. Then, a Soft Patch Feature Selection Network (SPFSN) is introduced to softly highlight the discriminative patches while suppressing the overabundant trial patches by adaptively producing the nonnegative weights for each patch based on its feature. Finally, the weighted features of patches are aggregated to a final image-level feature for final classification. Moreover, we apply the knowledge distillation [15] to transfer the classification knowledge learned from DPS^{2}L to EfficientNet-B3 for speeding up the inference time and consequently improving the practicability of the model in real-world applications. We evaluate our work on a large-scale Bituminous Pavement Disease Detection dataset named CQU-BPDD [13], consisting of 60,059 high-resolution pavement images that involve seven different diseases and normal pavement. Extensive experiments demonstrate the superiority of DPS^{2}L over other patch-based approaches in both pavement distress detection and recognition. The results also show that the prominent advantage of our knowledge distilled variant in inference efficiency.

The main contributions can be summarized as follows,

- We propose a novel patch-based pavement distress classification method named DPS^{2}L, which can adaptively select the most discriminative patches for softly achieving better label inference via dynamically weighting the features of patches. Extensive results on a large-scale dataset show that DPS^{2}L achieves better performances over other patch-based methods in both pavement distress detection and recognition tasks.
- We apply the knowledge distillation technique to transfer classification knowledge from a patch-based method to an image-based method, which can significantly speed up the inference step. To the best of our knowledge, this is the first attempt to utilize knowledge distillation for addressing pavement image analysis issues.

Methodology: Let $X = \{x_1, \cdots, x_n\}$ and $Y = \{y_1, \cdots, y_n\}$ be the collection of pavement images and their pavement labels, respectively. $y_i \in \mathbb{R}^{C\times 1}$ is a C-dimensional one-hot vector whose j-th element y_{ij} indicates whether the pavement image belongs to the j-th category or not where C is the number of categories. The pavement distress classification task aims to derive a classifier $F(\cdot)$ as a pavement distress detector or a recognizer, to label the pavement image correctly, $y_i \leftarrow F(x_i)$, where $F_{cls} \in \{F_{det}, F_{rec}\}$. In this letter, we present a novel pavement distress classification method named Deep Patch Soft Selective Learning (DPS^{2}L) which consists of four core steps, namely Patch Collection, Patch Encoding, Soft Patch Feature Selection, and Patch Aggregation, coupled with an additional step named Knowledge Distillation.

Patch Collection: Similar to WSPLIN-IP [14], we adopt Image Pyramid (IP) as our patch collection strategy, which assists the model in exploiting the scale space of images. In our implementation, we construct a three-layer image pyramid from top to down, and apply the non-overlapped sliding window operation $\tau(\cdot)$ to output a collection of 300×300 patches for each pavement image x_i,

$$P_i = \left\{\tau\left(x_i^l\right)\right\}_{l\in\{0,1,2\}} = \left\{p_i^1, \ldots, p_i^m\right\}, \tag{1}$$

where l indicates which layer of the image pyramid is used and p_i^m is the m-th patch of i-th pavement image. In our case, $m = 12$ (when $l = 2) + 4$ (when $l = 1) + 1$ (when $l = 0) = 17$.

Patch Encoding: Following many patch-based pavement distress classification approaches [13, 14], we use EfficientNet-B3 [16] for extracting the CNN features of patches,

$$Z_i = \text{GAP}\left(f\left(P_i\right)\right) = \left[z_i^1; \ldots; z_i^m\right] \in \mathbb{R}^{m\times L}, \tag{2}$$

where $f(\cdot)$ is mapping function of EfficientNet-B3 while GAP$(\cdot)$ denotes the global average pooling operation of each feature patch. Note, the patch encoder is not just limited to EfficientNet-B3, and can be flexibly replaced by any other CNN models.

