Complex Shadow Removal in Pavement Imagery: Leveraging Diffusion Models for Advanced Solutions

T.W. Muturi, Y. Adu-Gyamfi & D. Kesse
University of Missouri, Columbia, Columbia, Missouri, USA

ABSTRACT: Shadows introduce uneven illumination; obscuring crack details and causing shadow boundaries to be misinterpreted as cracks within crack detection algorithms. This study proposes a shadow removal algorithm leveraging conditional diffusion models to eliminate shadows in top-down and oblique-view pavement images. We introduce the Pavement Image Shadow Triplet Dataset (PISTD), based on the ISTD dataset for the task. Hyperparameter tuning and training are explored on 256x256 and 512x512-pixel images to determine the tradeoff between crack detection accuracy and reconstruction precision. Moreover, the two models are compared with state-of-the-art models across PSNR, RMSE, and F1 accuracy metrics. The proposed model achieves a 6% and 38% higher segmentation accuracy on top-down and oblique view images, respectively. Evaluation of the algorithm is performed with real-world shadow images with qualitative results on the images demonstrating the effectiveness of the approach.

1 INTRODUCTION

To streamline maintenance operations, pavement management systems (PMS) have been implemented to survey, analyze, and assist decision-makers in allocating resources for pavement maintenance (Al-Mansour et al., 2022). As an initial step in pavement management, data collection is critical to the decision-making process.

Various data collection methods such as manual surveys, cameras or laser scanners fitted to vehicles, portable platforms, and unmanned aerial vehicles (UAVs), have been employed. Recent efforts have seen the adoption of cameras fitted to portable devices such as cars, motorcycles, and bicycles for the collection of pavement distress information (Arya et al., 2022). This approach is lauded for its cost-effectiveness and efficiency in data collection. For instance, Mei and Gül (2020) reported spending approximately 350 USD on purchasing and mounting a GoPro to their vehicle for pavement distress data collection. Furthermore, the efficiency of this approach is also evident in the introduction of big data competitions that utilize camera data for pavement distress detection(Arya et al., 2022). However, the adoption of this cost-effective data collection method inevitably encounters the challenge of shadows within the images.

Shadows obscure crack patterns due to their similar intensity to crack regions (Zou et al., 2012). Additionally, boundary regions caused by shadows can be misinterpreted as cracks, leading to false positive detections (Pal et al., 2021). Furthermore, the uneven illumination caused by shadow regions results in inconsistencies in detection by machine learning and segmentation algorithms (Zou et al., 2012). Therefore, the elimination of shadow regions in these images is paramount for the accurate detection of pavement distress.

This article therefore aims to adopt a novel shadow removal algorithm in top-down and oblique-view pavement images captured with low-cost cameras. Specifically, we will (1) Compile a Pavement Image Shadow Triplet Dataset (PISTD) composed of top-down and oblique-view pavement distress images by performing image processing manipulation on shadow-free images using ISTD (Wang et al., 2017) mask images. (2) Train the latent diffusion model by Mei et al. (2023) to eliminate shadows in pavement images. Hyperparameter tuning of the training iterations and learning rate will be performed. Additionally, the model will be trained with 512x512 and 256x256 image sizes. (3) Compare model results with existing state-of-the-art models using Root Mean Squared Error (RMSE), Peak Signal-to-Noise Ratio (PSNR), and F1 score metrics. F1 score metrics are based on segmentation accuracy. (4) Test the zero-shot performance of the model for shadow removal.

2 LITERATURE REVIEW

Within literature, shadow removal across different scenes has been explored through image processing and deep learning approaches. Deep learning has emerged as a prevailing standard for shadow removal through popular algorithms such as the ShadowGAN (Hu et al., 2019), Stacked Conditional GAN (ST-CGAN) (Wang et al., 2017), ARGAN (Ding et al., 2019), and DC-ShadowNet (Jin et al., 2023). However, these models require a large amount of training data and, due to their architecture and adversarial training approach, they are computationally complex. This complexity can introduce training instability and poor generalizability (Mei et al., 2023).

Diffusion models have demonstrated improved stability over GANs in generative tasks (Rombach et al., 2022). Given this, in an initial attempt to perform instance shadow removal, Mei et al (2023) adopted the conditional diffusion model for shadow removal across different scenes. Although this approach outperforms state-of-the-art methods, it has not been specifically applied to pavement images, which present a new level of complexity due to the similarity in intensity between shadow regions and cracks.

