Medical Image Analysis 18 (2014) 591–604

Contents lists available at ScienceDirect

Medical Image Analysis

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

Weakly supervised histopathology cancer image segmentation and classification

Yan Xu [a,b], Jun-Yan Zhu [c], Eric I-Chao Chang [b], Maode Lai [d], Zhuowen Tu [e,*]

[a] State Key Laboratory of Software Development Environment, Key Laboratory of Biomechanics and Mechanobiology of Ministry of Education, Beihang University, China
[b] Microsoft Research Asia, No. 5 Danling Street, Haidian District, Beijing 10080, PR China
[c] Computer Science Division, University of California, Berkeley, USA
[d] Department of Pathology, School of Medicine, Zhejiang University, China
[e] Department of Cognitive Science, University of California, San Diego, CA, USA

ARTICLE INFO

Article history:
Received 7 October 2012
Received in revised form 30 December 2013
Accepted 28 January 2014
Available online 22 February 2014

Keywords:
Image segmentation
Classification
Clustering
Multiple instance learning
Histopathology image

ABSTRACT

Labeling a histopathology image as having cancerous regions or not is a critical task in cancer diagnosis; it is also clinically important to segment the cancer tissues and cluster them into various classes. Existing supervised approaches for image classification and segmentation require detailed manual annotations for the cancer pixels, which are time-consuming to obtain. In this paper, we propose a new learning method, multiple clustered instance learning (MCIL) (along the line of weakly supervised learning) for histopathology image segmentation. The proposed MCIL method simultaneously performs image-level classification (cancer vs. non-cancer image), medical image segmentation (cancer vs. non-cancer tissue), and patch-level clustering (different classes). We embed the clustering concept into the multiple instance learning (MIL) setting and derive a principled solution to performing the above three tasks in an integrated framework. In addition, we introduce contextual constraints as a prior for MCIL, which further reduces the ambiguity in MIL. Experimental results on histopathology colon cancer images and cytology images demonstrate the great advantage of MCIL over the competing methods.

1. Introduction

Histopathology image analysis is a vital technology for cancer recognition and diagnosis (Tabesh et al., 2007; Park et al., 2011; Esgiar et al., 2002; Madabhushi, 2009). High resolution histopathology images provide reliable information differentiating abnormal tissues from the normal ones. In this paper, we use tissue microarrays (TMAs) which are referred to histopathology images here. Fig. 1 shows a typical histopathology colon cancer image, together with a non-cancer image. Recent developments in specialized digital microscope scanners make digitization of histopathology readily accessible. Automatic cancer recognition from histopathology images thus has become an increasingly important task in the medical imaging field (Esgiar et al., 2002; Madabhushi, 2009). Some clinical tasks (Yang et al., 2008) for histopathology image analysis include: (1) detecting the presence of cancer (image classification); (2) segmenting images into cancer and non-cancer region (medical image segmentation); (3) clustering the tissue region into various classes. In this paper, we aim to develop an integrated framework to perform classification, segmentation, and clustering altogether.

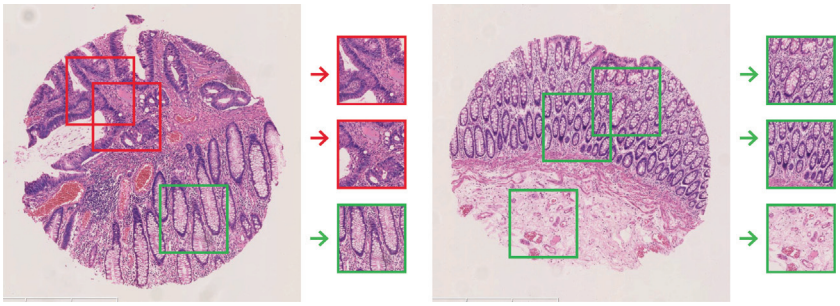
Several practical systems for classifying and grading cancer histopathology images have been recently developed. These methods are mostly focused on the feature design including fractal features (Huang and Lee, 2009), texture features (Kong et al., 2009), object-level features (Boucheron, 2008), and color graphs features (Altunbay et al., 2010; Ta et al., 2009). Various classifiers (Bayesian, KNN and SVM) are also investigated for pathological prostate cancer image analysis (Huang and Lee, 2009).

