Contexts-Constrained Multiple Instance Learning for Histopathology Image Analysis

Yan Xu[1,2], Jianwen Zhang[2], Eric I-Chao Chang[2], Maode Lai[4],
and Zhuowen Tu[2,3]

[1]State Key Laboratory of Software Development Environment,
Key Laboratory of Biomechanics and Mechanobiology of Ministry of Education,
Beihang University, China
[2]Microsoft Research Asia, China
[3]Lab of Neuro Imaging, Department of Neurology and Department of Computer
Science, UCLA, USA
[4]Department of Pathology, School of Medicine, Zhejiang University, China
E-Mail: zhuowent@microsoft.com

Abstract. Histopathology image analysis plays a very important role in cancer diagnosis and therapeutic treatment. Existing supervised approaches for image segmentation require a large amount of high quality manual delineations (on pixels), which is often hard to obtain. In this paper, we propose a new algorithm along the line of weakly supervised learning; we introduce context constraints as a prior for multiple instance learning (ccMIL), which significantly reduces the ambiguity in weak supervision (a 20% gain); our method utilizes image-level labels to learn an integrated model to perform histopathology cancer image segmentation, clustering, and classification. Experimental results on colon histopathology images demonstrate the great advantages of ccMIL.

1 Introduction

High resolution histopathology images provide critical information for cancer diagnosis and analysis [1]. Some clinical tasks for the histopathology image analysis may include [2]: (1) diagnosing the presence of cancer (image classification); (2) segmenting images into cancer and non-cancer cells (image segmentation); (3) clustering the tissue cells into various sub-classes. In this paper, we focus on the segmentation task but our integrated framework essentially is able to perform classification, segmentation, and clustering altogether.

Standard unsupervised image segmentation methods [3] may not work well for the histopathology cancer images due to their complicated patterns. Most of the existing supervised approaches [4] for tissue cell segmentations require detailed manual annotations; this task is not only time-consuming but also intrinsically ambiguous, even for well-trained experts. Recent development in weakly-supervised learning (WSL), more specifically multiple instance learning (MIL) [5], uses coarse-grained labeling to aid automatic exploration of fine-grained information. In MIL, a training set consists of bags (images in our case); each bag

consists of a number of instances (patches in our case); only bag-level labels are given in training; the training algorithm then automatically explores instance-level and bag-level models to best fit the given bag labels. Encouraging results have been observed in medical image classifications [6]; in other medical applications, a multiple instance learning approach was adopted in [7] to detect accurate pulmonary embolism among the candidates; in [8] a CAD system was proposed for polyp detection with the main focus on supervised learning features, which are then used in multiple instance regression; MIL-based histopathology image classification was tackled in [9]. However, none of the above methods were targeted for image segmentation, which is also different from an integrated framework of segmentation, clustering, and classification.

In this paper, we design an MIL-based histopathology image segmentation formulation by proposing a new framework, contexts-constrained multiple instance learning (ccMIL); it significantly reduces the ambiguity due to the independence assumptions in the standard multiple instance learning. We observe great improvement of ccMIL over competing methods. Our approach also differs from existing formulations in machine learning in the following aspects: latent conditional random fields algorithm [10] deals mostly with compositional components of object models but not for segmentation; MIL on structured data was proposed in [11] but we emphasize the contextual information of instances as a prior here; multiple clustered instance learning (MCIL) [12] adopts the clustering concept into MIL but it takes the assumption of independent instances; a context-based learning/segmentation framework was proposed in [13] but it is a fully supervised approach.

2   Methods

Rich contextual information has important significance for accurate segmentation [13]. ccMIL aims to take into consideration such contextual information to enhance the performance and achieve robustness. An integrated framework, multiple clustered instance learning (MCIL) [12], was recently proposed to perform simultaneous image-level classification, pixel-level segmentation and patch-level clustering. ccMIL inherits some aspects of MCIL but studies the contextual prior in the MIL training stage to reduce the intrinsic ambiguity due to the nature of weak supervision. We observe significant improvement of ccMIL over MCIL in experiments, e.g. over 20% gain. Fig. 1 illustrates the distinction between standard supervised learning, MIL, MCIL and ccMIL.

