Scribble-based Domain Adaptation via Co-segmentation

Reuben Dorent[1], Samuel Joutard[1], Jonathan Shapey[1,2,3], Sotirios Bisdas[3], Neil Kitchen[3], Robert Bradford[3], Shakeel Saeed[4], Marc Modat[1], Sébastien Ourselin[1] and Tom Vercauteren[1]

[1] School of Biomedical Engineering and Imaging Sciences, King's College London
reuben.dorent@kcl.ac.uk
[2] Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London
[3] National Hospital for Neurology and Neurosurgery, London
[4] UCL Ear Institute, University College London

Abstract. Although deep convolutional networks have reached state-of-the-art performance in many medical image segmentation tasks, they have typically demonstrated poor generalisation capability. To be able to generalise from one domain (e.g. one imaging modality) to another, domain adaptation has to be performed. While supervised methods may lead to good performance, they require to fully annotate additional data which may not be an option in practice. In contrast, unsupervised methods don't need additional annotations but are usually unstable and hard to train. In this work, we propose a novel weakly-supervised method. Instead of requiring detailed but time-consuming annotations, scribbles on the target domain are used to perform domain adaptation. This paper introduces a new formulation of domain adaptation based on structured learning and co-segmentation. Our method is easy to train, thanks to the introduction of a regularised loss. The framework is validated on Vestibular Schwannoma segmentation (T1 to T2 scans). Our proposed method outperforms unsupervised approaches and achieves comparable performance to a fully-supervised approach.

Keywords: Domain Adaptation · Weak supervision · Regularised loss

1 Introduction

Deep Neural Networks (DNNs) are achieving state-of-the-art performance for many medical image segmentation tasks. However, deep networks still lack in their generalisation capability when confronted with new datasets.

Domain adaptation (DA) approaches have been developed to ensure that networks trained on a source domain can be successfully used on a target domain. A first supervised solution consists of annotating (a sufficient number of) new images from the target domain and fine-tune a network initially trained on the source domain [8, 15]. Although easy to implement, stable during training,

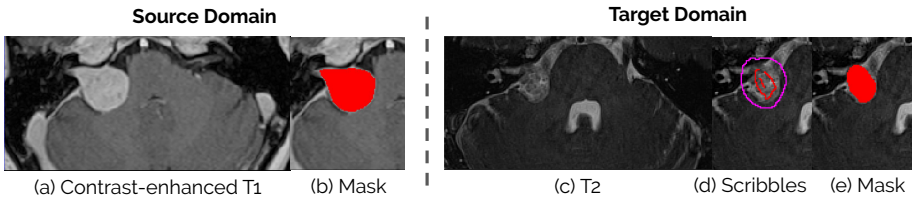

Fig. 1. Examples of Vestibular Schwanomma tumours. T1-c (a) and T2 (c) scans from the source and target domain are shown with their segmentation (b+e). Source masks (b) and target scribbles (e) are used at training stage.

and achieving satisfying performance, such supervised techniques may not be a practical option given the time and expertise required to manually segment additional medical images. For this reason, unsupervised methods, based on self-supervised learning and adversarial learning, have been proposed. Self-supervised techniques [18, 20] typically use pretext tasks to learn task-agnostic feature representations that are adapted to the target domain. Example of self-supervision includes optimising for prediction consistency across different strongly augmented versions of the same target data [18]. Although these techniques have shown promising results, they have only been tested on relatively similar source and target domain. Alternatively or concurrently, adversarial learning has been used to ensure that the learned feature representations are similar across the two domains via a discriminator network [23, 7, 18, 13, 6, 17]. Relying on a complex and unstable adversarial optimisation procedure based on many heuristics, successfully training these models is particularly challenging and time-consuming. Moreover, they are often limited to 2D models due to high memory requirements.

In parallel, efforts have been done to help clinicians segment medical images more efficiently. In particular, semi-automated segmentation has been shown to be a reliable option [25]. Based on efficient user interactions such as scribbles, DNN predictions are fine-tuned at an image-specific level [24]. Fine-tuning is performed for each new test image and is typically *forgotten* on purpose afterwards as the image-specific nature implies a poorer generalisation capability. Looking beyond single images to streamline the annotation task, weakly-supervised methods based on scribbles have been introduced. Networks trained using scribbles are used to perform inference on unseen and unlabelled data. A standard modelling approach is to rely on Conditional Random Fields (CRFs) with DNN outputs being used as unary weights [16, 10, 4]. The optimisation procedure typically alternates between proposing a one-hot crisp segmentation proposal extending from the scribbles (e.g. via a mean-field or graph-cut approach) and training the DNN with supervision provided by these proposals. A recent work [22] has shown that this two-step alternate optimisation can be efficiently approximated by a direct loss minimisation problem exploiting a regularised loss formulation.

