Computerized Medical Imaging and Graphics (2023)

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/media

BBox-Guided Segmentor: Leveraging Expert Knowledge for Accurate Stroke Lesion Segmentation Using Weakly Supervised Bounding Box Prior

Yanglan Ou[a,*], Sharon X. Huang[a], Kelvin K. Wong[b], Jonathon Cummock[b], John Volpi[c], James Z. Wang[a], Stephen T.C. Wong[b]

[a]*Data Science and Artificial Intelligence Area, College of Information Sciences and Technology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*
[b]*T.T. and W.F. Chao Center for BRAIN & Houston Methodist Cancer Center, Houston Methodist Hospital, Houston, Texas 77030, USA*
[c]*Eddy Scurlock Comprehensive Stroke Center, Department of Neurology, Houston Methodist Hospital, Houston, Texas 77030, USA*

ARTICLE INFO

Article history:

Keywords: Stroke, Lesion Segmentation, Weakly Supervised, Adversarial Learning, Efficient Annotation, Bounding Box

ABSTRACT

Stroke is one of the leading causes of death and disability in the world. Despite intensive research on automatic stroke lesion segmentation from non-invasive imaging modalities including diffusion-weighted imaging (DWI), challenges remain such as a lack of sufficient labeled data for training deep learning models and failure in detecting some small lesions. In this paper, we propose BB-Guided Segmentor, a method that significantly improves the accuracy of stroke lesion segmentation by leveraging expert knowledge. Specifically, our model uses a very coarse bounding box label provided by the expert and then performs accurate segmentation automatically. The small overhead of having the expert provide a rough bounding box leads to large performance improvement in segmentation, which is paramount to accurate stroke diagnosis. To train our model, we employ a weakly-supervised approach that uses a large number of weakly-labeled images with only bounding boxes and a small number of fully labeled images. The scarce fully labeled images are used to train a generator segmentation network, while adversarial training is used to leverage the large number of weakly-labeled images to provide additional learning signals. We evaluate our method extensively using a unique clinical dataset of 99 fully labeled cases (*i.e.*, with full segmentation map labels) and 831 weakly labeled cases (*i.e.*, with only bounding box labels), and the results demonstrate the superior performance of our approach over state-of-the-art stroke lesion segmentation models. We also achieve competitive performance as a SOTA fully supervised method using less than one-tenth of the complete labels. Our proposed approach has the potential to improve stroke diagnosis and treatment planning, which may lead to better patient outcomes.

1. Introduction

Stroke is the fifth most dominant cause of death in the United States (Xu et al., 2021) and the most common cause of severe disability (Adamson et al., 2004). According to the National Health and Nutrition Examination Survey, more than 7.8 million people suffer from stroke, and the number keeps increasing (Tsao et al., 2022). Diffusion-weighted magnetic resonance

*All correspondence should be addressed to Y. Ou, S.X. Huang, and K.K. Wong.
e-mail: yanglanou@psu.edu (Yanglan Ou), suh972@psu.edu (Sharon X. Huang), kwong@houstonmethodist.org (Kelvin K. Wong)

imaging (DWI) is a valuable tool in the diagnosis of vascular strokes in the brain, as it has shown superiority over other conventional imaging methods, such as computed tomography (CT) or ordinary magnetic resonance imaging (MRI), in the detection of small and early cerebral infarction (Crisostomo et al., 2003; Liu et al., 2014). For evaluation and treatment guidance based on DWI, accurate segmentation of acute ischemic lesions in DWI images is essential (Woo et al., 2019). In clinical practice, however, it can take a radiologist more than an hour to trace lesions in a brain scan manually. The laborious manual segmentation process has the risk of introducing bias too. To alleviate this burden and improve accuracy, automatic and reliable lesion segmentation systems are in urgent need.

Over the last few years, many deep learning methods have achieved great success in medical image segmentation (Kirillov et al., 2019; Ma et al., 2021; Chen et al., 2021; Dolz et al., 2020; Silva-Rodríguez et al., 2021), some are specifically applied to ischemic stroke lesion segmentation (Wu et al., 2019; Zhang et al., 2020a; Wong et al., 2022; Abramova et al., 2021; Qi et al., 2019; Zhou et al., 2019). However, most existing deep learning methods for segmentation are supervised, which would require large datasets precisely annotated by radiologists. Generally, for a 3D volume consisting of many 2D slices such as DWI data, the labels are obtained by asking radiologists or expert users to manually label the stroke region slice by slice, which is time-consuming and difficult to reproduce. Furthermore, there are often variations among labels provided by different experts due to variations in segmentation protocol and experience level.

