AUTOMATING THE SEGMENTATION OF X-RAY IMAGES WITH DEEP NEURAL NETWORKS

Alba Castrillo Perote (s230221), Andrea Matamoros Alonso (s233514),
Jesús Díaz Pereira (s233142), Fernando Augusto Marina Urriola (s233144)

Technical University of Denmark

ABSTRACT

This study focuses on the segmentation of tomographic X-ray images for solid oxide fuel cells. To address the time-consuming and error-prone manual segmentation, we built, trained, optimized, and compared two Deep Learning models, a U-Net and a Pix2Pix. In addition, we evaluated the performance of the model by progressively increasing the amount of training data. Results indicate superior performance of Pix2Pix, even with a small training dataset compared to U-Net, supporting the hypothesis that the presence of a discriminator can yield promising results. However, further investigation is required to correctly assess the generalization of these models on different types of images.

Link to the Github repository

***Index Terms*—** X-Ray, Synchrotron, Segmentation, Deep Learning, U-Net, Pix2Pix

1. INTRODUCTION

In the intricate realm of solid oxide fuel cells (SOFC) and electrolysis cells, the composition of their electrodes plays a pivotal role in ensuring optimal performance. In this context, composite cermet of nickel and yttria-stabilized zirconia is state-of-the-art [1]. However, a primary drawback associated with nickel utilization is that it can suffer from microstructural changes, ultimately leading to the degradation of the performance of the cell.

Changes in nickel's microstructure can be addressed through a wide range of imaging techniques, for example, ptychographic X-ray computed tomography (PXCT) [2]. This type of image consists of different voxels with different grayscale intensities, which are related to the attenuation of the material to X-rays given by properties such as their density. Before their analysis, these images need to be segmented, being manual segmentation or traditional methods such as histogram thresholding the conventional approaches. However, these techniques are time-consuming and prone to errors. Thus, advanced techniques such as Machine Learning, and Deep Learning have been proposed as an alternative to traditional segmentation methods [3, 4].

Thanks to DTU.

1.1. Motivation

After the previous facts have been stated, automating the segmentation process becomes imperative. To address this challenge, in this paper, we plan to leverage the power of deep neural networks to automate the segmentation of PXCT images, removing the need for human intervention. In particular, this paper outlines the utilization and comparison of two distinct neural network architectures to tackle this segmentation challenge.

2. MATERIALS AND METHODS

2.1. Data

The initial dataset includes 500 high-quality PXCT 2D grayscale images, each accompanied by corresponding manual segmentations. These images have a resolution of 512 x 512 pixels and were already post-processed. However, due to computational reasons, they were resized to a resolution of 256 x 256. For more information please contact: ***Salvatore De Angelis (sdea@dtu.dk)***.

2.2. Network Architectures

2.2.1. Pix2Pix

Generative Adversarial Networks (GANs) are a type of neural network introduced by Goodfellow et al. [5]. GANs consist of two distinct architectures, a generator and a discriminator, which are trained simultaneously through adversarial training. The generator creates synthetic data, while the discriminator evaluates its authenticity. The competition between these two networks results in the generator producing increasingly realistic data, making GANs particularly effective for image synthesis, style transfer, and data generation tasks.

In particular, Pix2Pix is a specific architecture that is based on conditional GANs (cGANs) and that was designed for image-to-image translation tasks [6]. Unlike traditional GANs that generate images from random noise, cGANS like Pix2Pix is composed of a generator that takes an input image and transforms it into a corresponding output image, while the discriminator evaluates the generated images and tries to distinguish them from real images in the output domain.

The Pix2Pix's discriminator works as a classifier, determining if the synthetic image is real or not; whereas the generator's objective function is described as a combination of two terms. The first term, represents the adversarial loss, pushing the generator to produce images that are convincing to the discriminator. The second term, is the L1 loss, ensuring that the generated images are structurally similar to the ground truth images. Thus, the Pix2Pix generator has two ways of updating its weights during training, one is by the internal circuit, and an external path is provided by the comparison of results between ground truth and fake images from the discriminator. Thus, the generator learns to produce segmented images that resemble more the target segmentations.

The Pix2Pix final objective function is given by:

$$\begin{aligned}\mathcal{L}_{\text{Pix2Pix}}(G, D) = \mathbb{E}_{x,y}\left[\log D(x, y)\right] \\ + \lambda \mathbb{E}_{x}\left[\|y - G(x)\|_1\right]\end{aligned} \tag{1}$$

Where the first term represents the adversarial loss, and the second term is the L1 Loss.

