Supplementary - CCT-Net: Category-Invariant Cross-Domain Transfer for Medical Single-to-Multiple Disease Diagnosis

Yi Zhou¹, Lei Huang², Tao Zhou³, and Ling Shao⁴

¹School of Computer Science and Engineering, Southeast University, Nanjing, China
²SKLSDE, Institute of Artificial Intelligence, Beihang University, Beijing, China
³School of Computer Science and Technology, Nanjing University of Science and Technology, China
⁴Inception Institute of Artificial Intelligence, Abu Dhabi, UAE

A. Differences with standard domain adaptation and transfer learning.

Our proposed scenario is different from the standard domain adaptation (DA) and transfer learning (TL) task settings. We do not simply classify our method into one of them. However, we want to spotlight this scenario because it happens very commonly in real medical applications, where pixel-level labels are only available for a few disease classes, while we want to exploit these limited data to improve recognition for more diseases both on segmentation and classification. We believe this scenario still falls into the generalized transfer learning field, because the domain knowledge and disease diagnosis ability (particularly for segmentation) is indeed transferred from the source domain to the target domain. Thus, we finalized our method name using the term of cross-domain transfer. Moreover, due to the setting difference, although most TL methods cannot be directly applied to our task, we carefully selected some extendable TL methods (DANN, MCD, ITL) in the classification task for comparison.

B. Comparison to weakly supervised learning methods.

In our scenario, the segmentation task in the target domain can be also considered as a weakly-supervised semantic segmentation problem with available image-level labels. We first report the results by fully-supervised methods in the upper part of Table 1 for reference, since researchers usually consider fully-supervised methods to be the "upperbound" in the performance. Then, we compare our CCTbased weakly-supervised fashion with three state-of-the-art weakly-supervised learning methods. In the lower part of Table 1, CAM denotes the baseline using the class activation map method, which is as same as the w/o-CCT denoted

Table 1. Comparison with weakly-supervised semantic segmentation methods on brain tumor segmentation. The two best results are in **red** and blue. **'w/o'='without'**, **'w'='with'**.

10° m red and blue. 10° = without , $w = with$.								
	Methods		Meningioma		Glioma		Pituitary Tumor	
Target Domain			Dice	Haus.	Dice	Haus.	Dice	Haus.
	Supervised	FCN-8s	0.824	6.17	0.621	11.30	0.785	5.09
		DL_V3+	0.851	5.65	0.638	10.73	0.812	4.73
		U-Net	0.875	5.20	0.647	10.61	0.830	4.49
		Att. U-Net	0.889	4.97	0.665	10.28	0.849	4.15
	ıklySupervised	CAM	0.715	8.55	0.531	15.26	0.660	6.93
		SEC [28]	0.748	7.91	0.554	14.73	0.692	6.50
		DSRG [25]	0.825	6.09	0.607	12.45	0.773	5.24
		IRNet [2]	0.836	5.94	0.625	11.12	0.791	4.88
		CCT-w-CWP_B1	0.873	5.21	0.653	10.55	0.830	4.46
	Nea	CCT-w-CWP_B2	0.883	5.05	0.668	10.22	0.841	4.27

in the original paper. It does not apply CCT but directly resizes the coarse map predictions (binarized by the threshold 0.2) obtained by the CWP-based DSE-1 to the size of the ground truths for evaluation. Seed, Expand and Constrain (SEC [28]) is a good weakly-supervised method, which trains a convolutional neural network (CNN) and applies CAM to produce pseudo ground-truth masks to train a fully convolutional network (FCN) against the generated seeds, against the image-level label, and a constraint loss against the CRF-processed maps. Deep seeded region growing (DSRG [25]) trains a CNN and applies region-growing on the generated CAMs to produce a pseudo ground-truth for training a FCN. Inter-pixel relation network (IRNet [2]) segments individual instances by performing the random walk from low-displacement field centroids in the CAM seeds up until the class boundaries as the pseudo groundtruths for training a FCN. As illustrated in the results, our model performs dominant over other methods.

C. How visually related the diseases should be?

Someone might be interested in how related should be visually for our approach to work well given the source and target domain data having different disease categories. Two reasons make our approach work well. First, most regions between the source and target domain basically share a similar image pattern (small domain shift doesn't matter). For example, an aligned eye region with blood vessels and an optic disk inside and orange background. Although lesion appearances of different diseases may vary, through the CCT, abnormal regions can still be feasibly segmented compared to healthy regions, given a few of classes' lesion masks for training. Second, the CWP-based CAM largely drops healthy regions and predicts a coarse map for each disease already. The CCT doesn't do segmentation from scratch. Moreover, the more classes with pixel-level labels the source domain has, the better the transfer can be done. We tried an extreme case that only 2 classes are available in source, the improvement on segmenting other diseases in target is very limited compared to w/o-CCT.

D. Acknowledgements.

This work was supported by the National Natural Science Foundation of China (62106043), and the Fundamental Research Funds for the Central Universities, China (2242021k30056). We also thank our collaborated physicians who provided us the datasets and annotations.