To mitigate these difficulties, we propose an annotation-efficient segmentation method with prior knowledge: we incorporate a loose bounding box provided by the expert into an automatic segmentation model. The motivation for this implementation is that bounding box annotation is not as time-consuming to get as pixel-level annotations. In clinical practice, considering the effort needed to redraw segmentation when an algorithm gives inaccurate results, the time investment in providing a bounding box that will boost the accuracy is worthwhile. Although having users provide bounding boxes is a simple and popular interaction paradigm considered by many existing interactive image segmentation frameworks (Zhang et al., 2020b; Yu et al., 2017; Wang et al., 2018), our approach is different in that it leverages a small number of fully-labeled cases (*i.e.*, with pixel-level annotation) and a large number of weakly labeled cases (*i.e.*, with only loose bounding box annotation) for training so that our model can be retrained and continuously improve itself as more labeled cases become available. This is especially important for improving the generalizability of the model and during model deployment phase when quick bounding box labels can be used to label more cases including those with rare occurring stroke lesions and different types of image artifacts.

In the literature, there have been some weakly supervised methods applied to semantic segmentation with box-level labels (Hsu et al., 2019; Remez et al., 2018; Han et al., 2020; Lee et al., 2021; Zhang et al., 2022), but they usually couple object detection and segmentation and the bounding boxes are used to train object detectors. In comparison, we directly use expert-provided bounding boxes as input to our segmentation network to improve model generalizability and segmentation accuracy. Furthermore, most existing methods utilizing box-level labels rely on global context to achieve accurate object detection. However, for stroke lesion segmentation, there is no clear context information to utilize because the locations of lesions are usually random, and the shape and extent of lesions vary greatly. To address the limitations of current methods including annotation-efficient methods as applied to ischemic stroke lesion segmentation, we present a weakly supervised segmentation method guided by bounding box input. Our model is trained with a small number of fully-labeled images and a large number of weakly-labeled images. It is based on a novel adversarial learning framework for segmentation, which consists of a segmentation network (*i.e.*, generator) that generates the segmentation prediction, and a discriminator network that outputs a confidence map indicating the assessed quality of the prediction at the pixel level. The generator takes as input a DWI channel, an exponential apparent diffusion map (eADC) channel, and a lesion bounding boxes map (in the form of a binary map). For the small amount of pixel-level labeled data, we make use of the ground truth label maps to update both the generator and discriminator networks. When using the weakly labeled data, the confidence map produced by the discriminator can be used to provide feedback signals to further refine the segmentation network.

To evaluate our framework, we conduct experiments on an ischemic stroke DWI dataset with 99 3D images fully labeled at the pixel level and 831 weakly labeled 3D images with lesion bounding boxes. Our framework achieves a much higher DSC of 91.77% in this setting as compared to the highest achieved by baseline methods at 87.33% DSC. To further evaluate our method in a real clinical scenario, our method is tested on a larger dataset that includes challenging cases and achieves competitive performance compared to a fully-supervised method.

In summary, the main contributions of our work are as follows. First, we propose BB-Guided Segmentor, a new weakly-supervised segmentation pipeline with bounding box input that learns a highly accurate segmentation model using only a small amount of fully-label images. Second, we introduce a novel adversarial framework that supports learning from weakly-labeled data and significantly improves segmentation accuracy. Third, our method demonstrates promising performance through evaluation, achieving comparable segmentation accuracy to a state-of-the-art fully-supervised approach while using less than one-tenth of the fully-labeled data.

2. Methodology

In this section, we introduce our proposed adversarial deep learning framework for segmenting ischemic stroke lesions. Considering that weak annotations such as lesion bounding boxes can provide useful localization information, we work in a weakly supervised setting to investigate the benefit of using additional bounding box information. In this scenario, a user only needs to provide a rough 3D bounding box for each lesion, and then our proposed method can automatically perform precise lesion segmentation.

2.1. Data Setting

The training has two modes: (1) We use a small amount of fully labeled data $\mathcal{F}$ where each data tuple $[x, b, y] \in \mathcal{F}$ consists of an input image x concatenating the Exponential ADC and DWI channels, the labeled binary mask of bounding boxes b of the lesions inside x, and the full lesion segmentation label map y; and (2) we use a large amount of weakly labeled data $\mathcal{W}$ where each data tuple $[x, b] \in \mathcal{W}$ only includes the input image x and the mask of lesion bounding boxes b.

2.2. Adversarial Learning with Fully Labeled Data

To fully leverage the fine-grained lesion boundary annotations in the small amount of fully labeled data $\mathcal{F}$, we use $\mathcal{F}$ inside an adversarial learning framework so that a trained critic (or discriminator) network can provide feedback to the segmentor (or generator) network based on pixel-wise confidence of the predicted segmentation. Fig. 1 (Top) shows an overview of our algorithm for training with fully labeled data.

Generator (segmentation network): As illustrated in Fig. 1 (Top), we follow the conditional generative adversarial network (CGAN (Mirza and Osindero, 2014)) formulation, where a generator G is used to map Gaussian noise z and input image x to the predicted probability map $\hat{y} = G(z, x, b)$.

