Automating the segmentation of X-ray images with Deep Neural Networks

Alba Castrillo Perote (s230221)[1], Andrea Matamoros Alonso (s233514)[1], Fernando Augusto Marina Urriola (s233144)[1] and Jesús Díaz Pereira (s233142)[1]

1 DTU, Technical University of Denmark

Introduction

In solid oxide fuel cells (SOFC) and electrolysis cells, electrode composition is crucial for optimal performance. The state-of-the-art solution is nickel and yttria-stabilized zirconia cermet, but nickel usage can lead to microstructural changes, degrading cell performance. To address these changes, imaging techniques like ptychographic X-ray computed tomography (PXCT) are used. However, manual or traditional segmentation methods are time-consuming and error-prone. Automating segmentation is crucial for keeping up with data acquisition rates.
This paper proposes using deep neural networks to automate PXCT image segmentation, eliminating the need for human intervention, by comparing two distinct neural network models, more specifically a U-Net and a conditional Generative Adversarial Network (cGAN), Pix2Pix.

Key points

- The training dataset comprises PXCT images and the results will be benchmarked against manually labeled images. For more information about the acquisition process and information on the dataset please contact: **Salvatore De Angelis (sdea@dtu.dk)**
- The **DICE coefficient** was specifically selected as the evaluation metric to compare the architecture performance due to its effectiveness in quantifying the overlap between segmented regions in the predicted and ground truth images.
- U-Net architecture, comprising an encoder and decoder, has demonstrated notable success in various image segmentation tasks. The **encoder** captures hierarchical features, while the **decoder** reconstructs the spatial information. During training, these components work collaboratively, with the encoder extracting essential features and the decoder refining the segmentation [1].
- cGANs consist of two distinct architectures, namely, a **generator and a discriminator**, which are trained simultaneously through adversarial training. The generator creates a synthetic image based on an input image, while the discriminator evaluates it [2].
- We conducted a comparative analysis of the amount of training data required to achieve acceptable results. This is particularly crucial as the training process involves manual segmentation as an initial step.

Materials

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.

