Semi-supervised Segmentation of Brain MRI Images

CS229 Project Proposal (Life Sciences)

Ali Mottaghi
Department of Electrical Engineering
Stanford University
mottaghi@stanford.edu

Serena Yeung
Department of Biomedical Data Science
Stanford University
syyeung@stanford.edu

Abstract

While progress in neural network architectures has led to significant recent progress in the task of semantic segmentation, the challenge of obtaining large amounts of labeled segmentation masks has limited broad use in real-world applications such as medical image analysis. This has led to an emerging body of work focused on semi-supervised segmentation where a larger quantity of unlabeled data can be used with a smaller quantity of labeled data to train segmentation models. Recent work on semi-supervised classification has demonstrated that when simple techniques such as consistency regularization are used effectively, the gains in performance can be dramatic. In this work, we explore effective use of consistency regularization for semi-supervised segmentation, and show that when we use a consistency loss for segmentation in conjunction with a discriminator selecting informative labeled images, the resulting model significantly outperforms the prior body of work on semi-supervised semantic segmentation across multiple standard benchmarks. The code for our implementation is available at https://github.com/samottaghi/brain-segmentation.

1 Introduction

Semantic segmentation has a wide range of real-world applications such as in medical diagnosis [1] [2] [3] and in autonomous vehicles [4]. In recent years deep learning approaches have become the mainstream in this area and they have shown significant improvement in the performance in this task [5] [6] [7]. However, one of the remaining important challenges for most of the current approaches is that they need a large densely annotated dataset of images and segmentation masks in order to perform well. This challenge may limit the real-world applications that these methods are applicable to. For example for the applications in medical diagnosis, only hospitals that can afford to collect a large dataset with the corresponding annotation can utilize these approaches to aid their physician for a more accurate diagnosis.

Semi-supervised learning can mitigate the problem of a small annotated dataset by incorporating a large unlabeled dataset. Since the unlabeled data is usually much easier to acquire, this approach can open significantly improve the performance of our model whiteout increasing the annotation burden. Many algorithms have been suggested in this area including adding a consistency regularization on the prediction of the model on unlabeled data [8] [9], and pseudo labeling in which the predictions of the model on some unlabeled images are used as ground-truth for further training the model [10] [11]. Recently, [12] proposed FixMatch which achieves state-of-the-art performance on the image classification task. This method utilizes most of the successful approaches in previous works in a simple yet efficient algorithm.

In this work, we propose SegMatch, a new semi-supervised learning model for semantic segmentation task. In our approach, we use a discriminator for selecting informative vs non-informative unlabeled images. A pseudo mask will be generated for each informative example, and the consistency regularization is enforced on different transformations on the unlabeled images. Our approach is agnostic to the segmentation model and proposed a new simple but effective way of utilizing unlabeled images in the training loop of any segmentation model which can significantly improve its performance when the number of labeled examples is very limited. In our experiments, we show that our model outperforms the previous state-of-the-art methods in semi-supervised semantic segmentation on ... and ... datasets.

We can summarise our main contributions in this works as:

1. Proposing a new model agnostic framework for semi-supervised semantic segmentation
2. Achieving the state-of-the-art performance in semantic segmentation task on ... and ... datasets by using a discriminator for selecting informative unlabeled examples.
3. Analysing the variants of our model, and studying how previous approaches can be reproduced in our framework.

The rest of the paper is organized as follows. In section 2, we review the current approaches for semi-supervised learning on semantic segmentation. We will propose our model and the algorithm for training it in section 3. Our experiments and ablations studies show the performance of our model compared to the previous model in section 4. Finally, we conclude in section 5.

