Pre-training Segmentation Models for Histopathology

Payden McBee[1]   PM2KB@VIRGINIA.EDU

[1] *Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA*

Nazanin Moradinasab[1]   NM4WU@VIRGINIA.EDU

Sana Syed[2]   SS8XJ@VIRGINIA.EDU

[2] *Department of Pediatrics, University of Virginia, Charlottesville, VA*

Donald E. Brown[1]   DEB@VIRGINIA.EDU

Abstract

In limited data settings, transfer learning has proven useful in initializing model parameters. In this work, we compare random initialization, pre-training on ImageNet, and pre-training on histopathology datasets for 2 model architectures across 4 segmentation histopathology datasets. We show that pre-training on histopathology datasets does not always significantly improve performance relative to ImageNet pre-trained weights for both model architectures. We conclude that unless larger labeled datasets or semi-supervised techniques are leveraged, ImageNet pre-trained weights should be used in initializing segmentation models for histopathology.

Keywords: Transfer learning, histopathology, segmentation.

1. Introduction and Related Works

Transfer learning is a technique where a model developed for a specific task can be reused as the initial model for a second task with limited labeled data. A common transfer learning approach for medical images is to start with the standard network architectures, e.g., VGG (Simonyan and Zisserman, 2014) and ResNet (He et al., 2016) pre-trained on the large-scale natural images such as ImageNet (Deng et al., 2009) and PASCAL VOC (Everingham et al., 2010), and then fine-tune them on medical images. The effectiveness of pre-trained deep convolutional neural networks (CNNs) with sufficient fine-tuning was investigated on four different medical imaging applications in Tajbakhsh et al. (2016). This study demonstrated that, in most cases, fine-tuning a pre-trained model achieved better performance and robustness in comparison to those trained from scratch. Similarly, Devan et al. (2019) demonstrated that transfer learning with ImageNet can significantly enhance model performance in detecting herpesvirus capsids in microscopy images, particularly when labeled data is limited. Conze et al. (2020) utilized a VGG-11 encoder pre-trained on ImageNet for the shoulder muscle MRI segmentation task. These results indicate that a CNN pre-trained on ImageNet learns features that are applicable to both natural and medical images. However, the gap in features between medical and natural images has motivated pre-training on medical datasets. Ray et al. (Ray et al., 2022) demonstrated an increase in performance and faster convergence for CNNs pre-trained on histopathology datasets relative to a model pre-trained on ImageNet. To the best of the authors' knowledge, the efficacy of utilizing a model pre-trained on natural images compared to a medical image pre-trained models for

nuclei segmentation tasks has not been investigated. Our contribution is to systematically compare segmentation performance using different pre-trained weights across 4 datasets derived from whole-slide-images: eosinophilic esophagitis (EoE) with 2,056 images, Crohn's disease (Crohns) with 800 images, Colorectal Nuclear Segmentation and the Pheno-types (CoNSeP) (Graham et al., 2019) with 660 images, and PanCancer Histology Dataset for Nuclei Instance Segmentation and Classification (PanNuke) (Gamper et al., 2019) with 7,901 images. We seek to answer the following: Does pre-training on histopathology datasets improve segmentation performance relative to encoders pre-trained on ImageNet?

2. Methods

For our analysis, we use HoVer-Net (Graham et al., 2019) and U-Net++ (Zhou et al., 2018) models. HoVer-Net has three separate task-specific decoders, which are used for nuclei detection, separation, and classification, respectively. U-Net++ has a single decoder to provide pixel classification. The Preact-ResNet50 is utilized as an encoder for both models. For HoVer-Net, we follow the exact hyper-parameters and training strategies presented in (Graham et al., 2019). For U-Net++, we train each model for 200 epochs and select the model that minimizes the validation binary cross-entropy (EoE and Crohns) or cross-entropy (CoNSeP and PanNuke) loss. We train and test each of the models for each of the 4 datasets given encoders with various pre-trained weights. We use the MoNuSAC ResNet50 encoder weights from (Graham et al., 2019), and the other weights are obtained by initializing a model with ImageNet and training it on a given histopathology dataset.