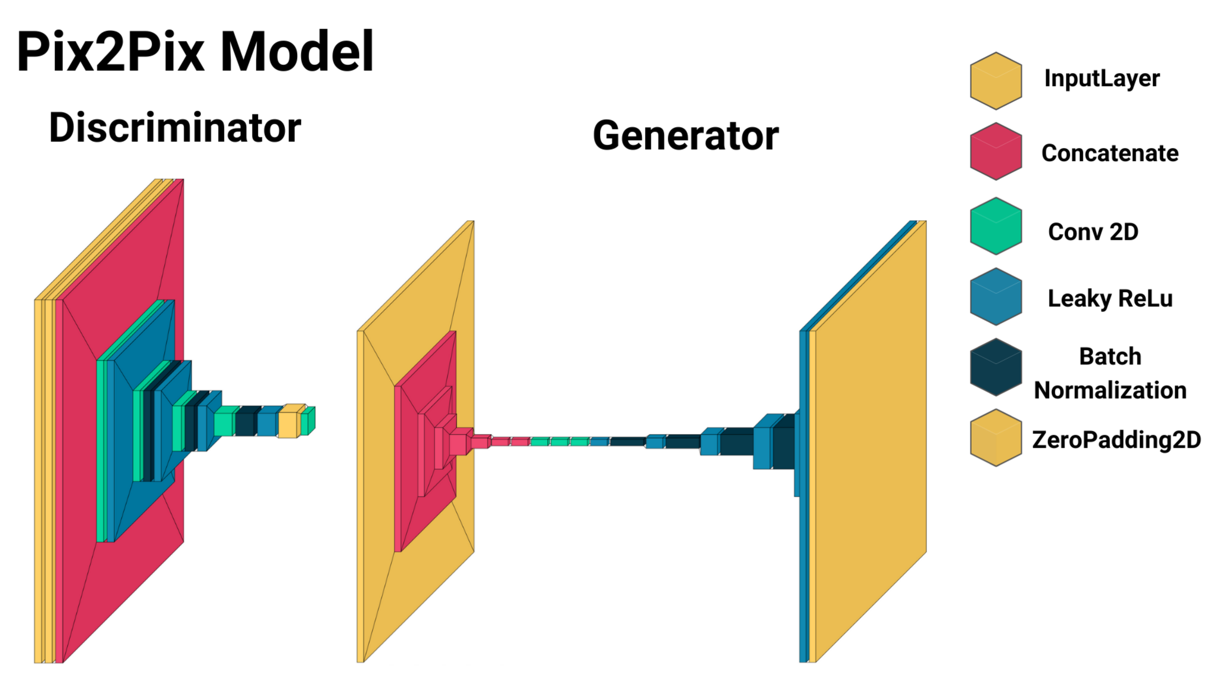

Figure 1: Pix2Pix architecture representation.

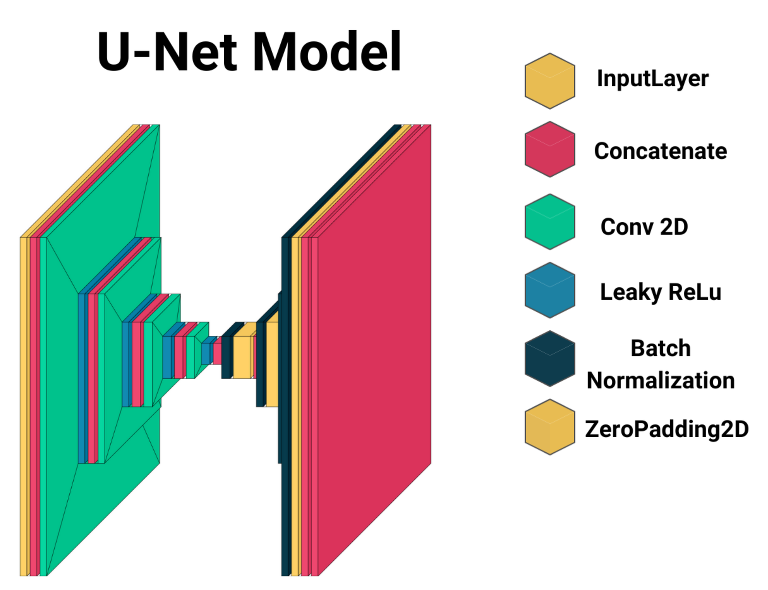

Figure 2: U-Net architecture representation.

Methods

PXCT images were first downsized to a resolution of 256 x 256 pixels due to computational limitations. Both Neural Network models were built and trained in Python 3 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. Lastly, training and validation sets were merged into a single training, and models were trained with different numbers of images and their performance was evaluated using the test set.

	Learning Rate	Batch Size	Epochs	B_1 *Weights*
U-Net	$1x10^{-3}$	32	35	-
Pix2Pix	$2x10^{-3}$	4	35	0.9

Table 1: Hyperparameters of both networks. Adam optimizer was used in the two networks.

