Using Synthetic Data and Artificial Intelligence for Optimizing Tillage Process Quality Measurement

M. Sc. **S. Schulpius**, AGCO GmbH, Wolfenbüttel, Germany;
M. Sc. **M. Graf**, AGCO GmbH, Marktoberdorf, Germany;
M. Sc. **T. Stirnkorb**, AGCO GmbH, Marktoberdorf, Germany;
Prof. Dr. **L. Frerichs**, Technische Universität Braunschweig, Braunschweig;
Prof. Dr.-Ing. **M. Geimer**, Karlsruhe Institute of Technology, Karlsruhe

Abstract

Automation in agriculture enhances efficiency and productivity. Taking tillage as an example, driving tasks such as steering and speed control are already highly automated, shifting the focus toward automating the tillage process itself. Measuring crop residue coverage - a key factor for erosion resistance, soil structure, and moisture - with semantic segmentation of camera images requires large, accurately annotated datasets. Manual annotation is time-consuming, error-prone, and challenging due to the fine structures of straw and the indistinct boundaries of soil aggregates. To overcome these issues, synthetic training data were generated using the modeling software Blender to model soil textures, residue distributions, and environmental conditions. Photorealism was subsequently enhanced through the machine learning method ControlNet. The approach was evaluated and tested using three datasets - real-world, Blender-generated, and ControlNet-generated - assessed with the mean Intersection over Union (mIoU) and Fréchet Inception Distance (FID) metrics. A semantic segmentation network, PIDNet, trained on real-world data, achieved an mIoU of 75.0 %. The network trained on the Blender dataset obtained 52.9 % due to limited realism. In contrast, ControlNet-generated data achieved 69.3 % with improved FID scores compared to the Blender dataset, indicating higher realism and superior model performance. Finally, after fine-tuning the segmentation model based on the ControlNet dataset with real data, an mIoU of 75.4 % was reached. These findings indicate that high-quality synthetic data can reduce annotation effort, minimize labeling errors, and, in some instances, outperform real data in training machine learning models.

1. Introduction

Efficient process control in tillage requires accurate measurement of key soil and surface parameters, including crop residue coverage, surface roughness, or aggregate size distribution. While operational tasks such as steering and speed regulation have largely been automated, the automation of process control itself remains an open challenge in many respects.

This contribution focuses on crop residue coverage, a parameter that influences erosion resistance, soil structure, and moisture. It refers to the proportion of soil surface covered by organic matter after tillage and can be measured using either offline or online methods. Offline approaches include the meterstick method [1], [2], comparison with images [3] or grid-based techniques [4]. Online approaches encompass various methods, such as those proposed by [5], [6], [7], [8], [9]. While offline methods are not suitable for real-time applications, online methods are more relevant for automation. Among these, techniques based on thresholds [5], or edge detection [6] tend to be less robust compared to machine learning-based methods. However, a significant limitation of machine learning approaches is their reliance on annotated training data.

The manual annotation process is time-consuming and challenging due to the difficulty of accurately annotating fine objects, such as straw stalks, and uncertainties caused by unclear fracture edges of soil aggregates. To overcome these challenges, this contribution utilizes synthetic data to facilitate data acquisition and enhance annotation. One advantage of this approach is the ability to generate specific scenarios and environmental conditions. A second advantage is the automated generation of annotated data, which ensures an error-free ground truth and accelerates the data generation.

Cieslak et al. [10] developed a method for generating synthetic training data for the segmentation of soybean plants and weeds in agricultural fields. In this approach, 3D models of soybean plants, grass weeds, and broadleaf weeds - procedurally generated by specialized software - are randomly arranged within a 3D scene. Using the GAN-Model Contrastive Unpaired Translation (CUT), the authors investigated the domain adaptation. They pointed out that using real data and synthetic data results in segmentation models that demonstrate at least equivalent performance to those that are trained only on real-world data. Models trained on synthetic and real-world data exhibit higher generalizability.

Another tool to reduce the domain gap between real and synthetic data is the machine learning method ControlNet [11]. ControlNet operates by taking a pre-trained text-to-image diffusion model and freezing its original weights. A parallel, trainable branch is then added, connected through zero-initialized convolutions, to introduce new conditional inputs without altering the base model. This design enables the integration of spatial guidance, such as edges, depth maps, or poses, with text prompts, allowing precise control over the generated image’s structure. ControlNet has been applied in various domains, for example, to create and modify material microstructures with precise control over shapes [12] or to generate realistic weed images to improve weed detection performance by 1.4 % compared to using only real images [13].

