UNBALANCED GRAPH-BASED TRANSDUCTION ON SUPERPIXELS FOR AUTOMATIC CERVIGRAM IMAGE SEGMENTATION

Sheng Huang[†], Mingchen Gao[‡], Dan Yang[†], Xiaolei Huang[§], Ahmed Elgammal[‡] and Xiaohong Zhang[†]

 [†]Ministry of Education Key Laboratory of Dependable Service Computing in Cyber Physical Society, Chongqing University, Chongqing, P.R.C, 400044
 [‡] Center for Infectious Disease Imaging, Department of Radiology and Imaging Sciences National Institutes of Health (NIH), Bethesda, MD 20892, USA
 [‡]Rutgers University, Piscataway, NJ, 08854, USA
 [§]Lehigh University, Bethlehem, PA, 18015, USA

ABSTRACT

We propose a novel medical image segmentation algorithm by transductively inferring the labels. In this approach, superpixels are first generated to incorporate the local spatial information and also to speed up the segmentation. The segmentation task can be deemed as an unbalanced superpixels labeling problem due to the fact that the region of interest is only a small fraction compared to the whole image. We present a new transductive learning-based algorithm called Class Averaging Graph-based Transduction (CAGT) to avoid the biased labeling caused by the imbalance. The proposed algorithm was applied to the automatic cervigram image segmentation to demonstrate it effectiveness.

Index Terms— Image Segmentation, Transductive Learning, Unbalanced Classification, Graph Learning, Semisupervised Learning

1. INTRODUCTION

Image segmentation plays a crucial role in many medical imaging applications by automatically detecting the regions of interest. For example, the accurate and automatic segmentation of tissue regions in cervigram images is useful to identify the precancerous regions. Specifically speaking, Acetowhite (AW) region, which is caused by the whitening of potentially malignant regions of the cervix epithelium is the most important observed area to segment. Accurate identification and segmentation of AW regions in cervigrams have significant implications for diagnosis and grading of cervical lesions. While the task of segmenting this area is important, it is particularly challenging due to high variability where tissue color distributions frequently overlap with different classes.

Accurate segmentation of medical images has many challenges, such as the common presence of image noise, nonuni-



Fig. 1. The flow diagram of our segmentation approach.

form object texture, variations in lighting, and various other artifacts in medical images. To address these challenges, many segmentation techniques have been developed, such as deformable models, level set, statistical shape, appearance model and graph-based segmentation [1]. Essentially, the image segmentation issue is a pixel labeling problem. However, there are several specific characteristics of the medical image segmentation task that makes it unique in comparison with the regular labeling problem: 1) The training data is limited, since manual labeling of the ground truth by experts is very time consuming and uneconomical; 2) The pixels of the image can be input as a batch; and 3) The positive and negative pixels are often highly unbalanced, since the region of interest often only represents a small fraction. For example, the number of negatives in our used cervigram images is twenty times more than the number of positives. Clearly, an ideal labelling algorithm for tackling the medical segmentation task should accommodate these three characteristics. Transductive learning is a good option for such medical image segmentation-based labeling, since it can accommodate the first two characteristics [2]. The main assumption behind the transductive learning model is that points that share more similarities are likely to have the same label. Although it already has been proved to be successful in various applications, it is still not wide-

This work has been partially funded by Fundamental Research Funds for the Central Universities (Grant No. CDJXS11181162) and the National Natural Science Foundation of China (Grant No. 91118005, 61173131 and 61402062).

ly used in medical image segmentation. As far as we know, the transductive learning only has been applied for prostate segmentation [3].

In order to accommodate the three aforementioned characteristics, we improve the conventional Graph-based Transduction (GT) algorithm as a new unbalanced transduction algorithm called Class Averaging Graph-based Transduction (CAGT). In CAGT, the classification loss of GT is averaged by classes to avoid the sample is labeled biased towards to the label with a high proportion. Moreover, Simple Linear Iterative Clustering (SLIC) [4] is applied to efficiently generate the superpixels from image for incorporating the local spatial information and speeding up the segmentation. Finally, the segmentation can be achieved by the CAGT-based superpixel labeling. The framework of our approach is illustrated in Figure 1. An archive of 60,000 images of the uterine cervix [5, 6] is employed for evaluating our work. The experimental results demonstrate the superiority of the proposed algorithm in comparison with the baseline methods.

We conclude three contributions of our works: as far as we know, our work is the first approach that applies the graphbased transduction to the cervigram image segmentation; We present a novel graph-based transduction algorithm for addressing the unbalanced labeling issue; Superpixel technique is leveraged for speeding up and incorporating the local spatial information.