2 Related Works

Semi-supervised Classification: Most of the early works in semi-supervised learning has focused on classification problem [13]. Two main ideas are usually used in this problem: 1) Consistency loss: which enforce the model's predictions on unlabeled data to be consistent under different transformation. [8] [9] 2) Pseudo labeling: which uses the most certain predictions of the model as a ground truth [10] [11]. Recently, [12] proposed FixMatch which set a new state-of-the-art in this area using these two techniques. In FixMatch, two transformations for each unlabeled image is generated using week-augmentation and strong-augmentations. If the model's prediction for weekly-augmented images is above some threshold τ a pseudo-label is generated and will be used as ground truth for the strongly-augmented version of the unlabeled images. The idea of using two transformations of unlabeled image and enforcing the consistency on predictions has also been applied in SimCLR [cite] which archives the state-of-the-art in self-supervised learning. Our work utilizes some of the ideas and techniques that are introduced in this area for the semantic segmentation problem.

We categorize the current approaches in semi-supervised segmentation based on the main idea that is used in them:

Consistency Regularization: Consistency relies on the fact that unlabeled data has the same distribution as the labeled data, therefore we expect the trained model to have a consistent and confident prediction on them as well as on labeled data. It has been applied in many approaches in semi-supervised segmentation such as [14] [15] [16]. In these works, different versions of an unlabeled image under different transformations are feed into the model and the consistency loss is enforced on the predicted masks of the model.

Attention Maps: Some of the work in semi-supervised segmentation relay on extracting segmentation masks from attention or saliency maps of the network such as [17] [18]. These approaches focus on the case where image-level labels are given for images without a ground truth mask. These labels, therefore, are used in training a classifier, and the parts of an image that the model attends to when predicting each label are used as segmentation masks. However, in this work, we assume that we do not have any form of supervision on unlabeled data, therefore a classification network cannot be trained on them.

Adversarial Training: The generative model has been widely used on different tasks in semi-supervised learning [19] [20]. The main idea in these works is to expect the model to have the same output distribution on both labeled and unlabeled examples. This approach has also been used in semi-supervised segmentation in [21] [22] [23], where given a segmentation mask the discriminator network decides whether it is a ground-truth mask or a prediction of the segmentation network. Here, we use the discriminator as a decision-maker on whether an unlabeled image is informative for the model or not as is will be described in the next section.

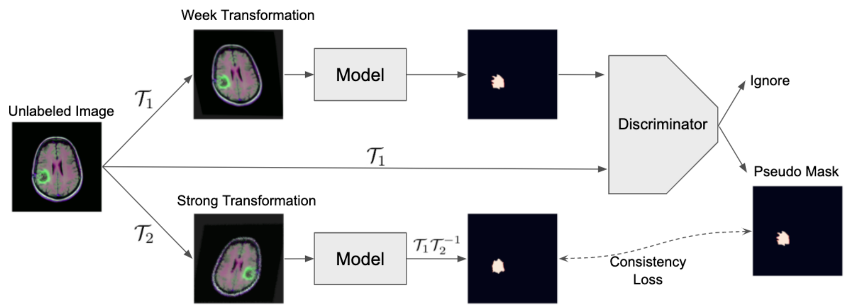

Figure 1: SegMatch. Our model produces a week transformation $\mathcal{T}_1$ and a strong transformation $\mathcal{T}_2$ of the image. Based on the predicted mask pf the week transformation and the transformed original images, discriminator decide about the informativeness of the unlabeled image. It only generates a pseudo mask if it found the mask and the image good matches and informative for the training of the model. Finally, the consistency loss is applied between the pseudo mask and the aligned prediction of the model on the strong transformation of the image.

3 Method

We propose a simple framework that utilizes both consistency regularization and pseudo labeling for the semantic segmentation. The main distinction of our model compared to the previous works in the literature is introducing a new pseudo labeling method that decides about the informativeness of each unlabeled image given both the image and the predicted mask, and only generates pseudo labels for the informative examples for further training the model. We also, combine our pseudo labeling scheme and consistency regularization method such that the consistency loss is only enforced on informative examples. In this way, our method is not trained on very noisy hard pseudo labels. As a result it reduces the variance in training and improves the overall performance.