For the architecture of the generator network G, we adopt the Patcher (Ou et al., 2022) as the backbone. Patcher is newly proposed for medical image segmentation, which is extended from the transformer and has a strong ability to incorporate both global and local contexts in learning. The Patcher encoder employs a cascade of four Patcher blocks to produce four feature maps with decreasing spatial dimensions and increasing receptive fields. Then it applies a Mixture of Experts (MoE) decoder to use a gating network to select a suitable set of features from the encoder to output the prediction. The inputs to the Patcher include an image x with size $256 \times 256 \times 2$, a Gaussian noise channel z, and a bounding box map b. The two-channel input image x is formed by the concatenation of Exponential ADC and DWI channels of the ischemic stroke clinic data.

Discriminator: As shown in Fig. 1 (Top), we also include a discriminator D to differentiate between the generated prediction $\hat{y}$ and the real label y for the input image x. The discriminator D is based on a fully convolutional network (FCN) (Long et al., 2015). We multiply the input image with each of the class probability maps $\hat{y}$ (or ground truth y), leading to an adversarial input with 2 channels. Then the discriminator takes the product of either the fake pair $(x, \hat{y})$ or the real pair (x, y). The symbol $\bigodot$ in the diagram means Hadamard (or pixel-wise) product. Instead of producing a single image-level probability as in classical GAN, our proposed discriminator produces a probability per pixel, thus giving rise to a 256×256 confidence map. The value at each pixel in the confidence map is between 0 and 1, where 0 and 1 represent fake and real, respectively. The confidence map allows us to evaluate the "goodness" of the segmentation at the pixel level, which will prove beneficial in our adversarial setting.

The adversarial loss $\mathcal{L}_{\text{CGAN}}$ is applied to each pixel of the confidence map and the average loss is used to train the discriminator.

The CGAN objective can be written as:

$$\begin{aligned}\mathcal{L}_{\text{CGAN}}(G, D) =& \mathbb{E}_{x,y}[\sum_i \log D(x_i, y_i)] \\ &+ \mathbb{E}_{x,b,z}[\sum_i \log(1 - D(x_i, \hat{y}_i))] \,.\end{aligned} \tag{1}$$

Note that our adversarial loss is pixel-wise, as x_i and y_i represent the pixel value and label of pixel i, respectively. $\hat{y}_i = G(z, x, b)_i$ is the predicted probability value at pixel i.

Besides the adversarial loss $\mathcal{L}_{\text{CGAN}}$, we also use the cross-entropy loss between the label y and the predicted probability map $\hat{y}$ as additional supervision to train the generator:

$$\mathcal{L}_{CE} = \sum_i -\left(y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)\right) \,. \tag{2}$$

The overall loss for a fully labeled pair (x, y) in $\mathcal{F}$ is defined as:

$$\mathcal{L}_{\mathcal{F}} = \mathcal{L}_{CE} + \lambda_{\text{CGAN}} \mathcal{L}_{\text{CGAN}} \,, \tag{3}$$

where λ_{CGAN} is a weighting coefficient.

2.3. Learning with Weakly Labeled Data

When the discriminator D is sufficiently trained with fully labeled data $\mathcal{F}$, we start weakly-supervised training with weakly-labeled data $\mathcal{W}$. Fig. 1 (Bottom) illustrates our framework using weakly labeled data.

During the training of the discriminator using fully labeled data, the confidence map for real pairs is expected to have high values (as close as possible to 1) at all pixels, while for fake pairs the values are expected to be low (as close as possible to 0). Using weakly labeled data, the input to the discriminator is a fake pair $(x, \hat{y})$, and the discriminator outputs a confidence map for this fake pair; we can use this confidence map, $D(x, \hat{y})$, to assess the confidence in the label prediction at each pixel for the image x. Then we obtain a mask M by binarizing the confidence map: $M = \mathbb{I}(D(x, \hat{y}) > T)$, where T is the confidence threshold. In our experiments, we set the value of the confidence threshold T to be 0.49. As mask M highlights the areas where the predicted segmentation is believed to be realistic by the discriminator, we can then formulate a loss:

$$\mathcal{L}_{\mathcal{W}} = -M \odot \left(y' \log \hat{y} + (1 - y') \log(1 - \hat{y})\right) \,, \tag{4}$$

where we use the binarized prediction y' as the pseudo label, and calculate the cross-entropy loss between y' and the predicted probability map $\hat{y}$ within the masked areas. Again, $\odot$ stands for the Hadamard (or pixel-wise) product.

An alternative approach to using a binary mask M is to directly multiply the confidence map $D(x, \hat{y})$ with the cross-entropy loss. However, in our experiments, this alternative approach did not give as good a performance as using M.

2.4. Inference

For testing and inference in the weakly supervised segmentation setting, we use the four-channel input consisting of the two-channel clinical data x, the Gaussian noise channel z, and the bounding box mask b for the generator. Given the input, the trained generator produces the probability map $\hat{y}$, which is then binarized to get the segmentation map y', as shown in Fig. 2 (Bottom).