2.1   Contexts-Constrained Multiple Instance Learning (ccMIL)

In ccMIL, learning examples are represented by a bag of instances. In our case, a histopathology image is a bag and each patch sampled from an image is an instance. Patches with cancer tissues are treated as positive instances and the ones without cancer tissues are negative. A bag is labeled as positive (cancer image) if the bag contains at least one positive instance.
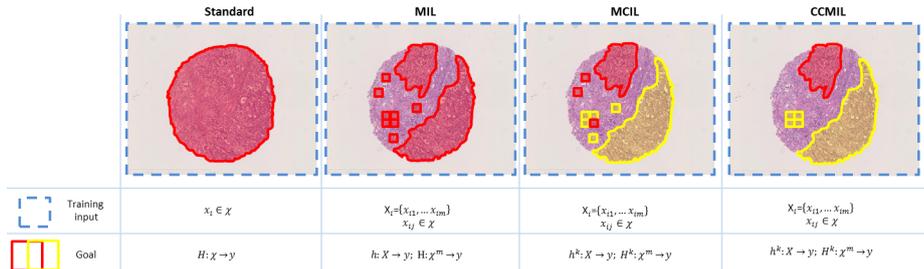
Fig. 1. Distinct learning goals between supervised learning, MIL, MCIL and ccMIL. ccMIL makes an important step over MIL by studying the contextual prior information among the instances to reduce the instance-level ambiguity due to weak-supervision.

We are given a training set $\mathcal{X}$, and x_i is the ith bag in $\mathcal{X}$. Each bag consists of a set of m instances *i.e.* $x_i = \{x_{i1}, \ldots, x_{im}\}$; each x_i is also associated with a label $y_i \in \mathcal{Y} = \{-1, 1\}$. Assume there are K clusters (cancer types), then each instance x_{ij} has a corresponding label $y_{ij}^k \in \mathcal{Y} = \{-1, 1\}$, $k \in \{1, \ldots, K\}$, that denotes whether the instance belongs to the kth cluster. If this instance belongs to one of the K clusters, that is $y_{ij}^k = 1$, then this instance is considered as positive. Note that, this y_{ij}^k is not known during training. A bag is labeled positive if at least one of its instances belongs to one of the K groups:

$$y_i = \max_j \max_k (y_{ij}^k). \tag{1}$$

The goal of ccMIL is to split the positive instance into K groups by learning K instance-level classifiers $h^k(x_{ij}) : \mathcal{X} \to \mathcal{Y}$ for K clusters, using only bag labels y_i, such that $\max_j \max_k h^k(x_{ij}) = y_i$.

We combine AnyBoost[14–16] framework, the same as MIL-Boost [15], to solve $h^k(x_{ij})$. First loss function $\mathcal{L}(h)$ (details are given in the next subsection) is introduced to find the optimal weak classifier response $h_t^k : \mathcal{X} \to \mathcal{Y}$ that most reduces the loss on the training data. We train h_t^k by minimizing the training data error weighted by $|w_{ij}^k|$: $h_t^k = argmin_h \sum_{ij} \mathbf{1}(h(x_{ij}^k) \neq y_i^k)|w_{ij}^k|$; while $w_{ij}^k \equiv -\frac{\partial \mathcal{L}(h)}{\partial h_{ij}^k}$. A differentiable *softmax* function $g_l(v_l)$ is given to approximate the max over $\mathbf{v} = \{v_1, \ldots, v_m\}$. It is defined as follows:

$$g_l(v_l) \approx \max_l(v_l) = v^*, \quad \frac{\partial_{g_l}(v_l)}{\partial v_l} \approx \frac{\mathbf{1}(v_i = v^*)}{\sum_l \mathbf{1}(v_l = v^*)}, \quad m = |\mathbf{v}|. \tag{2}$$