In this work, we propose a novel weakly-supervised domain adaptation method. The contributions of this work are four-fold. First, we introduce a new formula-

tion of domain adaptation as a co-segmentation problem. Secondly, we present a new structured learning approach to propagate information across domains. Thirdly, we show that alternating the proposal generation and network training can be approximated by directly minimising a regularised loss. Fourthly, we evaluate our framework on a challenging problem, unpaired cross-modality domain adaptation. Our method demonstrates the benefits of leveraging source data and obtained similar results compared to a fully-supervised approach.

2 Conditional Random Fields for Structured Predictions

In this section, we briefly present Conditional Random Fields (CRFs) for semantic segmentation and define some notations used in the remainder of this work. CRFs have been commonly used in image segmentation for their ability to produce structured predictions.

Let $\mathbf{Y}$ be the random variable representing the overall label assignment, i.e. the segmentation, of a random image $\mathbf{Y} \in \mathbb{R}^N$, where N is the number of voxels. For each voxel k, Y_k is an element of the set of C possible classes $\mathcal{L} = \{l_1, \dots, l_C\}$. The general idea of CRFs is to model the pair $(\mathbf{X}, \mathbf{Y})$ as a graph where the nodes (i.e. voxels) are associated with voxel-wise labels and the edges are associated with the similarity between the voxels. Specifically, a CRF is characterised by a Gibbs distribution $P(\mathbf{Y} = \hat{\mathbf{y}}|\mathbf{X}) \propto \exp(-E_I(\hat{\mathbf{y}}|\mathbf{X}))$. Here $E_I(\hat{\mathbf{y}}|\mathbf{X})$ is the Gibbs energy and represents the cost associated to the label configuration $\hat{\mathbf{y}} \in \mathcal{L}^N$. Given an observed image $\mathbf{x}$, the optimal segmentation $\hat{\mathbf{y}}^*$ minimises the assignment cost. In the fully-connected pairwise CRF model, the problem is defined as:

$$\hat{\mathbf{y}}^* = \underset{\hat{\mathbf{y}} \in \mathcal{L}^N}{\arg\min} \left\{ E_I\left(\hat{\mathbf{y}}|\mathbf{x}\right) = \sum_{k \in [\![1;N]\!]} \psi_u\left(\hat{y_k}|\mathbf{x}\right) + \sum_{k,l \in [\![1;N]\!]} \psi_p\left(\hat{y_k}, \hat{y_l}|\mathbf{x}\right) \right\} \tag{1}$$

where $\psi_u\left(\hat{y_k}|\mathbf{x}\right)$ and $\psi_p\left(\hat{y_k}, \hat{y_l}|\mathbf{x}\right)$ are the unary and pairwise potentials.

Partial annotations, such as scribbles, provide known class values for a subset of voxels. Since each voxel depends on its neighbours, the sparse annotation information can be propagated within the image. Let $\mathbf{y} = (y_i)_{i \in \Omega_a} \in \mathcal{L}^{|\Omega_a|}$ be a partial annotation, where Ω_a is the set of annotated voxels (i.e. the scribbles). The optimisation problem then becomes a constrained one:

$$\begin{aligned} \hat{\mathbf{y}}^* = \underset{\hat{\mathbf{y}} \in \mathcal{L}^N}{\arg\min} & \left\{ \sum_{k \in [\![1;N]\!]} \psi_u\left(\hat{y_k}|\mathbf{x}\right) + \sum_{k,l \in [\![1;N]\!]} \psi_p\left(\hat{y_k}, \hat{y_l}|\mathbf{x}\right) \right\} \\ \text{subject to} \ : \quad & \forall k \in \Omega_a, \ \hat{y_k} = y_k \end{aligned} \tag{2}$$

The problem is typically solved by graph-cut [3] for submodular problems or mean-field inference [14, 2] for the general case.

