Exploring the Potential of Synthetic Data for Bike Path Surface Classification using Diffusion Models

Valentin Göttisheim[1], Holger Ziekow[1], Peter Schanbacher[1], Oliver Taminé[1], Jochen Baier[1]
Djafar Ould-Abdelsam[2]

[1] Furtwangen University,
{goev, zie, spet, tao, bai}@hs-furtwangen.de
[2] Université de Haute Alsace
djafar.ould-abdeslam@uha.fr

Abstract. Road surface classification of bike paths enables image recognition applications for bike route planning, navigation optimization or path maintenance. However, acquiring and annotating real data can be costly and time-consuming. Synthetic data can overcome data scarcity and annotation costs. We use synthetic data, generated by Stable Diffusion to improve neural network performance on new or unseen surfaces. We compare model performance for different real-synthetic data ratios. Our results show that synthetic data decreases the amount real data needed and improves neural network performance in road surface classification on new surfaces.

Keywords: synthetic data; diffusion models, image recognition, surface classification.

1 Introduction

We analyze road surface classification in the context of an image recognition application that supports bike route planning, navigation optimization, path maintenance, and traffic safety (Baier et al., 2023). A challenge for the system is the classification of unseen road or path surfaces, as acquiring and annotating large-scale, diverse, and accurately labeled real data can be costly and time-consuming. Synthetic data is artificially generated data that mimics the characteristics of the real-world and offers a potential solution to address the limitations of data availability and annotation efforts. Unseen path surfaces are surfaces that are not present in the training data but appear in the real world. For example, a surface classifier trained on asphalt, concrete, and gravel may encounter difficulties to classify images of bricks, snow, other gravel types or lightings conditions which diverge from the images present in the training data resulting in a poor performance of the classifier. A possible solution to this problem is to enhance new path surfaces with synthetic images which cover a wider range of surfaces variations to improve the generalization and robustness of the classifier (He et al., 2023; Lu et al., 2023).

2 Related Work

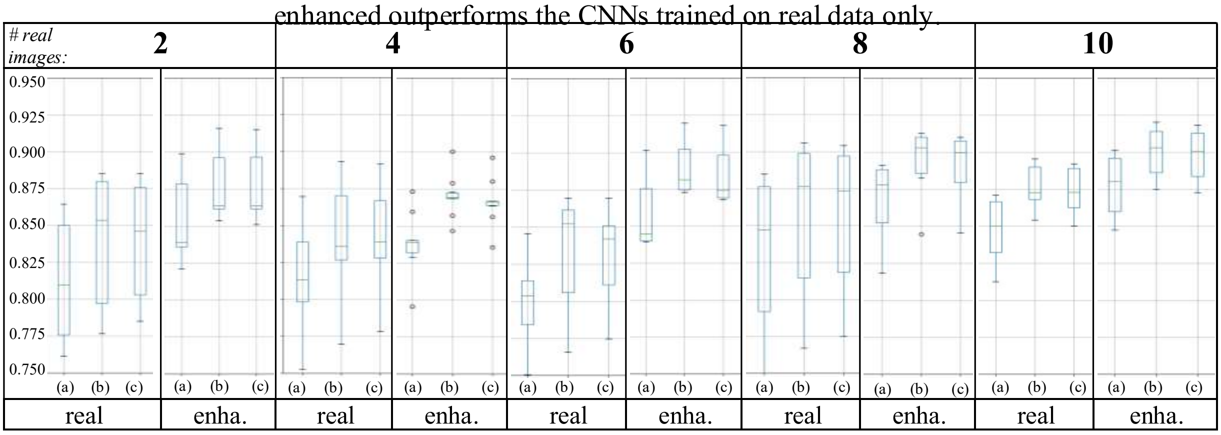
Bike path surface classification is a task of identifying the type or condition of road surfaces captured from images, videos or other sensors (e.g. Heidt and Dorer, 2021). Baier et al. (2023) proposed an approach for automatic analysis of bike paths. They used a convolutional neural network (CNN) to detect and classify road surfaces from camera images. Our work shares the experimental setup, however, they did not consider the use of synthetic data. Several methods have been proposed for synthetic image generation, such as generative adversarial networks (GANs) (Goodfellow et al., 2014) or latent diffusion models (LDMs) such as Stable Diffusion (Rombach et al., 2022). The general potential of using synthetic data from diffusion models has been shown by (Azizi et al., 2023) ImageNet data. In our work we use Stable Diffusion in the use case of path surface classification. We are focusing on easing the addition of new surfaces to a classification system for path surfaces, by expanding small real-world data samples with synthetic data.

Figure 1. "Original and Synthetic Data": Examples of input-image and generated synthetic images.

