Self Attention Network for Medical Image Segmentation

Naved Siddiqui

University at Buffalo
navedsid@buffalo.edu

Abstract

In this paper, Self-Attention Generative Adversarial Network (SAGAN) is implemented which allows attention-driven, long range dependency modeling for segmentation task of Whole Slide Image (WSI). In SAGAN, details can be generated using cues from all feature locations. The proposed SAGAN performs better than prior work[1] , the F1 score increases from 0.93 to 0.97 on the Mouse kidney samples with four classes. Moreover, we apply spectral normalization to the network for the regularization term and find that this improves training dynamics.

I. Introduction

Deep learning-based approaches became popular also in medical image domain due to the increasing computational power and availability of data. However, these methods still lack in robustness across different datasets and require a significant amount of annotated data. These limitations are even more substantial in the medical domain (relative to the natural domain) because the annotated data is highly heterogeneous and its size is relatively small. Image segmentation in medical images is a very challenging task due to large class imbalance problem. The minority class that is of primary concern which needs to be segmented has significantly lesser number of pixels in an image as compared to the majority class i.e. the background. It is of high importance that the minority class is accurately classified as well as majority class should not be classified as the minority class.

II. Related Work

GANs have achieved great success in various image generation tasks, including image-to-image translation (Isola et al.[2], 2017; Taigman et al.[3], 2016; Liu & Tuzel[4], 2016; Xue et al.[5], 2018; Park et al.[6], 2019), image super-resolution (Ledig et al.[7], 2017; Snderby et al.[8], 2017) and text-to-image synthesis (Reed et al.[9], 2016; Zhang et al.[10], 2016; Hong et al.[11], 2018). Despite this success, the training of GANs is known to be unstable and sensitive to the choices of hyperparameters. Several works have attempted to stabilize the GAN training dynamics and improve the sample diversity by designing new network architectures (Radford et al.[12], 2015; Zhang et al.[13], 2017; Karras et al.[14], 2018), modifying the learning objectives and dynamics (Arjovsky et al.[15], 2017; Salimans et al.[16], 2018; Metz et al.[17], 2017; Che et al.[18], 2016; Zhao et al.[19], 2016; Jolicoeur-Martineau[20], 2018), adding regularization methods (Gulrajani et al.[21], 2017; Miyato et al.[22], 2018) and introducing heuristic tricks (Salimans et al.[23], 2016; Odena et al.[24], 2016; Azadi et al.[25], 2018). Recently, Miyato et al. (Miyato et al.[26], 2018) proposed limiting the spectral norm of the weight matrices in the discriminator in order to constrain the Lip-

schitz constant of the discriminator function. Combined with the projection-based discriminator (Miyato & Koyama[26], 2018), the spectrally normalized model greatly improves class-conditional image generation on ImageNet.
Recently, attention mechanisms have become an integral part of models that must capture global dependencies (Bahdanau et al.[27], 2014; Xu et al.[28], 2015; Yang et al.[29], 2015; Gregor et al.[30], 2015; Chen et al.[31], 2017). In particular, self-attention (Cheng et al.[32], 2016; Parikh et al.[33], 2016), also called intra-attention, calculates the response at a position in a sequence by attending to all positions within the same sequence. Vaswani et al. (Vaswani et al.[34], 2017) demonstrated that machine translation models could achieve state-of-the-art results by solely using a self-attention model. Parmar et al. (Parmar et al.[35], 2018) proposed an Image Transformer model to add self-attention into an autoregressive model for image generation. Wang et al. (Wang et al.[36], 2017) formalized self-attention as a non-local operation to model the spatial-temporal dependencies in video sequences. Han Zhang et al. (Han Zhang et al.[37], 2018) proposed a SAGAN, which introduces a self-attention mechanism into convolutional GANs. However, this model is used primarily for image classification. Inspite of this progress, Self attention networks for image segmentation still has not been explored.

III. SAGAN for Image Segmentation

The discriminator of the model remains the same as of the original SAGAN paper. The main difference is in the generator model. The generator takes an image as an input and returns the segmented image as the output. The generator uses U-net architecture with ResNet layers. The input image and the output image is of the same size. The discriminator is used to predict if the segmented image given to it is a real image or an image from the generator. The github link for the code is provided here.

IV. Results

In this project the dataset used for image segmentation is the podocyte / non-podocyte dataset which are mouse WSI images from Brendon et al.[1]. This is a four class problem namely - background, glomeruli, podocyte and non-podocyte. Also, in each of the images in this dataset there is a huge imbalance in the classes of podocytes and non-podocytes as compared to the background. For such cases, F1 score gives a good estimate how well the network performs. In Brendon et al.[1], the F1 score achieved was 0.93. When using a SAGAN network for the same task, the model achieves a F1 score 0.9762. Some other evaluation metrics as well as per class F1 score is listed in the table below.

Evaluation Metric	Score
F1 score	0.9762
Accuracy	0.9925
IOU	0.954
Precision	0.9788
Recall	0.9736

It is important for us to have better values for the minority classes i.e. podocytes and non-podocytes. The below table shows the per class values for F1 score.

Evaluation Metric	F1 score
Podocyte	0.9607
Non-Podocyte	0.9574
Glomeruli	0.9896
Background	0.9969

The below table shows the per class values for IOU.

