Improving Pavement Distress segmentation with Diffusion-Based Generative AI

M.A. Talaghat[1], A. Golroo[1,2], M. Rasti[2]
[1]*Department of Civil and Environmental Engineering, Amirkabir University of Technology, Tehran, Iran*
[2]*Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland*

ABSTRACT: The lack of diverse and sufficiently labeled datasets in pavement management limits the performance and generalization of segmentation and detection models. Existing datasets often fail to capture real-world variability, such as different pavement distress types, leading to models that struggle with unseen scenarios and rare distress patterns. This study addresses these challenges by applying diffusion-based generative AI to mitigate data scarcity. Using a Denoising Diffusion Probabilistic Model (DDPM), synthetic images were generated to augment datasets, significantly enhancing model robustness and accuracy. The synthetic images achieved an SSIM of 0.92 and a diversity score of 0.85, resulting in notable improvements in the U-Net segmentation model, including an IoU increase from 0.81 to 0.93, an F1 Score rise from 0.85 to 0.95, and pixel-level accuracy improvement from 0.89 to 0.95. These findings underscore the potential of generative AI to support scalable and robust digital twin solutions for pavement management systems.

1 INTRODUCTION

Digital twin technology has revolutionized infrastructure management by enabling real-time monitoring and predictive maintenance through virtual replicas of physical systems, providing valuable insights for optimizing asset performance (Saghatforoush et al., 2022). In pavement management, digital twins integrate data from IoT sensors, historical records, and predictive models to proactively address damages, reducing costs and enhancing safety (Talaghat, 2024). However, the effectiveness of these systems depends on the quality and diversity of labeled datasets, which traditional methods fail to achieve due to data scarcity and generalization limitations (Kumar, 2024). To overcome this, diffusion-based generative AI models offer a robust solution, generating diverse, high-quality synthetic images that enhance segmentation accuracy and support scalable, robust digital twin applications in pavement management (Ho et al., 2020).

2 LITERATURE REVIEW

Several studies have highlighted the potential of diffusion models to outperform GANs and VAEs in stability and image quality (Nichol & Dhariwal, 2021). These capabilities enhance training datasets, improving the performance and generalization of machine learning models. Han et al. (2024) proposed CrackDiffusion, a two-stage framework that integrates diffusion models with U-Net, achieving superior IoU scores on public datasets. Cano-Ortiz et al. (2024) introduced RoadPainter, a semantic diffusion model that significantly improved segmentation efficiency by augmenting datasets with diverse synthetic crack images. Additionally, Yan et al. (2024) demonstrated how integrating diffusion models with Transformer-in-Transformer algorithms enhanced road surface friction coefficient detection, achieving accuracy improvement. These studies exemplify the transformative potential of diffusion-based generative AI in creating scalable, effective solutions for digital twins and infrastructure monitoring (Xu et al. 2024).

Generative AI in pavement distress detection and segmentation is underexplored, with challenges like data scarcity and computational complexity. Datasets like Crack500 lack diversity, limiting model training for rare distress types (Zhang et al., 2020). Traditional generative models like GANs face instability (Goodfellow et al., 2014). Diffusion models offer a robust solution by generating high-quality synthetic images that improve data diversity and segmentation accuracy while reducing computational demands (Shorten et al., 2019). This research addresses these gaps using DDPM to enhance dataset robustness and model scalability.

3 OBJECTIVE AND SCOPE

This study applies diffusion models to augment pavement distress data, enhancing crack segmentation via deep learning. Contributions include detecting complex cracks like alligator cracks, optimizing image augmentation for peak performance, and testing model robustness across domains. The scope involves using DDPM for augmentation and U-Net/DeepLab models, focusing on longitudinal, transverse, and alligator cracks in asphalt pavements.

4 METHODOLOGY

4.1 *Dataset preparation*

The dataset preparation for this study involved characterizing the Crack500 dataset and applying essential preprocessing techniques to ensure data quality. The Crack500 dataset includes various types of pavement distress, such as longitudinal, transverse, and alligator cracks, with annotations indicating the type and location of distress. Preprocessing steps included resizing images for model compatibility, normalizing pixel values, adjusting contrast and brightness, and converting images to grayscale. Data cleaning removed low-quality or corrupted images, while histogram equalization and Gaussian blur were applied to enhance contrast and reduce noise, ensuring a consistent and high-quality dataset for robust model training.

