



# Attention, Suggestion and Annotation: A Deep Active Learning Framework for Biomedical Image Segmentation

Haohan Li<sup>1</sup>(✉) and Zhaozheng Yin<sup>2</sup>

<sup>1</sup> Department of Computer Science, Missouri University of Science and Technology,  
Rolla, MO 65409, USA  
hl87c@umsystem.edu

<sup>2</sup> AI Institute, Department of Biomedical Informatics,  
Department of Computer Science,  
Department of Applied Mathematics and Statistics (Affiliated),  
Stony Brook University, Stony Brook, NY 11794, USA  
zyin@cs.stonybrook.edu

**Abstract.** Despite the great success, deep learning based segmentation methods still face a critical obstacle: the difficulty in acquiring sufficient training data due to high annotation costs. In this paper, we propose a deep active learning framework that combines the attention gated fully convolutional network (ag-FCN) and the distribution discrepancy based active learning algorithm (dd-AL) to significantly reduce the annotation effort by iteratively annotating the most informative samples to train the ag-FCN for the better segmentation performance. Our framework is evaluated on 2015 MICCAI Gland Segmentation dataset and 2017 MICCAI 6-month infant brain MRI Segmentation dataset. Experiment results show that our framework can achieve state-of-the-art segmentation performance by using only a portion of the training data.

## 1 Introduction

Automated image segmentation is a cornerstone of many image analysis applications. Recently, thanks to their representation power and generalization capability, deep learning models have achieved superior performance in many image segmentation tasks [1, 2]. However, despite the success, deep learning based segmentation still faces a critical hindrance: the difficulty in acquiring sufficient training data due to the high annotation cost. In biomedical image segmentation, this hindrance is even more severe for the reason that: (1) Only experts can provide precise annotations for biomedical image segmentation tasks, making crowd-computing quite difficult; (2) Biomedical images from high-throughput experiments contain big data of images, which require extensive workforces to provide pixel-level annotations; (3) Due to the dramatic variations in biomedical images (e.g., different imaging modalities and specimens), deep learning models need specific sets of training data to achieve good segmentation performances,

rather than using a general training dataset and transfer learning techniques to solve all kinds of segmentation tasks. Due to these reasons, a real-world biomedical image segmentation project may require thousands of annotation hours from multiple domain experts. Thus, there is a great need to develop annotation suggestion algorithms that can assist human annotators by suggesting the most informative data for annotation to accomplish the task with less human efforts.

## 1.1 Related Works

Despite the great success of deep learning in image segmentation tasks [3, 4], deep learning based segmentation algorithms still face a critical difficulty in acquiring sufficient training data due to the high annotation cost. To alleviate the annotation burden in image segmentation tasks, weakly supervised segmentation algorithms [5–7] have been proposed. However, how to select representative data samples for annotation is overlooked. To address this problem, active learning [8] can be utilized as an annotation suggestion to query informative samples for annotation. As shown in [9], by using active learning, good performance can be achieved using significantly less training data in natural scene image segmentation. However, this method is based on the pre-trained region proposal model and pre-trained image descriptor network, which cannot be easily acquired in the biomedical image field due to the large variations in various biomedical applications. A progressively trained active learning framework is proposed in [10], but it only focuses on the uncertainty and the representativeness of suggested samples in the unlabeled set and ignores the rarity of suggested samples in the labeled set, which can easily incur serious redundancy in the labeled set.

## 1.2 Our Proposal and Contribution

In this work, we propose a deep active learning framework, combining a new deep learning model and a new active learning algorithm, which iteratively suggests the most informative annotation samples to improve the model’s segmentation performance progressively.

