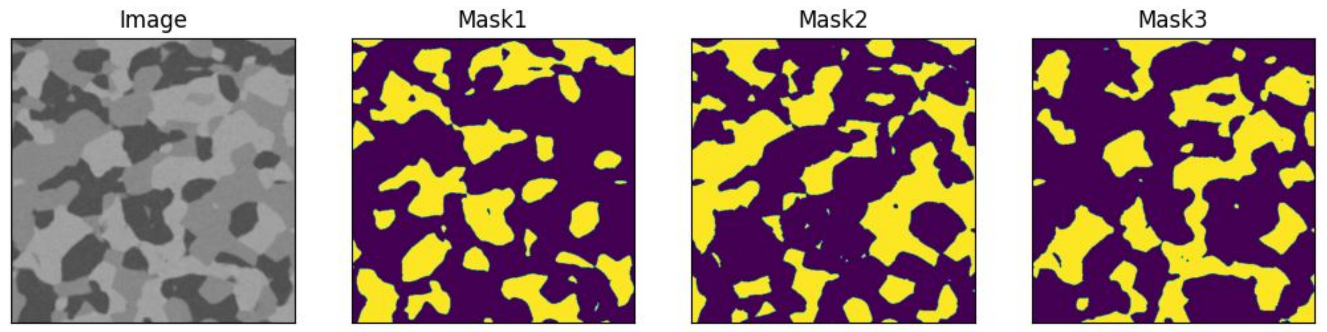
STUDIES ON SEGMENTATION OF X-RAY IMAGES WITH DEEP NEURAL NETWORKS

Metehan Gelgi (s232839), Qingwen Zeng (s232892), Sina Rahimian (s232248), Fatemeh Siar (s236644)
DTU Compute, Technical University of Denmark

INTRODUCTION

- **X-ray imaging** is pivotal in medical and interdisciplinary research.
- Manual segmentation poses challenges in terms of time and accuracy.
- Dataset of **500** images with 501*501 image size
- Deep Neural Networks (DNNs), with a specific focus on **ResNeXt-UNet** and **VGGNet-UNet** architectures.



METHODOLOGY

Data Augmentation

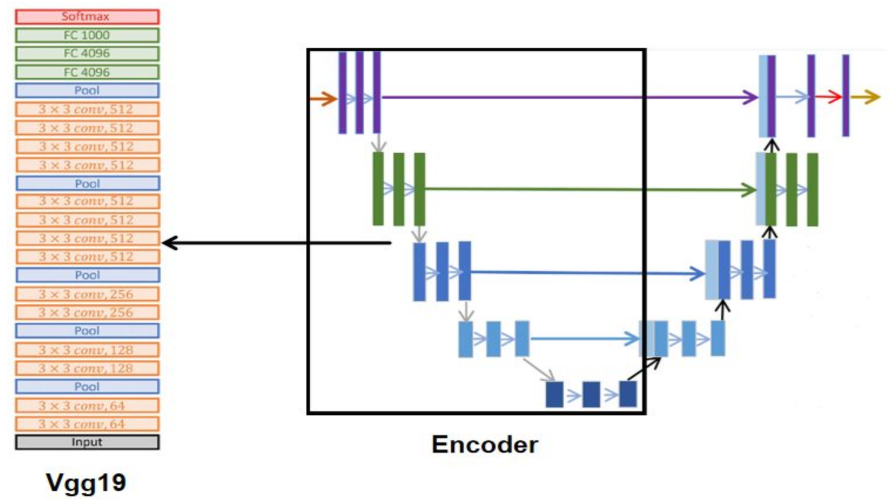
Reasons:

- limited dataset
- similar images
- push model to learn noised data

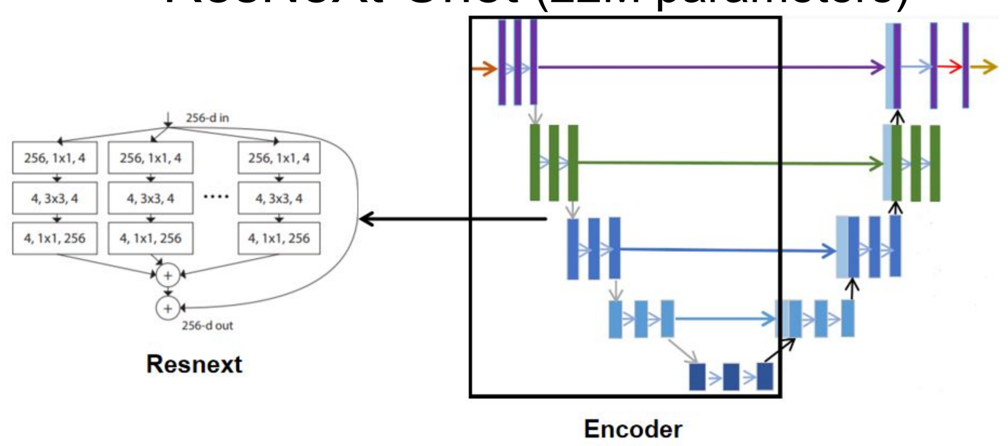
Methods: rotation,shifting, flipping, translation,scaling and **adding Gaussian noise, pixel saturation**

Models

VGG-Unet (20M parameters)



ResNeXt-Unet (22M parameters)



Loss and Metrics

Dice Loss:

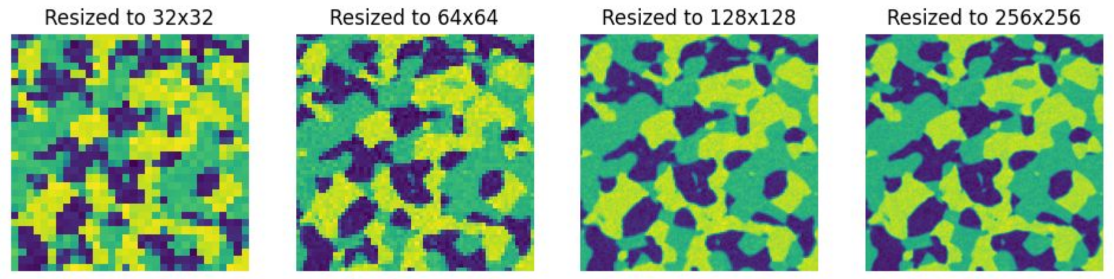
$$L_{dice} = 1 - \frac{\sum_1^N t_i y_i + \varepsilon}{\sum_1^N (t_i + y_i) + \varepsilon}$$

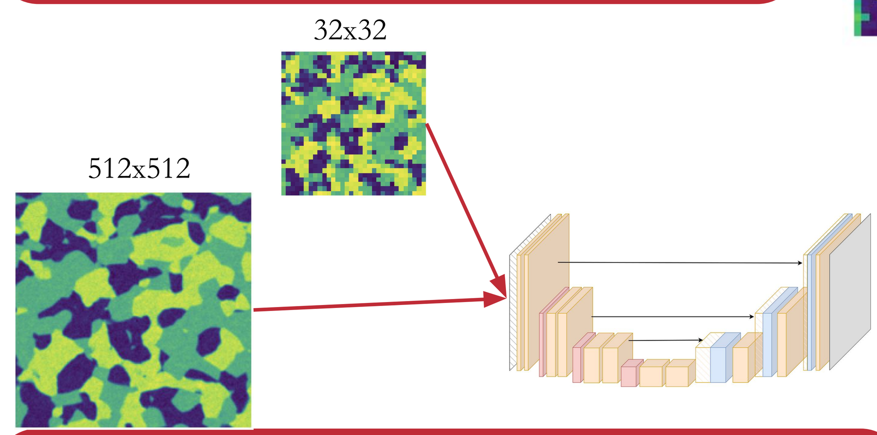
Metrics:

- F1 Score
- Accuracy
- IoU Score

STUDIES

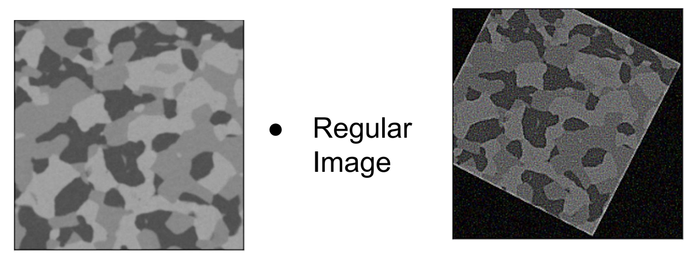
1 - Training Models with Smaller-Sized Images



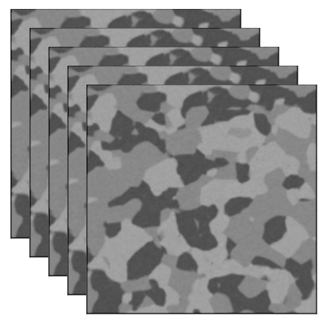


2 - Testing Models with Bigger Images (512x512)

3 - Test on Different Levels of Noised Images



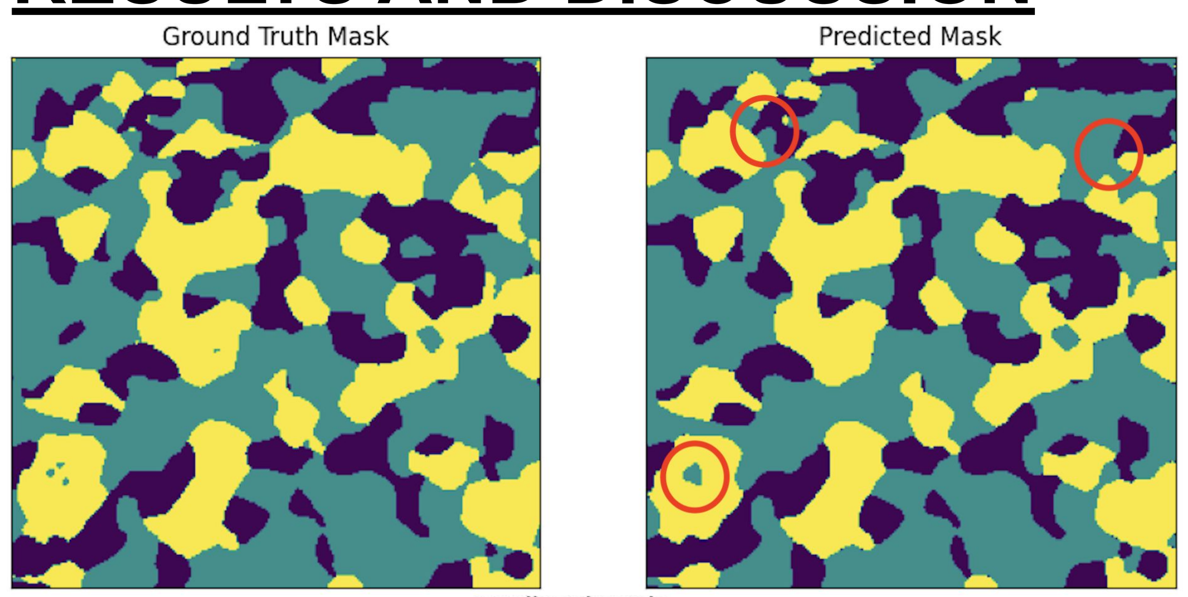
- Gaussian Noise
- Flipped
- Shifted
- Scaled
- Rotated
- Pixel Contrast and Saturation Changed
- and more..