There are a number of approximations for g. We choose GM model [16], that is $g_l(v_l) = (\frac{1}{m} \sum_l v_l^r)^{\frac{1}{r}}$, based on the experiment results. In order to optimize the loss function $\mathcal{L}$, we must get p_i, which is defined as the maximum over p_{ij}^k, the probability of an instance x_{ij} belonging to the kth cluster: $p_{ij}^k = \sigma(2h_{ij}^k)$, where $h_{ij}^k = h^k(x_{ij})$. Using the *softmax* g in place of the max, we can get p_i as:

$$p_i = g_j(g_k(p_{ij}^k)) = g_{jk}(p_{ij}^k) = g_{jk}(\sigma(2h_{ij}^k)), \sigma(v) = \frac{1}{1 + \exp(-v)}. \tag{3}$$

So, the weights w_{ij} can be written as $w_{ij}^k = -\frac{\partial \mathcal{L}(h)}{\partial h_{ij}^k} = -\frac{\partial \mathcal{L}(h)}{\partial p_i} \frac{\partial p_i}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k}$. Thus, we can train the weak classifier h_t^k by optimizing weighed error $|w_{ij}^k|$, and finally, get a strong classifier: $\mathbf{h}^k \leftarrow \mathbf{h}^k + \alpha_t{}^k h_t^k$, where α_t weighs the weak learners relative importance.

2.2   Loss Function and Solving Process of ccMIL

The key to ccMIL is a formulation for introducing the contextual information as a prior for MIL.

Now we define two functions $\mathcal{L}_A(h)$ and $\mathcal{L}_B(h)$ as:

$$\mathcal{L}_A(h) = -\sum_{i=1}^n w_i(\mathbf{1}(y_i = 1) \log p_i + \mathbf{1}(y_i = -1) \log(1 - p_i)), \ and \qquad (4)$$

$$\mathcal{L}_B(h) = \sum_{i=1}^n w_i \sum_{(j,m) \in E_i} v_{jm} \parallel p_{ij} - p_{im} \parallel^2, \qquad (5)$$

where w_i is the weight of the ith training data (the ith bag). E_i denotes the set of all the neighboring instance pairs in the ith bag. v_{jm} is the weight on a pair of instances (patches) j and m related to the distance (on the image, denoted as d_{jm}) between them. Higher weights are put on those closer pairs. In our experiment, we chose: $v_{jm} = \exp(-d_{jm})$.

Then, we can define loss function as:

$$\mathcal{L}(h) = \mathcal{L}_A(h) + \lambda \mathcal{L}_B(h). \qquad (6)$$

$\mathcal{L}_B(h)$ imposes an effective contextual constraints (in a way smoothness prior) over the instances to remove the ambiguity in training; it encourages the nearby image patches to share similar class types. λ is the weight of the additional item that reflects the importance of relationship between the current instance and its context (neighbors). The overall classification function obtained with the new formulation is thus robust to noise and able to achieve more accurate segmentation results.

According to the new loss function we compute the weight w_{ij}^k as following:

$$w_{ij}^k = -\frac{\partial \mathcal{L}(h)}{\partial h_{ij}^k} = -\frac{\partial \mathcal{L}(h)}{\partial p_i} \frac{\partial p_i}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k}. \qquad (7)$$

$$\frac{\partial \mathcal{L}(h)}{\partial h_{ij}^k} = \frac{\partial \mathcal{L}_A(h)}{\partial h_{ij}^k} + \lambda \frac{\partial \mathcal{L}_B(h)}{\partial h_{ij}^k} = \frac{\partial \mathcal{L}_A(h)}{\partial p_i} \frac{\partial p_i}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k} + \lambda \frac{\partial \mathcal{L}_B(h)}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k}. \qquad (8)$$

$$\frac{\partial \mathcal{L}_A(h)}{\partial p_i} = \begin{cases} -\dfrac{w_i}{p_i} & if \ y = 1; \\ \dfrac{w_i}{1 - p_i} & if \ y = -1, \end{cases} \qquad \frac{\partial \mathcal{L}_B(h)}{\partial h_{ij}^k} = w_i \sum_{(j,m) \in E_i} 2v_{jm}(p_{ij}^k - p_{im}^k).$$
$$\qquad (9)$$

Details of ccMIL are demonstrated in Algorithm 1. K is the number of cancer types, and T is the number of weak classifiers in Boosting.