Schulpius et al. [14] used ControlNet to enhance tramline detection using synthetic images. By fine-tuning the model with real-world data, they increased the segmentation accuracy to an mIoU of 83.28 %, representing a 1.61 % improvement over the baseline.

2. Method and Data

For the real-world dataset, images were collected during the tillage process using two different cultivators equipped with varying roller and tine configurations. The camera system was mounted behind the roller to capture the outcome – the quality – of the tillage process. The dataset includes two distinct categories of tillage practices: primary tillage in spring (organic matter: intermediate crops), and stubble cultivation after harvest (organic matter: wheat and corn). To ensure robustness and variability, additional images were captured using camera systems with different resolutions, including smartphone cameras. The dataset covers a wide range of conditions, with crop residue coverage in the photos varying from low (less than 10 %) to high (over 50 %). Overall, the real-world dataset included 115 training, 20 validation, and 15 test images with manually generated annotations, as shown in Figure 1.

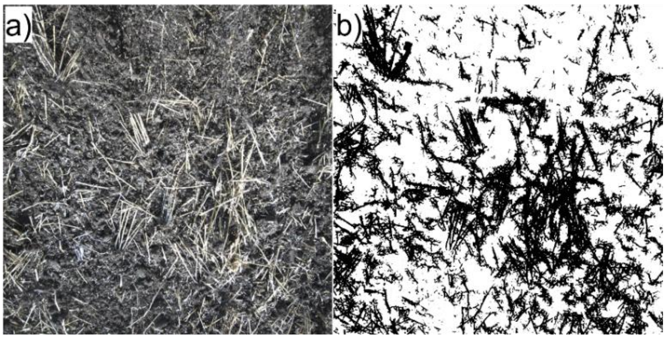

Figure 1: Exemplary image (a) and manual annotation mask (b) from the real-world dataset

Synthetic images were generated using the open-source 3D modeling software Blender [15]. For this purpose, the procedural pipeline BlenderProc [16] was used, enabling the creation of not only RGB images but also segmentation masks and other outputs. The synthetic images were derived from a 3D scene in Blender, consisting of a ground surface with randomly placed crops and crop residues. In total, five different surface types, five crop objects, and various lighting conditions were used. This effort resulted in a dataset consisting of 3100 images.

A domain gap is often observed between synthetic and real images. To address this, the machine learning method ControlNet was used to generate improved synthetic images using annotation masks created with Blender. The workflow for creating synthetic images with

Blender is shown in Figure 2. For training, 115 real images (1024x1024 px) with corresponding annotation masks were used. Furthermore, each image was paired with a text prompt, “Close-up image of straw on soil,” which is necessary for training. After training ControlNet, the model was used to generate 3100 new images based on 3100 annotation masks created in Blender and the same text prompt describing the desired output.