Soft Patch Feature Selection: A key design element of DPS^{2}L is its Soft Patch Feature Selection Network (SPFSN) for highlighting the discriminative patches while suppressing the overabundant trivial patches by adaptively producing the nonnegative weights for each patch based on its feature. Specifically, we simply use a two fully connected layers with the activation function of Gaussian Error Linear Units (GELU) to implement the SPFSN. Let $W_1 \in \mathbb{R}^{L\times\frac{L}{4}}$ and $W_2 \in \mathbb{R}^{\frac{L}{4}\times 1}$ be the weights of these two fully connected layers respectively. The patch scoring procedure can be mathematically represented as follows,

$$V_i = \text{softmax}(\text{GELU}\left(Z_i W_1\right) W_2) = \left[v_i^1; \ldots; v_i^t; \ldots; v_i^m\right] \in \mathbb{R}^{m\times 1}, \tag{3}$$

where softmax$(\cdot)$ is a softmax layer, V_i encodes the importance scores of all patches collected from the i-th pavement image, and $v_i^t \in (0, 1)$ is its t-th element indicates the importance of the t-th patch for classifying the i-th pavement image. These generated scores are used as weights for aggregating features in the next step, and then the patch feature selection can be conducted in a soft way.

Patch Aggregation and Image Classification: We aggregate the features of patches in a same image via combining the features of patches weighted by their corresponding scores, and then employ a classification network (or head) for obtaining the predicted label of image,

$$\tilde{y}_i = \text{softmax}\left(\mathcal{H}\left(V_i^\top Z_i\right)\right) \in \mathbb{R}^{C\times 1}, \tag{4}$$

where $\tilde{y}_i$ is the predicted label of the i-th image, and $\mathcal{H}(\cdot)$ is the mapping function of classification head. Cross-entropy is used to measure the discrepancy of the predicted label and the ground truth. Thereby, the optimal DPS^{2}L model can be obtained via addressing the following programming problem,

$$F := \{f, W_1, W_2, \mathcal{H}\} \leftarrow \arg\min -\frac{1}{n}\sum_{i=1}^{n} y_i^\top \log \tilde{y}_i. \tag{5}$$

Once all these mapping functions and parameters are learned, we can use them to infer the pavement distress label of any pavement image.

Knowledge Distillation: In order to incorporate the advantage of image-based approach in the inference speed, the recently popular Knowledge Distillation (KD) [15] is introduced to transfer the classification knowledge learned from our model (the teacher model) to a simpler image-based student model. Let S_t and S_s be the logits of the teacher and the student models respectively. Following [15], we employ the Kullback-Leibler divergence KL$(\cdot,\cdot)$ to measure the consistency of the classification knowledge between the teacher and the student models,

$$\mathcal{L}_k = \text{KL}\left(\text{softmax}\left(S_s/\eta\right), \text{softmax}\left(S_t/\eta\right)\right), \tag{6}$$

where η is the tunable temperature for the distillation, and here we empirically fix it to 5. The KD loss will be used as a regularization for aiding the student model to acquire the classification knowledge learned by the teacher model.

Table 1. The pavement distress detection performances of different methods. $T \rightarrow S$ indicates distillation of classification knowledge from the model T to the model S. ↑ indicates that the larger value means the better performance. The bold is the best performance.

Detectors(**DET**)	AUC ↑	P@R=95% ↑	Throughput (imgs/s) ↑
HOG+PCA+SVM [9]	77.7%	28.4%	-
LBP+PCA+SVM [17]	82.4%	30.3%	-
ResNet-50 [18]	90.5%	35.3%	-
VGG-19 [19]	94.2%	45.0%	-
Inception-v3 [20]	93.3%	42.3%	-
Effi-B3 [16]	95.4%	51.1%	**91**
IOPLIN [13]	97.4%	67.0%	55
DPS^{2}L	**97.9%**	72.3%	42
DPS^{2}L → Effi-B3	96.9%	67.3%	**91**
DPS^{2}L → DPS^{2}L	**97.9%**	**74.6%**	42

Table 2. Comparison on pavement distress recognition.

Recognizers(**REC**)	Top-1 ↑	F_1 ↑	Throughput (imgs/s) ↑
RGB + RF [21]	30.5%	-	-
HOG + SVM [9]	31.8%	-	-
ResNet-50 [18]	71.2%	61.5%	-
VGG-16 [19]	74.6%	65.0%	-
Inception-v3 [20]	77.6%	69.8%	-
Effi-B3 [16]	78.6%	70.3%	**91**
WSPLIN-IP [14]	83.7%	73.0%	42
DPS^{2}L	83.7%	75.4%	42
DPS^{2}L → Effi-B3	81.6%	73.7%	**91**
DPS^{2}L → DPS^{2}L	**84.0%**	**76.1%**	42