3.1 Problem Definition

Semantic segmentation is the problem of classifying each pixel in an image to one of the classes. We assume that we are given a labeled dataset $\mathcal{D}_L = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N_L}$ and an unlabeled dataset $\mathcal{D}_U = \{(x^{(i)})\}_{i=1}^{N_U}$. Here $x^{(i)} \in \mathbb{R}^d$ is an image and $y^{(i)} \in \{1, 2, \ldots, k\}^d$ is the corresponding mask, where each pixel in the image should be classified as on of the k classes in the mask. The task is to minimize the cross-entropy loss between the ground truth and predicted mask, using both labeled and unlabeled images. In this paper, we use 'label' and 'mask' interchangeably in the context of semantic segmentation problem.

3.2 SegMatch

Our model, SegMatch, is inspired by FixMatch model where week and strong transformations of unlabeled images are produced and the consistency loss in enforced between the predictions when the model is certain about its predictions (The max probability is above some threshold τ). However, in our method instead of relying on softmax probabilities outputted by the model as a measure of the confidence and informativeness of unlabeled images, we use a discriminator. This discriminator takes the predicted masks and the original images and generates a pseudo mask if it is informative for the model.

Our method is shown in Figure 1. As it is shown in this figure, for each unlabeled image we produce a week transformation $\mathcal{T}_1$ and a strong transformation $\mathcal{T}_2$ of the image and feed them to our model M. Given the prediction of the model on weekly transformed image and the and the transformed image itself, the discriminator D generates the pseudo masks for informative ones. Finally, the cross-entropy loss is applied to the generated pseudo masks and the aligned prediction of the model on the strongly-transformed version. As a result, the consistency loss is only applied when discriminator decides an unlabeled image is informative.

3.2.1 Transformations

We use two forms of transformation in our algorithm.

Week Transformations ($\mathcal{T}_1$): Week transformations are used to augment the data without generating out-of-distribution data. In our experiments, we use random horizontal flip and random horizontal and vertical translation (shift) as week transformations.

Strong Transformation ($\mathcal{T}_2$): Strong transformations are intended to enforce the consistency between model's prediction. We expect our model to have a consistent prediction over the transformed versions of the input. This acts as a source of supervision for the model on unlabeled data. In our experiments, we use two augmentation strategies for strong transformation: 1) RandAugment [24] which randomly selects transformations for each batch of data from a set of predefined augmentations. 2) CTAugment [25] which learns the magnitude of each transformation based on the predictions of the model during the training. CTAugment slowly increases the magnitude of the transformations based on how close the model's prediction is to the true label, as a result, it does not need an additional validation set for tuning. The complete list of transformations and their associated parameters that are used in our experiments are listed in the appendix.

3.3 Discriminator

We add a discriminator to our model in order to guide the consistency loss. More specifically, we want to only enforce the consistency loss on the unlabeled images that are informative for the model to be trained on. In fact, some predictions of the model on unlabeled images are noisy on which training the model may hurt the final performance. Therefore by adding a discriminator, only informative unlabeled images are utilized in the training loop. So pseudo labels are not generated for the images on which the model does not perform robustly and confidently.

The discriminator takes not only the prediction but also the corresponding image to decide whether the prediction is a good fit for the image or not. We expect the discriminator to be class-agnostic. This means that given the images and the predicted mask, discriminator can determine if the mask is a good fit for the images without knowing its class. This is an ability that human beings can easily do based on the prior information on the shape of the objects. So they can tell if a mask is a good fit for an image even if they don't know the objects.

We use a Convolutional Neural Network (CNN) for our discriminator module and we train it on both labeled and unlabeled data for understanding whether a mask is a good fit for the images or not. We train $\widetilde{D}$ with the following loss function that classifies each image-mask as informative vs non-informative:

$$\mathcal{L}_{\widetilde{D}} = \sum_{\substack{(x_l, y_l) \in \mathcal{D}_L \\ (x_u) \in \mathcal{D}_U}} \log(1 - D(x_l, y_l)) + \log(D(x_u, M(x_u))) \quad (1)$$

Then the output of the discriminator will be $D(X, Y) = \hat{Y}$ only if $\widetilde{D}(X, Y) > \tau$, where $\hat{Y} = argmax(Y)$ is the pseudo mask and τ is a hyper-parameter for threshold.