2. METHODOLOGY

We present an image segmentation approach that segments the medical image via labeling its superpixels. Figure 1 illustrates our framework. It has three steps. The first step is the efficient generation of the superpixels using Simple Linear Iterative Clustering (SLIC) [4]. The second step is the representation of superpixel. The third step is the unbalanced binary transduction for superpixel labelling.

2.1. Superpixel Generation and Representation

SLIC, which is a recently popular superpixel generation algorithm in computer vision community, is adopted for superpixel generation [4]. It efficiently generates the high quality superpixels via simply performing the k-means clustering based on the location and color information of pixels. We consider the mean LAB color vector of pixels in the same superpixel as the representation of superpixel.

$$x_i = \sum_{t \in i} \frac{p_t}{n_i} \tag{1}$$

where x_i is the representation of superpixel *i*, p_t is the LAB color vector of the pixel *t* and n_i is the number of pixels in superpixel *i*.

In the training dataset, we only have the labels of pixels while the labels of superpixels are unknown. We label these superpixels by computing the proportions of positive and negative pixels in the superpixels. Let us consider the positive label as 1 while the negative label as 0. The label assignment of the superpixel i can be denoted as follows

$$l_i = \begin{cases} 1, \ \mathcal{P}_i > \mathcal{P}_{all} \\ 0, \ otherwise \end{cases}$$
(2)

where \mathcal{P}_i is the proportion of the positive pixels in the superpixel *i* and \mathcal{P}_{all} is the one of the whole training dataset. This strategy can avoid the biased labeling when the data is extremely unbalanced.

There are two main reasons for us to use superpixels instead of pixels. The first reason is that it can greatly speed up the segmentation, since the computational complexity of the subsequent step is cubic with respect to the number of samples however SLIC can efficiently reduce this number. The second reason is that the superpixels can improve the robustness of our model, since the generation of superpixel incorporates the local spatial information which can benefit the segmentation particular when the ground truth is not perfectly labeled.

2.2. Unbalanced Binary Transduction

Now, the segmentation task is translated as a superpixel labeling task. As we have mentioned in Section 1, there are three characteristics of the medical image segmentations specific labeling problem. A good suggestion for such labeling is Graph-based Transduction (GT) [2], since it performs well in the small training data case and can jointly label the samples via fully considering the unlabeled data. In other words, it can perfectly accommodate the first two characteristics. In this section, we intend to propose a new GT-based labeling algorithm for accommodating the third characteristic that the data is highly unbalanced.

Given the superpixels $\{x_t\}_{t \in i}$ of the image i, we will achieve the segmentation by performing the Graph-based Transduction (GT). The main idea of the graph-based transduction is to transductively infer the labels of the superpixels based on the similarities between each two superpixels. Therefore, the core of the GT is the construction of the graph Laplacian which encodes such similarities. Here we leverage the heat kernel scheme [7] to measure this similarity between each two superpixels $w_{ij} = \exp\left(-\frac{||x_i - x_j||^2}{\sigma}\right)$ where σ is a positive for scaling the Euclidean distance between the superpixels x_i and x_j . w_{ij} is regarded as the (i,j)-th element of the affinity matrix W. Note, the affinity matrix is constructed from the superpixels of both training images and the target segmentation image, since the unlabeled superpixels from the same image are jointly labeled.

According to the Laplacian Eigenmapping [8], the normalized graph Laplacian can be computed as follows

$$L = D^{-1/2} (D - W) D^{-1/2} = I - D^{-1/2} W D^{-1/2}$$
(3)

where D is a diagonal matrix and $D_{ii} = \sum_j w_{ij}$. I is the identity matrix.

The graph-based transductive label inference can be naively deemed as a regularized normalize cut in which the collection of graph cuts are defined as the classifiers [2, 9]. Such graph cuts should not only minimize the loss of the similarities among superpixels but also minimize the classification loss of training data.

$$\hat{F} = \arg\min_{F} (F^{T}LF + \lambda ||F - Y||^{2})$$

$$\Rightarrow \quad \hat{f}_{i} = \arg\min_{f_{i}} \sum_{i \in \{pos, neg\}} (f_{i}^{T}Lf_{i} + \lambda ||f_{i} - y_{i}||^{2})$$
(4)

where $F = [f_{pos}, f_{neg}]$ is the collection of graph cuts (sub-classifier functions, $f_i(\cdot), i \in \{pos, neg\}$) and $Y = [y_{pos}, y_{neg}]$ is the collection of label vectors (label functions, $y_i(\cdot), i \in \{pos, neg\}$). λ is a positive to reconcile these two losses. In $y_{pos}, y_{pos}(i) = 1$ or -1 if the superpixel feature x_i has been labeled as positive or negative respectively, and 0 if it is unlabeled. In y_{neg} , the assignment of the $y_{neg}(i)$ is the inversion of the assignment of the $y_{pos}(i)$.