3 PROPOSED METHODOLOGY

In this research, a diffusion network is employed for shadow removal due to its robustness compared to other methods. The algorithm proposed by Mei et al. (2023) is selected over other diffusion techniques as it introduces a learned latent feature space used for conditioning the diffusion model and integrates noise features within the diffusion network to mitigate the risk of converging to local optima during training. The adopted algorithm can be delineated into three critical modules: (1) the conditional diffusion module, (2) the latent encoder module, and (3) the noise fusion module.

3.1 *Conditional Diffusion Module*

As an extension of ordinary diffusion models, conditional diffusion models incorporate additional information or conditions, such as class labels, text descriptions, or other modalities during training. Fundamentally, diffusion models simulate the process of data corruption (forward diffusion process) and its reversal (reverse diffusion process), enabling the generation of new data samples by reversing the diffusion process. In the context of this research, the forward process corrupts the shadowed image, while the reverse process generates the shadow-free image.

3.2 *Latent Encoder*

In this research, a novel latent feature space is employed to guide the model. Given the shadow mask and the shadowed image, the latent encoder generates a latent feature vector, which is subsequently used as the condition within the diffusion model. A U-Net model with an architecture similar to that of the diffusion network is adopted.

3.3 *Noise Fusion Model*

Posterior collapse refers to a phenomenon where the learned latent representations fail to encode meaningful information about the input data. Observed within diffusion models, this happens when the training procedure of generative models falls into a trivial local optimum. Posterior collapse is especially undesirable in shadow removal due to the complexity of different shadows within the models. This article, therefore, adopts Mei et al. (2023) approach, which involves fusing the learned embeddings and features within the diffusion network to overcome posterior collapse.

4 DATASET

Within the domain of shadow removal, two primary benchmark datasets are commonly referenced: the ISTD (Wang et al., 2017) and AISTD/ISTD+ (Le & Samaras, 2019). The ISTD and AISTD datasets include triplets of shadow images, shadow-free images, and shadow masks, covering a variety of scenes such as walls, grass surfaces, and sidewalks of different colors. However, shadow removal on pavements presents unique challenges as pavement distresses can have intensities similar to those within the shadowed regions.

This article therefore introduces the Pavement Image Shadow Triplet Dataset (PISTD). To construct this dataset, shadow masks from the ISTD dataset are applied to pavement images with varying degrees of opacity, location and orientation. To better emulate real-world shadow conditions, these masks are also enlarged to cover more extensive regions within the images. Top-down shadow images are sourced from the CFD (Shi et al., 2016), CrackTree200 (Zou et al., 2012), and DSPS24 datasets. Additionally, to replicate angled data collection, crack-free images were manually collected by the authors. To further augment this dataset, crack-free images from the EdmCrack600 (Q. Mei & Gül, 2020) dataset were also included. Figure 1 presents an example of shadow image, shadow mask and shadow-free image. The entire dataset was divided into training and testing sets in a 4:1 ratio to facilitate robust evaluation.

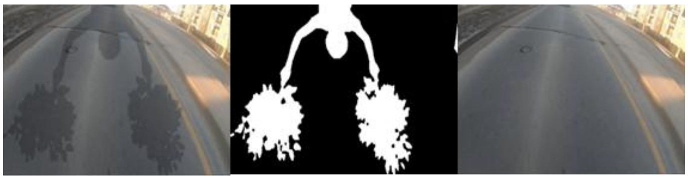

Figure 1. PISTD sample.

5 TRAINING

Hyperparameter tuning was performed with results revealing a learning rate of 1.0e-5, training over 300,000 iterations with a gaussian noise schedule steps of 1,000 achieving optimum performance. Furthermore, a comparison of training on different image sizes (256x256 vs 512x512) showed that an increase in image size results in a decrease in PSNR and an increase in RMSE.

Give a batch size of 1, trained over an NVIDIA RTX 3090 GPU with 24GB of memory compute, took approximately 81 hours with a 512x512 image size.

6 RESULTS

6.1 *Comparison with the state-of-the-art*

The performance metrics of the model are benchmarked against three state-of-the-art shadow removal models: DC-ShadowNet (Jin et al., 2023), SG-ShadowNet (Wan et al., 2022), and G2R-ShadowNet (Qu et al., 2017). These comparative models utilize Generative Adversarial Networks (GANs) with distinct variations in their architectures, loss functions, and training methodologies

As observed in Table 1, our model, trained on 256x256 images, achieves PSNR and RMSE metric scores comparable to state-of-the-art models when tested on top-down view images. Although these values represent a slight decrease compared to the best-performing SG-ShadowNet, our approach produces smoother, shadow-free images (Figure 2 last column). The slightly lower RMSE and PSNR are attributed to the noise introduced by our model; however, this noise does not adversely affect segmentation accuracy, as demonstrated by an F1 score of 59.3% (Table 1).