Recent works have proposed to combine the strengths of deep learning and structured learning via CRFs [24, 16, 10, 4]. The common idea consists in defining the unary potentials ψ_u with a neural network f_θ parameterised by the weights

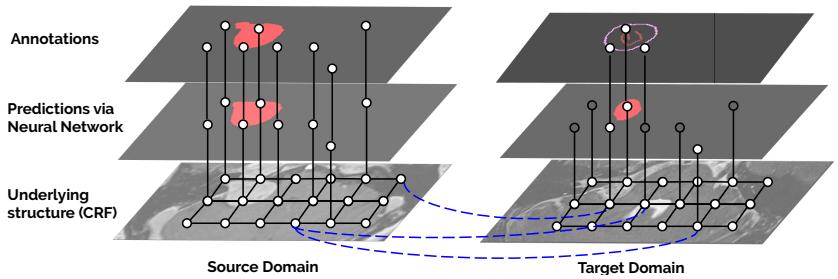

Fig. 2. Overview of the proposed graphical model. Each image voxel is a node. Annotations impose constraints on the predictions using a neural network f_θ. All the nodes are connected together within each image (image-specific CRF) and between images (domain adaptation; blue dashed lines). Only a few of these connections are shown. Although only two images are represented, all the images are connected to each other within and between domains.

θ. Existing methods typically alternate between proposal generation, i.e. solving (2), and network parameters learning with supervision from these proposals. Recently this alternate optimisation has been replaced by a direct optimisation via a regularised loss [22], thereby reducing the optimisation complexity, the computational cost during training and at inference, and avoiding learning from synthetically generated labels. The formulation in [22] reads:

$$\arg\min_{\theta} \Big\{ \sum_{k \in \Omega_a} H(y_k, p_k) + R(\mathbf{p}_\theta) \Big\} \tag{3}$$

where $\mathbf{p}_\theta = f_\theta(\mathbf{x})$ is the softmax output of the network, H is the cross-entropy and R is a regularisation term that encourages spatial and image intensity consistency. In the next section, we provide more details about this regularisation term and we show how we adapt it for domain adaptation purposes.

3 The Scribble Domain Adaptation Model

Group co-segmentation formulation In weakly-supervised domain adaptation, we are given a source domain $\mathcal{D}_s = \{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}$ of n_s fully-labelled samples and a target domain $\mathcal{D}_t = \{(\mathbf{x}_i^t, \mathbf{y}_i^t)\}$ of n_t partially annotated samples. We denote $\Omega_{i,a}$ the set of annotated voxels for each image $\mathbf{x}_i^s$ ($\Omega_{i,a}$ representing the entire image) or $\mathbf{x}_i^t$ ($\Omega_{i,a}$ representing scribbles). Figure 1 shows an example of scribbles used in this work.

The overall objective is to predict accurate segmentation for the target data using a neural network f_θ. Since the annotations are partial on the target domain, we use a graphical model (a CRF) to include prior contextual information and perform structured predictions. This allows for propagating the partial annotation information within a particular image. Given data from the target domain, we aim to minimise each image-specific Gibbs energy E_I, as defined in (1).

However, this basic formulation does not include the other important source of information we have access to: The fully-annotated data from the source domain.

Inspired by co-segmentation [11, 9], we extend the image-specific CRF to a dataset-level CRF. Specifically, in addition to including typical image-specific pairwise potentials, each node (i.e. voxel) of each image is connected to every nodes of every other images, as shown in Figure 2. The annotation information is then propagated between images, including from the fully-annotated images to the partially-annotated images. Consequently, knowledge is transferred from the source domain to the target domain, i.e. domain adaptation is performed. For this reason, we denote E_{DA} the proposed energy term associated to pairs of images. Our proposed optimisation problem can be defined as follows:

$$\begin{aligned} &\underset{\theta,(\hat{\mathbf{y}}_\mathbf{i})_i \in \mathcal{L}^{S\times N}}{\arg\min} \Big\{ \sum_{i\in[\![1;S]\!]} \big(E_I(\hat{\mathbf{y}}_\mathbf{i}|\mathbf{x}_\mathbf{i}) + \sum_{j\in[\![1;S]\!], j\neq i} E_{DA}(\hat{\mathbf{y}}_\mathbf{i}, \hat{\mathbf{y}}_\mathbf{j}|\mathbf{x}_\mathbf{i}, \mathbf{x}_\mathbf{j})\big)\Big\} \\ &\text{subject to } : \quad \forall i \in [\![1;S]\!],\ \forall k \in \Omega_a,\ \hat{y}_{i,k} = y_{i,k} \end{aligned} \tag{4}$$

where $S = n_s + n_t$ is the total number of scans, and indices i, j correspond to image index while k, l are voxels index. Note that the constraints impose that the proposals for the source training data are exactly their fully-annotated masks.