4.2 *Diffusion Model Architecture*

The diffusion model architecture is central to generating high-quality synthetic images that augment existing datasets for pavement distress detection. Built on the DDPM (UNet2DModel), this architecture works by progressively adding Gaussian noise to input images and then learning to reverse this process, gradually restoring the images to their original state. This iterative denoising approach allows the model to generate highly realistic images that closely resemble real-world pavement distress conditions. The model utilizes a U-net structure, which includes encoder-decoder blocks and skip connections to retain fine spatial details during the downsampling and upsampling stages. This ensures the preservation of crucial features, such as crack patterns and textures, which are essential for accurate segmentation. By enhancing dataset diversity and improving image quality, this model supports more robust training of pavement distress segmentation models, ultimately improving their performance and generalization. The effectiveness of the generated images is measured using metrics such as Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), and diversity score, confirming high image quality and variability, which boost model performance and generalization.

4.3 *Segmentation model framework*

This study uses two state-of-the-art segmentation models, U-Net (ResUNet with ResNet34 Backbone) and DeepLab (DeepLabV3+ with ResNet101 Backbone), to develop a robust pavement distress detection system. U-Net, with its encoder-decoder structure and skip connections, is effective for tasks requiring precise localization, while DeepLab employs atrous convolution and spatial pyramid pooling to capture multi-scale context. Both models were configured with a uniform input size of 128x128 pixels and trained using cross-entropy and Dice loss functions, optimized by the Adam optimizer. Performance was evaluated using metrics such as IoU, F1 Score, pixel-level accuracy, and boundary precision, to assess the effectiveness of the models when trained on augmented datasets from diffusion models. These models aim to improve the robustness and accuracy of pavement distress segmentation.

4.4 *Comparative analysis framework*

The comparative analysis framework evaluates performance improvements from integrating generative AI into pavement distress segmentation models. This involves pre- and post-generative AI performance benchmarking using metrics like IoU, F1 Score, and pixel-level accuracy, followed by statistical validation through paired t-tests, confidence intervals, and cross-validation. The baseline performance of U-Net and DeepLab was first assessed using the original Crack500 dataset (Including 250 segmented image), revealing model strengths such as U-Net's performance on alligator cracks and DeepLab's robustness under variable lighting. After augmenting the dataset with synthetic images generated by the diffusion model, the models were retrained and reassessed using the same performance metrics, with additional metrics like diversity and realism scores to assess synthetic data quality. This framework demonstrates the effectiveness of generative AI in enhancing model performance and generalization.

5 RESULTS

5.1 *Generated Image Characteristics*

The diffusion model enhances dataset quality and model performance by progressively refining noisy synthetic images, revealing basic patterns by epoch 50, clearer structures by epoch 100, detailed cracks by epoch 150, and real-world pavement distress complexity by epoch 200.

Figure 1 provides a visual comparison of original pavement distress images and synthetic images generated by the diffusion model.

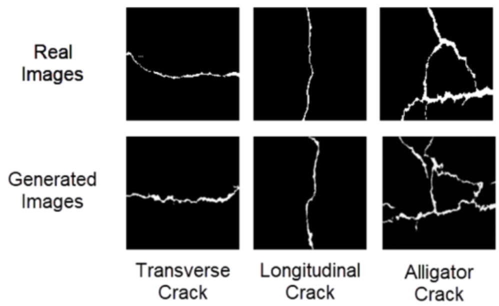

Figure 1. Comparison of real distress and synthetic images.

This side-by-side comparison illustrates how well the synthetic images replicate key visual features of real pavement cracks, including crack shape, width, alignment, and continuity. Additionally, the synthetic images introduce variability, which helps improve model robustness by capturing a wider range of crack patterns and distress types.

The evaluation of these synthetic images was conducted using several quantitative metrics to assess their quality and realism, as detailed in Table 1. These metrics are as follows:

Table 1. Synthetic image metrics.

Metric	Value	Interpretation
SSIM	0.92	High structural similarity to real-world images
PSNR	32.5 dB	Low noise and high-quality synthetic images
Diversity Score	0.85	High variability in crack types and patterns
Realism Score	0.90	Realistic appearance matching real-world data

These metrics confirm that the synthetic images maintain structural fidelity, add diversity, and closely resemble real-world data, enhancing dataset quality and improving segmentation model performance.

5.2 *Segmentation performance metrics*

The performance of U-Net and DeepLab models was evaluated using several metrics before and after dataset augmentation with synthetic images. Table 2 presents the performance improvements, including IoU, F1 Score, pixel-level accuracy, and boundary precision, with significant gains observed after augmentation. Both models showed notable performance enhancements with 1000 synthetic images, with IoU scores increasing by 8-9%, F1 Scores improving by 8%, and pixel-level accuracy rising by 4%. However, beyond 1000 augmented images, performance declined, suggesting diminishing returns and potential overfitting.