From a different angle, there is a rich body of literature on supervised approaches for image detection and segmentation (Viola and Jones, 2004; Shotton et al., 2008; Felzenszwalb et al., 2010; Tu and Bai, 2010). However, supervised approaches require a large amount of high quality annotated data, which are labor-intensive and time-consuming to obtain. In addition, there is intrinsic ambiguity in the data delineation process. In practice, obtaining the very detailed annotation of cancerous regions from a histopathology image could be a challenging task, even for expert pathologists.

Unsupervised learning methods (Duda et al., 2001; Loeff et al., 2005; Tuytelaars et al., 2009), on the other hand, ease the burden

* Corresponding author. Tel.: +1 858 822 0908.
E-mail addresses: xuyan04@gmail.com (Y. Xu), junyanz@eecs.berkeley.edu (J.-Y. Zhu), echang@microsoft.com (E.I.-C. Chang), lmd@zju.edu.cn (M. Lai), ztu@ucsd.edu (Z. Tu).

http://dx.doi.org/10.1016/j.media.2014.01.010
1361-8415/



(a) cancer image (b) non-cancerimage

Fig. 1. Example histopathology colon cancer and non-cancer images: (a) positive bag (cancer image) and (b) negative bag (non-cancer image). Red rectangles: positive instances (cancer tissues). Green rectangles: negative instances (non-cancer tissues). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of having manual annotations, but often at the cost of inferior results.

In the middle of the spectrum is the weakly supervised learning scenario. The idea is to use coarsely-grained annotations to aid automatic exploration of fine-grained information. The weakly supervised learning direction is closely related to semi-supervised learning in machine learning (Zhu, 2008). One particular form of weakly supervised learning is multiple instance learning (MIL) (Dietterich et al., 1997) in which a training set consists of a number of bags; each bag includes many instances; the goal is to learn to predict both bag-level and instance-level labels while only bag-level labels are given in training. In our case, we aim at automatically learning image models to recognize cancers from weakly supervised histopathology images. In this scenario, only image-level annotations are required. It is relatively easier for a pathologist to label a histopathology image than to delineate detailed cancer regions in each image.

In this paper, we develop an integrated framework to classify histopathology images as having cancerous regions or not, segment cancer tissues from a cancer image, and cluster them into different types. This system automatically learns the models from weakly supervised histopathology images using multiple clustered instance learning (MCIL), derived from MIL. Many previous MIL-based approaches have achieved encouraging results in the medical domain such as major adverse cardiac event (MACE) prediction (Liu et al., 2010), polyp detection (Dundar et al., 2008, 2006, 2011), pulmonary emboli validation (Raykar et al., 2008), and pathology slide classification (Dundar et al., 2010). However, none of the above methods aim to perform medical image segmentation. They also have not provided an integrated framework for the task of simultaneous classification, segmentation, and clustering.

We propose to embed the clustering concept into the MIL setting. The current literature in MIL assumes single cluster/model/classifier for the target of interest (Viola et al., 2005), single cluster within each bag (Babenko et al., 2008; Zhang and Zhou, 2009; Zhang et al., 2009), or multiple components of one object (Dollár et al., 2008). Since cancer tissue clustering is not always available, it is desirable to discover/identify the classes of various cancer tissue types; this results in patch-level clustering of cancer tissues. The incorporation of clustering concept leads to an integrated system that is able to simultaneously perform image segmentation, image-level classification, and patch-level clustering.

In addition, we introduce contextual constraints as a prior for cMCIL, which reduces the ambiguity in MIL. Most of the previous MIL methods make the assumption that instances are distributed independently, without considering the correlations among instances. Explicitly modeling the instance interdependencies (structures) can effectively improve the quality of segmentation. In our experiment, we show that while obtaining comparable results in classification, cMCIL improves the segmentation significantly (over 20%) compared MCIL. Thus, it is beneficial to explore the structural information in the histopathology images.

2. Related work

Related work can be roughly divided into two broad categories: (1) approaches for histopathology image classification and segmentation and (2) MIL methods in machine learning and computer vision. After the discussion about the previously work, we show the contributions of our method.

2.1. Existing approaches for histopathology image classification and segmentation

Classification. There has been rich body of literature in medical image classification. Existing methods for histopathology image classification however are mostly focused on the feature design in supervised settings. Color graphs were used in Altunbay et al. (2010) to detect and grade colon cancer in histopathology images; multiple features including color, texture, and morphologic cues at the global and histological object levels were adopted in prostate cancer detection (Tabesh et al., 2007); Boucheron et al. proposed a method using object-based information for histopathology cancer detection (Boucheron, 2008). Some other work is focused on classifier design: for instance, Doyle et al. developed a boosted Bayesian multi-resolution (BBMR) system for automatically detecting prostate cancer regions on digital biopsy slides, which is a necessary precursor to automated Gleason grading (Artan et al., 2012). In Monaco et al. (2010), a Markov model was proposed for prostate cancer detection in histological images.