Algorithm 1 ccMIL

  Input: Bags $\{X_1, \ldots, X_n\}, \{y_1, \ldots, y_n\}, K, T$
  Output: $\mathbf{h}^1, \ldots, \mathbf{h}^K$
  for $t = 1 \rightarrow T$ **do**
    for $k = 1 \rightarrow K$ **do**
      Compute weights $w_{ij}^k = -\frac{\partial \mathcal{L}(h)}{\partial h_{ij}^k} = -\left(\frac{\partial \mathcal{L}_A(h)}{\partial p_i} \frac{\partial p_i}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k} + \lambda \frac{\partial \mathcal{L}_B(h)}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k}\right)$
      Train weak classifiers h_t^k using weights $|w_{ij}^k|$
      $h_t^k = argmin_h \sum_{ij} \mathbf{1}(h(x_{ij}^k) \neq y_i^k)|w_{ij}^k|$
      Find α_t via line search to minimize $\mathcal{L}(., \mathbf{h}^k, .)$
      $\alpha_t^k = argmin_\alpha \mathcal{L}(., \mathbf{h}^k + \alpha h_t^k, .)$
      Update strong classifiers $\mathbf{h}^k \leftarrow \mathbf{h}^k + \alpha_t^k h_t^k$
    end for
  end for

---

3   Experiments

ccMIL is a general approach for common cancer types, including colon, prostate, and breast cancer. Without loss of generality, colon histopathology images are chosen in our experiments to illustrate its effectiveness. We collected the image dataset in Department of Pathology of Zhejiang University in September 2010. The images are obtained from the Nano Zoomer 2.0HT digital slice scanner produced by Hamamatsu Photonics with a magnification factor of 40. In this dataset, 30 non-cancer (NC) images and 53 cancer images are included. The cancer images can be medically divided into four cancer types according to their morphological characteristics. These four cancer types are Moderately or well differentiated tubular adenocarcinoma (MTA), Poorly differentiated tubular adenocarcinoma (PTA), Mucinous adenocarcinoma (MA), and Signet-ring carcinoma (SRC). To ensure the ground truth of the image dataset, images are carefully studied and labeled by experts. Specifically, each image is independently labeled by two pathologists, the third pathologist moderates their discussion until they get an agreement on the result. All images are labeled as cancer images or non-cancer images. For cancer images, cancer tissues are further annotated and corresponding cancer type is identified for the evaluation.