2.2.2. U-Net

U-Nets are Convolutional Neural Networks (CNNs) whose structure follows a characteristic *U* shape, conferred by the contracting and the expansive symmetric paths. The contracting path initiates with a series of convolutions alternated with max-pooling layers. This process reduces spatial dimensions and augments the number of feature maps, taking the network to its bottleneck, where the input is represented as a 1-D vector.

The expansive path operates in reverse, employing alternating convolutional and up-convolutional layers to increase spatial dimensions and decrease feature map numbers, ultimately restoring the input to its original shape. The final layer has a softmax activation function, which assigns a probability to each class label for each pixel. Each convolution has a ReLu activation function, except for the last one, and the max-poolings and up-convolutions are performed with a 2x2 kernel.

The original purpose for U-Nets was the segmentation of biomedical images [7]. U-nets were revolutionary due to the presence of the skip connections present in their structure. These connections join the pair of symmetric steps in the contracting and expansive paths and allow the network to concatenate feature maps, preserving fine-grained details and localization information.

2.2.3. Training

Both Neural Network models were built in Python 3 using TensorFlow Keras and trained using Kaggle's GPU T4. The initial dataset of 500 images was divided into train, validation, and test (70%, 20%, and 10%). The hyperparameters of each architecture were optimized using the validation set. The test set was used for final evaluation and model comparison. The DICE coefficient, as seen in Equation 2 is defined by the intersection over the union, was used to evaluate the segmentation performance of both, U-Net and Pix2Pix models [8].

$$\text{Dice-score} = \frac{2 \times |X \cap Y|}{|X| + |Y|} \tag{2}$$

Table 1 summarizes the optimal hyperparameters obtained on the validation set for both architectures, U-Net and Pix2Pix. The hyperparameters include learning rate, batch size, epochs, and specific weights associated with the Adam optimizer.

Network	Learning Rate	Batch Size	Epochs	B1 Weights
Pix2Pix	2×10^{-3}	4	35	0.9
U-Net	1×10^{-3}	32	35	-

Table 1. Optimal hyperparameters for U-Net and Pix2Pix. Adam optimizer was used in both networks.

For the U-Net architecture, the optimal configuration involves a learning rate of 1x10^{-3}, a batch size of 32, and training for 35 epochs. On the other hand, the Pix2Pix architecture demonstrated superior performance with a learning rate of 2x10^{-3}, a smaller batch size of 4, and 35 training epochs. Additionally, the optimal B1 weights for the Adam optimizer in this case were determined to be 0.9; B1 weights for the U-Net's Adam optimizer were not applied.

Finally, training and validation sets were merged into a single training dataset and were subsequently trained with different numbers of images. Their performance was evaluated using the same test set. This approach aimed to systematically explore the influence of dataset volume on model efficacy. We trained with different proportions of data, ranging from 10% to 100% of the new training set, increasing by 10% in each iteration. By doing this we tried to discern the minimum dataset size that ensures a satisfactory model performance, given the associates costs of data collection for training.

3. RESULTS

U-Net and Pix2Pix models were trained and optimized using the training and validation datasets respectively. Finally, both models were evaluated with the internal test dataset.

Figure 1 displays random examples of the segmentation results from both models on the test dataset. Each row shows the original image, the ground truth, and the segmentation results from each network. As illustrated, both models were able to segment the images with a high degree of quality, producing similar results to the ground truth.

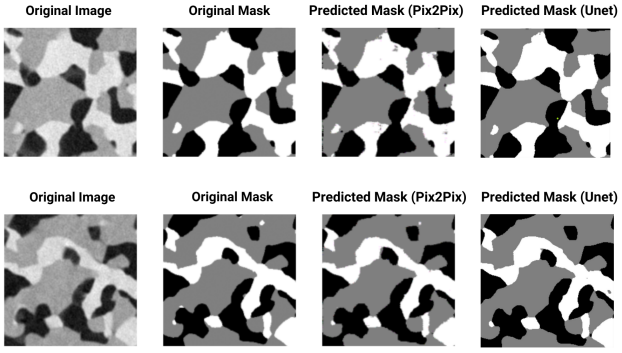

Fig. 1. Test of segmentation results of U-Net and Pix2Pix model for PXCT images segmentation, compared with ground truth.

Quantitative results obtained from the test set, including the Dice coefficient's mean and standard deviation from each model are reported in Table 2. The results show that Pix2Pix achieved a slightly higher mean Dice Coefficient than U-Net (0.987 vs. 0.905) indicating that it performed better on average.