Segmentation. A number of supervised approaches for medical image segmentation have also been proposed before, for example on histopathology images (Kong et al., 2011) and vasculature retinal images (Soares et al., 2006). Structured data has also been taken into consideration in the previous work. Wang and Rajapakse (2006) presented a conditional random fields (CRFs) model to fuse contextual dependencies in functional magnetic resonance imaging (fMRI) data to detecting brain activity. A CRF-based segmentation method was also proposed in Artan et al. (2010) for localizing prostate cancer from multi-spectral MR images.

2.2. MIL and its applications

Compared with fully supervised methods, multiple instance learning (MIL) (Dietterich et al., 1997) has its particular advantages in automatically exploiting the fine-grained information and

reducing efforts in human annotations. In the machine learning community, many MIL methods have been developed in recent years such as Diverse Density (DD) (Maron and Lozano-Pérez, 1997), Citation-kNN (Wang et al., 2000), EM-DD (Zhang and Goldman, 2001), MI-Kernels (Gärtner et al., 2002), SVM-based methods (Andrews et al., 2003), and ensemble algorithms MIL-Boost (Viola et al., 2005).

Although first introduced in the context of drug activity prediction (Dietterich et al., 1997), the MIL formulation has made significant success in the area of computer vision, such as visual recognition (Viola et al., 2005; Babenko et al., 2008; Galleguillos et al., 2008; Dollár et al., 2008), weakly supervised visual categorization (Vijayanarasimhan and Grauman, 2008), and robust object tracking (Babenko et al., 2011). Zhang and Zhou (2009) proposed a multiple instance clustering (MIC) method to learn the clusters as hidden variables to the instances. Zhang et al. (2009) further formulated the MIC problem under the maximum margin clustering framework. MIC however is designed for datasets that have no negative bags and it assumes each bag containing only one cluster. Babenko et al. (2008) assumed a hidden variable, pose, to each face (only one) in an image. In our case, multiple clusters of different cancer types might co-exist within one bag (histopathology image). In addition, segmentation cannot be performed. In Dollár et al. (2008), object detection was achieved by learning individual component classifiers and combining these into an overall classifier, which also differs from our work. Multiple components were learned for a single object class. However, we have multiple instances and multiple classes within each bag in our work.

The MIL assumption was integrated into multiple-label learning for image/scene classification in Zhou and Zhang (2007), Zha et al. (2008), and Jin et al. (2009) and for weakly supervised semantic segmentation in Vezhnevets and Buhmann (2010). Multi-class labels were given as supervision in their methods; in our method, multiple clusters are hidden variables to be explored in a weakly supervised manner.

The MIL framework has also been adopted in the medical imaging domain with the focus mostly on the medical diagnosis (Fung et al., 2007). In Liu et al. (2010), an MIL-based method was developed to perform medical image classification; in Liang and Bi (2007), pulmonary embolisms among the candidates were screened by an MIL-like method; a computer aided diagnosis (CAD) system (Lu et al., 2011) was developed for polyp detection with the main focus on learning the features, which were then used for multiple instance regression; an MIL approach was adopted for cancer classification in histopathology slides (Dundar et al., 2010). However, these existing MIL approaches were designed for medical image diagnosis and none of them perform segmentation. Moreover, to the best of our knowledge, the integrated classification/segmentation/clustering task has not been addressed, which is the key contribution of this paper.

2.3. Our contributions

Although several tasks in computer vision and medical domain have been shown to benefit from the MIL setting, we find that the cancer image classification/segmentation/clustering task is a well-suited medical imaging application for the MIL framework. We propose a new learning method, multiple clustered instance learning (MCIL), along the line of weakly supervised learning. The proposed MCIL method simultaneously performs image-level classification (cancer vs. non-cancer image), medical image segmentation (cancer vs. non-cancer tissues), and patch-level clustering (different classes). We embed the clustering concept into the MIL setting and derive a principled solution to perform the above three tasks in an integrated framework. Furthermore, we demonstrate the importance of contextual information by varying the weight of contextual model term. Finally, we try to answer the following question: is time-consuming and expensive pixel-level annotation of cancer images necessary to build a practical working medical image analysis system, or could the weaker but much cheaper image-level supervision achieve the same accuracy and robustness?