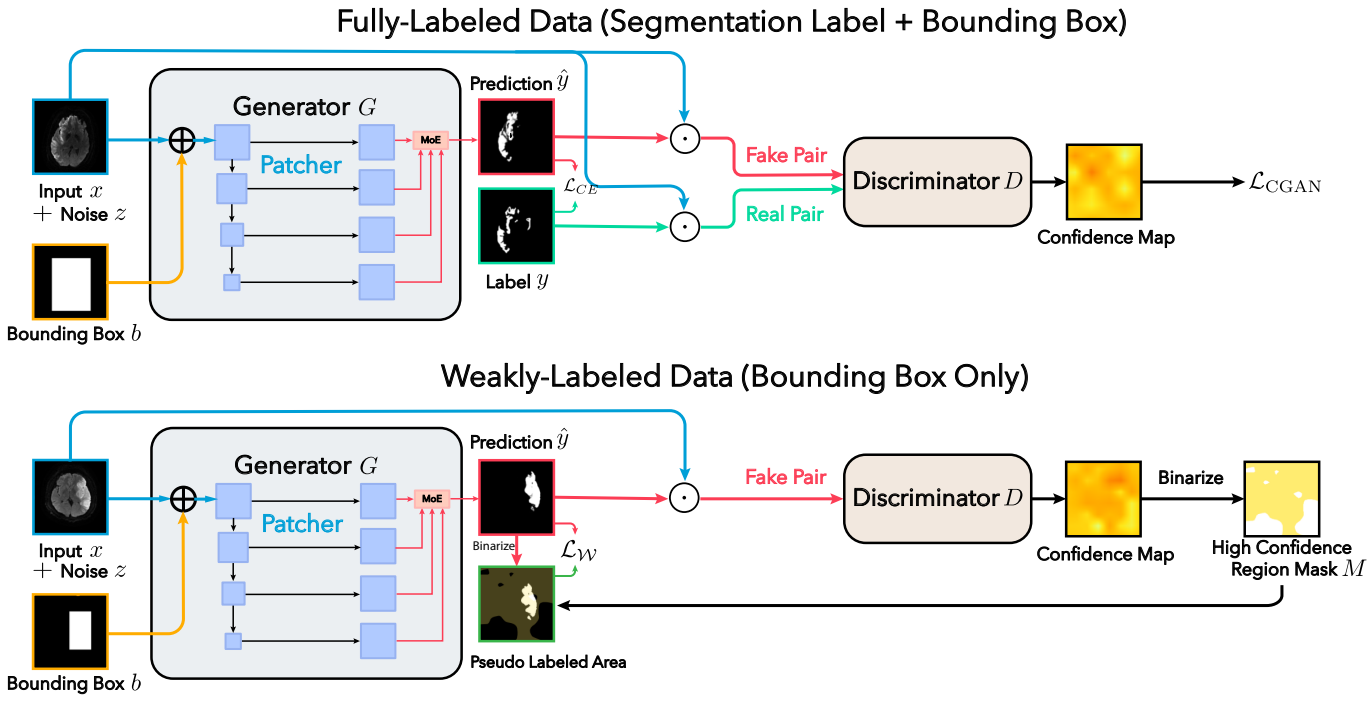

Fig. 1. An overview of our weakly supervised approach applied to stroke lesion segmentation, where we use both labeled and weakly labeled data. Top: For fully labeled data, we use the ground truth label y to train the generator G and discriminator D. Bounding box information is integrated into the framework via the ConvNet bounding box encoder. Bottom: For weakly labeled data, we use the confidence map produced by the trained discriminator to generate a masked region in which the cross-entropy loss is calculated using binarized segmentation map y' as the pseudo label.

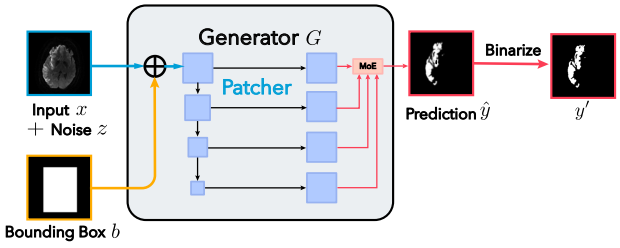

Fig. 2. An overview of the inference process during testing.

3. Experiments

3.1. Experimental Setup

3.1.1. Dataset

The clinical data we use to evaluate our model is provided by the Houston Methodist Hospital, Houston, Texas. 99 cases are fully labeled data, with pixel-wise segmentation label maps as the annotation. These cases are acute ischemic stroke cases sampled from a larger set of cases. The 99 sampled cases have large (n = 42) and small (n = 57) infarct sizes, and have an equal distribution of samples from stroke with the left or right middle cerebral artery (MCA), posterior cerebral artery (PCA), and anterior cerebral artery (ACA) origins. The cases contain a mix of 1.5T and 3.0T scans. Certain cases even have a mix of MCA and ACA in a single scan. Besides, there are cortical stroke and subcortical stroke, and acute and subacute ischemia represented in the cases. The acute and subacute ischemic infarcts are manually segmented by three experts based on the diffusion-weighted imaging (DWI) and the exponential apparent diffusion map (eADC). We plan to release this set of 99 fully labeled cases so that other researchers can evaluate their algorithms on this dataset and compare with ours.