Evaluation Metric	IOU
Podocyte	0.9244
Non-Podocyte	0.9182
Glomeruli	0.9794
Background	0.9939

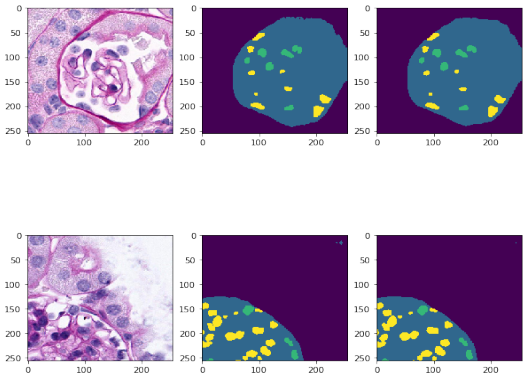
The below table shows the per class values for precision.

Evaluation Metric	Precision
Podocyte	0.97
Non-Podocyte	0.958
Glomeruli	0.9917
Background	0.9955

The below table shows the per class values for recall.

Evaluation Metric	Recall
Podocyte	0.9516
Non-Podocyte	0.9568
Glomeruli	0.9876
Background	0.9984

Some of the results from the model is as follows:



References

[1] Brendon Lutnick, Brandon Ginley, Darshana Govind, Sean D. McGarry, Peter S. LaViolette, Rabi Yacoub, Sanjay Jain, John E. Tomaszewski, Kuang-Yu Jen, and Pinaki Sarder. An integrated iterative annotation technique for easing neural network training in medical image analysis. *Nature Machine Intelligence*, 1(2):112âĂŞ119, Feb 2019.

[2] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks, 2016.

[3] Yaniv Taigman, Adam Polyak, and Lior Wolf. Unsupervised cross-domain image generation, 2016.

[4] Ming-Yu Liu and Oncel Tuzel. Coupled generative adversarial networks, 2016.

[5] Yuan Xue, Tao Xu, Han Zhang, L. Rodney Long, and Xiaolei Huang. Segan: Adversarial network with multi-scale l1 loss for medical image segmentation. *Neuroinformatics*, 16(3-4):383âĂŞ392, May 2018.

[6] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization, 2019.

[7] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network, 2016.

[8] Casper Kaae SÃ¿nderby, Jose Caballero, Lucas Theis, Wenzhe Shi, and Ferenc HuszÃąr. Amortised map inference for image super-resolution, 2016.

[9] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. Generative adversarial text to image synthesis, 2016.

[10] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks, 2016.

[11] Seunghoon Hong, Dingdong Yang, Jongwook Choi, and Honglak Lee. Inferring semantic layout for hierarchical text-to-image synthesis, 2018.

[12] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks, 2015.

[13] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris Metaxas. Stackgan++: Realistic image synthesis with stacked generative adversarial networks, 2017.

[14] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks, 2018.

[15] Martin Arjovsky, Soumith Chintala, and LÃľon Bottou. Wasserstein gan, 2017.

[16] Tim Salimans, Han Zhang, Alec Radford, and Dimitris N. Metaxas. Improving gans using optimal transport. *ArXiv*, abs/1803.05573, 2018.

[17] Luke Metz, Ben Poole, David Pfau, and Jascha Sohl-Dickstein. Unrolled generative adversarial networks. *ArXiv*, abs/1611.02163, 2016.

[18] Tong Che, Yanran Li, Athul Paul Jacob, Yoshua Bengio, and Wenjie Li. Mode regularized generative adversarial networks. *ArXiv*, abs/1612.02136, 2016.

[19] Junbo Jake Zhao, Michaël Mathieu, and Yann LeCun. Energy-based generative adversarial network. *ArXiv*, abs/1609.03126, 2016.

[20] Alexia Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard gan. *ArXiv*, abs/1807.00734, 2018.

[21] Ishaan Gulrajani, Faruk Ahmed, Martín Arjovsky, Vincent Dumoulin, and Aaron C. Courville. Improved training of wasserstein gans. In *NIPS*, 2017.

[22] Takeru Miyato and Masanori Koyama. cgans with projection discriminator. *ArXiv*, abs/1802.05637, 2018.

[23] Tim Salimans, Ian J. Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *ArXiv*, abs/1606.03498, 2016.

[24] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *ICML*, 2016.

[25] Samaneh Azadi, Catherine Olsson, Trevor Darrell, Ian J. Goodfellow, and Augustus Odena. Discriminator rejection sampling. *ArXiv*, abs/1810.06758, 2018.

[26] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. *ArXiv*, abs/1802.05957, 2018.

[27] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2014.

[28] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1316–1324, 2017.

[29] Zichao Yang, Xiaodong He, Jianfeng Gao, Li Deng, and Alexander J. Smola. Stacked attention networks for image question answering. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21–29, 2015.

[30] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, and Daan Wierstra. Draw: A recurrent neural network for image generation. *ArXiv*, abs/1502.04623, 2015.

[31] Xi Chen, Nikhil Mishra, Mostafa Rohaninejad, and Pieter Abbeel. Pixelsnail: An improved autoregressive generative model. *ArXiv*, abs/1712.09763, 2017.

[32] Jianpeng Cheng, Li Dong, and Mirella Lapata. Long short-term memory-networks for machine reading. In *EMNLP*, 2016.

[33] Ankur P. Parikh, Oscar Täckström, Dipanjan Das, and Jakob Uszkoreit. A decomposable attention model for natural language inference. In *EMNLP*, 2016.

[34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS*, 2017.

[35] Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Noam Shazeer, Alexander Ku, and Dustin Tran. Image transformer. In *ICML*, 2018.

[36] Xiaolong Wang, Ross B. Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7794–7803, 2017.

[37] Han Zhang, Ian J. Goodfellow, Dimitris N. Metaxas, and Augustus Odena. Self-attention generative adversarial networks. *ArXiv*, abs/1805.08318, 2018.