Although the motivation seems to be straightforward, it is challenging to design a framework that can perfectly integrate a deep learning model into an active learning process due to the following challenges: (1) The deep learning model should have a good generalization capability so that it can produce reasonable results when little training data are available in the active learning process; (2) The deep learning model should perform well when using the entire training set so that it can provide a good upper-bound of the performance for the active learning framework; (3) The active learning algorithm should be able to make judicious annotation suggestions based on the limited information provided by a not-well-trained deep learning model in the early training stage. To overcome these three challenges, we design a deep active learning framework with two major components: (1) the attention gated Fully Convolutional Network (**ag-FCN**) and (2) the distribution discrepancy based active learning algorithm (**dd-AL**):

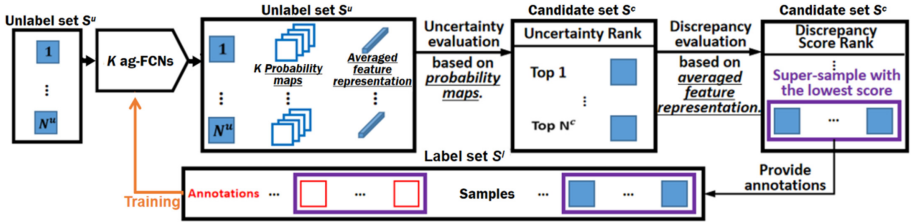


Fig. 1. The workflow of our deep active learning framework.

- **Attention:** For the first and second challenges, we design a novel ag-FCN that uses attention gate units (AGUs) to automatically highlight salient features of the target content for accurate pixel-wise predictions. In addition, both of the ag-FCN and the AGU are built using bottleneck designs to significantly reduce the number of network parameters while maintaining the same number of feature channels at the end of each residual module. This design ensures the good generality of the proposed ag-FCN.
- **Suggestion and Annotation:** For the third challenge, we design the dd-AL to achieve the final goal of the iterative annotation suggestion: decreasing the distribution discrepancy between the labeled set and the unlabeled set<sup>1</sup>. If the discrepancy between these two sets is small enough, which means their distributions are similar enough, the classifier trained on the labeled set can achieve similar performance compared to the classifier trained on the entire training dataset with all samples annotated. Therefore, besides the uncertainty metric, dd-AL also evaluates each unlabeled sample’s effectiveness in decreasing the distribution discrepancy between the labeled set and the unlabeled set after we annotate it, which is further evaluated by the representativeness and rarity metrics.

## 2 Method

Figure 1 shows the workflow of our deep active learning framework. In each annotation suggestion stage, first we pass each unlabeled sample through  $K$  ag-FCNs to obtain its  $K$  segmentation probability maps and the corresponding averaged feature representation. Then, dd-AL selects the most informative unlabeled samples based on their uncertainties to the currently-trained ag-FCNs and effectiveness in decreasing the data distribution discrepancy between the labeled and unlabeled set. Finally, the small set of suggested samples are annotated for fine-tuning the ag-FCNs. We conduct this annotation suggestion process iteratively until satisfied.

<sup>1</sup> In this paper, the labeled set and unlabeled set refer to the labeled and unlabeled portions of a training dataset, respectively.

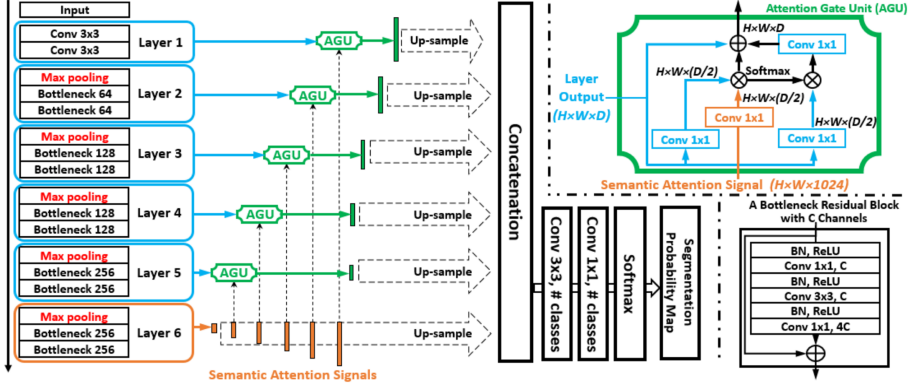


Fig. 2. The architecture of the semantic attention guided fully convolutional network.