3. Experiments and Results

Table 1 shows the average performance of the U-Net++ and HoverNet models over 3 runs across the various pre-trained weights for EoE, Crohns, PanNuke, and CoNSeP. We put the maximum performance for each test set and model across the pre-trained weights in bold and put a star if optimal performance is statistically significant. Notably, the models pre-trained on histopathology and the models pre-trained on ImageNet do not have differences that are statistically significant, except for HoVer-Net pre-trained on MoNuSAC for PanNuke, where $p = 0.052499$ from a Welch's t-test comparing it with the ImageNet pre-trained model performance. This indicates that pre-training on these histopathology datasets does not increase the segmentation performance relative to ImageNet weights. The randomly initialized weights are lower for all datasets except U-Net++ for EoE, suggesting that some kind of pre-training is useful. Furthermore, the number of epochs trained when using a model initialized with ImageNet weights is comparable to models pre-trained on histopathology, being significantly larger only for U-Net++ on PanNuke. Thus, there is no set of consistently optimal pre-trained weights, and the ImageNet weights provide the same or better performance than weights from a model pre-trained on multiple histopathology datasets. Also, the time for training for models with ImageNet pre-trained encoder is comparable to models pre-trained with histopathology.

Table 1: Model Performance

Model	Pre-Trained Weights	Crohns Dice	Epochs	EoE Dice	Epochs
		Test Dataset			
U-Net++	Random	0.55 ± 0.033	84	**0.62** ± 0.009	83
U-Net++	ImageNet	**0.572** ± 0.006	31	0.615 ± 0.02	60
U-Net++	MoNuSAC	0.554 ± 0.009	109	0.612 ± 0.015	103
U-Net++	CoNSeP	0.565 ± 0.019	30	0.618 ± 0.018	63
U-Net++	PanNuke	0.554 ± 0.031	11	0.599 ± 0.001	65
U-Net++	EoE	0.557 ± 0.026	21	-	-
U-Net++	Crohns	-	-	0.606 ± 0.016	89
HoVer-Net	Random	0.389 ± 0.192	74	0.572 ± 0.007	80
HoVer-Net	ImageNet	**0.609** ± 0.012	90	0.621 ± 0.004	93
HoVer-Net	MoNuSAC	0.6 ± 0.011	83	**0.624** ± 0.002	97
		PanNuke		**CoNSeP**	
U-Net++	Random	0.552 ± 0.014	73	0.65 ± 0.011	129
U-Net++	ImageNet	0.571 ± 0.013	127	0.669 ± 0.008	82
U-Net++	MoNuSAC	0.54 ± 0.012	69	0.667 ± 0.009	109
U-Net++	CoNSeP	0.57 ± 0.02	50	-	-
U-Net++	PanNuke	-	-	**0.678** ± 0.012	62
U-Net++	EoE	**0.593** ± 0.033	114	0.656 ± 0.039	83
U-Net++	Crohns	0.577 ± 0.017	92	0.664 ± 0.022	81
HoVer-Net	Random	0.555 ± 0.007	78	0.403 ± 0.048	67
HoVer-Net	ImageNet	0.585 ± 0.003	94	0.67 ± 0.007	97
HoVer-Net	MoNuSAC	**0.602*** ± 0.008	98	**0.679** ± 0.016	83
HoVer-Net	CoNSeP	0.589 ± 0.003	95	-	-
HoVer-Net	PanNuke	-	-	0.644 ± 0.05	78

4. Conclusion

We showed that training a model with ImageNet pre-trained weights did not have significantly different performance than pre-training on multiple histopathology datasets for 2 state-of-the-art medical segmentation models, the U-Net++ and HoVer-Net. This is likely due in part to the relatively small size of the datasets used in pre-training. Small datasets do not allow the model to learn a diversity of features, even when they come from the target domain. Furthermore, the number of training epochs to minimize the validation loss did not increase for the models pre-trained with ImageNet relative to those trained on histopathology. We conclude that, unless an abundant amount of histopathology data is available, pre-training on relatively small histopathology datasets is not likely to increase performance or decrease training time relative to an ImageNet baseline.

Acknowledgments

References

Pierre-Henri Conze, Sylvain Brochard, Valérie Burdin, Frances T Sheehan, and Christelle Pons. Healthy versus pathological learning transferability in shoulder muscle mri segmentation using deep convolutional encoder-decoders. *Computerized Medical Imaging and Graphics*, 83:101733, 2020.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

K Shaga Devan, Paul Walther, Jens von Einem, Timo Ropinski, Hans A Kestler, and Clarissa Read. Detection of herpesvirus capsids in transmission electron microscopy images using transfer learning. *Histochemistry and cell biology*, 151:101–114, 2019.

Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338, 2010.

Jevgenij Gamper, Navid Alemi Koohbanani, Ksenija Benes, Ali Khuram, and Nasir Rajpoot. Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification. In *European Congress on Digital Pathology*, pages 11–19. Springer, 2019.

Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58:101563, 2019.

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*, pages 770–778, 2016.

Indranil Ray, Geetank Raipuria, and Nitin Singhal. Rethinking imagenet pre-training for computational histopathology. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 3059–3062, 2022. doi: 10.1109/EMBC48229.2022.9871687.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312, 2016.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation, 2018. URL https://arxiv.org/abs/1807.10165.