Image-specific Gibbs energy We used a standard formulation of the unary and pairwise potentials for the image-specific energy E_I defined in (2). Similarly to [16, 24, 4], the DNN f_θ is used to compute the unary potentials:

$$\forall k \in [\![1;N]\!], \quad \psi_u(\hat{y}_{i,k}|\mathbf{x}_\mathbf{i}) = -\log P_\theta(\hat{y}_{i,k}|\mathbf{x}_\mathbf{i}) = H(\hat{y}_{i,k}, p_{i,k;\theta}) \tag{5}$$

where $\mathbf{p}_{\mathbf{i};\theta} = f_\theta(\mathbf{x}_\mathbf{i})$ is the probability given by softmax output of the DNN and H the cross-entropy. For the image-specific pairwise potentials, we follow the typical choice of using the Potts model and a bilateral filtering term:

$$\forall k,l \in [\![1;N]\!],\ \psi_p\left(\hat{y}_{i,k}, \hat{y}_{i,l}|\mathbf{x}\right) = [\hat{y}_{i,k} \neq \hat{y}_{i,l}] \exp\Big(-\frac{(x_{i,k} - x_{i,l})^2}{2\sigma_\alpha^2} - \frac{d(x_{i,k}, x_{i,l})^2}{2\sigma_\beta^2}\Big)$$

where $d(.,.)$ denotes the Euclidean distance between the pixel locations. By denoting W_i the affinity matrix of an image $\mathbf{x}_\mathbf{i}$ [22], E_i can be relaxed as:

$$\sum_{k,l\in[\![1;N]\!]} \psi_p\left(\hat{y}_{i,k}, \hat{y}_{i,l}|\mathbf{x}_\mathbf{i}\right) = \hat{\mathbf{y}}_\mathbf{i}{}^T W_i(1 - \hat{\mathbf{y}}_\mathbf{i}) \triangleq R_I(\hat{\mathbf{y}}_\mathbf{i}) \tag{6}$$

Domain adaptation Gibbs energy The DA Gibbs energy E_{DA} only involves pairwise potential associated with voxels from different images. By minimising E_{DA}, we expect to assign similar labels to voxels with similar visual features representation across the datasets. Since the domains are shifted, the image intensity distributions are different between the two domains. Consequently the

image intensity cannot be used as features. Instead, we propose to use the features extracted from the DNN. Specifically, we used the output of the penultimate convolution, i.e. just before the softmax regression. The domain adaptation cost is then defined as:

$$E_{DA}(\hat{\mathbf{y}}_\mathbf{i}, \hat{\mathbf{y}}_\mathbf{j} | \mathbf{x}_\mathbf{i}, \mathbf{x}_\mathbf{j}) = \sum_{k,l \in [\![1;N]\!]} [\hat{y}_{i,k} \neq \hat{y}_{j,l}] \exp\Big(- \frac{(g_{i,k} - g_{j,l})^2}{2\sigma_\gamma^2} \Big) \tag{7}$$

where $\mathbf{g}_i = g_\theta(\mathbf{x}_i)$. Note that the spatial position is not taken into account here. Again, the domain adaptation Gibbs energy can be relaxed as:

$$E_{DA}(\hat{\mathbf{y}}_\mathbf{i}, \hat{\mathbf{y}}_\mathbf{j} | \mathbf{x}_\mathbf{i}, \mathbf{x}_\mathbf{j}) = [\hat{\mathbf{y}}_\mathbf{i}, \hat{\mathbf{y}}_\mathbf{j}]^T W_{i-j} (1 - [\hat{\mathbf{y}}_\mathbf{i}, \hat{\mathbf{y}}_\mathbf{j}]) \triangleq R_{DA}(\hat{\mathbf{y}}_\mathbf{i}, \hat{\mathbf{y}}_\mathbf{j}; \theta) \tag{8}$$

4 Optimization via a Regularised Loss

In this section, we propose a method to optimise the parameters θ of the DNN. Similarly to [22], we show that the optimization problem can be approximated with a regularised loss. Let $\mathbf{p}^\mathbf{t}_{\mathbf{i};\theta} = f_\theta(\mathbf{x}^\mathbf{t}_\mathbf{i})$ and $\mathbf{p}^\mathbf{s}_{\mathbf{i};\theta} = f_\theta(\mathbf{x}^\mathbf{s}_\mathbf{i})$ be the outputs of the network for a target domain image $\mathbf{x}^\mathbf{t}_\mathbf{i}$ and a source domain image $\mathbf{x}^\mathbf{s}_\mathbf{i}$. We denote $H_{\Omega_a}(\mathbf{u}, \mathbf{v}) = \sum_{k \in \Omega_a} H(u_k, v_k)$. By combining (4), (6) and (8), the optimisation problem is defined as:

$$\begin{aligned} \underset{\theta, (\hat{\mathbf{y}}_\mathbf{i})_{i \in [\![1;n]\!]}}{\arg\min} \Big\{ & \sum_i \big(H_\Omega \left(\hat{\mathbf{y}}_\mathbf{i}{}^\mathbf{s}, \mathbf{p}^\mathbf{s}_{\mathbf{i};\theta} \right) + R_I(\hat{\mathbf{y}}_\mathbf{i}{}^\mathbf{s}) \big) \\ & + \sum_i \big(H_\Omega \left(\hat{\mathbf{y}}_\mathbf{i}{}^\mathbf{t}, \mathbf{p}^\mathbf{t}_{\mathbf{i};\theta} \right) + R_I(\hat{\mathbf{y}}_\mathbf{i}{}^\mathbf{t}) \big) + \sum_{i,j} R_{DA}\big(\hat{\mathbf{y}}_\mathbf{i}, \hat{\mathbf{y}}_\mathbf{j}; \theta\big) \Big\} \\ \text{subject to} \ : \quad & \forall i \in [\![1; S]\!], \ \forall k \in \Omega_a, \ \hat{y}_{i,k} = y_{i,k} \end{aligned} \tag{9}$$

By adding a null negative entropy term $-\sum_i H\left(\hat{\mathbf{y}}_\mathbf{i}{}^\mathbf{t}, \hat{\mathbf{y}}_\mathbf{i}{}^\mathbf{t}\right) = 0$ and integrating the constraints directly in the formulation, (9) can be rewritten as:

$$\underset{\theta}{\arg\min} \Big\{ \sum_{i,j} u(\mathbf{p}_{\mathbf{i},\theta}) + \mathbb{1}_{\mathbf{x}_\mathbf{i} \in \mathcal{D}^t} \min_{\hat{\mathbf{y}}_\mathbf{i}{}^\mathbf{t}} \left\{ KL(\hat{\mathbf{y}}^\mathbf{t}_\mathbf{i}, \mathbf{p}_{\mathbf{i};\theta}) + R(\hat{\mathbf{y}}_\mathbf{i}{}^\mathbf{t}, \hat{\mathbf{y}}_\mathbf{j}{}^\mathbf{t}) \right\} \Big\} \tag{10}$$

where $u(\mathbf{p}_{\mathbf{i},\theta}) = H_{\Omega_{a,i}}(\mathbf{y}_\mathbf{i}, \mathbf{p}_{\mathbf{i},\theta})$, $R(\hat{\mathbf{y}}_\mathbf{i}{}^\mathbf{t}, \hat{\mathbf{y}}^\mathbf{t}_\mathbf{j}) = R_I(\hat{\mathbf{y}}_\mathbf{i}{}^\mathbf{t}) + R_{DA}(\hat{\mathbf{y}}_\mathbf{i}{}^\mathbf{t}, \hat{\mathbf{y}}_\mathbf{j}{}^\mathbf{t})$ and KL denotes the Kullback–Leibler divergence. Given that full annotations are provided for the source domain, the inner minimisation with respect to the proposals, $\hat{\mathbf{y}}^\mathbf{s}_\mathbf{i}$, only relates to the target data $(\mathbf{x}^t_i, \mathbf{y}^t_i)$.

The inner problem corresponds to minimising a divergence between the network output $\mathbf{p}^\mathbf{t}_{\mathbf{i};\theta}$ and the proposal $\hat{\mathbf{y}}_\mathbf{i}{}^\mathbf{t}$ together with a regularisation term. This discrepancy is null if the proposal is equal to the network output. We thus expect the optimal proposal to be close to the network output, i.e. $\hat{\mathbf{y}}_\mathbf{i}{}^{\mathbf{t}*} \approx \mathbf{p}^\mathbf{t}_{\mathbf{i};\theta}$. We assume that equality stands, which allows us to reformulate the problem as:

$$\underset{\theta}{\arg\min} \Big\{ \mathcal{L}(\theta) = \sum_{\substack{(\mathbf{x}_\mathbf{i}, \mathbf{y}_\mathbf{i}) \\ (\mathbf{x}_\mathbf{j}, \mathbf{y}_\mathbf{j})}} H_{\Omega_{a,i}}(\mathbf{y}^\mathbf{t}_\mathbf{i}, \mathbf{p}_{\mathbf{i};\theta}) + \mathbb{1}_{\mathbf{x}_\mathbf{i} \in \mathcal{D}^t} (R_I(\mathbf{p}_{\mathbf{i};\theta}) + R_{DA}(\mathbf{p}_{\mathbf{i};\theta}, \mathbf{p}_{\mathbf{j};\theta})) \Big\} \tag{11}$$