Table 2. Quantitative results of segmentation models.

Metric	IoU	F1 Score	Pixel-Level Accuracy	Boundary Precision	Dice Coefficient
U-Net (Original)	0.81	0.85	0.89	0.79	0.84
U-Net (+50)	0.84	0.87	0.9	0.81	0.86
U-Net (+150)	0.87	0.9	0.92	0.84	0.89
U-Net (+250)	0.9	0.93	0.93	0.87	0.92
U-Net (+1000)	0.93	0.95	0.95	0.9	0.94
U-Net (+5000)	0.91	0.94	0.93	0.88	0.92
DeepLab (Original)	0.84	0.87	0.91	0.81	0.87
DeepLab (+50)	0.86	0.89	0.92	0.83	0.89
DeepLab (+150)	0.89	0.92	0.94	0.86	0.92
DeepLab (+250)	0.92	0.95	0.95	0.88	0.94
DeepLab (+1000)	0.94	0.96	0.96	0.9	0.95
DeepLab (+5000)	0.92	0.94	0.94	0.87	0.93

Figure 2 demonstrates the improved training and validation curves for U-Net models trained with 1000 synthetic images, showing faster convergence and reduced overfitting.

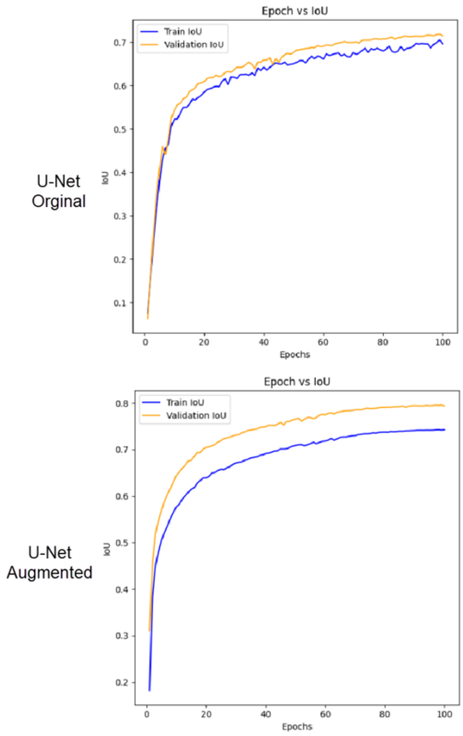

Figure 2. Training and validation curve of u-net model.

Further validation on additional datasets (GAPs, DeepCrack, CrackSegNet, and CFD) showed that both models maintained strong generalization capabilities, achieving high IoU, F1 Scores, and accuracy across different pavement distress types and conditions (see Table 3).

Table 3. Performance of models in percentage on different datasets with and without augmentation.

Dataset	Model	IoU	F1 Score	Pixel-Level Accuracy	Boundary Precision	Dice Coefficient
GAPs	U-Net	86 (75)	89 (80)	93 (85)	88 (78)	90 (82)
	DeepLab	88 (76)	91 (81)	94 (86)	90 (79)	92 (83)
DeepCrack	U-Net	85 (74)	88 (79)	92 (84)	87 (77)	89 (81)
	DeepLab	87 (75)	90 (80)	93 (85)	89 (78)	91 (82)
CrackSegNet	U-Net	84 (73)	87 (78)	91 (83)	86 (76)	88 (80)
	DeepLab	86 (74)	89 (79)	92 (84)	88 (77)	90 (81)
CFD	U-Net	83 (72)	86 (77)	90 (82)	85 (75)	87 (79)
	DeepLab	85 (73)	88 (78)	91 (83)	87 (76)	89 (80)

These findings demonstrate that the dataset augmentation using synthetic images significantly improves segmentation performance and generalization, validating the effectiveness of the generative AI approach in enhancing pavement distress detection models.

5.3 *Performance enhancement areas*

The integration of synthetic images generated by the diffusion model led to significant improvements in pavement distress segmentation, focusing on accuracy, boundary precision, and generalization.

Figure 3 shows a comparison of real images and predicted images from both original and augmented models, illustrating the enhanced accuracy and clarity of crack segmentation with the augmented dataset.