Earlier conference versions of our approach were presented in Xu et al. (2012b,a). Here, we further illustrate that: (1) the MCIL method could be applied to analyze image types other than histopathology, such as cytology images, (2) additional features such as gray-level co-occurrence matrix (GLCM) are added to this paper, and (3) a new subset of histopathology images has been created in this experiment. In this paper, we focus on colon histopathology image classification, segmentation and clustering. However, it is noted that our MCIL formulation is general and it can be adopted to other image modalities.

3. Methods

We follow the general definition of bags and instances in the multiple instance learning (MIL) formulation (Dietterich et al., 1997).

In this paper, the ith histopathology image is considered as a bag x_i; the jth image patch densely sampled from an image corresponds to an instance x_{ij}. A patch of cancer tissue is treated as a positive instance ($y_{ij} = 1$) and a patch without any cancer tissues is a negative instance ($y_{ij} = -1$). The ith bag is labeled as positive (cancer image), namely $y_i = 1$, if this bag contains at least one positive instance. Similarly, in histopathology cancer image analysis, a histopathology image is diagnosed as positive by pathologists as long as a small part of image is considered as cancerous. Fig. 1 shows the definition of positive/negative bags and positive/negative instances.

An advantage brought by MIL is that if an instance-level classifier is learned, the image segmentation task then can be directly performed; bag-level (image-level) classifier can also be obtained.

In the following sections, we first give the overview of the MIL literature, especially recent gradient decent boosting based MIL approaches; then we introduce the formulation for the proposed method, MCIL, which integrates the clustering concepts into the MIL setting; properties of MCIL with various variations are provided. In addition, we introduce contextual constraints as a prior for MCIL, resulting in context-constrained multiple clustered instance learning (cMCIL). Fig. 2 and Algorithm 1 show the flow diagram of our algorithms. The inputs include both cancer images and noncancer images. Cancer images are used to generate positive bags (red circles) and noncancer images are used to generate negative bags (green circles). Within each bag, each image patch represents an instance. cMCIL/MCIL is used as a multiple instance learning framework to perform learning. The learned models generate several classifiers for patch-level cancer clusters. Red, yellow, blue and purple colors represent different cancer types while green represents the noncancer patches. The overall image-level classification (caner vs. non-cancer) can be obtained based on the prediction from the patch-level classification.

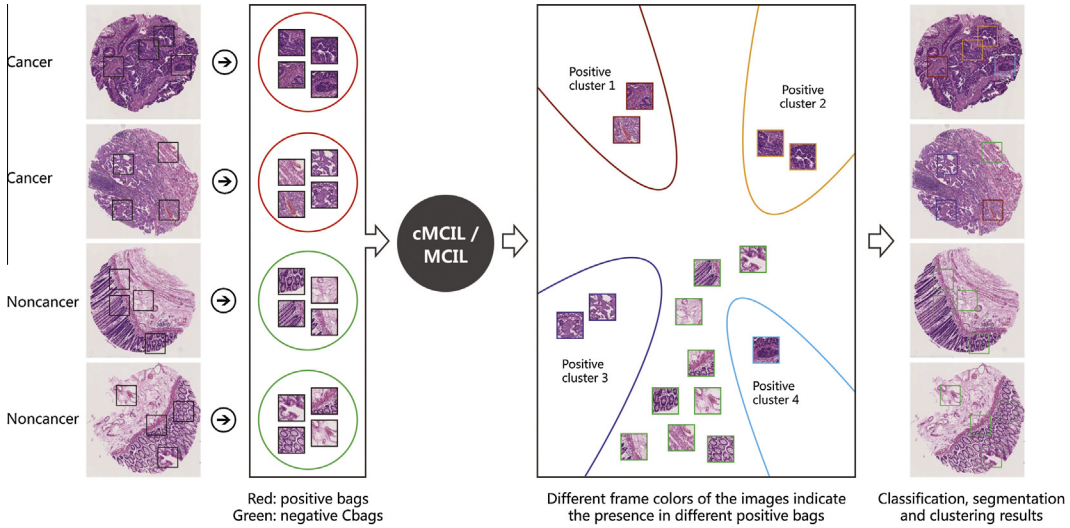


Fig. 2. Flow diagram of our algorithms. The inputs include both cancer images and noncancer images. Cancer images are used to generate positive bags (red circles) and noncancer images are used to generate negative bags (green circles). Within each bag, each image patch represents an instance. cMCIL/MCIL is used as a multiple instance learning framework to perform learning. The learned models generate several classifiers for patch-level cancer clusters. Red, yellow, blue and purple colors represent different cancer types while green represents the noncancer patches. The overall image-level classification (caner vs. non-cancer) can be obtained based on the prediction from the patch-level classification. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Algorithm 1. Algorithm

Input: Colon histopathology images
Output: Image-level classification models for cancer vs. noncancer and patch-level classification models for different cancer classes
Step 1: Extract patches from colon histopathology images.
Step 2: Generate bags for models using extracted patches.
Step 3: Learn models in a multiple instance learning framework (MCIL/cMCIL).
Step 4: Obtain image segmentation and patch clustering simultaneously.