Results

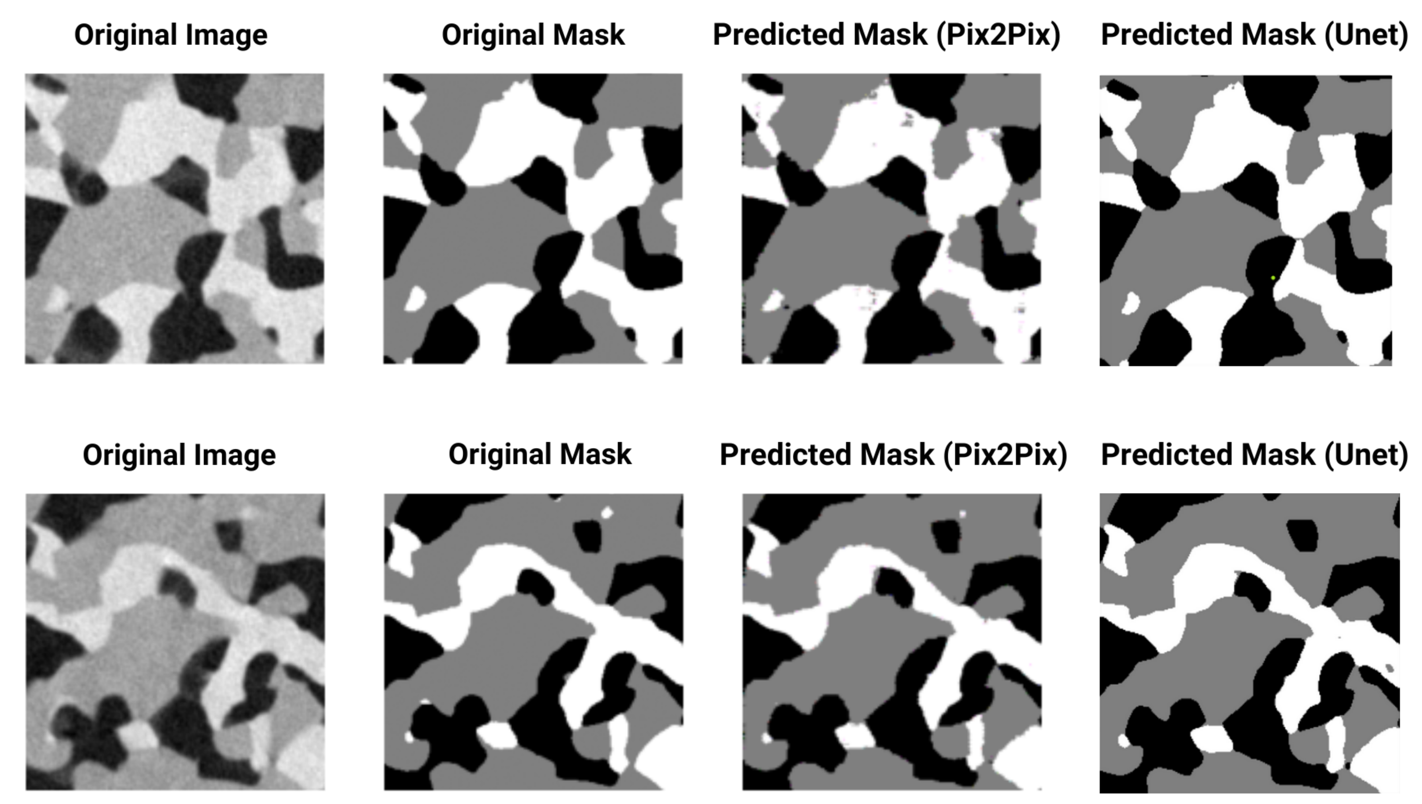

Figure 3: Test of segmentation results of U-Net and Pix2Pix model for PXCT images segmentation, compared with ground truth.