3 Experiment Design and Results

We tested the hypothesis that synthetic data can enhance the training for new bike paths, in particular with a small number of real images. We compare the performance of two CNNs: one trained on real data and other ones trained on real data plus synthetic data. We evaluate both types on a new bike path dataset. We test different numbers of real images as basis for synthetic data, ranging from 0 to 10 per class, and analyze the performance gain through synthetic data. We collected a new bike path dataset similar to Baier et al., (2023). The dataset contained 4 different paths with 300 images per path, for a total of 1200 real images. For 10 real images of the new dataset, we generate 400 synthetic variations for each image using the common Stable Diffusion image-to-image 768-v-ema (Rombach et al., 2022) method. We use input prompts ("photo of a paved path, concrete, asphalt, Canon EOS R3") and a negative prompt ("painting, digital art") to match the variations of the real data samples. Figure 1 shows examples. Note that we do not evaluate how realistic the synthetic images are. Rather we focus on the improvement of synthetic images on model performance. We constructed 6 training sets (referred to as "real") with each class containing exactly 0, 2, 4, 6, 8 or 10 real new bike path images in addition to the original data set (20444 samples). Further, we created 6 additional data sets by adding the corresponding synthetic images for each real image (referred to as "enhanced"). For example, for 2 real world images per class we added all synthetic variations resulting in 22048 images (20444 original samples +2·2 real world images +1600 synthetic images) in total. As a common

Figure 2. "Performance Synthetic Enhanced vs Real Data": CNNs trained on synthetic enhanced (enha.) and on purely real data (real) with number of real images used (# real images). Reported f1-scores: (a) macro, (b) micro, (c) weighted. The comparison of performances shows the synthetic enhanced outperforms the CNNs trained on real data only.



CNN architecture we use one augmentation layer, three convolutional layers with ReLu activation and max-pooling layer, a dropout and two fully connected layers. The augmentation involves random horizontal flipping, zooming and change in brightness. We trained the CNN with subsequent settings: 25 epochs, batch size 32, learning rate of 0.001, sparse cross-entropy loss, Adam-optimizer and accuracy as metric. For evaluation we used the model with the best validation loss. We trained the CNN on the original training dataset and evaluated it on the test set of the original dataset achieving an f1-score of 0.97. However, on the test set of the new bike path data, the f1 dropped to 0.83, showing that the CNN trained on the original dataset performs worse on the new bike paths. To test our hypothesis, we trained the CNN on the remaining final real world and synthetic enhanced datasets and compare the performance on the new bike path test set. We repeated the experiment 10 times. The results are visualized in Figure 2 which shows the boxplots of the weighted, macro and micro scores of f1 versus the amount of real data needed for different models on the bike path test set. Here, we only discuss the micro f1-score for simplicity. The classifiers trained on synthetic enhanced data outperform those trained on real data only (see Figure 2). The positive effect of synthetic data increases with the number of real images used. For example, with 2 real images the median f1 are 0.868 (real data) and 0.875 (enhanced data) compared to 10 images with 0.873 (real data) and 0.91 (enhanced data). We further verified the performance on the original test set during the experiments which showed only small variations of about an f1-score of 0.97 for all trained classifiers.

4 Discussion and Conclusion

We generate synthetic data for bike path surface classification using diffusion models. We compared two CNNs trained on real data and real data plus synthetic data and evaluated them on a new bike path dataset. We showed that enhancing real path images with synthetic data improves the classifier performance and therefore can enhance training for new bike paths surfaces. Our work has some limitations and challenges for future work. First, our synthetic data generation method relies on text prompts and model parameters, which may be subject to further optimization. E.g., the generated asphalted paths could be improved using other inputs. Second, our experiment was conducted on a small dataset of bike path images, which may limit the generalization and robustness. However, our work showed a promising approach, demonstrating the of use synthetic data to improve the CNN performance, especially on new surfaces.

References

1. Baier, J., Taminé, O., Deyringer, J., Mahn, N., Dittrich, I., Alili, B.: A Multi-Stage AI-based Approach for Automatic Analyzation of Bike Paths: Stage 1 – Road Surface Detection, TFC23: Towards AI-Aided Invention and Innovation, 2023
2. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y.: Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680), 2014
3. He, R., Sun, S., Yu, X., Xue, C., Zhang, W., Torr, P., Bai, S., & Qi, X.: Is synthetic data from generative models ready for image recognition? arXiv preprint arXiv:2210.07574, 2023.
4. Lu, Y., Shen, M., Wang, H., & Wei, W.: Machine learning for synthetic data generation: A review. arXiv preprint arXiv:2302.04062, 2023.
5. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B.: High-resolution image synthesis with latent diffusion models. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

6. Azizi, S., Kornblith, S., Saharia, C., Norouzi, M., & Fleet, D. J.: Synthetic data from diffusion models improves imagenet classification. arXiv preprint arXiv:2304.08466, 2023.
7. Heidt, Johannes, and Klaus Dorer. "Classification and Prediction of Bicycle-Road-Quality using IMU Data." In Artificial Intelligence: Application in Life Sciences and Beyond, pp. 138-149. 2021.