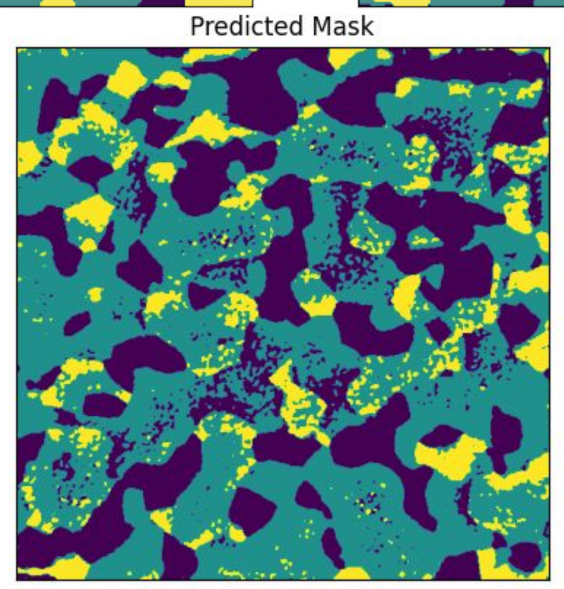


50 / 100 / 300

4 - Training Models with Less Data

RESULTS AND DISCUSSION





*Results are evaluated with Dice Loss, Accuracy, IoU Score and F1 Score. Only Accuracy Scores are included for explanation purposes

Table. Results for Different Image Sizes*

Model	Train Image Size	Training Time (Epoch) in secs	STUDY 1 Same Sized Images with Train (not Augmented)	STUDY 2 Testing with 512x512 Sized Images (not Augmented)	STUDY 3 Testing with 512x512 Sized Images (Augmented)
ResNeXt-UNet	32x32	7	85.42%	70.16%	71.67%
ResNeXt-UNet	64x64	8	91.08%	79.32%	71.01%
ResNeXt-UNet	128x128	8	**95.44%**	**94.19%**	89.49%
ResNeXt-UNet	256x256	11	**98.50%**	**98.72%**	**98%**
VGGNet-UNet	32x32	5	86.69%	79.83%	78.39%
VGGNet-UNet	64x64	5	**94.10%**	90.31%	87.13%
VGGNet-UNet	128x128	6	**97.38%**	**96.85%**	90.73%
VGGNet-UNet	256x256	10	**98.70%**	**98.65%**	**98.36%**

- Overall VGGNet-UNet outperformed ResNeXt-UNet in given scenarios even if it is less complex models. Less complex models should be selected for this kind of tasks.
- Training Image size is one of the main factors affects accuracy, it should be set according to requirements and Resources (time, computational power).
- Transfer learning proved crucial, enabling computationally lightweight models for testing on larger datasets.
- Models should be tested with noised data to evaluate performance of models correctly
- Train Dataset Size study demonstrated that this task can be handed in with less data which makes training faster and gets similar accuracy

Table. **(STUDY 4)** Results for Different Train Sizes*

Model	Train Size	Testing with 512x512 Sized Images (not Augmented)	Testing with 512x512 Sized Images (Augmented)
ResNeXt-UNet	50 (256x256)	90.32%	88.85%
ResNeXt-UNet	100 (256x256)	**98.27%**	**97.34%**
ResNeXt-UNet	300 (256x256)	**98.65%**	**98%**
VGGNet-UNet	50 (256x256)	91.55%	91.64%
VGGNet-UNet	100 (256x256)	**98.67%**	**97.56%**
VGGNet-UNet	300 (256x256)	**98.72%**	**98.36%**

References

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical im- age segmentation," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th In- ternational Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, 2015, pp. 234– 241.
[2] Mohammad Shahjahan Majib, Md Mahbubur Rahman, TM Shahriar Sazzad, Nafiz Imtiaz Khan, and Samrat Ku- mar Dey, "Vgg-scnet: A vgg net-based deep learning framework for brain tumor detection on mri images," IEEE Access, vol. 9, pp. 116942–116952, 2021.
[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[4] Hugo Touvron, Andrea Vedaldi, Matthijs Douze, and Herv´e J´egou, "Fixing the train-test resolution discrep- ancy," CoRR, vol. abs/1906.06423, 2019.

Acknowledgments

This work is done as a part of 02456 Deep Learning, DTU Compute. This project is supervised by Salvatore De Angelis, Peter Stanley Jørgensen and Luke Besley.