## 2.1 Attention Gated Fully Convolutional Network

Based on recent advances of deep neural networks such as the fully convolutional network (FCN, [3]) and non-local network [11], we propose an attention gated fully convolutional network (ag-FCN) that can not only conduct accurate image segmentation but also be suitable for active learning. Compared with the original FCN, our ag-FCN, shown in Fig. 2, has three main improvements:

**Attention Gate Units:** We propose the Attention Gate Unit (AGU) to fuse the high-level semantic features to low- and mid-level features. AGU exploits the high-level semantic information as soft-attentions that lead low- and mid-level features to focus on target areas and highlight the feature activations that are relevant to the target instance. Hence, AGU ensures that the ag-FCN can conduct accurate segmentation on object instances with high variabilities.

**Feature Fusion Strategy:** Compared with the conventional skip-connections that progressively merge low-level features to the up-sampling process of high-level features [3], the feature fusion strategy in the ag-FCN considers each layer’s attentive features (with semantic attention) as an up-sampling seed. All seeds will be progressively up-sampled to the input image size, and then be concatenated for generating smooth segmentation results.

**Bottleneck Residual Modules:** In ag-FCN, we replace most convolutional layers by bottleneck residual modules to significantly reduce the number of parameters while maintaining the same receptive field size and feature channels at the end of each module. This design reduces the training cost with less parameters (i.e., suitable for the iterative active learning) and maintains ag-FCN’s generalization capability.

These three improvements of our ag-FCN are essential when combining deep neural networks and active learning. By using our AGUs and feature fusion strategy, the ag-FCN can achieve state-of-the-art segmentation performance using all

training data, which provides a good upper-bound performance for our framework. By using the bottleneck residual blocks, the ag-FCN can have good generalization capability even when very little training data are available, and facilitate the iterative active learning with small sets of network parameters.

## 2.2 Distribution Discrepancy Based Active Learning Algorithm

In general, our distribution discrepancy based active learning algorithm (dd-AL) suggests samples for annotation based on two criteria: (1) the uncertainty to the segmentation network and (2) the effectiveness in decreasing the distribution discrepancy between the labeled set and unlabeled set. Since parallelly evaluating these two criteria of each unlabeled sample is computational expensive, dd-AL conducts the annotation suggestion process in two sequential steps. As shown in Fig. 1, first, dd-AL selects  $N^c$  samples with the highest uncertainty scores from the unlabeled set as candidate samples. Secondly, among these  $N^c$  candidate samples, dd-AL selects a subset of them that have the highest effectiveness in decreasing the distribution discrepancy between the labeled and unlabeled set.

**Evaluating a Sample’s Uncertainty:** In the first step of dd-AL, to evaluate the uncertainty of each unlabeled sample, we adopt the bootstrapping strategy that trains  $K$  ag-FCNs, each of which only uses a subset of the suggested data for training in each annotation suggestion stage, and calculates the disagreement among these  $K$  models. Specifically, in each annotation suggestion stage, for each unlabeled sample  $s^u$  whose spatial dimension is  $h \times w$ , we first use  $K$  ag-FCNs to generate  $K$  segmentation probability maps of  $s^u$ . Then, we compute an uncertainty score  $u_k^{s^u}$  of the  $k$ -th ( $k \in [1, K]$ ) segmentation probability map of  $s^u$  by using the Best-versus-Second-Best (BvSB) strategy:

$$u_k^{s^u} = \frac{1}{h \times w} \sum_{i=1}^{h \times w} (1 - |p_{k,i}^{best} - p_{k,i}^{second}|), \quad (1)$$

where  $p_{k,i}^{best}$  and  $p_{k,i}^{second}$  denote the probability values of the most-possible class and second-possible class of the  $i$ -th pixel on  $s^u$ , respectively, predicted by the  $k$ -th ag-FCN.  $(1 - |p_{k,i}^{best} - p_{k,i}^{second}|)$  denotes the pixel-wise BvSB score, where a larger score indicates more uncertainty. In Eq. 1, the uncertainty score of  $s^u$  estimated by the  $k$ -th ag-FCN is the average of the BvSB scores of all pixels in this image. We compute the final uncertainty score of  $s^u$  by averaging the uncertainty scores predicted by all  $K$  ag-FCNs:

$$u_{final}^{s^u} = \frac{1}{K} \sum_{k=1}^K u_k^{s^u}. \quad (2)$$

Then, we rank all the unlabeled samples based on their final uncertainty scores and select the top  $N^c$  samples with the highest uncertainty scores as the candidate set  $S^c$  for the second step of dd-AL.