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

Table 2: Internal Test Results for Dice Coefficient

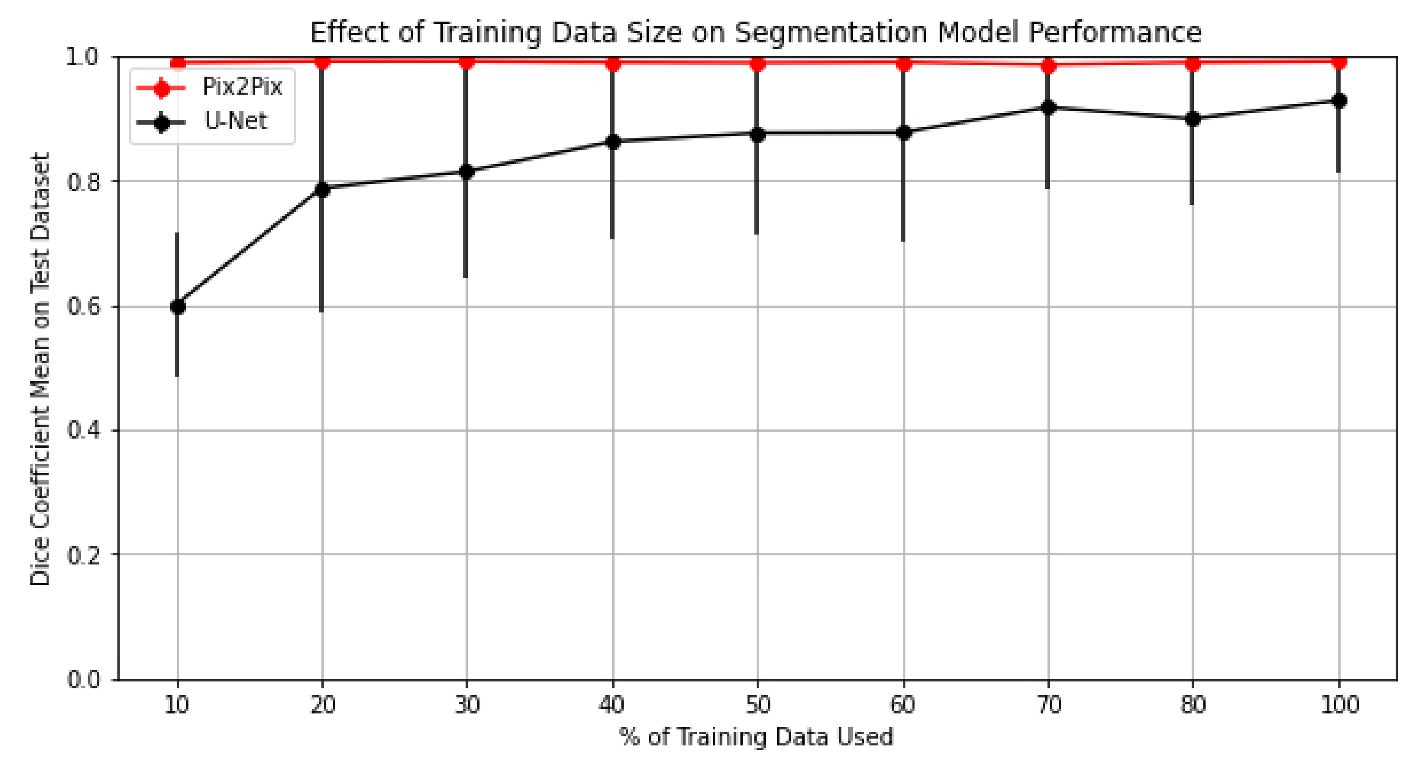

Figure 4: Impact of training data when evaluating U-Net (black) and Pix2Pix (red) for PXCT images segmentation in the test set.

Discussion

This study focused on the comparative analysis for PXCT images 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 to the discriminator integration, as demonstrated in previous segmentation tasks [3]. 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 [4]. In addition, Pix2Pix yielded significantly superior overall results when using a reduced number of training images.
However, the dataset's high internal correlation, involving slices from volume images, may influence these results. Whether these neural network architectures can similarly enhance the segmentation of different PXCT images remains to be investigated.

Conclusions and Future Work

Our project demonstrated that the use of Deep Learning may be beneficial in reducing the manual segmentation workload and time and improving the overall process which possibility of being applied in both industry and research. More specifically, the incorporation of Pix2Pix in combination with residual blocks enhances the model's capability to produce sharper and more contextually accurate segmentations, addressing the limitations of traditional U-Net architectures.
Future research may explore further refinements of the Pix2Pix and U-Net architecture and its applicability to other types of PXCT images, paving the way for continued advancements in both industry and academia.

Acknowledgements

The authors would like to thank S. De Angelis and DTU for providing the data used in this project and his guidance along the whole process.

References

[1] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, et al. U-net: deep learning for cell counting, detection, and morphometry. *Nature methods*, 16(1):67–70, 2019.

[2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. (arXiv:1611.07004), Nov. 2018. URL `http://arxiv.org/abs/1611.07004`. arXiv:1611.07004 [cs].

[3] D. Popescu, M. Deaconu, L. Ichim, and G. Stamatescu. Retinal blood vessel segmentation using pix2pix gan. In *2021 29th Mediterranean Conference on Control and Automation (MED)*, pages 1173–1178. IEEE, 2021.

[4] Z. Zhang, Q. Liu, and Y. Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749â753, May 2018. ISSN 1545-598X, 1558-0571. doi: 10.1109/LGRS.2018.2802944. arXiv:1711.10684 [cs].

s230221@dtu.dk, s233514@dtu.dk, s233144@dtu.dk, s233142@dtu.dk