We then discuss other simpler choices for the discriminator in our model:

Identity: In this case, our approach will be equivalent to only applying Consistency Loss, where the model is enforced to make a consistent prediction on transformations of unlabeled data. Therefore, we have $D(x, y) = y$

Mean Confidence: In this scenario, our approach will be similar to segmentation's version of FixMatch, where the model only utilizes the unlabeled data that it is certain about their prediction. Here we have $\widetilde{D}(x, y) = \bar{y}$ where $\bar{y}$ the mean confidence over the mask.

Confidence Map: If we decide about each pixel individually, the discriminator outputs a confidence map. Therefore, only parts of the unlabeled image that the model is confident about will be used in the training. Here the discriminator is defined as $D(x_i, y_i) = \hat{y}_i$ if $y_i > \tau$ where x_i and y_i are corresponding pixels of image and mask.

3.4 Training Algorithm

The full training algorithm for our model is presented in Algorithm 1.

Algorithm 1 SegMatch: Semi-supervised Semantic Segmentation

Input: Labeled pool $\mathcal{D}_L$, Unlabeled pool $\mathcal{D}_U$, Initialized weights for segmentation module θ_M, and discriminator module θ_D, Week transformations $\mathcal{T}_1$, and strong transformation $\mathcal{T}_2$

repeat

  Sample labeled batch $\{(x_l^{(b)}, y_l^{(b)}) : b \in (1, \ldots, B)\} \sim \mathcal{D}_L$

  Sample unlabeled batch $\{x_u^{(b)} : b \in (1, \ldots, \mu B)\} \sim \mathcal{D}_U$

  Compute labeled loss as $\mathcal{L}_s = \frac{1}{B} \sum_b H(M(\mathcal{T}_1(x_l^{(b)})), \mathcal{T}_1(y_l^{(b)}))$

  Generate pseudo labels for unlabeled batch $\hat{y}_u^{(b)} = D(M(\mathcal{T}_1(x_u^{(b)}))$

  Compute unlabeled loss as $\mathcal{L}_u = \frac{1}{\mu B} \sum_b \mathbb{1}_{\{\hat{y}_u^{(b)}\}} H(\mathcal{T}_1 \mathcal{T}_2^{-1} M(\mathcal{T}_2(x_u^{(b)})), \hat{y}_u^{(b)})$

  Update weights of segmentation module θ_M by optimizing $\mathcal{L}_s + \lambda_u \mathcal{L}_u$

  Compute discriminator loss $\mathcal{L}_D$ using Equation 1

  Update weights of discriminator module θ_D by optimizing $\mathcal{L}_D$

until Converge

return Trained segmentation module θ_M, and discriminator module θ_D

4 Experiments

We implement our method on brain tumor image segmentation task for different proportions of labeled/unlabeled data. Our dataset [5] containers MRI scans and their corresponding mask for 110 patients. We use 100 scans for training and the remaining 10 scans for the validation. Here we utilize U-Net model [1] which has been shown to be efficient in segmentation tasks. Also, we use the dice score as the metric for evaluating the model. Figure 2 shows the performance of the model on these baselines for brain tumor segmentation task. As it is shown in the figure, FixMatch approach, when it is applied to segmentation task, outperforms other baselines. It also performs significantly better that the supervised case especially, when the number of labeled patients is small.

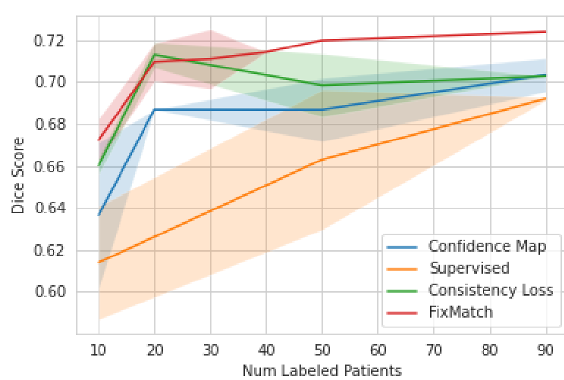

Figure 2: Performance of variants of the model

5 Conclusion

In this work, we propose a new algorithm called SegMatch for semi-supervised semantic segmentation task. Our approach uses consistency loss for incorporating unlabeled data inside the training, however, unlike previous approaches we introduce a discriminator which is responsible for discriminating informative vs non-informative unlabeled and only producing pseudo labels for informative ones. We showed the performance of our model on brain MRI Segmentation task.