We combine all the images to generate three different subsets: $binary$, $multi1$, and $multi2$. Each subset contains 60 different histopathology images. $binary$ contains 30 non-cancer and 30 MTA cancer images. It is used to test the capability of cancer image detection. $multi1$ and $multi2$ mean two or more types of cancer images as well as non-cancer images are contained. They can reveal the ability of pixel-level segmentation. In particular, $multi1$ consists of 30 NC, 15 MTA, 9 PTA and 6 SRC; $multi2$ consists of 30 NC, 13 MTA, 9 PTA and 8 MA. Settings are made as following. First we down-sample the images by 5 times, and then extract 64×64 patches from each image. The parameters in algorithm are set as: $r = 20$, $K = 4$, $T = 200$. r controls sharpness and accuracy in GM model of $softmax$ function. The λ used in the loss function is set to 0.01 according to the results of cross validation. We assume the initial distribution is uniform so that
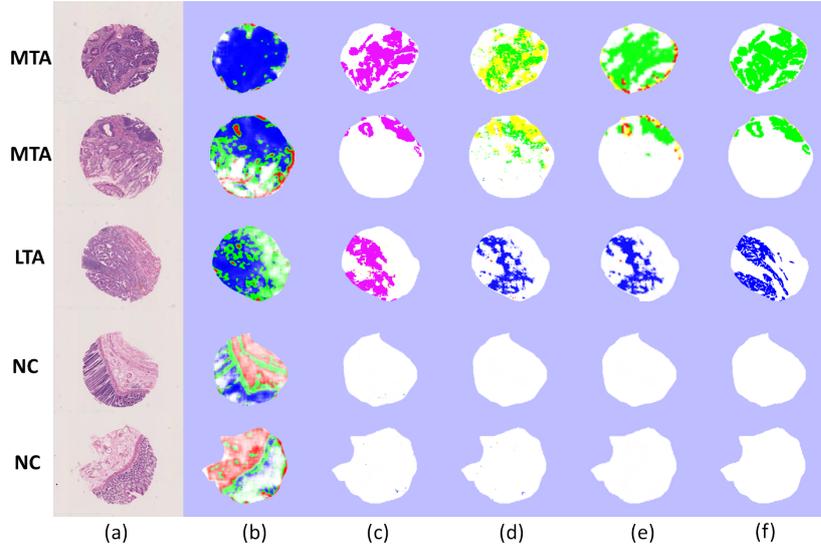
Fig. 2. Image Types: (a): The original images. (b): The instance-level results (pixel-level segmentation and patch-level clustering) for image-level supervision + K-means, (c): pixel-level full supervision, (d): MCIL, (e): ccMIL, (f): The instance-level ground truth labeled by three pathologists. Different colors stand for different types of cancer tissues. Cancer Types: from top to bottom: MTA, MTA, PTA, NC, and NC.

the prior weight w_i for the ith bag is set as the same value. Our method is not focusing on feature design, so generic features for object classification are used here, including $L^*a^*b^*$ Color Histogram, Local Binary Pattern, and SIFT. The weak classifier we use is Gaussian function. Experimental results are reported in a 5-fold cross validation. All the methods in the following experiments are conducted under the same experimental settings.

Pixel-level Segmentation. We tested subset $multi2$ with different methods to measure pixel-level segmentation. Fig. 2 shows the segmentation results. ccMIL significantly improves results by reducing the intrinsic training ambiguity compared to other weakly supervised methods. For example, ccMIL can correctly recognizes noises and small isolated areas in cancer images and achieve cleaner boundaries, which can be observed from the segmented results of MTA and PTA cancer images in the figure. Moreover, due to the guidance of contextual information, ccMIL reduces the possibility of extracting noises as a positive instance from a non-cancer image and further improves the accuracy of cancer detection.

For the quantitative evaluation of the segmentations, F-measure is used here to evaluate the segmentation. the F-measure values of image-level supervision, MCIL and ccMIL are 0.312, 0.601 and 0.717. ccMIL improves F-measure by 20%, compared with the closet competing method.
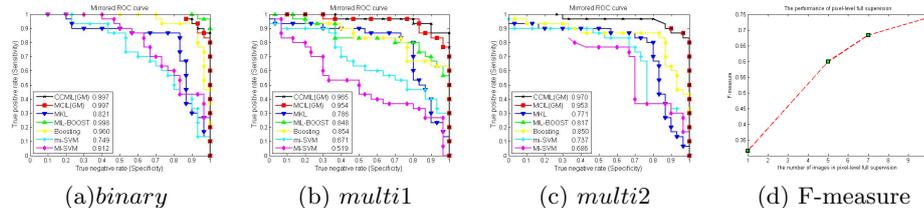
| (a)*binary* | (b) *multi*1 | (c) *multi*2 | (d) F-measure |

Fig. 3. ROC curves for different learning methods (*binary*, *multi*1 and *multi*2) and (d) the segmentation F-measure of pixel-level full supervision.