Table 2 presents the quantitative comparisons of the models when tested on angled view images. As shown, our model outperforms all other approaches on the RMSE and PSNR metrics when trained on 256x256-sized images. Similarly to top-down view images, while other models result in shadow boundaries, our proposed approach achieves smoother shadow removal (Figure 2). However, as in the top-down view results, the proposed model trained on 512x512 images exhibits a color shift, adversely affecting the RMSE and PSNR values. Despite this color shift, the model effectively removes all shadows within the images, attaining a 61.4% F1 score for segmentation, a 74% increase from the shadow image (Table 2).

Table 1. Quantitative comparison results in top-down view images

Method	RMSE	PSNR	F1 score
DC-Shadow Net	9.8	28.9	55.5
SG-Shadow Net	**7.4**	**31.9**	49.3
G2R-ShadowNet	10.9	30.1	39.5
Ours-256	9.5	30.0	39.5
Ours-512	14.4	26.0	**59.3**
Shadow Image	_	_	43.7

Table 2. Quantitative comparison results in oblique view images.

Method	RMSE	PSNR	F1 score
DC-Shadow Net	8.1	30.6	44.2
SG-Shadow Net	8.8	30.1	37.2
G2R-ShadowNet	12.3	28.2	32.9
Ours-256	**7.8**	**31.0**	47.3
Ours-512	11.3	28.1	**61.4**
Shadow Image	_	_	33.7

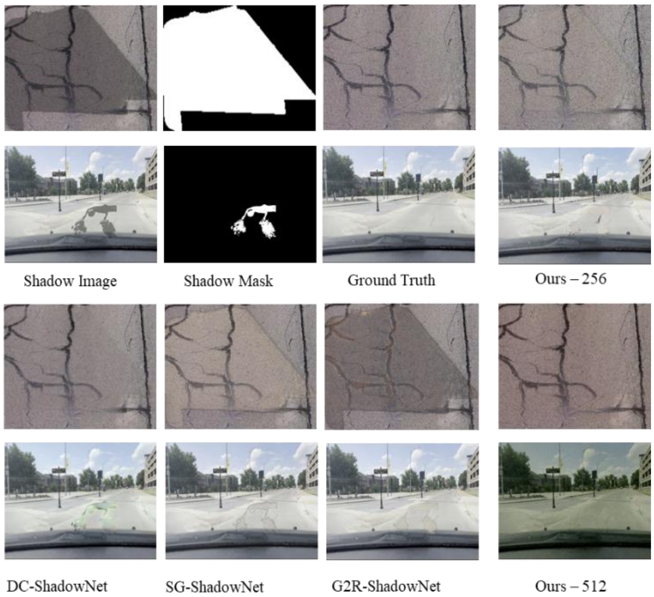

Figure 2. Visual comparison of shadow removal algorithms.

6.2 *1.1 Zero-shot Results*

This section aims to evaluate the proposed approach's performance with real-world shadows. As shown in Figure 3, the images presented are captured at an oblique angle, representing the current trend in data collection studies. Since original masks for these images do not exist, a SAM model was trained and used to develop shadow masks for the evaluation. As observed, given the SAM shadow mask, the proposed model eliminates shadows with high accuracy, particularly in the model trained with 256x256-pixel images. In contrast, the model trained with 512x512-pixel images experiences a color shift and struggles with a larger shadow boundary.

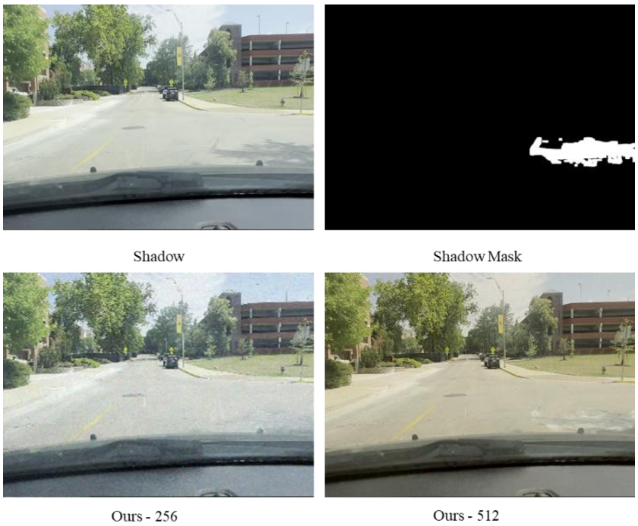

Figure 3. Zero shot results of the proposed approach.