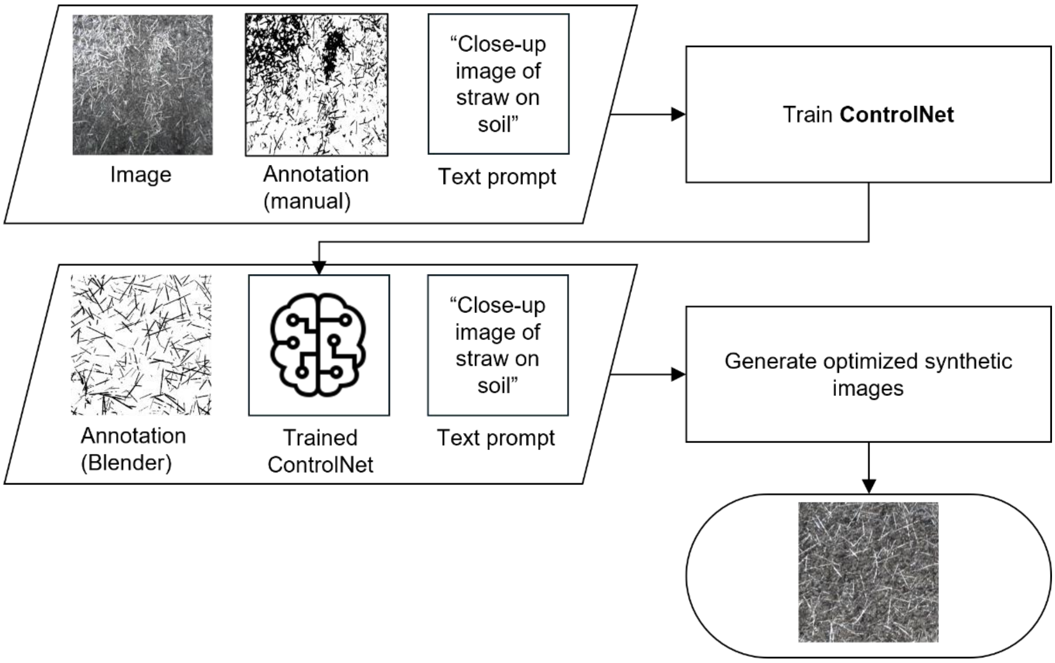
“Close-up image of straw on soil”
Image
Annotation (manual)
Text prompt
Train ControlNet
“Close-up image of straw on soil”
Annotation (Blender)
Trained ControlNet
Text prompt
Generate optimized synthetic images

Figure 2: Workflow to generate images with ControlNet

To evaluate the quality of the different datasets, two metrics were used: mean Intersection over Union (mIoU) for assessing semantic segmentation performance, and Fréchet Inception Distance (FID) for measuring the realism of the generated images.

For the mIoU evaluation, the semantic segmentation network PIDNet [17] was used. The PIDNet architecture is specifically designed for real-time semantic segmentation. It effectively balances segmentation accuracy and computational efficiency, rendering it particularly suitable for deployment in resource-constrained environments. The network was trained under four distinct conditions. Initially, the network was trained separately on three datasets: the real-world dataset, the Blender dataset, and the ControlNet dataset. To further enhance performance, the model trained on the ControlNet dataset was subsequently fine-tuned using real-world images. Fine-tuning a semantic segmentation model in this context denotes adapting a pre-trained network to a target domain using domain-specific data, such as real-world images. The segmentation task was defined as two classes: background and organic matter. The organic matter class encompassed all plant-derived materials without further distinction, including harvested crop residues such as straw, emerging crop plants, and weeds. All datasets were evaluated using a consistent validation set comprising 20 manually annotated real-world images, which were used during training to monitor model performance. Additionally, final performance was assessed on an independent test set consisting of 15 real-world images. In total, eight mIoU scores were computed: two for each of the three primary training datasets (real-world, Blender, ControlNet) and two for the fine-tuned ControlNet-based model.

An exemplary external real-world dataset of 70 images was used as a reference distribution for the FID computation. Each of the three training datasets - real-world, Blender-generated, and ControlNet-generated - was compared against this external real-world dataset to obtain corresponding FID scores. Lower FID scores indicate a higher visual similarity to the real-world reference data, and thus better generative quality.

3. Results

Figure 3 presents example images from each dataset: (a) real-world, (b) Blender-generated, (c) ControlNet-generated image. These examples illustrate the varying degrees of realism and complexity across datasets.

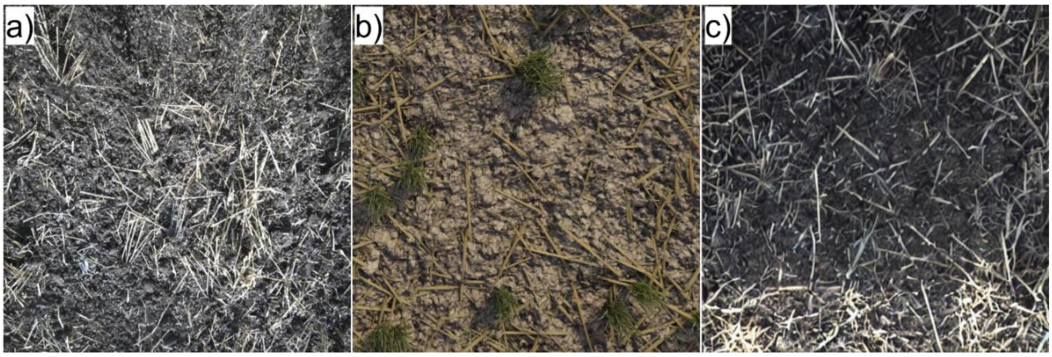

Figure 3: Example images of the datasets: a) real-world dataset; b) Blender dataset; c) ControlNet dataset