According to Equation 4, we can know that such regular graph-based transduction model measures the classification loss by sample. However, the superpixel labeling task can be considered as a typical unbalanced classification task, since the number of negative samples highly exceeds the one of positives, for example, in our case, the negative samples are more than twenty times of the positives. Clearly, the previous classification loss computation strategy is easy to promote the classification results biased towards to the negative samples which leads to a poor image segmentation.

To address this problem, we measure the classification loss by class instead of by sample. In that case, the classification loss of each class is treated equally. We consider the mean of the classification losses of samples as the the classification loss of a class. So, Equation 4 can be further modified as follows

$$\hat{f}_{i} = \arg\min_{f_{i}} \sum_{i} \left\{ f_{i}^{T} L f_{i} + \lambda \sum_{c \in C} \sum_{j \in c} \frac{[f_{i}(j) - y_{i}(j)]^{2}}{m_{c}} \right\}$$
$$\Rightarrow \quad \hat{F} = \arg\min_{F} \{F^{T} L F + \lambda M (F - Y) (F - Y)^{T}\} \quad (5)$$

where $C = \{pos, neg, unlabeled\}$ is the collection of class labels (or the statuses of samples) and m_c is the number of samples in class c. M is a diagonal matrix called class averaging matrix.

$$M = diag\left(\underbrace{\frac{1}{m_{pos}}, \cdots, \underbrace{\frac{1}{m_{neg}}, \cdots, \underbrace{\frac{1}{m_{unlabeled}}, \cdots}}_{m_{pos}}, \underbrace{\frac{1}{m_{unlabeled}}}_{m_{unlabeled}}\right)$$
(6)

We name this new graph-based transduction model Class Averaging Graph-based Transduction (CAGT). This model can be efficiently solved by the regularized least square. The optimal F is as follows

$$F = \frac{\lambda MY}{(L + \lambda M)} \tag{7}$$

Finally, the labeling of *i*-th superpixel can be accomplished by assigning it to the *j*-th class that satisfies

$$\hat{j} = \arg\max_{j} f_j(i), \ j \in \{pos, neg\}$$
(8)



Fig. 2. The performances of GT and CAGT in our synthesized dataset.

3. EXPERIMENTS

We first use a synthetic dataset to verify the effectiveness of our algorithm in unbalanced labeling case. In this dataset, there are 200 positive samples and 2800 negative samples. We use all positives and randomly select different amounts of negatives to construct several experimental datasets which suffer from different degrees of imbalance in distribution of classes. For example, 200 positives and 1000 negatives can construct a dataset whose ratio of negative to positive is 1000/200 = 5. In these datasets, we apply the two-fold cross validation to study the influence of the imbalance of the data distribution to the performances of GT and CAGT. Following [10], Dice Similarity Coefficient (DSC), which is defined as $(2 \times \text{True Positive})/(2 \times \text{True Positive} + \text{False Positive} +$ False Negative), is employed as the evaluation metric. DSC is a comprehensive measurement of the coincidence of the labeling result with the ground truth.

Figure 2(a) shows the distribution of data. The labeling performances of GT and CAGT are plotted in Figure 2(b) while the relative improvements of CAGT over GT, which is defined as (DSC of CAGT / DSC of GT - 1) under different ratios of negative to positive, are reported in Figure 2(c). From the results, we can know that CAGT consistently outperforms in the unbalanced case and also can get a similar or even better performance in the balanced case. Comprehensively speaking, the relative improvement of CAGT over GT is increasing along with the rising of the degree of imbalance in the data distribution. This phenomenon verifies that CAGT is more robust to the imbalance of data.

