PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Generative modeling of histology tissue reduces human annotation effort for segmentation model development

Brendon Lutnick, Nicholas Lucarelli, Pinaki Sarder

Brendon Lutnick, Nicholas Lucarelli, Pinaki Sarder, "Generative modeling of histology tissue reduces human annotation effort for segmentation model development," Proc. SPIE 12471, Medical Imaging 2023: Digital and Computational Pathology, 124711Q (6 April 2023); doi: 10.1117/12.2655282

SPIE.

Event: SPIE Medical Imaging, 2023, San Diego, California, United States

Generative Modeling of Histology Tissue Reduces Human Annotation Effort for Segmentation Model Development

Brendon Lutnick[a], Nicholas Lucarelli[b], Pinaki Sarder*[c]

[a]Department of Pathology and Anatomical Sciences, University at Buffalo – The State University of New York, Buffalo, NY, USA; [b]Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA; [c]Division of Nephrology, Hypertension, and Renal Transplantation, Department of Medicine, University of Florida, Gainesville, FL, USA.

*Address correspondence to:
Pinaki Sarder, PhD; Pinaki.sarder@medicine.ufl.edu; (352) 273-6018

ABSTRACT

Segmentation of histology tissue whole side images is an important step for tissue analysis. Given enough annotated training data, modern neural networks are capable of accurate reproducible segmentation; however, the annotation of training datasets is time consuming. Techniques such as human-in-the-loop annotation attempt to reduce this annotation burden, but still require vast initial annotation. Semi-supervised learning—a technique which leverages both labeled and unlabeled data to learn features—has shown promise for easing the burden of annotation. Towards this goal, we employ a recently published semi-supervised method, datasetGAN, for the segmentation of glomeruli from renal biopsy images. We compare the performance of models trained using datasetGAN and traditional annotation and show that datasetGAN significantly reduces the amount of annotation required to develop a highly performing segmentation model. We also explore the usefulness of datasetGAN for transfer learning and find that this method greatly enhances the performance when a limited number of whole slide images are used for training.

Keywords: Generative Adversarial Network, semi-supervision, DatasetGAN, histology, segmentation

1. INTRODUCTION

The segmentation of histological tissue structures from whole slide images (WSIs) is often an important first step for further downstream analysis and is therefore well explored in the literature.[1-6] The use of convolutional neural networks (CNNs) is currently considered the state of the art; however, the generation of annotated training sets for segmentation of WSIs is time consuming and requires domain-level expertise.[7] Semi-supervised learning has been successfully used to reduce the amount of annotation needed to generate performant models by utilizing information from unlabeled training images. While the adoption of semi-supervised learning to medical image datasets has been slower than in the field of computer science, this approach is ideal for histology data, where unlabeled data is plentiful.

Recently, several approaches for utilizing generative adversarial network (GAN) architectures for semi-supervised segmentation training have been proposed.[8,9] One promising method is Nvidia's recently published DatasetGAN,[9] which proposes training a simple decoder network to predict semantic labels for synthetic images produced by the generator. Using this method, the state-of-the-art GAN architecture StyleGAN[10] is first trained without labels, to produce realistic synthetic images. Through this unsupervised training, StyleGAN learns a low dimensional representation of the training dataset in the latent code which is mapped to synthetic images by the generator. A decoder network is then trained to produce an accurate pixel-wise labeling of the synthetic images given a handful of labeled images. This allows the GAN to function as a labeled data factory, capable of producing large and heterogenous labeled datasets which can be used to train a supervised architecture such as DeepLabV3+.[11] DatasetGAN has shown compelling results for segmentation of natural images and is ideally suited for application to WSIs.