Experimental Results and Discussion: A large-scale bituminous pavement distress dataset named CQU-BPDD [13] is used for evaluation. This dataset involves seven different types of distress along with the normal pavement images with unbalanced quantity. The pavement images in this dataset are all acquired by the professional pavement inspection vehicle in the wild. Its training set contains 5,140 distressed pavement images involving all distresses and 5,000 normal pavement images, while its testing set has 11589 diseased pavement images and 38330 normal pavement images. We evaluate our methods on two pavement distress classification tasks, namely pavement distress detection, and pavement distress recognition. We follow the experimental settings of [13] and [14] for implementing the pavement distress detection and recognition respectively. Note, we only use the coarse-grained label (distress and non-distress) in the pavement distress detection task. Following the conventions, we adopt Area Under Curve (AUC) as the comprehensive evaluation metric in the detection task. Since people always pay more attention to the distressed samples over the normal ones, we also employ P@R=95% as another metric, which indicates the precision when the corresponding recall is equal to 95%. About the pavement distress recognition, we apply Top-1 accuracy and Marco F_1 score to measure the recognition performances. Due to inherent data distribution imbalance, F_1 can better reflect the discriminating powers of different models. Moreover, throughput is applied to evaluate the efficiency of the model in inference. A larger throughput means more images can be processed by the model in a second.

Pavement Distress Detection: Table 1 shows the pavement distress detection performances of our method in comparison with some classical shallow learning [8, 9, 17], deep learning [16, 18–20] and patch-based approaches [13, 14]. Three conclusions can be summarized from results. The first one is that DPS^{2}L outperforms all baselines in AUC and P@R=95%. For example, DPS^{2}L gets 0.5% more AUC and 5.3% more P@R=95% respectively but consumes around 20% more time for inference in comparison with IOPLIN, which is both the runner-up method

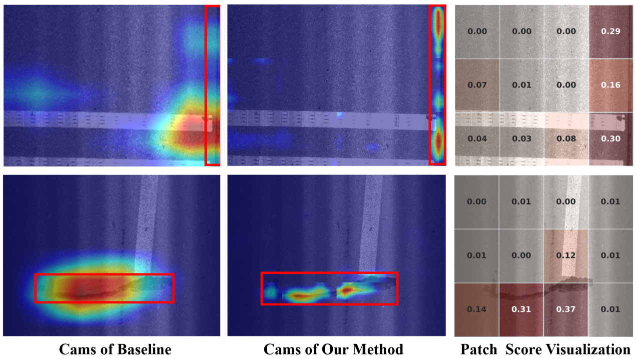

Fig 2 *The visualization examples of the features and the patch scores produced by our method. We employ CAM for feature visualization and the baseline is EfficientNet-B3 [16] here. The red box indicates the distressed area.*

Table 3. The ablation study of DPS^{2}L.

Method	**DET** (P@R=95%)	**REC** (F_1)
Effi-B3 [16]	51.1%	70.3%
DPS^{2}L w/o SPFSN	67.4%	71.7%
DPS^{2}L	72.3%	75.4%
DPS^{2}L → DPS^{2}L	**74.6%**	**76.1%**

and also a patch-based method. The second one is that the deep learning-based approaches outperform the shallow learning-based approaches, while the patch-based approach also consistently performs better than the image-based approaches. The third conclusion is to apply knowledge distillation to transfer the classification knowledge from DPS^{2}L to a more efficient image-based approach enables significantly speeding up the inference. For example, the knowledge distillation speeds up the inference efficiency of DPS^{2}L by a factor of 1.7 times. Meanwhile, the model can obtain a similar classification performance as IOPLIN. Another interesting phenomenon is that performing knowledge distillation to DPS^{2}L itself can also further boost the classification performance (+ 2.3% in P@R=95%). We attribute this to the fact that the knowledge distillation can be a regularization to alleviate the overfitting problem.

Pavement Distress Recognition: Table 2 records the pavement distress recognition performances of different methods. Similar conclusions as the ones in pavement distress detection can be summarized according to the results. The sample distribution of over categories is imbalanced within the CQU-BPDD dataset. In such a manner, the top-1 accuracy cannot decently reflect the real recognition abilities of models, since it is always biased toward the classification result of categories, which have more samples. F_1 score is a more fair classification evaluation metric. The results show that DPS^{2}L gets 2.4% more performance gains in F_1 while attaining the same efficiency in inference over WSPLIN-IP, which is the runner-up recognition model and also a patch-based approach.