**Evaluating a Sample’s Effectiveness in Decreasing Discrepancy:** In the second step of dd-AL, we aim to annotate a subset of the candidate set  $S^c$  that can achieve the smallest distribution discrepancy between the labeled set and unlabeled set after the annotation. After several annotation suggestion stages, if the distributions of the labeled set and unlabeled set are similar enough, the classifier trained on the labeled set can achieve similar performance compared to the classifier trained on the entire dataset with all samples annotated.

In each annotation suggestion stage, we define  $S^l$  as the labeled set with  $N^l$  samples and  $S^u$  as the unlabeled set with  $N^u$  samples. We use the  $i$ -th candidate sample  $s_i^c$  in  $S^c$ , where  $i \in [1, N^c]$ , as a reference data point to estimate the data distributions of the unlabeled set  $S^u$  and the labeled set  $S^l$ , and compute a distribution discrepancy score  $d_i^c$  that represents the distribution discrepancy between  $S^u$  and  $S^l$  after annotating  $s_i^c$ :

$$d_i^c = \frac{1}{N^l + 1} \sum_{j=1}^{N^l+1} Sim(s_i^c, s_j^l) - \frac{1}{N^u - 1} \sum_{j=1}^{N^u-1} Sim(s_i^c, s_j^u). \quad (3)$$

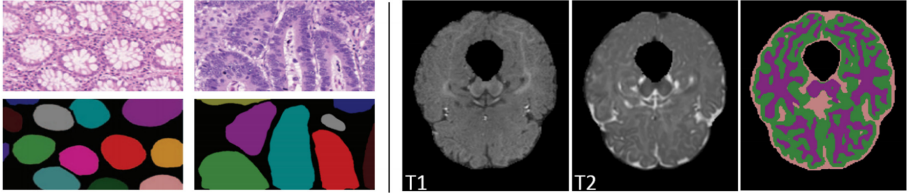
In Eq. 3, the first term represents the data distribution of the labeled set  $S^l$  estimated by  $s_i^c$ , where  $Sim(s_i^c, s_j^l)$  represents the *cosine similarity* between  $s_i^c$  and the  $j$ -th sample  $s_j^l$  in the labeled set  $S^l$  in the high-dimensional feature space<sup>2</sup>. The second term in Eq. 3 represents the data distribution of the unlabeled set  $S^u$  estimated by  $s_i^c$ , where  $Sim(s_i^c, s_j^u)$  represents the *cosine similarity* between  $s_i^c$  and the  $j$ -th sample  $s_j^u$  of the unlabeled set  $S^u$  in the high-dimensional feature space. After we compute the distribution discrepancy scores for all candidate samples in  $S^c$ , the candidate sample with the lowest score can be chosen as the most informative sample for annotation.

To accelerate the annotation suggestion process, we prefer to suggest multiple samples for the annotation in each stage instead of suggesting one sample at a time. However, directly ranking the candidate samples in an ascending order based on their distribution discrepancy scores and suggesting the top ones is inaccurate. Since the distribution discrepancy of the labeled and unlabeled sets is computed based on annotating one sample at a time.

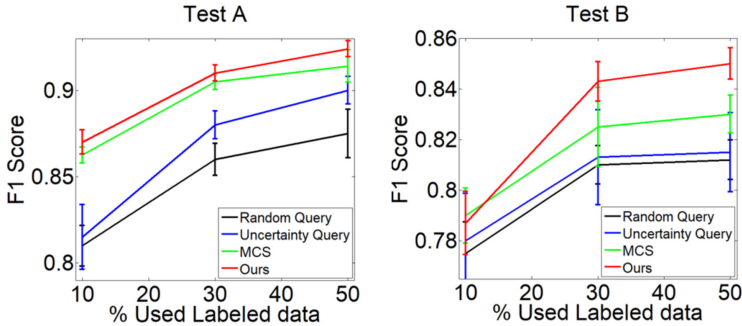
To address this problem, we propose the idea of super-sample  $s^{super}$ , which is a  $m$ -combination of the candidate set  $S^c$  with  $N^c$  samples. In total, there are  $\binom{N^c}{m}$  possible super-samples that can be generated from  $S^c$ . The feature representation of each super-sample is the average of the feature representations of the  $m$  samples within it. Thus, we can rewrite the distribution discrepancy score computation in Eq. 3 into a super-sample version as:

$$d_q^{super} = \frac{1}{N^l + m} \sum_{j=1}^{N^l+m} Sim(s_q^{super}, s_j^l) - \frac{1}{N^u - m} \sum_{j=1}^{N^u-m} Sim(s_q^{super}, s_j^u), \quad (4)$$