Besides the 99 cases of fully labeled data, we have 831 weakly labeled cases, where we collected 3D bounding boxes for all lesions. In our experiments, we use all 831 weakly labeled cases (20,663 slices) and 67 of the 99 fully labeled cases (1,652 slices) for training. We then use the remaining 32 fully labeled cases for validation and testing; 20 (499 slices) of these cases are used for validation and 12 cases (300 slices) are used for testing. The 12 cases used for testing were carefully chosen to make sure the stroke size, location, and type are nicely balanced in the testing set.

3.1.2. Implementation Details

Our implementation is using PyTorch (Paszke et al., 2019). All experiments are conducted on two NVIDIA RTX 6000 GPUs with 24 GB memory. When training the segmentation network, we use the Stochastic Gradient Descent (SGD) optimization method with a learning rate of 0.01, momentum 0.9, and weight decay $5e-4$. For the discriminator network, the Adam optimizer (Kingma and Ba, 2014) is adopted with a learning rate of $1e-4$ and weight decay of $5e-5$. To ensure that our program runs smoothly on the GPUs, we train our models with a mini-batch of 2 samples per GPU. The first 50 epochs of training are run in a fully supervised mode. After the 50 epochs, we incorporate weakly labeled data into our training, and it converged after 200 epochs.

3.1.3. Baselines and Metrics

We compare our method against three weakly-supervised segmentation baselines: ELN (Kwon and Kwak, 2022), TCSM (Li et al., 2020) and AdvSemiSeg (Hung et al., 2018). We implement baselines on our dataset using publicly released code. For a fair comparison, we also provide a bounding box map as input for all baselines. We use two common evaluation metrics–DSC and IoU–for measuring lesion segmentation accuracy to provide a quantitative comparison on all lesions. Since case-level performance makes more sense in clinical practice and that is what doctors care more about, we report case-level results as well, in addition to slice-level results.

3.1.4. Field Testing using Challenging Cases with Bounding Box Variations

In order to better assess the generalization ability of our method and its robustness to variations in bounding box definition, we further test it on 44 more challenging cases. Two expert users are randomly assigned to cases for which they draw the bounding boxes loosely to include all the stroke lesions and the labeling time per case is recorded, which mimics a real application scenario. Most of these challenging cases include many small and scattered lesions and are low performing with the roto-translation equivariant Group UNet (GUNet) model trained in a fully-supervised manner using 700 fully-labeled cases (Wong et al., 2022). The 700 fully-labeled cases used by the GUNet are from the 831 cases described in Sec. 3.1.1; note that our approach only uses the weak labels (*i.e.*, bounding boxes) of these cases in the proposed weakly-supervised framework while the GUNet (Wong et al., 2022) takes a fully-supervised approach that uses pixel-level labels for the 700 training cases. The 44 testing cases are not in the set of 831 cases. We show segmentation results from our method on the 44 challenging cases and also compare them to results from the GUNet fully-supervised method.

3.2. Results

Table 1 shows a segmentation performance comparison between our method and several other baseline methods. One can see that our weakly supervised method (4th row) outperforms other methods when all baselines are also given the bounding box map as an input channel.

We also perform an ablation study (Table 2) to verify the contributions of individual components of our weakly supervised method. We evaluate the performance of: (1) removing the bounding box map (*i.e.*, a two-channel input excluding the bounding box channel), (2) removing the pseudo label supervision on weakly labeled data (*i.e.*, excluding Eq. 4), (3) removing the discriminator (*i.e.*, excluding adversarial loss Eq. 1), (4) without using the weakly labeled data. Results in the 1st to 4th rows show that the main contributor to performance gain is from bounding box information. Comparing the results shown in the 1st row and 5th row of Table 2, we notice that the dice score (DSC) of our weakly supervised method (using bounding boxes) is 7% higher than without using bounding boxes. Even though the bounding boxes provided in the weakly labeled data are loose around most lesions, it still provides useful information that can be integrated into the generator to improve model performance. In clinical practice, considering the effort needed to redraw segmentation when an algorithm gives inaccurate results, the time investment in providing rough bounding boxes that will lead to a 7% accuracy gain should be worth it. Also, as shown in the 3rd row and 5th row of Table 2, the discriminator is another major contributor to the good performance. It is evident that our adversarial critic design successfully encourages the generator to learn to generate more accurate segmentation.

Fig. 3 shows qualitative testing results by our weakly supervised method. It can be seen that the masks produced by our method (the last column) are the closest to the ground truth. Especially in the third sample (in the 3rd row), our weakly supervised method successfully detected and segmented the small lesions, while the other methods either missed those lesions or did not predict the lesion boundary accurately.