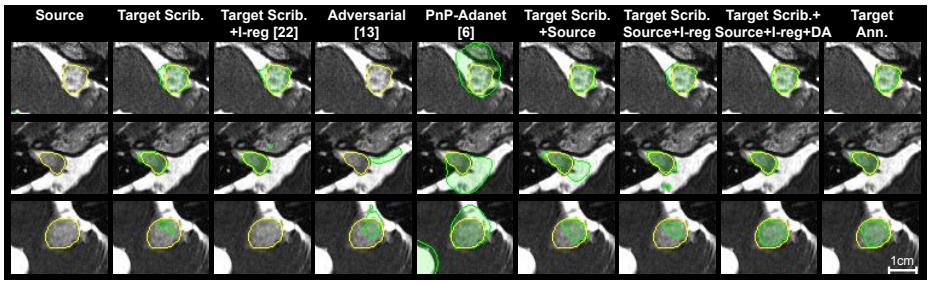

Fig. 3. Qualitative evaluation of different networks for Vestibular Schwannoma segmentation on T2 scans. Segmentation results (green curves) and the ground truth (yellow curves) are shown.

The parameters θ are directly optimised via a stochastic gradient descent. The high-dimensional filtering method proposed by [1] is used to reduce the quadratic complexity of the computation of R_I and R_{DA} to a linear one.

5 Experiments

Experimental setup. We conducted experiments on Vestibular Schwanomma (VS), a benign brain tumour arising from the vestibulocochlear nerve, the main nerve connecting the brain and inner ear. Current MR protocols include contrast-enhanced T1-weighted (T1-c) and high-resolution T2 scans. T1-c are generally currently used for segmenting the tumour as offering a better contrast, see Figure 1. However, T2 imaging could be a reliable, safer and lower-cost alternative to T1-c [5, 21].

In this work, we propose to segment VS images using T2 images only as input. The source domain data corresponds to 150 T1-c scans with the full set of annotations and the target domain training data corresponds to 30 T2 scans with scribble annotations only. Specifically, on average 1% of the T2 scans and 7% of the tumour has been annotated. 4 T2 scans and 20 T1 scans were used as validation set. For testing, 50 T2 scans (target domain) have been manually fully segmented. Images had an in-plane resolution of $0.4 \times 0.4 \times \text{mm}^2$, a slice thickness of $1.0-1.5$mm and were cropped manually with a bounding box of size $100 \times 50 \times 50\text{mm}^3$, covering the full axial brain length as shown in Figures 1a,1c.

Implementation details. Our models were implemented in PyTorch using TorchIO [19] [5]. A 2.5D U-Net was used for all our experiments, similar to [26]. A PyTorch GPU implementation of the high-dimensional filtering [12] was employed. We used the Adam optimizer with weight decay 10^{-5}. At each iteration, two images from the source domain and two images from the target domain are

[5] Code available at: https://github.com/KCL-BMEIS/ScribbleDA

Table 1. Quantitative evaluation of different networks for Vestibular Schwannoma segmentation. I-reg: The image-specific regularised loss proposed by [22]. DA: Our proposed Domain Adaptation regularised loss.

	Test on Source		Tets on Target	
Method / Training	Dice (%)	ASSD (mm^3)	Dice (%)	ASSD (mm^3)
Source	**93.7 (3.3)**	0.3 (0.5)	28.2 (33.0)	13.8 (10.8)
Target Scrib	46.9 (33.8)	10.9 (10.9)	77.6 (17.9)	2.1 (3.0)
Target Scrib+I-reg [22]	58.4 (29.5)	9.0 (8.5)	76.9 (18.8)	1.4 (2.0)
Source+Adversarial [13]	87.8 (8.9)	1.6 (1.5)	9.3 (18.9)	24.9 (16.2)
PnP-Adanet [6]	79.3 (15.2)	3.4 (3.1)	27.3 (21.1)	13.3 (4.2)
Target Scrib+Source	92.4 (4.6)	0.4 (0.4)	75.1 (18.6)	2.7 (4.5)
Target Scrib+Source+I-reg	93.2 (3.7)	0.3 (0.5)	76.7 (17.9)	1.6 (2.5)
Target Scrib+Source+I-reg+DA	93.3 (4.0)	**0.2 (0.2)**	**83.4 (10.4)**	**0.8 (0.8)**
Target Ann.	63.7 (33.9)	8.3 (11.2)	81.6 (13.1)	1.8 (2.8)

randomly selected and fed to the network. The initial learning rate 5.10^{-4} was reduced by a factor of 5 whenever the moving average of the validation loss has not improved in the last 5 epochs and training was stopped after no improvements in the last 10 epochs. Rotation, scaling and white noise augmentation were applied during training.