For comparing with supervised approaches, we implemented two methods: (1) one utilizes supervision in the image level by treating all the pixels in the positive and negative bags as positive and negative instances respectively, and (2) one with the full pixel-level supervision (require laborious labeling work). The advantage of ccMIL over the image-level supervision is proved by both segmented images and F-measure evaluation. As for full pixel-level supervision, in order to compare the performance, we used varying numbers $(1, 5, 7, 10)$ of images of pixel-level fully supervision, and calculated the corresponding values of F-measure. The figure is plotted in Fig. 3.d, from which, it can be concluded that ccMIL is able to achieve comparable results (the value of F-measure is around 0.7).

Patch-level Clustering. We also obtained the clustering results of the same test data mentioned in pixel-level segmentation and the results are shown in Fig. 2. ccMIL achieves less noisy clustering results than MCIL. Also, it revises the error caused by MCIL, which can be observed from the results obtained from the two MTA images in Fig. 2.

Image-level Classification. Bag-level classification, that is cancer and non-cancer images classification, is compared in this experiment also. Seven methods, namely MI-SVM, mi-SVM, Boosting, MIL-BOOST, MKL (multiple kernel learning as a widely used image categorization technique), MCIL, and ccMIL, are compared in this experiment with the same features and parameters (we do not put all the references due to the space limit). Fig. 3 shows the receiver operating characteristic (ROC) curves in the three subsets. The results demonstrate the practicality of ccMIL.

4 Conclusion

We have introduced the contexts constraints to the multiple instance learning framework for segmentation and observe significant improvement (20%) over the closest competing method. In addition, ccMIL is able to perform segmentation, clustering, and classification in a principled framework while achieving comparable results in segmentation with full pixel-level supervision approaches.

References

1. Gurcan, M., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N., Yener, B.: Histopathological image analysis: A review. IEEE Reviews in Biomedical Engineering **2** (2009) 147 –171
2. Yang, L., Tuzel, O., Meer, P., Foran, D.: Automatic image analysis of histopathology specimens using concave vertex graph. In: MICCAI. (2008)
3. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Transactions on Image Processing **10**(2) (2000) 266–277
4. Madabhushi, A.: Digital pathology image analysis: opportunities and challenges. Imaging in Medicine **1**(1) (2009) 7–10
5. Maron, O., Lozano-Pérez, T.: A framework for multiple-instance learning. In: NIPS. (1997)
6. Liu, Q., Qian, Z., Marvasty, I., Rinehart, S., Voros, S., Metaxas, D.: Lesion-specific coronary artery calcium quantification for predicting cardiac event with multiple instance support vector machines. In: MICCAI. (2010)
7. Liang, J., Bi, J.: Computer aided detection of pulmonary embolism with tobogganing and mutiple instance classification in ct pulmonary angiography. In: IPMI. (2007)
8. Lu, L., Bi, J., Wolf, M., Salganicoff, M.: Effective 3d object detection and regression using probabilistic segmentation features in ct images. In: CVPR. (2011)
9. Dundar, M., Badve, S., Raykar, V., Jain, R., Sertel, O., Gurcan, M.: A multiple instance learning approach toward optimal classification of pathology slides. In: ICPR. (2010)
10. Quattoni, A., Wang, S., Morency, L., Collins, M., Darrell, T.: Hidden conditional random fields. IEEE Trans. PAMI **29**(10) (2007) 1848–1852
11. Zhang, D., Liu, Y., Si, L., Zhang, J., Lawrence, R.D.: multiple instance learning on structured data. In: NIPS. (2011)
12. Xu, Y., Zhu, J.Y., Chang, E., Tu, Z.: Multiple clustered instance learning for histopathology cancer image segmentation, classification and clustering. In: CVPR. (2012)
13. Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3d brain image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **21**(10) (2010) 1744–1757
14. Mason, L., Baxter, J., Bartlett, P., Frean, M.: Boosting algorithms as gradient descent. In: NIPS, MIT Press (2000)
15. Viola, P.A., Platt, J., Zhang, C.: Multiple instance boosting for object detection. In: NIPS. (2005)
16. Babenko, B., Dollár, P., Tu, Z., Belongie, S.: Simultaneous learning and alignment: Multi-instance and multi-pose learning. In: Workshop of RealFaces. (2008)