Field testing results on challenging cases are summarized in Table 3. The average time to label a case with bounding boxes is 30 seconds for expert user1 (ten years experience) and 45 seconds for expert user2 (three years experience). We can see that without the help of bounding boxes, both DSC and IoU decrease a lot compared with using the bounding boxes (Line 2 vs. Line 1). Specifically, the DSC and IoU drop 31% and 27% respectively for case-level results. The large difference with and without bounding boxes validates the contribution from bounding box input.

In Table 3, we also show comparison results between our method and the fully-supervised method GUNet (Wong et al., 2022) (Line 3). GUNet is an automatic segmentation model trained using 700 fully labeled cases and it has been put into service in clinical practice. Our model trained on less than 1/10 of the full labels (*i.e.*, using 67 fully-labeled cases) with the bounding box prior is able to reach the performance of the fully supervised GUNet (trained using 700 fully-labeled cases). By requiring much less fully-labeled cases for training, our approach is easier to be adopted by different imaging centers. Experts at various imaging centers can provide bounding box labels more efficiently, and more case variations can be learned inexpensively to improve the performance of the existing model through retraining.

4. Discussion

In our adversarial learning setting, the generators take the concatenation of input image x and random noise z to generate the predicted segmentation map. One may be curious about how the segmentation results would differ given the same input image but different Gaussian noise. To answer this question, we performed 10 testings with the same input image x, but different noise channels z that were generated from 10 random seeds. The mean and standard deviation (std) of segmentation metrics from the 10 testings are summarized in Table 4. One can observe that the DSC and IoU are both stable with very small std, which means our learned models are robust to random noise and can generate stable segmentation results.

Our approach should be very useful in scenarios where there is limited annotation data. Since harmonization of images

Table 1. Segmentation performance comparison between different models

Method	slice-level		case-level	
	DSC (%)	IoU (%)	DSC (%)	IoU (%)
ELN (Kwon and Kwak, 2022)	83.27	78.42	78.95	72.86
TCSM (Li et al., 2020)	84.45	79.09	80.32	73.30
AdvSemiSeg (Hung et al., 2018)	87.33	82.08	82.95	77.01
BB-Guided Segmentor (Ours)	**91.77**	**84.05**	**86.97**	**77.56**

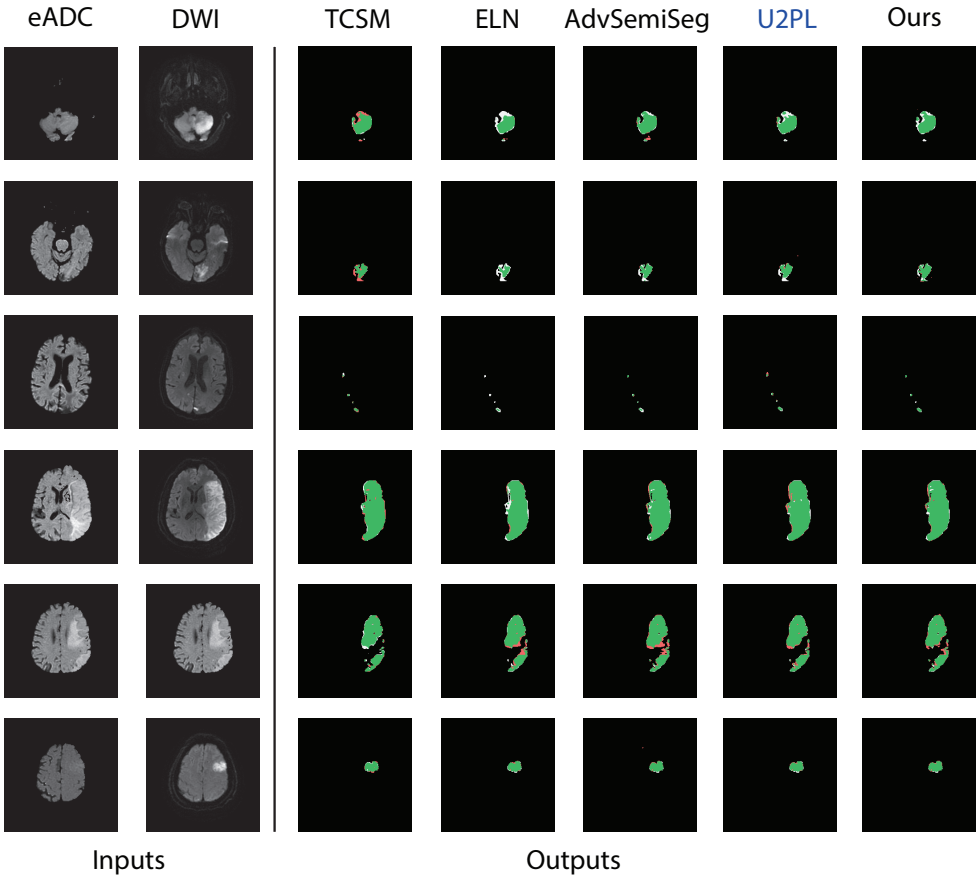

Fig. 3. Visualization of stroke lesion segmentation. We highlight correct predictions (green), false positives (red), and false negatives (white).

