Semantic Segmentation of Medical Image Datasets

Chinmaya Lad
Department of Computer Science
University at Buffalo
50288442
clad@buffalo.edu

Abstract

Semantic Segmentation has a wide variety of applications in Deep Learning. However, performing segmentation on medical images is a very tricky field. Unlike other applications that have an abundance of training data, the medical domain consists of a limited amount of properly labeled datasets. For multi-class problems, the samples are highly imbalanced, which makes it difficult to train efficient models. We propose using adversarial loss alongside the segmentation network. Using adversarial loss increases the accuracy and F1 score, providing a better-segmented image.

I. Introduction

Medical imaging is rapidly advancing with the invention of numerous imaging instruments like X-rays, CAT scans, Ultrasounds. These imaging technologies help the physician to view the structural anatomy of a particular part of the body to make a diagnosis and suggest proper treatment. Imaging is mainly used for identifying the affected region.

Improvements in Computer Vision and Deep Learning approach has lead to remarkable applications in object classification and image segmentation. Many models emerged that performed the task of segmentation and classification very efficiently. The use of adversarial loss for training a network showed a new way to training more efficient models with less supervised data. These approaches are of extensive use in Medical Images to help the physician detect an anomaly.

Using Deep Learning models in the Medical domain has proven useful and tons of research is being done. The main problem for model training in this domain is the scarcity of properly labeled data. Even if medical imaging has advanced, labeling of data requires immense domain expertise and time. Moreover, imbalance in labels in this domain seriously impedes the efficiency of the model. The important thing we need to remember in the medical domain is that mistakes can be costly, and training a model with limited data is a challenging task.

We propose to use adversarial loss during the training of the segmentation model. Our approach based on the paper [1] which uses DeepLab V2 [2] as a segmentation network and performs Human-in-the-Loop training. We train the DeepLab V3 [3] alongside adversarial loss provided by suing Multiscale L1 Loss as proposed in [4].

II. Related Work

A. An integrated iterative annotation technique for easing neural network training in medical image analysis

Semantic segmentation in medical image dataset is a difficult task. This paper provides us a technique of training segementation network efficiently. The paper suggest to use human-in-the-loop training method in which after each training iteration the images are sent to a human expert to correct them. Using these corrected annotation the network is retrained. This process is repeated until we get a proper functioning model.

B. DeepLab

DeepLab V3 is a semantic segmentation network developed by google. It is a state-of-the-art segmentation network that performs feature detection at multiple scales. DeepLab V3 has been proven to be a very efficient segmentation model and has obtained best results on *PASCAL VOC* training dataset. The network employs Atrous Convolutions with variable dilation rates and Spatial Pyramid Pooling which have been proven to improve efficiency of the model .

DeepLab V3 uses Atrous Convolution in place for Deconvolutions in order to recover spatial resolution. Atrous Convolution is significantly faster and efficient than Deconvolutions. Deeplab use ResNet blocks for feature extraction. Atrous convolution with $rate > 1$ is applied after block3 when output $stride = 16$.

Atrous Spatial Pyramid Pooling proposes four parallel atrous convolutions with different atrous rate are applied on top of feature map, which is effective to re-sample features at different scales for accurately and efficiently classifying regions [3]. ASPP effectively caputres multi-scale information. Global Average Pooling is then applied for context information and the features are then bilinearly upsampled to desired spatial dimension.

C. Multiscale L1 Loss

Multiscale L1 Loss is mentioned in paper [4] which proposed using a segmentor critic network to train for semantic segmentation of medical images. The critic is different from vanilla GAN discriminator which has a scalar output at the end. The critic provides a multiscale L1 loss function that forces segmentor and critic both to learn global and local