To test the feasibility of using DatasetGAN in the development of models for segmentation of WSIs, we employed it to train a model for glomeruli segmentation from renal tissue biopsies. We tested the effectiveness of this method against

Medical Imaging 2023: Digital and Computational Pathology, edited by John E. Tomaszewski,
Aaron D. Ward, Proc. of SPIE Vol. 12471, 124711Q · © 2023 SPIE
1605-7422 · doi: 10.1117/12.2655282

traditional annotation and training using Histo-cloud,[4] our recently developed pipeline for cloud-based WSI segmentation. Models created with DatasetGAN require very minimal annotation effort, making them an ideal way to initialize network parameters for transfer learning. We also tested the ability of the model trained with DatasetGAN annotations to improve further supervised training.

2. METHODS

2.1 Baseline StyleGAN Training

The first step to using DatasetGAN to produce synthetic segmented images is training the base GAN model used to generate the images. For this training, StyleGAN was trained using 2455 renal biopsy WSIs scanned at 40x magnification. This dataset came from various institutions and included periodic acid schiff (PAS), hematoxylin and eosin (H&E), and trichrome staining. To efficiently accomplish this training using such a large dataset, we used our Histo-fetch pipeline[12] to randomly select patches from the WSIs on the fly during training. StyleGAN was trained using 2 Nvidia Quadro P5000 RTX GPUs for 17,560,000 steps using a learning rate of 0.003 for the final resolution level 1024x1024.

2.2 Manual Image Labeling

To train DatasetGAN to produce fake images, a limited number of synthetic images needed to be labeled by hand. To do this, the previously trained StyleGAN was used to produce a large number of synthetic renal tissue patches. Twenty-four of these patches containing glomeruli were selected and the glomeruli locations were annotated. To train DatasetGAN using the hardware available to us, we reduced the resolution of the StyleGAN output to 256x256 pixels. This allowed us to train the DatasetGAN encoder to produce fake annotations corresponding to the generated images. To generate the final training set used to train the DeepLabV3+ segmentation network,[11] we used our DatasetGAN model to generate a training set of 50,000 image patches and annotations.

2.3 DatasetGAN Training

To train the model using the 50,000 synthetic patches produced by DatasetGAN, we used the official implementation of DeepLab and encoded the patches into tfrecord files for training. This training was done using 2 Nvidia Quadro P5000 RTX GPUs with a batch size of 12, and polynomial learning rate of starting at 0.007, for 30,000 steps. The training used a segmentation model pretrained on the ImageNet dataset[13] for transfer learning.

To test how this strategy compares to training using a hand annotated dataset, five Control models were trained using various amounts of hand annotated training data. The first Control model was trained using 4 annotated WSIs, we refer to this model as Control-4. The second Control model added 5 additional WSIs to the previous training set and is referred to as Control-9.

The training of the Control-4 and Control-9 models was performed using Histo-cloud.[4] Training was done using the same hyperparameters as the DatasetGAN model training, and also used the pretrained ImageNet model for transfer learning parameter initialization. The patches used for training were 512x512 and were extracted from the WSIs at a variety of downsampled scales including 1X, 2X, 3X, and 4X downsampled with respect to the native slide resolution. The details of the Histo-Cloud method of on-the-the-fly patch selection are further described in our previous work.[4]

The Transfer-learning model was trained by repeating the training using the Control-4 dataset, but instead of initializing the network parameters using the ImageNet pretrained model,[11] parameters from the DatasetGAN trained model were used.

The dataset generated by DatasetGAN included 50,000 synthetic image patches that were 256x256 pixels at an effective 10X down-sampled resolution with respect to the 40X WSIs used for the other training. The Control-4 dataset contained four WSIs containing 76 annotated glomeruli. This dataset contained 2 PAS-stained and 2 trichrome-stained slides with manually annotated glomeruli. The Control-9 dataset added 5 additional WSIs to the control set including 3 trichrome and 2 H&E-stained slides. All the slides in the control datasets were from a single institution. The holdout dataset contained 23 slides, with 10 PAS, 8 trichrome, and 5 H&E stained WSIs. The majority of these slides were selected from the same institution as the control training sets, but one from each stain originated from a separate institution. In total, the holdout dataset contained 486 glomeruli with hand-annotated boundaries as ground truth.