Network	Mean Dice Coefficient	Mean Standard Deviation
Pix2Pix	0.987	0.001
U-Net	0.905	0.136

Table 2. Internal Test Results for Dice Coefficient

Figure 2 shows the effect of the training data size on the performance of Pix2Pix and U-Net. We trained the models with different percentages of the available training data (dynamic training), ranging from 10% to 100%, and evaluated their performance on the test dataset using the Dice Coefficient metric. The reason for the dynamic training is to try to obtain more robust models. The results showed that both models improved their performance as more training data was used, but Pix2Pix achieved a higher Dice coefficient with less data compared to U-Net. This evidence shows that Pix2Pix appears to be a more robust network when trained with less data compared to the U-Net.

4. DISCUSSION

This study focused on the comparative analysis for PXCT image segmentation using standard U-Net and Pix2Pix. Our hypothesis suggested that a cGAN architecture would provide better results than a more general U-Net model thanks

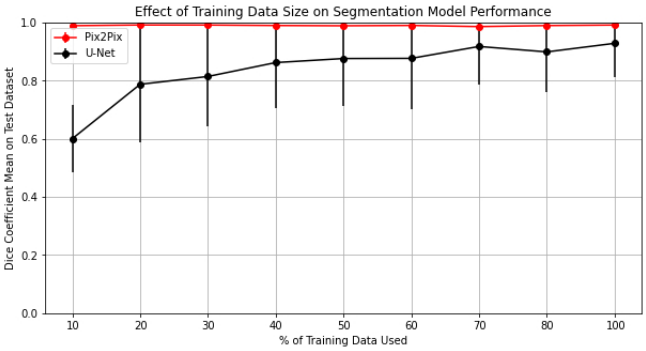

Fig. 2. Dice coefficient across varying training set sizes for both the U-Net and Pix2Pix models.

to the discriminator integration, as demonstrated in previous segmentation tasks [9]. Results from our internal test dataset supported this hypothesis. Notably, residual blocks were not used for the U-Net architecture but for the Pix2Pix Generator, which has been demonstrated to improve the segmentation results [10]. We hypothesize that the use of residual blocks may be useful for reusing low-level features such as corners and edges. This might be beneficial for these types of images with intricate patterns.

In addition, Pix2Pix yielded significantly superior overall results when using a reduced number of training images. Therefore, one of the main advantages of using Pix2Pix for PXCT image segmentation is that it requires less manual annotation of the training images, which can be a time-consuming and error-prone task. Our results show that Pix2Pix can achieve high segmentation quality with only 10% of the training images, which can significantly reduce the workload and cost of preparing the data.

While our study provides valuable insights, future research avenues should explore the robustness of these findings. Due to the dataset's high internal correlation, involving slices from volume images, these results may be too optimistic. Whether these neural network architectures can similarly enhance the segmentation of different PXCT images remains to be investigated. One possible solution that we suggest is to train the models using the 2D images from a single volume and evaluate how well these models generalize to the remaining volume slices. In addition, we propose conducting sensitivity analyses under diverse imaging conditions and assessing model performance in the presence of noise in PXCT data, as the images were already processed, and thus, asses reliability in real-world scenarios.

5. REFERENCES

[1] Salvatore De Angelis, Peter Stanley Jørgensen, Esther Hsiao Rho Tsai, Mirko Holler, Kosova Kreka, and Jacob R Bowen, "Three dimensional characterization of nickel coarsening in solid oxide cells via ex-situ ptychographic nano-tomography," *Journal of Power Sources*, vol. 383, pp. 72–79, 2018.

[2] Martin Dierolf, Andreas Menzel, Pierre Thibault, Philipp Schneider, Cameron M Kewish, Roger Wepf, Oliver Bunk, and Franz Pfeiffer, "Ptychographic x-ray computed tomography at the nanoscale," *Nature*, vol. 467, no. 7314, pp. 436–439, 2010.

[3] Efim V Lavrukhin, Kirill M Gerke, Konstantin A Romanenko, Konstantin N Abrosimov, and Marina V Karsanina, "Assessing the fidelity of neural network-based segmentation of soil xct images based on pore-scale modelling of saturated flow properties," *Soil and Tillage Research*, vol. 209, pp. 104942, 2021.

[4] Sadegh Karimpouli and Pejman Tahmasebi, "Segmentation of digital rock images using deep convolutional autoencoder networks," *Computers & geosciences*, vol. 126, pp. 142–150, 2019.

[5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[6] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, Eds., Cham, 2015, pp. 234–241, Springer International Publishing.

[8] Lee R Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[9] Dan Popescu, Mihaela Deaconu, Loretta Ichim, and Grigore Stamatescu, "Retinal blood vessel segmentation using pix2pix gan," in *2021 29th Mediterranean Conference on Control and Automation (MED)*. IEEE, 2021, pp. 1173–1178.

[10] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, May 2018, arXiv:1711.10684 [cs].