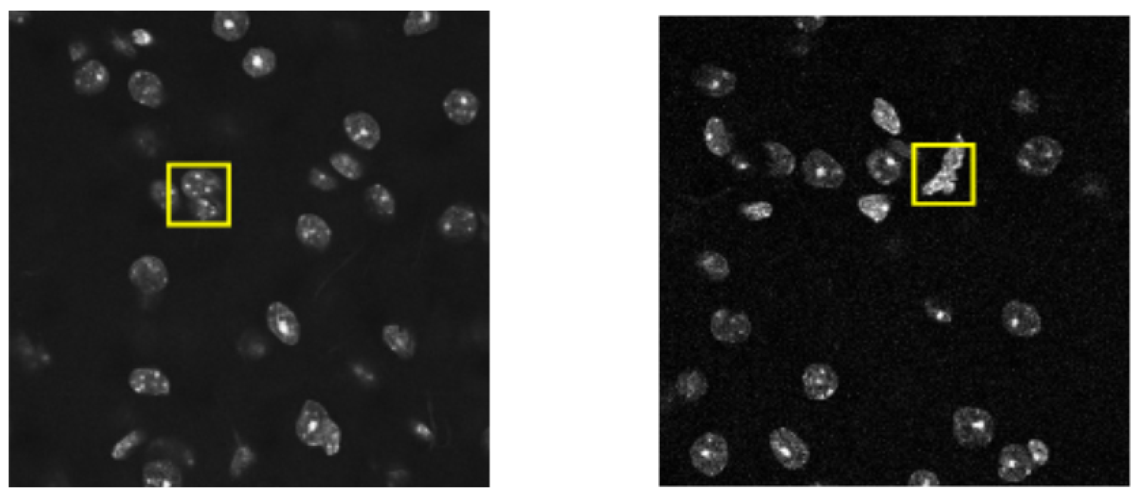
Exploring the Emergent Abilities of SwinIA for Object Segmentation

Dzvenymyra Yarish
Yuliia Siur
Yaroslav Prytula

Institute of Computer Science, University of Tartu

Introduction

In this work, we explore the emergent abilities of the transformer-based model SwinIA [1]. We focus on the semantic segmentation task. We show that the model primarily developed for the denoising task exhibits a promising performance on the image segmentation use case on several cell datasets. We compare the original model with other models, such as Noise2Same on 3 popular cell microscopy datasets.



Example images from FMD Two-photon mice denoising dataset

Methodology

To perform segmentation, we extract multi-dimensional feature maps before the final projection layer.
Then, we use K-Means clustering algorithm for pixel categorization into two groups: object and background. Clustering is performed on extracted feature maps of shape *DxHxW* with *k=2* for binary segmentation. K-Means is insensitive to class labels, so clusters' correspondence to object or background is determined based on scoring in the evaluation phase.

Experiments

1. *Out-of-domain evaluation*

We assess adaptability of Noise2Same and SwinIA models (originally trained for denoising on Two-Photon mice dataset [3]) to unseen data.

2. *Domain specific evaluation*

Models undergo self-supervised training on SevenCellLines dataset. Segmentation capabilities are evaluated on the same dataset, offering insights into performance within a familiar domain.