3. RESULTS

The performance of the resulting segmentation models was compared using a holdout dataset containing 23 WSIs. The DatasetGAN model was trained with a synthetic dataset of 50,000 image patches generated using DatasetGAN. Examples of synthetic data generated by DatasetGAN are shown in **Figure 1**. While the unsupervised training of the GAN was done using 2455 WSIs, only 24 synthetic glomeruli were hand annotated. Examples of the annotated synthetic glomeruli are shown in Figure 1. When tested on the holdout dataset, the DatasetGAN model produced a Mathews correlation coefficient (MCC) of 0.68.

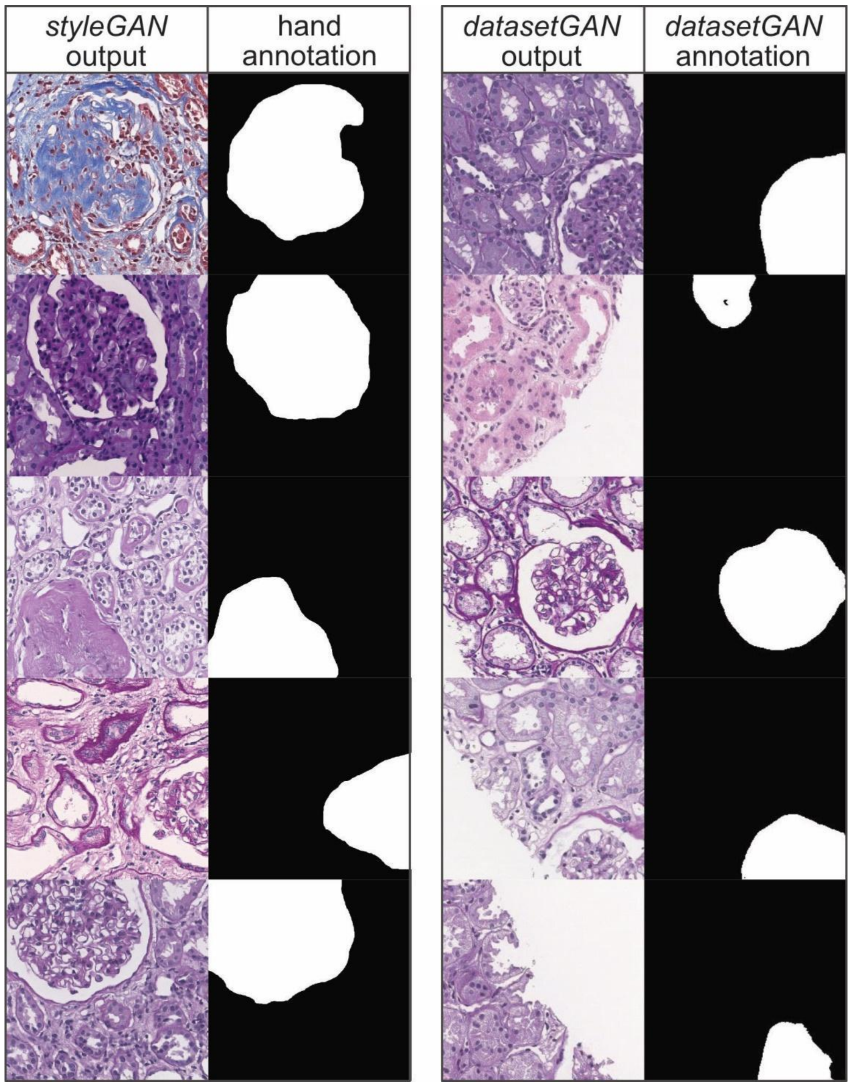

Figure 1. Examples of DatasetGAN Inputs and Outputs. The panels on the left are examples of the hand annotations used to train DatasetGAN. Twenty-four selected StyleGAN outputs were annotated for glomeruli by hand. The panels on the right are examples of labeled images produced by DatasetGAN. The DatasetGAN outputs are generated by the pre-trained StyleGAN branch, and semantic annotations are automatically produced by the decoder branch of DatasetGAN. Due to memory constraints when training DatasetGAN, the outputs are produced at 1/4th the size of the originally trained StyleGAN (256x256 pixels instead of 1024x1024).