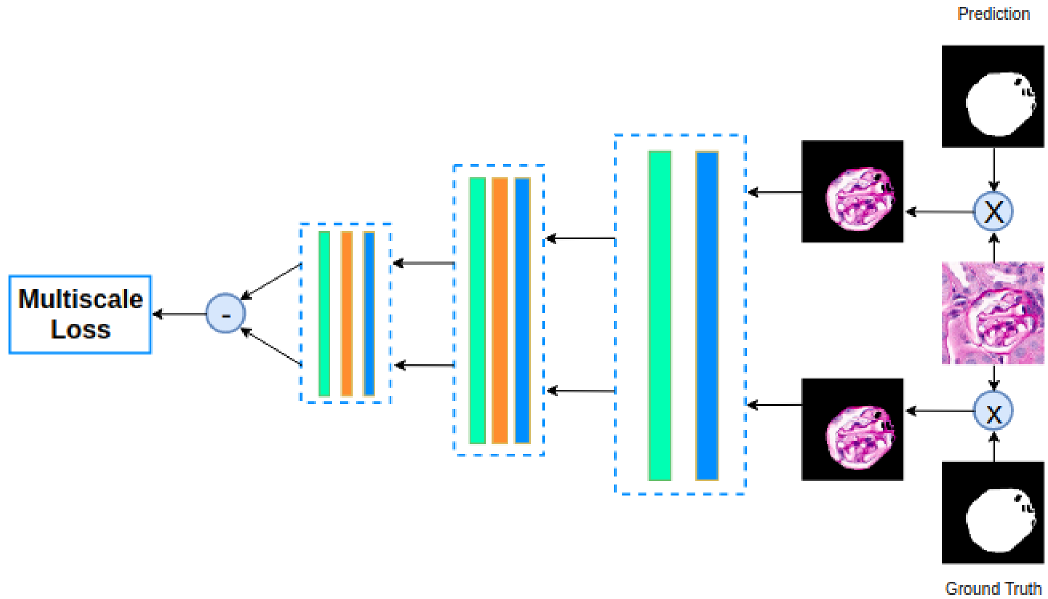

Fig. 1. Critic Architecture

features that capture long-range and short-range spatial relationships between pixels.

III. Background

In this paper we propose use of state-of-the-art DeepLabV3 network alongside critic with Mulitscale L1 Loss for training on Podocyte dataset. DeepLab V3 uses atrous spatial pyramid pooling which allow the network to classify images at multiple scales. The DeepLab is trained with Cross Entropy between the Ground Truth and Output of the network. Our main intuition is that training the DeepLab network along side an adversarial loss, will help the segmentor network learn the features faster and better. We train the segmentor on weighted sum of Cross Entropy Loss and Multiscale L1 Loss as shown equation 1. The critic is a deep convolutional network that takes in Ground Truth and Predicted result and outputs contatenation feature maps of all the layers. Moreover, rather than passing only the Ground Truth and predicted images we perform masking i.e. we mask the input image with the the predicted label and also mask the input image with ground truth label as show in Fig.2 and then calculate the Multiscale L1 Loss (l_{mae}) between them.

$$L = \lambda_{ce} l_{CE} + \lambda_{mae} l_{mae} \tag{1}$$

$$l_{mae}(x, x') = \frac{1}{L} \sum_{i=1}^{L} \left\| f_c^i(x) - f_c^i(x') \right\| \tag{2}$$

IV. Experiments

We perform the training on Podocyte Dataset provided in [1]. We trained the model on 890, 256x256 podocyte images and 100 test images. The ground truth has 3 labels that denotes glomureli, podocyte and non-podocyte. The model is trained for 25K iterations.

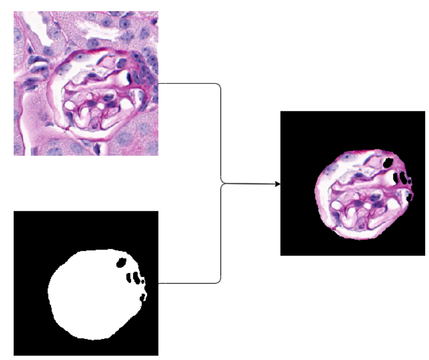

Fig. 2. Masking Input image with Label map

The segmentor network outputs a segmented image for dimension HxWxC where each channel represents the label pixel map. We train two different versions of model. The first one consist of one critic that take in all channels of segmented image and second one consisting of four critics each per label. In this way each critic learns different channels.

The images in the dataset are largely imbalanced with almost 60% of pixels being labeled 0 as background and 25% being labeled 1 as glomureli. Due to this large imbalance we use a technique proposed in [5] that scales the error for each class separated and then performs the backpropagation. We scale the error of an label with the marginal probability of that label in the entire dataset.

V. Evaluation Metric

We use F1 score and Intersection over Unions (IoUs) as our evaluation metrics. The F1 score is calculated without using