3.1. Review of the MIL method

We give a brief introduction to the MIL formulation and focus on boosting-based (Mason et al., 2000) MIL approaches (Viola et al., 2005; Babenko et al., 2008), which serve as the building blocks for our proposed MCIL.

In MIL, we are given a training set consisting of n bags: $\mathcal{X}^m = \{x_1, \ldots, x_n\}$. x_i is the ith bag, and m denotes the number of instances in each bag, *i.e.* $x_i = \{x_{i1}, \ldots, x_{im}\}$ where $x_{ij} \in \mathcal{X}$ and $\mathcal{X} = \mathbb{R}^d$ (although each bag may have different number of instances, for clarity of notation, we use m for all the bags here). Each x_i is associated with a label $y_i \in \mathcal{Y} = \{-1, 1\}$. It is assumed that each instance x_{ij} in the bag x_i has a corresponding label $y_{ij} \in \mathcal{Y}$, which in fact is not given as supervision during the training stage. As mentioned before, a bag is labeled as positive if at least one of its m instances is positive and a bag is negative if all its instances are negative. In the binary case, the assumption can be expressed as:

$$y_i = \max_j(y_{ij}), \tag{1}$$

where max is essentially equivalent to an OR operator since for $y_{ij} \in \mathcal{Y}$, $\max_j(y_{ij}) = 1 \Longleftrightarrow \exists j$, s.t. $y_{ij} = 1$.

The goal of MIL is to learn an instance-level classifier $h(x_{ij}) : \mathcal{X} \to \mathcal{Y}$. A bag-level classifier $H(x_i) : \mathcal{X}^m \to \mathcal{Y}$ could be built with the instance-level classifier:

$$H(x_i) = \max_j h(x_{ij}). \tag{2}$$

To accomplish this goal, MIL-Boost (Viola et al., 2005) was proposed by combining the MIL cost functions and the AnyBoost framework (Mason et al., 2000). The general idea of AnyBoost (Mason et al., 2000) is to minimize the loss function $\mathcal{L}(h)$ via gradient descent on the h in the function space. The classifier h is written in the form of h_t as:

$$h(x_{ij}) = \sum_{t=1}^{T} \alpha_t h_t(x_{ij}), \tag{3}$$

where α_t weighs the weak learners' relative importances.

To find the best h_t, we proceed with two steps: (1) computing the weak classifier response and (2) selecting the weak classifier from available candidates which achieves the best discrimination. We consider h as a vector with components $h_{ij} \equiv h(x_{ij})$. To find the optimal weak classifier in each phase, we compute $-\frac{\partial \mathcal{L}}{\partial h_{ij}}$, which is a vector with components $w_{ij} \equiv -\frac{\partial \mathcal{L}}{\partial h_{ij}}$. Since we are limited in the choice of h_t, we train the weak classifier h_t by minimizing the training error weighted by $|w_{ij}|$, using the follow formula: $h_t = \operatorname{argmin}_h \sum_{ij} \mathbf{1}(h(x_{ij}) \neq y_i)|w_{ij}|$.

The loss function, a function over h, defined in the MIL-Boost (Viola et al., 2005; Babenko et al., 2008) is a standard negative log likelihood expressed as:

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

where $\mathbf{1}(\cdot)$ is an indicator function. The bag probability $p_i \equiv p(y_i = 1|x_i)$ is defined in terms of h. w_i is introduced here as the prior weight of the i^{th} training sample.

A differentiable approximation of the max, namely *softmax* function, is then used. For m variables $\{v_1, \ldots, v_m\}$, the idea is to approximate the max over $\{v_1, \ldots, v_m\}$ by a differentiable function $g_l(v_l)$, which is defined as:

$$g_l(v_l) \approx \max_l(v_l) = v^*, \tag{5}$$

$$\frac{\partial g_l(v_l)}{\partial v_l} \approx \frac{\mathbf{1}(v_i = v^*)}{\sum_l \mathbf{1}(v_l = v^*)}. \tag{6}$$

Table 1
Four softmax approximations $g_l(v_l) \approx \max_l(v_l)$.