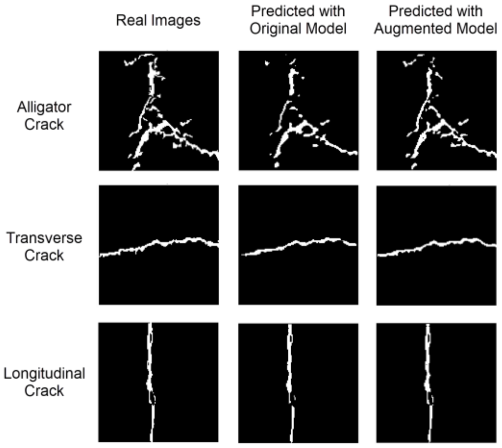

Figure 3. Comparison of real image with predicted using original and augmented models.

Boundary precision improved from 0.79 to 0.9, enhancing crack boundary delineation and reducing false positives/negatives. The diffusion model required more resources, taking 2 hours for training and 8 hours for inference with 1000 images. In comparison, the augmented U-Net and DeepLab models showed faster training and inference but demanded more computational power than their baseline versions.

Despite the benefits of using diffusion-based generative AI for pavement distress segmentation, several limitations exist. The approach is computationally intensive, requiring significant resources for training and image generation. There's also a risk of overfitting when using excessive synthetic data, which can reduce generalization. Initial datasets may contain biases that limit the diversity of synthetic images. Additionally, limited validation on diverse real-world conditions affects generalizability, and synthetic images may not fully capture the complexity of real-world data.

6 CONCLUSION

This study highlights the impact of diffusion-based generative AI in enhancing pavement distress segmentation. Synthetic images with high SSIM (0.92), diversity (0.85), and realism (0.90) led to significant performance gains. For U-Net, IoU improved from 0.81 to 0.93, F1 Score from 0.85 to 0.95, and pixel-level accuracy from 0.89 to 0.95. DeepLab saw similar improvements, with IoU rising from 0.84 to 0.94. Boundary precision improved by 9-11%, and generalization on unseen data increased, with IoU rising from 0.73 to 0.85. Despite high computational demands, the integration of synthetic data led to enhanced model performance and generalization. These findings offer valuable implications for infrastructure maintenance, safety, and cost reduction, with potential for broader applications. Future research should focus on real-time implementation, dataset expansion, and AI optimization.

ACKNOWLEDGEMENT

This paper is supported by Business Finland via Project 3992/31/2023 DigiPave.

REFERENCES

Cano-Ortiz, S., Sainz-Ortiz, E., Iglesias, L. L., del Árbol, P. M. R., & Castro-Fresno, D. (2024). Enhancing pavement crack segmentation via semantic diffusion synthesis model for strategic road assessment. *Results in Engineering*, 23, 102745.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672-2680.

Han, C., Yang, H., Ma, T., Wang, S., Zhao, C., & Yang, Y. (2024). CrackDiffusion: A two-stage semantic segmentation framework for pavement crack combining unsupervised and supervised processes. *Automation in Construction*, 160, 105332.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 6840-6851.

Kumar, A. V. (2024). Pavement surface condition assessment: A state-of-the-art research review and future perspective. *Innovative Infrastructure Solutions*, 9(470). Retrieved from Springer.

Nichol, A., & Dhariwal, P. (2021). Improved denoising diffusion probabilistic models. *Proceedings of the International Conference on Machine Learning*, 139, 8162–8171.

Saghatforoush, E., Abbasianjahromi, H., Talaghat, M. A., & Kahvandi, Z. (2022). Development of a Management System to Improve the Energy Efficiency of Public Buildings by Integrating IoT and BIM. *Lean Construction Journal*, 82-104

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60.

Talaghat, M. A., Golroo, A., Kharbouch, A., Rasti, M., Heikkilä, R., & Jurva, R. (2024). Digital twin technology for road pavement. *Automation in Construction*, 168, 105826.

Xu, H., Omitaomu, F., Sabri, S., Zlatanova, S., Li, X., & Song, Y. (2024). Leveraging generative AI for urban digital twins: a scoping review on the autonomous generation of urban data, scenarios, designs, and 3D city models for smart city advancement. *Urban Informatics*, 3(1), 29.

Yan, Z., Yue, L., Luo, W., & Sun, J. (2025). Real-time detection of road surface friction coefficient: A new framework integrating diffusion model and Transformer in Transformer algorithms. *Alexandria Engineering Journal*, 113, 620-632.

Zhang, Z., Liu, Q., & Wang, Y. (2020). Crack500: A dataset for pavement crack detection. *Data in Brief*, 30, 105409.