Concerning the regularisation terms, a typical value of α was chosen (15). Similar results results were obtained for different values of β ($\{0.5, 0.05, 0.005\}$), the ones reported correspond to $\beta = 0.05$. In order to reduce the computational complexity, only two channels were used to compute the pairwise distance in the DA regularisation term. Specifically, at each training iteration, two channels were chosen randomly among the total number of channels (here 48). γ was set up to 0.1. Domain adaptation regularisation was introduced after a few epochs (70). Finally, we observed large improvements by using the Dice loss instead of the cross-entropy, thus we reported scores with the Dice loss.

Model Comparison Firstly, we studied each component of our method independently. As a baseline, we trained a model on the target scribbles only (Target Scrib) and with the regularised loss [22] (Target Scrib+I-reg). Then the source data was used during training without (Target Scrib+I-reg+Source) and with (Target Scrib+I-reg+Source+DA) the cross-modality DA regularisation. Secondly, we compared our method with a fully-supervised approach trained using the same 30 T2 scans with the full set of annotations (Target Ann.). Thirdly, we compared our approach with two well-established unsupervised DA methods based on adversarial learning [13] and designed specifically for cross-modality DA [6]. Quantitative results are reported in Table 1 using the Dice and average symmetric surface distance (ASSD) between segmentation results and the ground truth. Examples of outputs are presented in Figure 3.

Results Firstly, the ablation study shows that adding the cross DA regularisation brings significant improvements on the target domain compared to the other models trained using the target scribbles. Interestingly, including the source data during training only leads to improvements when the DA regularisation is employed. This shows the effectiveness of our DA method. Moreover, note that our technique didn't degrade the performance on the source domain. Secondly, our method obtained comparable performance to a fully-supervised model. Thus, scribble-based DA is a reliable option for performing supervised DA. Thirdly, both unsupervised methods failed on our problem. Since the inner brain and tumour appearance vary greatly between the contrast-enhanced T1 and T2 scans, our problem is too challenging for unsupervised approaches, highlighting the need for supervision.

6 Conclusion

This paper proposes a novel approach for weakly-supervised domain adaptation. Based on co-segmentation and structured learning, we introduced a new formulation for domain adaptation with scribbles. Our approach is mathematically grounded, easy to implement, new and relies on reasonable assumptions. We validated our method on challenging experiments: unpaired cross-modality brain lesion segmentation. Our model achieved comparable performance to a model trained on a fully-annotated data and outperformed existing unsupervised techniques. This work shows that scribbles is a reliable option for performing domain adaptation.

Acknowledgement This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) [NS/A000049/1] and Wellcome Trust [203148/Z/16/Z]. TV is supported by a Medtronic / Royal Academy of Engineering Research Chair [RCSRF1819\7\34].

References

1. Adams, A., Baek, J., Davis, M.A.: Fast High-Dimensional Filtering Using the Permutohedral Lattice. Computer Graphics Forum (2010)
2. Baque, P., Bagautdinov, T., Fleuret, F., Fua, P.: Principled parallel mean-field inference for discrete random fields. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
3. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient n-d image segmentation. Int. J. Comput. Vision **70**(2), 109–131 (2006)
4. Can, Y.B., Chaitanya, K., Mustafa, B., Koch, L.M., Konukoglu, E., Baumgartner, C.F.: Learning to segment medical images with scribble-supervision alone. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. pp. 236–244. Springer International Publishing, Cham (2018)
5. Coelho, D.H., Tang, Y., Suddarth, B., Mamdani, M.: Mri surveillance of vestibular schwannomas without contrast enhancement: Clinical and economic evaluation. The Laryngoscope **128**(1), 202—209 (2018)