Table 2. Ablation study results of our model.

Method	slice-level		case-level	
	DSC (%)	IoU (%)	DSC (%)	IoU (%)
w/o bbox	84.93	82.93	84.95	74.86
w/o pseu	91.36	83.69	86.52	76.84
w/o Discriminator	85.71	82.80	84.65	74.56
w/o $\mathcal{W}$	91.25	83.93	86.93	77.55
BB-Guided Segmentor (Ours)	**91.77**	**84.05**	**86.97**	**77.56**

Table 3. Field testing results on challenging cases.

Method	slice-level		case-level	
	DSC (%)	IoU (%)	DSC (%)	IoU (%)
w/o bbox	25.19	19.27	30.30	21.77
w/bbox	58.29	46.38	61.89	48.21
GUNet (Wong et al., 2022)	59.37	49.67	60.95	47.49

Table 4. Mean and standard deviation (std) of slice-level segmentation metrics from 10 testings with different random noise z for generators.

Measurement	DSC	IoU
mean	91.77	83.73
std	3.09e-4	2.98e-4

across imaging centers remains an open challenge, in an imaging center with limited resources, our approach can be adopted where clinicians only need to fully label a small number of cases imaged at that center, and then quickly provide bounding box weak labels during production phase to get highly accurate segmentation results. Only requiring rough bounding box weak labels, our method can achieve performance similar to a fully-supervised method that requires many more fully-labeled cases.

One direction for future work relates to deep domain adaptation (DDA). The idea is to adapt our models trained on MRI images to segmenting stroke lesions from non-contrast CT images. Considering that most US emergency departments do not have MRI available for acute stroke (Birenbaum et al., 2011) while CT is readily available, adapting our methodology and trained models to segmentation from CT through domain adaptation can make the framework more widely applicable and also reduce the need to collect a large number of annotated CT scans to train CT segmentation models.

5. Conclusion

We introduced an adversarial learning framework with bounding-box based prior for ischemic stroke lesion segmentation. Our novel weakly supervised segmentation model incorporates a bounding box prior that enhances the segmentation network using weak labels, *i.e.*, loose bounding boxes around stroke lesions. We proposed an adversarial learning framework, which leverages a trained discriminator network to provide feedback to the segmentor (or generator) network based on confidence of the predicted segmentation at the pixel-level. Our experiments on clinical datasets have shown that both the bounding box prior and the adversarial learning framework significantly improve segmentation accuracy. We hope the use of bounding box-guided segmentation in our work can provide a new perspective on weakly supervised architectures for medical imaging especially in model deployment stage to increase model performance by leveraging large number of bounding boxes from rare occurring cases. We plan to extend our framework to more practical scenarios, including making it robust to spurious bounding boxes and generalizing it to other imaging modalities such as CT. Our work may have implications for other areas of medical image analysis that rely on weakly supervised learning, and we hope our approach will inspire further research.

References

Abramova, V., Clèrigues, A., Quiles, A., Figueredo, D.G., Silva, Y., Pedraza, S., Oliver, A., Lladó, X., 2021. Hemorrhagic stroke lesion segmentation using a 3d u-net with squeeze-and-excitation blocks. Computerized Medical Imaging and Graphics 90, 101908.

Adamson, J., Beswick, A., Ebrahim, S., 2004. Is stroke the most common cause of disability? Journal of Stroke and Cerebrovascular Diseases 13, 171–177.

Birenbaum, D., Bancroft, L.W., Felsberg, G.J., 2011. Imaging in acute stroke. Western Journal of Emergency Medicine 12, 67–76.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y., 2021. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 .

Crisostomo, R.A., Garcia, M.M., Tong, D.C., 2003. Detection of diffusion-weighted MRI abnormalities in patients with transient ischemic attack: correlation with clinical characteristics. Stroke 34, 932–937.

Dolz, J., Desrosiers, C., Wang, L., Yuan, J., Shen, D., Ayed, I.B., 2020. Deep cnn ensembles and suggestive annotations for infant brain mri segmentation. Computerized Medical Imaging and Graphics 79, 101660.

Han, S., Hwang, S.I., Lee, H.J., 2020. A weak and semi-supervised segmentation method for prostate cancer in trus images. Journal of Digital Imaging 33, 838–845.

Hsu, C.C., Hsu, K.J., Tsai, C.C., Lin, Y.Y., Chuang, Y.Y., 2019. Weakly supervised instance segmentation using the bounding box tightness prior, in: Advances in Neural Information Processing Systems, pp. 6582–6593.

Hung, W.C., Tsai, Y.H., Liou, Y.T., Lin, Y.Y., Yang, M.H., 2018. Adversarial learning for semi-supervised semantic segmentation. arXiv preprint arXiv:1802.07934 .