We apply our work to an archive of 60,000 images of the uterine cervix [6] for cervigram image segmentation. In this archive, the training data is divided into five small training subsets with ten images in each. In these subsets, the data is



(a) GroundTruth

(b) rLR (c) rSLR (d) NNC Fig. 3. Best Segmentation results of different methods. (f) CAGT (ours)

highly unbalanced. The ratios of negative to positive are 17, 39, 31, 27 and 32 in subset1, subset2, subset3, subset4 and subset5 respectively at the superpixel level. The superpixels of the training images and the given test image are used for constructing the graph where each superpixel represents a n-ode in this graph. By considering the trade off between the accuracy and the efficiency, we set the maximal size of the superpixel to 5^2 pixels in SLIC. The simple Nearest Neighbour Classifier (NNC), the conventional Graph-based Transduction (GT) [2], robust Logistic Regression (rLR) [10] and robust Sparse Logistic Regression (rSLR) [10] are employed for comparison.

Table 1. The accuracies of pixel labeling.

Train Sets	DSC				
	subset1	subset2	subset3	subset4	subset5
NNC	0.1082	0.6409	0.2111	0.5051	0.1549
rLR [10]	0.4359	0.6337	0.1269	0.3339	0.3391
rSLR [10]	0.4218	0.5458	0.1079	0.3635	0.4047
GT [2]	0.3449	0.6832	0.2657	0.7704	0.4627
CAGT	0.7701	0.7687	0.7457	0.7990	0.7990

Table 1 reports the labeling performances of different approaches. Note, these labeling accuracies are **based on the** pixels. The pixels are labeled based on the obtained superpixel labels. From the observations, it is clear that CAGT outperforms the compared methods and significantly improve the performance of GT. More specifically, the performance gain of CAGT over GT are 0.4252, 0.0855, 0.4800, 0.0286 and 0.3363 on subset1, subset2, subset3, subset4 and subset5 respectively. Moreover, another interesting phenomenon we can observe from the table is that CAGT is more robust to the choices of the training samples in comparison with other methods. Figure 3 shows the best segmentation results of different approaches for an image. Similar as the results reported in Table 1, the observations demonstrate that our work can achieve the highest quality of segmentation in comparison with the compared approaches. In our experiments, the parameters of all methods are well tuned. We set $\mu = 1$ and $\lambda = 10^{-3}$ for our model.

4. CONCLUSION

We presented a new solution for medical image segmentation which divided the image into dozens of superpixels first and then labeled the superpixels by graph-based transduction. Since the segmentation is regarded as a unbalanced binary classification task, we address this problem via proposing a new graph-based transduction method named Class Averaging Graph-based Transduction (CAGT) which measures the classification losses by classes instead of by samples. A lot of meaningful works can be done to further improve our model. For example, the superpixel representation in this paper is just the mean LAB color vector which is weak to distinguish the positive and negative superpixels. Some more advanced features maybe can be applied to better capture the characteristics of superpixels. Moreover, different features can be fused to improve the quality of segmentation.

(e) GT

5. REFERENCES

- Ahmed Elnakib, Georgy Gimelfarb, Jasjit S Suri, and Ayman El-Baz, "Medical image segmentation: A brief survey," in Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies, pp. 1–39. Springer, 2011.
- [2] Olivier Duchenne, J-Y Audibert, Renaud Keriven, Jean Ponce, and Florent Ségonne, "Segmentation by transduction," in *CVPR*, 2008, pp. 1–8.
- [3] Yinghuan Shi, Shu Liao, Yaozong Gao, Daoqiang Zhang, Yang Gao, and Dinggang Shen, "Prostate Segmentation in CT Images via Spatial-Constrained Transductive Lasso," in *CVPR*, 2013, pp. 2227–2234.
- [4] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE PAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [5] Edward Kim, Wei Wang, Hongsheng Li, and Xiaolei Huang, "A parallel annealing method for automatic color cervigram image segmentation," in *MICCAI-GRID HPC Workshop*, 2009.
- [6] Shiri Gordon, Gali Zimmerman, Rodney Long, Sameer Antani, Jose Jeronimo, and Hayit Greenspan, "Content analysis of uterine cervix images: initial steps towards content based indexing and retrieval of cervigrams," in *Medical Imaging*, 2006, pp. 61444U–61444U.
- [7] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, and Hong-Jiang Zhang, "Face recognition using laplacianfaces," *IEEE PAMI*, vol. 27, no. 3, pp. 328–340, 2005.
- [8] Mikhail Belkin and Partha Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [9] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *NIPS*, 2006, pp. 1601–1608.
- [10] Mingchen Gao, Junzhou Huang, Xiaolei Huang, Shaoting Zhang, and Dimitris N Metaxas, "Simplified labeling process for medical image segmentation," in *MICCAI*, pp. 387–394. 2012.