7 CONCLUSION

This study presents a novel approach for shadow removal in pavement imagery using a diffusion model conditioned on encoded mask and shadow image latent. To achieve this, a Pavement Shadow Triplet Dataset (PSTD) was introduced by combining masks from the ISTD dataset with different pavement images at varying opacity. Hyperparameter tuning during model training showed that a learning rate of 1e-5 and training over 300,000 epochs resulted in the best model accuracy. A comparison of different image sizes revealed the model trained with 256x256-pixel images exhibited a superior performance in both quantitative metrics and qualitative assessments, providing smoother shadow-free images without the shadow boundaries observed in other models. The model trained with 512x512-pixel images, while effective in shadow removal, exhibited a color shift, highlighting the complexity introduced by larger image sizes. A comparison of segmentation accuracy shows the model trained on 512x512-pixel images attains the highest F1 score on both angled and top-down view pavement images. This was attributed to the increased resolution associated with larger images. The authors also note the segmentation accuracies presented within the results reflect lower accuracies than those within literature. This is because the model adopted was not trained on our dataset, thus the relative performance is emphasized. The zero-shot capability of our model was validated through the accurate removal of shadows in real-world oblique view images, further showcasing the proposed approach's robustness and applicability in practical scenarios. However, the authors note the limitation of the approach in the introduction of the color shift in 512x512 images and computational cost introduced in training. Future work should focus on addressing these limitations.

REFERENCES

Al-Mansour, A., Lee, K.-W. W., & Al-Qaili, A. H. (2022). Prediction of Pavement Maintenance Performance Using an Expert System. *Applied Sciences*, *12*(10), Article 10. https://doi.org/10.3390/app12104802

Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Omata, H., Kashiyama, T., & Sekimoto, Y. (2022). *Crowdsensing-based Road Damage Detection Challenge (CRDDC-2022)* (arXiv:2211.11362). arXiv. https://doi.org/10.48550/arXiv.2211.11362

Ding, B., Long, C., Zhang, L., & Xiao, C. (2019). *ARGAN: Attentive Recurrent Generative Adversarial Network for Shadow Detection and Removal* (arXiv:1908.01323). arXiv. https://doi.org/10.48550/arXiv.1908.01323

Hu, X., Jiang, Y., Fu, C.-W., & Heng, P.-A. (2019). Mask-ShadowGAN: Learning to Remove Shadows from Unpaired Data. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2472–2481. https://doi.org/10.1109/ICCV.2019.00256

Jin, Y., Sharma, A., & Tan, R. T. (2023). *DC-ShadowNet: Single-Image Hard and Soft Shadow Removal Using Unsupervised Domain-Classifier Guided Network* (arXiv:2207.10434). arXiv. https://doi.org/10.48550/arXiv.2207.10434

Le, H., & Samaras, D. (2019). *Shadow Removal via Shadow Image Decomposition* (arXiv:1908.08628). arXiv. https://doi.org/10.48550/arXiv.1908.08628

Mei, K., Figueroa, L., Lin, Z., Ding, Z., Cohen, S., & Patel, V. M. (2023). *Latent Feature-Guided Diffusion Models for Shadow Removal* (arXiv:2312.02156). arXiv. https://doi.org/10.48550/arXiv.2312.02156

Mei, Q., & Gül, M. (2020). A cost effective solution for pavement crack inspection using cameras and deep neural networks. *Construction and Building Materials*, *256*, 119397. https://doi.org/10.1016/j.conbuildmat.2020.119397

Pal, M., Palevičius, P., Landauskas, M., Orinaitė, U., Timofejeva, I., & Ragulskis, M. (2021). An Overview of Challenges Associated with Automatic Detection of Concrete Cracks in the Presence of Shadows. *Applied Sciences*, *11*(23), Article 23. https://doi.org/10.3390/app112311396

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). *High-Resolution Image Synthesis with Latent Diffusion Models* (arXiv:2112.10752). arXiv. https://doi.org/10.48550/arXiv.2112.10752

Shi, Y., Cui, L., Qi, Z., Meng, F., & Chen, Z. (2016). Automatic Road Crack Detection Using Random Structured Forests. *IEEE Transactions on Intelligent Transportation Systems*, *17*(12), 3434–3445. IEEE Transactions on Intelligent Transportation Systems. https://doi.org/10.1109/TITS.2016.2552248

Wang, J., Li, X., Hui, L., & Yang, J. (2017). *Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal* (arXiv:1712.02478). arXiv. https://doi.org/10.48550/arXiv.1712.02478

Zou, Q., Cao, Y., Li, Q., Mao, Q., & Wang, S. (2012). CrackTree: Automatic crack detection from pavement images. *Pattern Recognition Letters*, *33*(3), 227–238. https://doi.org/10.1016/j.patrec.2011.11.004