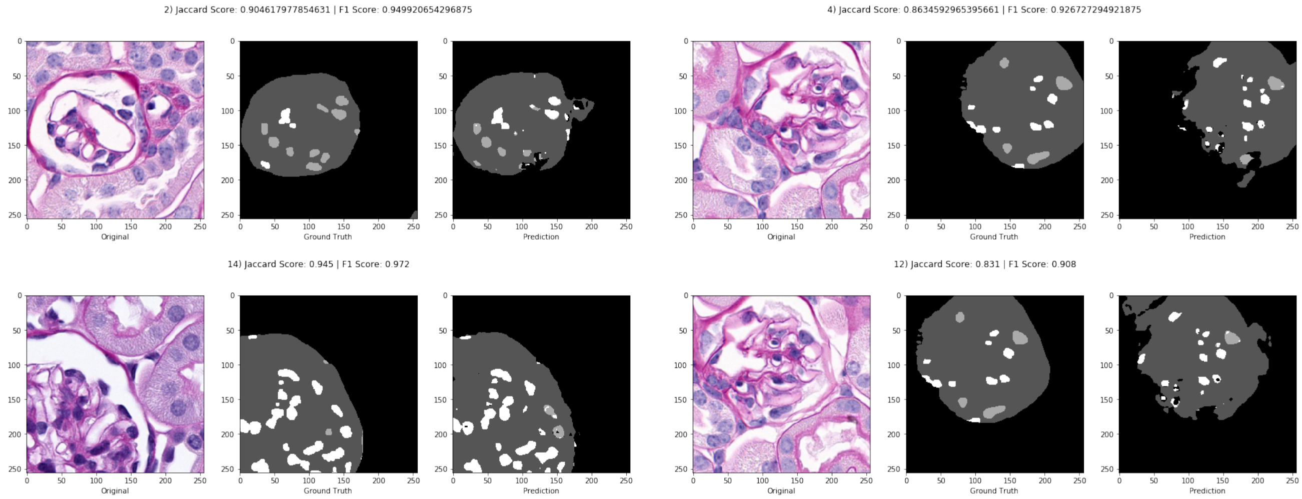

Fig. 3. Segmentation results of our model

class 0 background as it is 66% of the total labeled pixels.

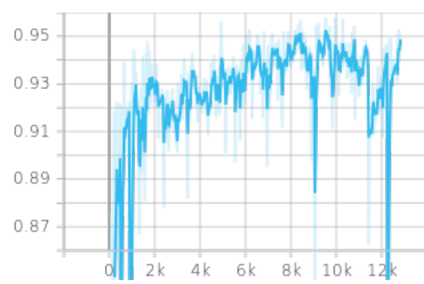

Fig. 4. F1 Score during training iterations

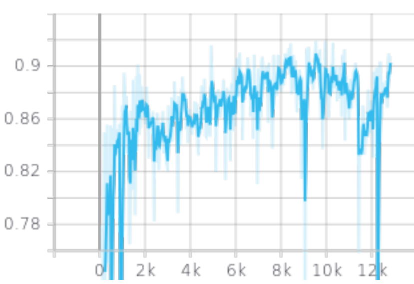

Fig. 5. Intersection over Union during training iterations

Apart from overall F1 score and IoUs we also calculate the metrics separately for each class to understand how model performs for different classes as you can see in Table I. As we can see that the model peforms well for class 0,1,3 but has considerable low metric for class 2.

VI. Results

We train the model with weighted sum of cross entropy loss and multiscale L1 loss. As shown in Figure 3 the label maps are colored in terms of grayscale with each shade denoting a class label. The background and glomureli of the predicted labels are visibly similar to the ground truth labels. We can observe that the model detects the nuclei perfectly however it struggles in classifying whether the nuclei belongs to podocyte or non-podocyte.

Label	Type	F1 Score	IoUs
0	Background	0.955	0.916
1	Glomureli	0.741	0.687
2	Podocyte	0.150	0.094
3	Non Podocyte	0.517	0.408

TABLE I
F1 Score and IoU per Labels.

The model with multiple critics does not perform well on this dataset and does not produce the result. The model trained with class weights for handling class imbalance problem decreases the F1 Score and is considerable inefficient than the original model. You can find some of the results of failed cases of the model in Figure 6. In some cases the model fails to detect proper nuclei and in some cases it fails to detect the glomureli.

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):112–119, Feb 2019.

[2] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *CoRR*, abs/1606.00915, 2016.

[3] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017.

[4] Yuan Xue, Tao Xu, Han Zhang, L. Rodney Long, and Xiaolei Huang. Segan: Adversarial network with multi-scale l_1 loss for medical image segmentation. *CoRR*, abs/1706.01805, 2017.

[5] Robin Chan, Matthias Rottmann, Fabian Hüger, Peter Schlicht, and Hanno Gottschalk. Application of decision rules for handling class imbalance in semantic segmentation. *CoRR*, abs/1901.08394, 2019.

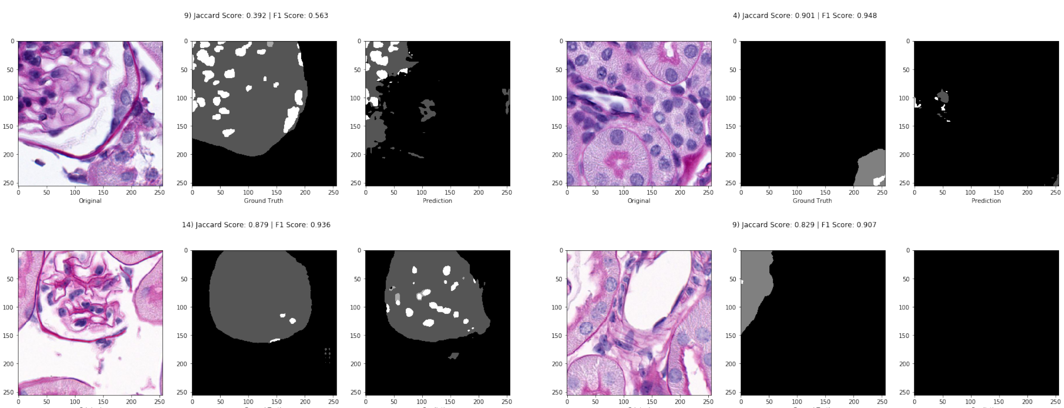

Fig. 6. Failed cases of segmentation model