References

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[2] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L Rubin, and Bradley J Erickson. Deep learning for brain mri segmentation: state of the art and future directions. *Journal of digital imaging*, 30(4):449–459, 2017.

[3] Alexander Kalinovsky and Vassili Kovalev. Lung image ssgmentation using deep learning methods and convolutional neural networks. 2016.

[4] Maoke Yang, Kun Yu, Chi Zhang, Zhiwei Li, and Kuiyuan Yang. Denseaspp for semantic segmentation in street scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3684–3692, 2018.

[5] Mateusz Buda, Ashirbani Saha, and Maciej A Mazurowski. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in biology and medicine*, 109:218–225, 2019.

[6] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. Ce-net: context encoder network for 2d medical image segmentation. *IEEE transactions on medical imaging*, 38(10):2281–2292, 2019.

[7] Ashish Sinha and Jose Dolz. Multi-scale self-guided attention for medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 2020.

[8] Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In *Advances in neural information processing systems*, pages 1163–1171, 2016.

[9] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*, 2016.

[10] Geoffrey J McLachlan. Iterative reclassification procedure for constructing an asymptotically optimal rule of allocation in discriminant analysis. *Journal of the American Statistical Association*, 70(350):365–369, 1975.

[11] H Scudder. Probability of error of some adaptive pattern-recognition machines. *IEEE Transactions on Information Theory*, 11(3):363–371, 1965.

[12] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *arXiv preprint arXiv:2001.07685*, 2020.

[13] Xiaojin Jerry Zhu. Semi-supervised learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2005.

[14] Amy Zhao, Guha Balakrishnan, Fredo Durand, John V Guttag, and Adrian V Dalca. Data augmentation using learned transformations for one-shot medical image segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8543–8553, 2019.

[15] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. *arXiv preprint arXiv:2003.09005*, 2020.

[16] Gerda Bortsova, Florian Dubost, Laurens Hogeweg, Ioannis Katramados, and Marleen de Bruijne. Semi-supervised medical image segmentation via learning consistency under transformations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 810–818. Springer, 2019.

[17] Bingfeng Zhang, Jimin Xiao, Yunchao Wei, Mingjie Sun, and Kaizhu Huang. Reliability does matter: An end-to-end weakly supervised semantic segmentation approach. *arXiv preprint arXiv:1911.08039*, 2019.

[18] Yunchao Wei, Huaxin Xiao, Honghui Shi, Zequn Jie, Jiashi Feng, and Thomas S Huang. Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7268–7277, 2018.

[19] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In *Advances in neural information processing systems*, pages 2234–2242, 2016.

[20] Zihang Dai, Zhilin Yang, Fan Yang, William W Cohen, and Russ R Salakhutdinov. Good semi-supervised learning that requires a bad gan. In *Advances in neural information processing systems*, pages 6510–6520, 2017.

[21] Wei-Chih Hung, Yi-Hsuan Tsai, Yan-Ting Liou, Yen-Yu Lin, and Ming-Hsuan Yang. Adversarial learning for semi-supervised semantic segmentation. *arXiv preprint arXiv:1802.07934*, 2018.

[22] Sudhanshu Mittal, Maxim Tatarchenko, and Thomas Brox. Semi-supervised semantic segmentation with high-and low-level consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[23] Nasim Souly, Concetto Spampinato, and Mubarak Shah. Semi supervised semantic segmentation using generative adversarial network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5688–5696, 2017.

[24] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical data augmentation with no separate search. *arXiv preprint arXiv:1909.13719*, 2019.

[25] David Berthelot, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Kihyuk Sohn, Han Zhang, and Colin Raffel. Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. *arXiv preprint arXiv:1911.09785*, 2019.