<sup>2</sup> The encoding part of each ag-FCN can be utilized as a feature extractor. Given an input image to  $K$  ag-FCNs, the average of outputs of Layer 6 in these ag-FCNs can be viewed as a high-dimensional feature representation of the input image.



**Fig. 3.** Some qualitative results of our framework on GlaS dataset (left) and iSeg dataset (right, pink: Cerebrospinal Fluid; purple: White Matter; green: Gray Matter) using only 50% training data. (Color figure online)



**Fig. 4.** Comparison using limited training data of GlaS dataset. FCN-MCS [10] is an active learning algorithm only considering uncertainty and representativeness.

where  $d_q^{super}$  denotes the distribution discrepancy score of the  $q$ -th super-sample  $s_q^{super}$  in the candidate set  $S^c$ . Then, the super-sample with the lowest distribution discrepancy score will be suggested, where the  $m$  samples within this super-sample will be the final suggested samples in this annotation suggestion stage. Finally, these samples will be annotated for fine-tuning the ag-FCNs.

The suggestion is to find the super-sample with the lowest distribution discrepancy score in Eq. 4. In other words, dd-AL aims to suggest samples that can minimize the first term in Eq. 4, which is equivalent to minimizing the similarity between suggested samples and the labeled set  $S^l$ . Therefore, the proposed dd-AL ensures the **high rarity** of suggested samples in the labeled set. Also, in Eq. 4, dd-AL aims to suggest samples that can maximize the second term, which is equivalent to maximizing the similarity between suggested samples and the unlabeled set  $S^u$ . Therefore, the proposed dd-AL can also ensure the **high representativeness** of suggested samples regarding to the unlabeled set.

### 3 Experiment

**Dataset.** As the same as [10, 14–16], we use the 2015 MICCAI gland segmentation dataset (GlaS, [12]) and the training set of 2017 MICCAI infant brain

**Table 1.** Comparing accuracies with state-of-the-art methods on GlaS dataset [12].

Method	F1 score		ObjectDice		ObjectHausdorf	
	Test A	Test B	Test A	Test B	Test A	Test B
CUMedNet [15]	0.912	0.716	0.897	0.781	45.418	160.347
TCC-MSFCN [16]	0.914	0.850	0.913	0.858	39.848	93.244
MILD-Net [14]	0.920	0.820	0.918	0.836	<b>39.390</b>	103.070
FCN-MCS [10]	0.921	0.855	0.904	0.858	44.736	96.976
Ours (full training data)	<b>0.938</b>	<b>0.866</b>	<b>0.929</b>	<b>0.870</b>	40.548	<b>93.103</b>
FCN-MCS [10] (50% training data)	0.913	0.832	0.901	0.836	–	–
ag-FCN-MCS (50% training data)	0.915	0.842	0.911	0.850	43.363	95.796
FCN-dd-AL (50% training data)	0.920	0.836	0.912	0.844	43.143	96.131
Ours (50% training data)	<b>0.924</b>	<b>0.850</b>	<b>0.918</b>	<b>0.866</b>	<b>41.988</b>	<b>95.331</b>

segmentation dataset (iSeg, [13]) to evaluate the effectiveness of our deep active learning framework on different segmentation tasks. The GlaS contains 85 training images and 80 testing images (Test A: 60; Test B: 20). The training set of iSeg contains T1- and T2-weighted MR images of 10 subjects. We augment the training data with flipping and elastic distortion. In addition, the original image (volume) is cropped by sliding windows into image patches (cubes), each of which is considered as a sample in the annotation suggestion process. There are 27200 samples generated from GlaS and 16380 samples generated from iSeg.