	$g_l(v_l)$	$\partial_{g_i}(v_l)/\partial v_i$	Domain
NOR	$1-\prod_l(1-v_l)$	$\frac{1-g_i(v_l)}{1-v_i}$	$[0,1]$
GM	$\left(\frac{1}{m}\sum_l v_l^r\right)^{\frac{1}{r}}$	$g_l(v_l)\frac{v_l'^{-1}}{\sum_l v_l'}$	$[0,\infty]$
LSE	$\frac{1}{r}\ln\frac{1}{m}\sum_l \exp(rv_l)$	$\frac{\exp(rv_i)}{\sum_l \exp(rv_l)}$	$[-\infty,\infty]$
ISR	$\frac{\sum_l v_l'}{1+\sum_l v_l'}, v_l'=\frac{v_l}{1-v_l}$	$\left(\frac{1-g_l(v_l)}{1-v_i}\right)^2$	$[0,1]$

Note that for the rest of the paper, $g_l(v_l)$ indicates a function g on all variables v_l indexed by l, not merely on one variable v_l. There are a number of approximations for g. We summarize 4 models used here in Table 1: noisy-or (NOR) (Viola et al., 2005), generalized mean (GM), log-sum-exponential (LSE) (Ramon and Raedt, 2000), and integrated segmentation and recognition (ISR) (Keeler et al., 1990; Viola et al., 2005). The parameter r controls the sharpness and accuracy in the LSE and GM models *i.e.* $g_l(v_l) \to v^*$ as $r \to \infty$.

The probability bag x_i is defined as p_i, which is computed from the maximum over the probability $p_{ij} \equiv p(y_{ij}=1|x_{ij})$ of all the instances x_{ij}. Using the *softmax* g to approximate $\max_j p_i$ is defined as:

$$p_i = \max_j(p_{ij}) = g_j(p_{ij}) = g_j(\sigma(2h_{ij})), \tag{7}$$

where $h_{ij}=h(x_{ij})$, and $\sigma(v)=\frac{1}{1+\exp(-v)}$ is the sigmoid function. Note that $\sigma(v)\in[0,1]$ and $\frac{\partial\sigma}{\partial v}=\sigma(v)(1-\sigma(v))$.

Then the weight w_{ij} and the derivative $\frac{\partial\mathcal{L}}{\partial h_{ij}}$ could be written as:

$$w_{ij} = -\frac{\partial\mathcal{L}}{\partial h_{ij}} = -\frac{\partial\mathcal{L}}{\partial p_i}\frac{\partial p_i}{\partial p_{ij}}\frac{\partial p_{ij}}{\partial h_{ij}}. \tag{8}$$

w_{ij} is obtained by taking three derivatives:

$$\frac{\partial\mathcal{L}}{\partial p_i} = \begin{cases} -\frac{w_i}{p_i} & \text{if } y=1; \\ \frac{w_i}{1-p_i} & \text{if } y=-1. \end{cases} \tag{9}$$

$$\frac{\partial p_i}{\partial p_{ij}} = \begin{cases} \frac{1-p_i}{1-p_{ij}} & \text{NOR;} \quad p_i\frac{(p_{ij})^{r-1}}{\sum_j (p_{ij})^r} \quad \text{GM;} \\ \frac{\exp(rp_{ij})}{\sum_j \exp(rp_{ij})} & \text{LSE;} \quad \left(\frac{1-p_i}{1-p_{ij}}\right)^2 \quad \text{ISR.} \end{cases} \tag{10}$$

$$\frac{\partial p_{ij}}{\partial h_{ij}} = 2p_{ij}(1-p_{ij}). \tag{11}$$

Once we obtain h_t, the weight α_t can be found via a line search, which aims to minimize $\mathcal{L}(h)$. Finally, we combine multiple weak learners into a single strong classifier *i.e.* $\mathbf{h} \leftarrow \mathbf{h} + \alpha_t h_t$. Algorithm 2 illustrates the details of MIL-Boost. The parameter T is the number of weak classifiers in AnyBoost (Mason et al., 2000).