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

Kirillov, A., He, K., Girshick, R., Rother, C., Dollár, P., 2019. Panoptic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9404–9413.

Kwon, D., Kwak, S., 2022. Semi-supervised semantic segmentation with error localization network. arXiv preprint arXiv:2204.02078 .

Lee, J., Yi, J., Shin, C., Yoon, S., 2021. Bbam: Bounding box attribution map for weakly supervised semantic and instance segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2643–2652.

Li, X., Yu, L., Chen, H., Fu, C.W., Xing, L., Heng, P.A., 2020. Transformation-consistent self-ensembling model for semisupervised medical image segmentation. IEEE Transactions on Neural Networks and Learning Systems 32, 523–534.

Liu, J., Li, M., Wang, J., Wu, F., Liu, T., Pan, Y., 2014. A survey of mri-based brain tumor segmentation methods. Tsinghua Science and Technology 19, 578–595.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440.

Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., Martel, A.L., 2021. Loss odyssey in medical image segmentation. Medical Image Analysis 71, 102035.

Mirza, M., Osindero, S., 2014. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 .

Ou, Y., Yuan, Y., Huang, X., Wong, S.T.C., Volpi, J., Wang, J.Z., Wong, K., 2022. Patcher: Patch transformers with mixture of experts for precise medical image segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 475–484.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. PyTorch: An imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems, pp. 8024–8035.

Qi, K., Yang, H., Li, C., Liu, Z., Wang, M., Liu, Q., Wang, S., 2019. X-Net: Brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 247–255.

Remez, T., Huang, J., Brown, M., 2018. Learning to segment via cut-and-paste, in: Proceedings of the European Conference on Computer Vision, pp. 37–52.

Silva-Rodríguez, J., Colomer, A., Naranjo, V., 2021. Weglenet: A weakly-supervised convolutional neural network for the semantic segmentation of gleason grades in prostate histology images. Computerized Medical Imaging and Graphics 88, 101846.

Tsao, C.W., Aday, A.W., Almarzooq, Z.I., Alonso, A., Beaton, A.Z., Bittencourt, M.S., Boehme, A.K., Buxton, A.E., Carson, A.P., Commodore-Mensah, Y., et al., 2022. Heart disease and stroke statistics—2022 update: A report from the American Heart Association. Circulation 145, e153–e639.

Wang, G., Li, W., Zuluaga, M.A., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., et al., 2018. Interactive medical image segmentation using deep learning with image-specific fine tuning. IEEE Transactions on Medical Imaging 37, 1562–1573.

Wong, K.K., Cummock, J.S., Li, G., Ghosh, R., Xu, P., Volpi, J.J., Wong, S.T., 2022. Automatic segmentation in acute ischemic stroke: Prognostic significance of topological stroke volumes on stroke outcome. Stroke , 10–1161.

Woo, I., Lee, A., Jung, S.C., Lee, H., Kim, N., Cho, S.J., Kim, D., Lee, J., Sunwoo, L., Kang, D.W., 2019. Fully automatic segmentation of acute ischemic lesions on diffusion-weighted imaging using convolutional neural networks: comparison with conventional algorithms. Korean Journal of Radiology 20, 1275–1284.

Wu, O., Winzeck, S., Giese, A.K., Hancock, B.L., Etherton, M.R., Bouts, M.J., Donahue, K., Schirmer, M.D., Irie, R.E., Mocking, S.J., et al., 2019. Big data approaches to phenotyping acute ischemic stroke using automated lesion segmentation of multi-center magnetic resonance imaging data. Stroke 50, 1734–1741.

Xu, J., Murphy, S.L., Kochanek, K.D., Arias, E., 2021. Deaths: Final data for

2019. National Vital Statistics Reports 70.

Yu, H., Zhou, Y., Qian, H., Xian, M., Wang, S., 2017. Loosecut: Interactive image segmentation with loosely bounded boxes, in: Proceedings of the IEEE International Conference on Image Processing (ICIP), IEEE. pp. 3335–3339.

Zhang, D., Song, K., Xu, J., Dong, H., Yan, Y., 2022. An image-level weakly supervised segmentation method for no-service rail surface defect with size prior. Mechanical Systems and Signal Processing 165, 108334.

Zhang, L., Song, R., Wang, Y., Zhu, C., Liu, J., Yang, J., Liu, L., 2020a. Ischemic stroke lesion segmentation using multi-plane information fusion. IEEE Access 8, 45715–45725.

Zhang, S., Liew, J.H., Wei, Y., Wei, S., Zhao, Y., 2020b. Interactive object segmentation with inside-outside guidance, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12234–12244.

Zhou, Y., Huang, W., Dong, P., Xia, Y., Wang, S., 2019. D-unet: a dimension-fusion u shape network for chronic stroke lesion segmentation. IEEE/ACM transactions on computational biology and bioinformatics 18, 940–950.