**Implementation Details.** In our experiments, we train 3 ag-FCNs ( $K = 3$ ). For each annotation stage, we select top 16 uncertain samples in the first step of dd-AL ( $N^c = 16$ ), and our super-sample size is 12 ( $m = 12$ ) in the second step of dd-AL. At the end of each stage, ag-FCNs will be fine-tuned with all available labeled data.

**Experiments on GlaS.** First, we compared our ag-FCNs using all training data with state-of-the-art methods. As shown in Table 1, our ag-FCNs achieve very competitive segmentation performances (best in five columns), which shows the effectiveness of our ag-FCN in producing accurate pixel-wise predictions. Secondly, to validate our entire framework (ag-FCNs and dd-AL), we simulate the annotation suggestion process by only providing the suggested samples and their annotations to the ag-FCNs for training. For fair comparison, we follow [10] to consider the annotation cost as the number of annotated pixels and set the annotation cost budget as 10%, 30% and 50% of the overall labeled pixels. Note, though the suggested samples are generated from the original image data by using data augmentation techniques, the annotation cost budget is based on the annotation of the original image data. Our framework is compared with (1) Random Query: randomly selecting samples; (2) Uncertainty Query: suggesting samples only considering uncertainties; (3) FCN-MCS, an active learning algorithm that only considers the uncertainty and representativeness proposed in [10]. We follow [10] to randomly divide the GlaS training set into ten folds,



**Table 2.** Analyzing the effect of changing the super-sample size on GlaS Test A.

Super-sample size	w/o	4	6	8	10	<b>12</b>	14
F1-score	0.918	0.920	0.920	0.919	0.923	<b>0.924</b>	0.919
Total training time	13.3 h	8.2 h	7.1 h	6.3 h	5.1 h	<b>4.4 h</b>	2.5 h

**Table 3.** Comparing computation cost with FCN-MCS [10], the current best annotation suggestion algorithm on GlaS dataset. Note, though our GPUs are different from [10], V100 is only 2.9x faster than P100 on training typical deep learning benchmarks on average [17]. So, the lead cause of reducing the computation cost is our newly-designed framework, not the GPU hardware.

Methods	Time cost per stage	# of stages	Total training time	GPU
FCN-MCS [10]	10 min	2000	333.3 h	4 Tesla P100
Ours	20 s	800	4.4 h	4 Tesla V100

each of which is used as the initial training data for one experiment. The average results are reported. As shown in Fig. 4, our framework is consistently better than the other three query methods. Thirdly, we conduct ablation studies on our framework (ag-FCN and dd-AL) by replacing our dd-AL by the active learning algorithm MCS proposed in [10] (shown as ag-FCN-MCS in Table 1) and replacing our ag-FCN by the FCN proposed in [10] (shown as FCN-dd-AL in Table 1). As shown in Table 1, both ag-FCN-MCS and FCN-dd-AL outperform the deep active learning framework FCN-MCS proposed in [10], and our framework obtains the best performance among all the four methods using 50% training data, which reveals that the boosted performance of our framework is due to both our ag-FCN and dd-AL. Fourthly, we study the effect of changing the size of the super-sample. As shown in Table 2, compared with our framework without using super-samples, our framework using super-sample size 12 largely improves the training time by 9 h and the F1-score by 0.6% on the GlaS Test A, which validates the effectiveness of our super-sample version of the distribution discrepancy score computation. Fifthly, in addition to outperforming the current best annotation suggestion algorithm [10] on biomedical image segmentation in terms of accuracy (Table 1), our framework is more efficient (Table 3).

**Experiments on iSeg.** We also extend our ag-FCN into the 3D version (3D-ag-FCN)<sup>3</sup> and test our framework (3D-ag-FCN and dd-AL) on the training set of iSeg using 10-fold cross-validation (9 subjects for training, 1 subject for testing, repeat 10 times). As shown in Table 4, our framework still achieves competitive performances even using only 50% training data. Figure 3 shows some qualitative examples on the two datasets.

<sup>3</sup> We replace all 2D operations with 3D operations (e.g., 2D conv  $\rightarrow$  3D conv, etc.).

**Table 4.** Comparison with state-of-the-art on iSeg dataset [13]. (DICE: Dice Coefficient; MHD: Modified Hausdorff Distance; ASD: Average Surface Distance).