Ablation Study: Table 3 reports the performances of DPS^{2}L under different settings across different pavement distress classification tasks. EfficientNet-B3 [16] is the baseline of our method. The results of the first two rows validate that the patch-based method is often more discriminative than the simple image-based approach. The results of the second and third rows verify the importance of soft patch feature selection in pavement distress classification. The SPFSN improves the performances of DPS^{2}L by 4.9% in P@R=95%, and 3.7% in F_1 respectively. The results of the last row indicate the performance of DPS^{2}L can be further boosted via self-distillation.

Visualization: Figure 2 visualizes the learned patch scores and the Class Activation Map (CAM) of a given pavement image. The results visually validate that DPS^{2}L enables adaptively generating the patch scores to well highlight the most discriminative patches, and more accurately capturing the features of the distressed area over the baseline.

Conclusion: In this letter, we present a novel patch-based approach named DPS^{2}L for pavement distress classification. DPS^{2}L partitions the pavement image into patches, and then employs a Soft Patch Feature Selection Network (SPFSN) to score every patch based on its feature for highlighting the discriminative patches while suppressing the trivial patches. Finally, the produced scores are considered as weights to aggregate the patch features for accomplishing the classification task. Moreover, we apply knowledge distillation to transfer the classification knowledge from DPS^{2}L to EfficientNet-B3 for speeding up the inference. Extensive experimental results on CQU-BPDD validate the effectiveness of DPS^{2}L and its variants.

Conflict of interest: The authors declare no conflict of interest.

Acknowledgments: This work was supported in part by the National Natural Science Foundation of China under Grant 62176030, and the Natural Science Foundation of Chongqing under Grant cstc2021jcyj-msxmX0568.

References

1. Huang, Y., Xu, B.: Automatic inspection of pavement cracking distress. Journal of Electronic Imaging 15(1), 013017 (2006)
2. Tsai, Y.C., Kaul, V., Mersereau, R.M.: Critical assessment of pavement distress segmentation methods. Journal of transportation engineering 136(1), 11–19 (2010)
3. Coenen, T.B., Golroo, A.: A review on automated pavement distress detection methods. Cogent Engineering 4(1), 1374822 (2017)
4. Jia, G., et al.: Sample generation of semi-automatic pavement crack labelling and robustness in detection of pavement diseases. Electronics Letters 55(23), 1235–1238 (2019)
5. Yuan, G., et al.: Curseg: A pavement crack detector based on a deep hierarchical feature learning segmentation framework. IET Intelligent Transport Systems (2022)
6. Xiang, X., Zhang, Y., El Saddik, A.: Pavement crack detection network based on pyramid structure and attention mechanism. IET Image Processing 14(8), 1580–1586 (2020)
7. Wang, B., Zhang, Q., Zhao, W.: Fast concrete crack detection method via l2 sparse representation. Electronics Letters 54(12), 752–754 (2018)
8. Cortes, C., Vapnik, V.: Support-vector networks. Machine learning 20(3), 273–297 (1995)
9. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition, vol. 1, pp. 886–893. Ieee (2005)
10. Dong, H., et al.: Automatic inspection and evaluation system for pavement distress. IEEE Transactions on Intelligent Transportation Systems (2021)
11. Gopalakrishnan, K., et al.: Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. Construction and building materials 157, 322–330 (2017)
12. Wang, W., et al.: Road damage detection and classification with faster r-cnn. In: 2018 IEEE international conference on big data (Big data), pp. 5220–5223. IEEE (2018)
13. Tang, W., et al.: An iteratively optimized patch label inference network for automatic pavement distress detection. IEEE Transactions on Intelligent Transportation Systems (2021)
14. Huang, G., et al.: Weakly supervised patch label inference network with image pyramid for pavement diseases recognition in the wild. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7978–7982. IEEE (2021)
15. Hinton, G., et al.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 2(7) (2015)
16. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning, pp. 6105–6114. PMLR (2019)
17. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE transactions on pattern analysis and machine intelligence 28(12), 2037–2041 (2006)
18. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778. (2016)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
20. Szegedy, C., et al.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826. (2016)
21. Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)