To quantitatively assess the segmentation performance and visual realism of the datasets, both the mIoU and FID metrics were evaluated. When trained on the real-world dataset, PIDNet achieved an mIoU of 73.6 % on the validation set and 75.0 % on the test set. In comparison, training on the Blender-generated dataset resulted in substantially lower mIoU scores of 50.5 % (validation) and 52.9 % (test). The ControlNet-generated dataset yielded improved results, with mIoU scores of 71.2 % (validation) and 69.3 % (test). Fine-tuning the ControlNet-based model with real-world data further increased performance, reaching 74.1 % on the validation set and 75.4 % on the test set.

To complement the segmentation evaluation, FID scores were computed against an external real-world reference dataset. The real-world dataset achieved the lowest FID score of 88.29. The Blender-generated dataset exhibited an FID of 257.3, while the ControlNet-generated data achieved an FID of 225.4.

4. Discussion

The presented results highlight distinct differences in segmentation performance across the evaluated datasets. The model trained on the real-world dataset achieved a test mIoU of 75.0 %, serving as the baseline for comparison. The Blender-generated dataset resulted in a test mIoU of 52.9 %, indicating a pronounced domain gap. This gap is attributable to the reduced photorealism of the synthetic images, as illustrated in Figure 3, and is further supported by the high Fréchet Inception Distance (FID) score of 257.3, which reflects substantial visual dissimilarity from real-world images.

The ControlNet-generated dataset yielded a test mIoU of 69.3 %, representing a considerable improvement over the Blender dataset and approaching the performance of the real-world baseline. Its lower FID score of 225.4 suggests enhanced visual realism, which contributes to improved feature learning for organic matter segmentation. Fine-tuning the ControlNet-based model with real-world data further increased the test mIoU to 75.4 %, surpassing the baseline and demonstrating the benefit of combining synthetic and real data.

Despite these improvements, the FID score of the ControlNet dataset remains higher than that of the real-world dataset (88.3), indicating that a domain gap still exists. This may be attributed to the synthetic nature of the ControlNet images, which, despite improved realism, do not fully capture the complexity and variability present in real-world scenes. In particular, the spatial distribution and appearance of organic matter in the synthetic data are constrained by the underlying Blender annotations and generation process.

These findings emphasize the importance of both visual realism and dataset diversity in training semantic segmentation models. While high-quality synthetic data can enhance model performance, especially when combined with real-world samples, it cannot yet fully replace the variability and richness of real-world data.

5. Conclusion

This study demonstrates that the realism, quantity, and diversity of training data have a substantial impact on semantic segmentation performance when evaluated on real-world imagery. The PIDNet model trained on the real-world dataset achieved a test mIoU of 75.0 %. In comparison, the Blender-generated dataset resulted in a test mIoU of 52.9 %, highlighting the limitations of synthetic data with reduced visual realism. The ControlNet-generated dataset improved performance to 69.3 %, indicating better alignment with real-world features. Further fine-tuning of the ControlNet-based model using real-world data increased the test mIoU to 75.4 %, slightly surpassing the baseline. These results suggest that ControlNet offers a promising approach for generating realistic training data, particularly for structures that are difficult to annotate manually, such as fine organic material like straw. For other types of organic matter, additional real-world data may be required, as the current approach may be tailored to straw-specific scenarios. Future work should focus on enhancing the semantic and structural realism of synthetic images, increasing dataset diversity, and evaluating generalization across broader agricultural contexts.

6. References

[1] J. E. Adams and G. F. Arkin, “A Light Interception Method for Measuring Row Crop Ground Cover,” *Soil Science Soc of Amer J*, vol. 41, no. 4, pp. 789–792, Jul. 1977, doi: 10.2136/sssaj1977.03615995004100040037x.