Results

1. Out-of-domain

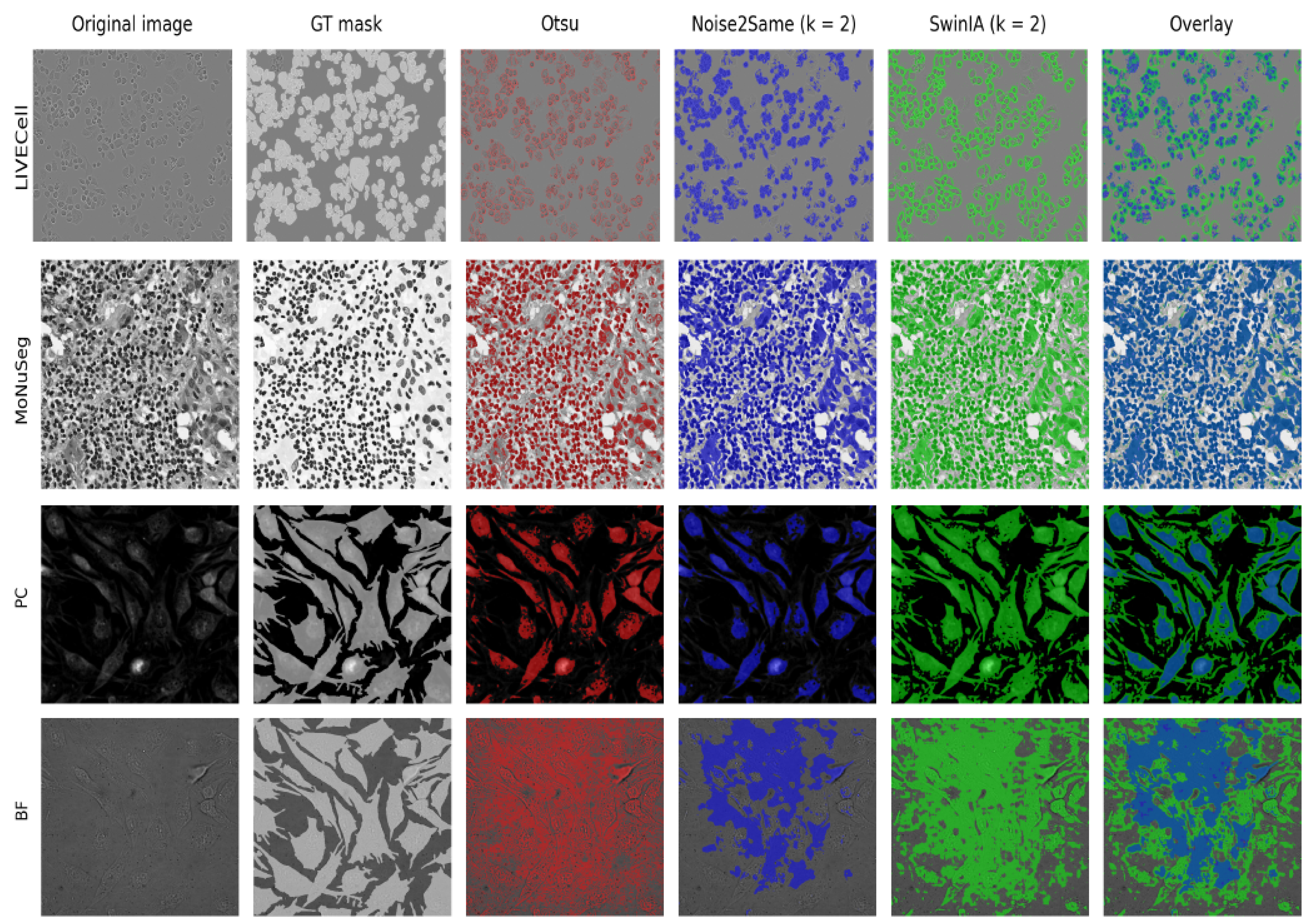
Model	Precision	Recall	F1	IoU
Otsu	**0.887**	0.834	0.487	0.344
Noise2Same	0.39	0.914	0.497	0.361
SwinIA	0.386	**0.987**	**0.512**	**0.382**

Method	Precision	Recall	F1	IoU
Otsu	0.979	**0.578**	**0.713**	**0.568**
Noise2Same	**0.992**	0.443	0.601	0.441
SwinIA	0.988	0.467	0.629	0.464

Dataset	Model	Precision	Recall	F1	IoU
PC	*Otsu*	0.989	0.339	0.503	0.338
	Noise2Same	**0.993**	0.395	0.563	0.394
	SwinIA	0.931	**0.806**	**0.863**	**0.761**
BF	*Otsu*	0.511	0.595	0.544	0.375
	Noise2Same	0.542	0.628	0.566	0.405
	SwinIA	**0.553**	**0.648**	**0.592**	**0.424**

2. Domain specific

Model	Precision	Recall	F1	IoU
Noise2Same	0.99	0.32	0.48	0.319
SwinIA	0.976	0.588	0.733	0.58



Binary masks produced from models' final features on different cell modalities from LIVECell, MoNuSeg, and SevenCellLines

Conclusion

In this work, we evaluated the zero-shot segmentation abilities of SwinIA model. We showed that it exhibits superior performance across 3 datasets in comparison to another popular denoising model, Noise2Same. This confirms that SwinIA learns meaningful features, equally good in global and local contexts. Model's capacity as a universal self-supervised feature extractor and a possible replacement for the existing backbones will be studied in future works.

References

[1] Mikhail Papkov and Pavel Chizhov. Swinia: Self-supervised blind-spot image denoising with zero convolutions. ArXiv, abs/2305.05651, 2023.
[2] Yaochen Xie, Zhengyang Wang, and Shuiwang Ji. Noise2same: Optimizing a self-supervised bound for image denoising. ArXiv, abs/2010.11971, 2020.
[3] Yide Zhang, Yinhao Zhu, Evan L. Nichols, Qingfei Wang, Siyuan Zhang, Cody J. Smith, and Scott S. Howard. A poisson-gaussian denoising dataset with real fluorescence microscopy images. CVPR, 2018 .

1632
UNIVERSITY OF TARTU
Institute of Computer Science