Method	White matter			Gray matter			Cerebrospinal fluid		
	DICE	MHD	ASD	DICE	MHD	ASD	DICE	MHD	ASD
3D-Unet [18]	0.896	5.39	0.44	0.907	5.90	0.38	0.944	13.86	0.15
VoxResNet [19]	0.899	<b>5.20</b>	0.44	0.906	5.48	0.38	0.943	13.93	0.15
3D-SDenseNet [20]	0.910	5.92	<b>0.39</b>	0.916	5.75	<b>0.34</b>	0.949	13.64	0.13
Ours (full training data)	<b>0.928</b>	5.21	0.40	<b>0.921</b>	<b>5.44</b>	0.35	<b>0.960</b>	<b>13.26</b>	<b>0.12</b>
Ours (50% training data)	0.920	5.33	0.44	0.914	5.61	0.37	0.951	13.58	0.17

## 4 Conclusion

To significantly alleviate the burden of manual labeling in the biomedical image segmentation task, we propose a deep active learning framework that consists of: (1) an attention gated fully convolutional network (ag-FCN) that achieves state-of-the-art segmentation performances when using the full training data and (2) a distribution discrepancy based active learning algorithm that progressively suggests informative samples to train the ag-FCNs. Our framework achieves state-of-the-art segmentation performance by only using a portion of the annotated training data on two MICCAI challenges.

## References

1. Kamnitsas, K., Ledig, C., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **36**, 61–78 (2017)
2. Liao, F., Liang, M., et al.: Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-or network. *IEEE Trans. Neural Netw. Learn. Syst.* **30**(11), 3484–3495 (2019)
3. Long, J., Shelhamer, E., et al.: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440 (2015)
4. He, K., Gkioxari, G., et al.: Mask R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969 (2017)
5. Papandreou, G., Chen, L.C., et al.: Weakly-and semi-supervised learning of a DCNN for semantic image segmentation. *arXiv, arXiv preprint arXiv:1502.02734* (2015)
6. Xiao, H., Wei, Y., et al.: Transferable semi-supervised semantic segmentation. In: *AAAI Conference on Artificial Intelligence* (2018)
7. Hong, S., Noh, H., et al.: Decoupled deep neural network for semi-supervised semantic segmentation. In: *Advances in Neural Information Processing Systems*, pp. 1495–1503 (2015)
8. Settles, B.: *Active learning literature survey*. University of Wisconsin-Madison Department of Computer Sciences (2009)
9. Dutt Jain, S., Grauman, K.: Active image segmentation propagation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2864–2873 (2016)

10. Yang, L., Zhang, Y., et al.: Suggestive annotation: a deep active learning framework for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 399–407 (2017)
11. Wang, X., Girshick, R., et al.: Non-local neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7794–7803 (2018)
12. Sirinukunwattana, K., Pluim, J.P., et al.: Gland segmentation in colon histology images: the glas challenge contest. *Med. Image Anal.* **35**, 489–502 (2017)
13. Wang, L., Nie, D., et al.: Benchmark on automatic six-month-old infant brain segmentation algorithms: the iSeg-2017 challenge. *IEEE Trans. Med. Imaging* **38**(9), 2219–2230 (2019)
14. Graham, S., Chen, H., et al.: MILD-Net: minimal information loss dilated network for gland instance segmentation in colon histology images. *Med. Image Anal.* **52**, 199–211 (2019)
15. Graham, S., Chen, H., et al.: DCAN: deep contour-aware networks for accurate gland segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2487–2496 (2016)
16. Ding, H., Pan, Z., et al.: Multi-scale fully convolutional network for gland segmentation using three-class classification. *Neurocomputing* **380**, 150–161 (2020)
17. TESLA V100 Performance Guide. <https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/tesla-product-literature/v100-application-performance-guide.pdf>
18. Çiçek, Ö., Abdulkadir, A., et al.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 424–432 (2016)
19. Chen, H., Dou, Q., et al.: VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage* **170**, 446–455 (2018)
20. Bui, T.D., Shin, J., et al.: Skip-connected 3D DenseNet for volumetric infant brain MRI segmentation. *Biomed. Signal Process. Control* **54**, 101613 (2019)