The control Control-4 and Control-9 models both performed worse than the DatasetGAN model, resulting in an MCC of 0.42 and 0.61 despite receiving more than 3X and 6X the annotation effort, respectively. To see if this performance would improve with the further addition of training data, we also trained three additional Control models using 20, 41, and 83 training WSIs. All three of these models outperformed the DatasetGAN model with MCC values of 0.87, 0.88, and 0.91 respectively. These data are tabulated in **Table 1**.

Finally, to test the effectiveness of transfer learning using DatasetGAN we retrained the DatasetGAN model using the Control-4 dataset. This process significantly improved the performance of the DatasetGAN model resulting in an MCC of 0.78.

Table 1. Holdout segmentation performance using different training sets.

Network	**Annotated Gloms**	**Training Slides**	**MCC**
Control-4	76	4	0.42
Control-9	161	9	0.61
Control-20	343	20	0.87
Control-41	672	41	0.88
Control-83	1296	83	0.91
DatasetGAN	24	2455	0.68
Transfer-learning *(DatasetGAN + Control)*	100	2455+4	0.78

4. CONCLUSIONS

Our tests show that the performance of a model trained using DatasetGAN synthetic images outperforms models trained with a limited number of traditionally annotated histological structures. In our experiments, we found that DatasetGAN easily outperforms traditional training even when more than six times the amount of annotation effort is used to produce the training set.

While the desired segmentation performance may not be achievable using DatasetGAN annotations alone, our results show that reusing the model parameter's learned training with the synthetic data drastically improved performance when compared to hand annotation alone.

REFERENCES

[1] Bueno, G., et al., Glomerulosclerosis identification in whole slide images using semantic segmentation. Computer Methods and Programs in Biomedicine, 2020. 184: p. 105273.

[2] de Bel, T., et al. Stain-transforming cycle-consistent generative adversarial networks for improved segmentation of renal histopathology. in Proceedings of the 2nd International Conference on Medical Imaging with Deep Learning; Proceedings of Machine Learning Research. 2019.

[3] Ginley, B., et al., Computational segmentation and classification of diabetic glomerulosclerosis. 2019. 30(10): p. 1953-1967.

[4] Lutnick, B., et al., A user-friendly tool for cloud-based whole slide image segmentation, with examples from renal histopathology. bioRxiv, 2021: p. 2021.08.16.456524.

[5] Van Eycke, Y.-R., A. Foucart, and C. Decaestecker, Strategies to reduce the expert supervision required for deep learning-based segmentation of histopathological images. Frontiers in Medicine, 2019. 6: p. 222.

[6] Lutnick, B., et al., An integrated iterative annotation technique for easing neural network training in medical image analysis. 2019. 1(2): p. 112.

[7] Szolovits, P., R.S. Patil, and W.B. Schwartz, Artificial intelligence in medical diagnosis. Annals of internal medicine, 1988. 108(1): p. 80-87.

[8] Mittal, S., M. Tatarchenko, and T. Brox, Semi-supervised semantic segmentation with high-and low-level consistency. IEEE transactions on pattern analysis and machine intelligence, 2019.

[9] Zhang, Y., et al. DatasetGAN: Efficient labeled data factory with minimal human effort. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[10] Karras, T., S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

[11] Chen, L.-C., et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. in Proceedings of the European conference on computer vision (ECCV). 2018.

[12] Lutnick, B., et al., Histo-fetch--On-the-fly processing of gigapixel whole slide images simplifies and speeds neural network training. arXiv preprint arXiv:2102.11433, 2021.

[13] Deng, J., et al. Imagenet: A large-scale hierarchical image database. in 2009 IEEE conference on computer vision and pattern recognition. 2009. Ieee.