6. Dou, Q., Ouyang, C., Chen, C., Chen, H., Glocker, B., Zhuang, X., Heng, P.A.: Pnp-adanet: Plug-and-play adversarial domain adaptation network with a benchmark at cross-modality cardiac segmentation. ArXiv (2018)
7. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-Adversarial Training of Neural Networks, pp. 189–209. Springer International Publishing, Cham (2017)
8. Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., Guttmann, C.R.G., de Leeuw, F.E., Tempany, C.M., van Ginneken, B., Fedorov, A., Abolmaesumi, P., Platel, B., Wells, W.M.: Transfer learning for domain adaptation in mri: Application in brain lesion segmentation. In: MICCAI 2017. pp. 516–524. Springer International Publishing, Cham (2017)
9. Hochbaum, D.S., Singh, V.: An efficient algorithm for co-segmentation. In: 2009 IEEE 12th International Conference on Computer Vision. pp. 269–276 (2009)
10. Ji, Z., Shen, Y., Ma, C., Gao, M.: Scribble-based hierarchical weakly supervised learning for brain tumor segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. pp. 175–183. Springer International Publishing, Cham (2019)
11. Joulin, A., Bach, F., Ponce, J.: Discriminative clustering for image co-segmentation. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 1943–1950 (2010)
12. Joutard, S., Dorent, R., Isaac, A., Ourselin, S., Vercauteren, T., Modat, M.: Permutohedral attention module for efficient non-local neural networks. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. pp. 393–401. Springer International Publishing, Cham (2019)
13. Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V., Simpson, J., Kane, A., Menon, D., Nori, A., Criminisi, A., Rueckert, D., Glocker, B.: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: Information Processing in Medical Imaging. pp. 597–609. Springer International Publishing, Cham (2017)
14. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. In: Advances in Neural Information Processing Systems 24, pp. 109–117. Curran Associates, Inc. (2011)
15. Kushibar, K., Valverde, S., González-Villà, S., Bernal, J., Cabezas, M., Oliver, A., Lladó, X.: Supervised domain adaptation for automatic sub-cortical brain structure segmentation with minimal user interaction. Scientific Reports **9**(1), 6742 (2019)
16. Lin, D., Dai, J., Jia, J., He, K., Sun, J.: Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
17. Mahmood, F., Chen, R., Durr, N.J.: Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. IEEE Transactions on Medical Imaging **37**(12), 2572–2581 (2018)
18. Orbes-Arteaga, M., Varsavsky, T., Sudre, C.H., Eaton-Rosen, Z., Haddow, L.J., Sørensen, L., Nielsen, M., Pai, A., Ourselin, S., Modat, M., Nachev, P., Cardoso, M.J.: Multi-domain adaptation in brain mri through paired consistency and adversarial learning. In: Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data. pp. 54–62. Springer International Publishing, Cham (2019)
19. Pérez-García, F., Sparks, R., Ourselin, S.: TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. arXiv:2003.04696 (2020)

20. Perone, C.S., Ballester, P., Barros, R.C., Cohen-Adad, J.: Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. NeuroImage **194**, 1 – 11 (2019)
21. Shapey, J., Wang, G., Dorent, R., Dimitriadis, A., Li, W., Paddick, I., Kitchen, N., Bisdas, S., Saeed, S.R., Ourselin, S., Bradford, R., Vercauteren, T.: An artificial intelligence framework for automatic segmentation and volumetry of vestibular schwannomas from contrast-enhanced t1-weighted and high-resolution t2-weighted mri. Journal of Neurosurgery JNS pp. 1 – 9 (2019)
22. Tang, M., Perazzi, F., Djelouah, A., Ben Ayed, I., Schroers, C., Boykov, Y.: On regularized losses for weakly-supervised cnn segmentation. In: The European Conference on Computer Vision (ECCV) (2018)
23. Tzeng, E., Hoffman, J., Darrell, T., Saenko, K.: Adversarial discriminative domain adaptation. In: Computer Vision and Pattern Recognition (CVPR) (2017)
24. Wang, G., Li, W., Zuluaga, M.A., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., Vercauteren, T.: Interactive medical image segmentation using deep learning with image-specific fine tuning. IEEE Transactions on Medical Imaging **37**(7), 1562–1573 (2018)
25. Wang, G., Zuluaga, M.A., Li, W., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., Vercauteren, T.: Deepigeos: A deep interactive geodesic framework for medical image segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence **41**(07), 1559–1572 (2019)
26. Wang, G., Shapey, J., Li, W., Dorent, R., Demitriadis, A., Bisdas, S., Paddick, I., Bradford, R., Zhang, S., Ourselin, S., Vercauteren, T.: Automatic segmentation of vestibular schwannoma from t2-weighted mri by deep spatial attention with hardness-weighted loss. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. pp. 264–272. Springer International Publishing, Cham (2019)