[2] D. P. Shelton and P. J. Jasa, “Estimating Percent Residue Cover Using the Line-Transect Method.” Accessed: Jun. 11, 2024. [Online]. Available: https://extensionpubs.unl.edu/publication/1085/html/view

[3] J. Brunotte and B. Ortmeier, *Fächer zur Bestimmung des Bodenbedeckungsgrades durch organische Rückstände*. 2007. Accessed: Jun. 11, 2024. [Online]. Available: https://www.openagrar.de/receive/timport_mods_00004395

[4] H.-H. Voßhenrich, J. Brunotte, and B. Ortmeier, “Gitterrastermethode mit Strohindex zur Bewertung der Stroheinarbeitung,” *LANDTECHNIK*, pp. 328-329 Seiten, Dec. 2005, doi: 10.15150/LT.2005.1255.

[5] F. Pforte, “Entwicklung Eines Online-Messverfahrens Zur Bestimmung Des Bodenbedeckungsgrades Bei Der Stoppelbearbeitung Zu Mulchsaatverfahren,” phdthesis, Universität Kassel, Kassel, 2010.

[6] A. Ribeiro, J. Ranz, X. P. Burgos-Artizzu, G. Pajares, M. J. Sanchez del Arco, and L. Navarrete, “An Image Segmentation Based on a Genetic Algorithm for Determining Soil Coverage by Crop Residues,” *Sensors*, vol. 11, no. 6, pp. 6480–6492, Jun. 2011, doi: 10.3390/s110606480.

[7] P. Riegler-Nurscher, J. Prankl, and M. Vincze, “Tillage Machine Control Based on a Vision System for Soil Roughness and Soil Cover Estimation,” in *Computer Vision Systems*, D. Tzovaras, D. Giakoumis, M. Vincze, and A. Argyros, Eds., Cham: Springer International Publishing, 2019, pp. 201–210. doi: 10.1007/978-3-030-34995-0_19.

[8] Y. Liu *et al.*, “Straw Segmentation Algorithm Based on Modified UNet in Complex Farmland Environment,” *CMC*, vol. 66, no. 1, pp. 247–262, 2020, doi: 10.32604/cmc.2020.012328.

[9] M. Schmidt, “AI-Based Tillage Job Quality Assessment for Advanced Machine Automation in Agriculture,” in *AgEng-LAND.TECHNIK 2022: International Conference on Agricultural Engineering*, in VDI-Berichte, vol. 2406. Berlin, Nov. 2022, pp. 567–572.

[10] M. Cieslak *et al.*, “Generating Diverse Agricultural Data for Vision-Based Farming Applications,” in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2024, pp. 5422–5431. doi: 10.1109/CVPRW63382.2024.00551.

[11] L. Zhang, A. Rao, and M. Agrawala, “Adding Conditional Control to Text-to-Image Diffusion Models,” Nov. 26, 2023, *arXiv*: arXiv:2302.05543. doi: 10.48550/arXiv.2302.05543.

[12] Y. Zhang, T. Long, and H. Zhang, “Stable diffusion for the inverse design of microstructures,” Sep. 27, 2024, *arXiv*: arXiv:2409.19133. doi: 10.48550/arXiv.2409.19133.

[13] B. Deng and Y. Lu, “Weed Image Augmentation by ControlNet-Added Stable Diffusion,” in *Synthetic Data for Artificial Intelligence and Machine Learning: Tools, Techniques,*

and Applications II, K. E. Manser, C. De Melo, R. M. Rao, and C. L. Howell, Eds., National Harbor, United States: SPIE, Jun. 2024, p. 25. doi: 10.1117/12.3014145.

[14] S. Schulpius, J. Schattenberg, and L. Frerichs, “Training neural networks for tramline detection in an autonomous driving tractor using synthetic images,” in *LAND.TECHNIK 2024*, in VDI-Berichte, vol. 2444. Düsseldorf: VDI Verlag, 2024.

[15] B. O. Community, *Blender - a 3D modeling and rendering package*. Stichting Blender Foundation, Amsterdam: Blender Foundation, 2018. [Online]. Available: http://www.blender.org

[16] M. Denninger *et al.*, “BlenderProc,” Oct. 25, 2019. doi: 10.48550/arXiv.1911.01911.

[17] J. Xu, Z. Xiong, and S. P. Bhattacharyya, “PIDNet: a real-time semantic segmentation network inspired by PID controllers,” in *2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 2023, pp. 19529–19539. doi: 10.1109/CVPR52729.2023.01871.