Algorithm 2. MIL-Boost

Input: Bags $\{x_1,\ldots,x_n\},\{y_1,\ldots,y_n\},T$
Output: h
for $t=1\to T$ **do**
  Compute weights $w_{ij} = -\frac{\partial\mathcal{L}}{\partial p_i}\frac{\partial p_i}{\partial p_{ij}}\frac{\partial p_{ij}}{\partial h_{ij}}$
  Train weak classifiers h_t using weights $|w_{ij}|$
  $h_t = \underset{h}{\text{argmin}} \sum_{ij} \mathbf{1}(h(x_{ij})\neq y_i)|w_{ij}|$
  Find α_t via line search to minimize $\mathcal{L}(h)$
  $\alpha_t = \underset{\alpha}{\text{argmin}}\ \mathcal{L}(h+\alpha h_t)$
  Update strong classifiers $h \leftarrow h+\alpha_t h_t$
end for

3.2. Multiple cluster assumption

Multiple cancer subtypes with different morphological characteristics might co-exist in a histopathology image. The single model/cluster/classifier in the previous MIL method is not capable of taking the different types into consideration. A key component of our work is to embed clustering into the MIL setting to classify the segmented regions into different cancer subtypes. Although there are many individual classification, segmentation and clustering approaches in the medical imaging and computer vision community, none of these algorithms meet our requirement since they are designed for doing only one of the three tasks. Here we simultaneously perform three tasks in an integrated system under weakly supervised learning framework.

We integrate the clustering concept into the MIL setting by assuming the existence of hidden variable $y_{ij}^k \in \mathcal{Y}$ which denotes whether the instance x_{ij} belongs to the kth cluster. If an instance belongs to one of K clusters, this instance is considered as a positive instance; if at least one instance in a bag is labeled as positive, the bag is considered as positive. This forms the MCIL assumption, which is formulated as:

$$y_i = \max_j \max_k\left(y_{ij}^k\right). \tag{12}$$

Again the max is equivalent to an OR operator where $\max_k\left(y_{ij}^k\right)=1 \iff \exists k,\ \text{s.t. } y_{ij}^k=1$.

Based on this multiple cluster assumption, next we discuss the proposed MCIL method. The differences among fully supervised learning, MIL, and MCIL are illustrated in Fig. 3. The goal of MCIL is to discover and split the positive instances into K groups by learning K instance-level classifiers $h^k(x_{ij}):\mathcal{X}\to\mathcal{Y}$ for K clusters, given only bag-level supervision y_i. The corresponding bag-level classifier for the kth cluster is then $H^k(x_i):\mathcal{X}^m\to\mathcal{Y}$. The overall image-level classifier is denoted as $H(x_i):\mathcal{X}^m\to\mathcal{Y}$:

$$H(x_i) = \max_k H^k(x_i) = \max_k \max_j h^k(x_{ij}) \tag{13}$$

3.3. The MCIL method

In this section, based on the previous derivations, we give the full formulation of our MCIL method. The probability $p_i \equiv p(y_i=1|x_i)$ now is computed as the *softmax* of the probability $p_{ij}\equiv p(y_{ij}=1|x_{ij})$ of all the instances x_{ij}; the p_{ij} is obtained as the *softmax* of $p_{ij}^k = p^k(y_{ij}=1|x_{ij})$, which measures the probability of the instance x_{ij} belonging to the kth cluster. Thus, using the *softmax* g in place of the *max* in Eq. (12) we compute the bag probability as:

$$p_i = g_j(p_{ij}) = g_j(g_k(p_{ij}^k)) \tag{14}$$

$$g_j\left(g_k\left(p_{ij}^k\right)\right) = g_{jk}\left(p_{ij}^k\right) = g_k\left(g_j\left(p_{ij}^k\right)\right) \tag{15}$$

$$p_i = g_{jk}\left(\sigma\left(2h_{ij}^k\right)\right), \tag{16}$$

where $h_{ij}^k = h^k(x_{ij})$. Again, the function of $g_k(p_{ij}^k)$ can be deduced from Table 1; it indicates a function g which takes all p_{ij}^k indexed by k; similarly, $g_{jk}\left(p_{ij}^k\right)$ could be understood as a function g including all p_{ij}^k indexed by k and j. Verification of this equation is shown in Remark 1 in Appendix A.

The next step is to compute w_{ij}^k with derivative: $w_{ij}^k = -\frac{\partial\mathcal{L}}{\partial h_{ij}^k}$. Using the chain rule we get:

$$w_{ij}^k = -\frac{\partial\mathcal{L}}{\partial h_{ij}^k} = -\frac{\partial\mathcal{L}}{\partial p_i}\frac{\partial p_i}{\partial p_{ij}^k}\frac{\partial p_{ij}^k}{\partial h_{ij}^k}. \tag{17}$$

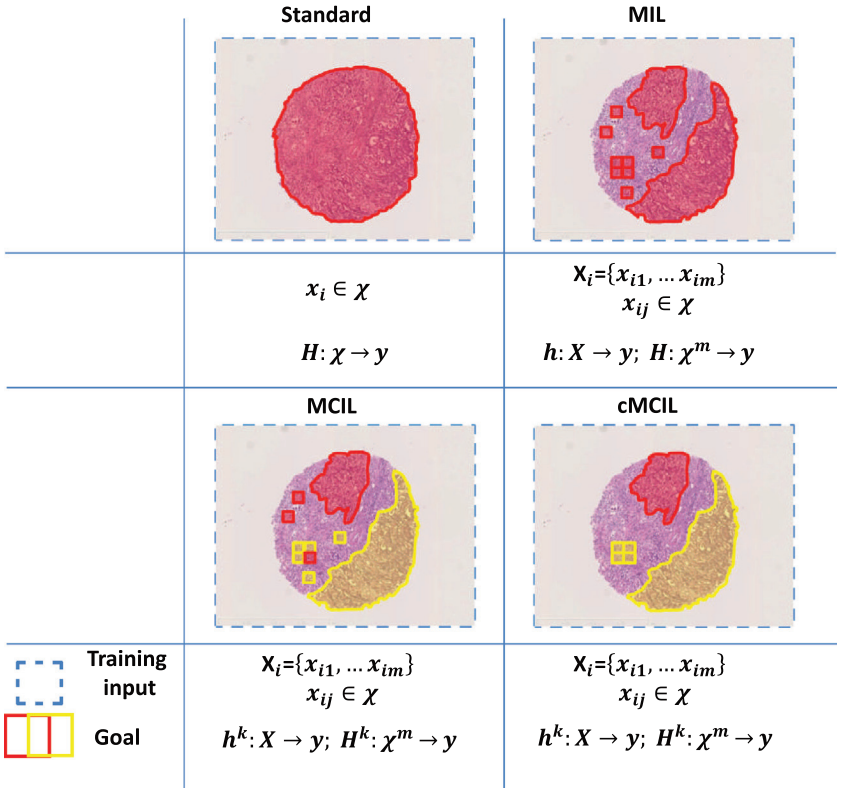


Fig. 3. Distinct learning goals between (a) Standard supervised learning, (b) MIL, (c) MCIL and (d) cMCIL. MCIL and cMCIL could perform image-level classification ($(x_i \to \{-1,1\})$), patch-level segmentation ($x_{ij} \to \{-1,1\}$) and patch-level clustering $\left(x_{ij} \to \left\{y_{ij}^1, \ldots, y_{ij}^K\right\}, y_{ij}^k \in \{-1,1\}\right)$. cMCIL studies the contextual prior information among the instances within the framework of MCIL and correctly recognizes noises and small isolated areas. Red and yellow squares and regions represent different type of cancer tissues. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
MCIL w_{ij}^k/w_i with different *softmax* functions.

w_{ij}^k/w_i	$y_i = -1$	$y_i = 1$
NOR	$-2p_{ij}^k$	$\frac{2p_{ij}^k(1-p_i)}{p_i}$
GM	$-\frac{2p_i}{1-p_i}\frac{\left(p_{ij}^k\right)^r-\left(p_{ij}^k\right)^{r+1}}{\sum_{j,k}\left(p_{ij}^k\right)^r}$	$2\frac{\left(p_{ij}^k\right)^r-\left(p_{ij}^k\right)^{r+1}}{\sum_{j,k}\left(p_{ij}^k\right)^r}$
LSE	$-\frac{2p_{ij}^k\left(1-p_{ij}^k\right)}{1-p_i}\frac{\exp\left(rp_{ij}^k\right)}{\sum_{j,k}\exp\left(rp_{ij}^k\right)}$	$\frac{2p_{ij}^k\left(1-p_{ij}^k\right)}{p_i}\frac{\exp\left(rp_{ij}^k\right)}{\sum_{j,k}\exp\left(rp_{ij}^k\right)}$
ISR	$-\frac{2\mathcal{X}_{ij}^k p_i}{\sum_{j,k}\mathcal{X}_{ij}^j},\ \mathcal{X}_{ij}^k=\frac{p_{ij}^k}{1-p_{ij}^k}$	$\frac{2\mathcal{X}_{ij}^k(1-p_i)}{\sum_{j,k}\mathcal{X}_{ij}^j},\ \mathcal{X}_{ij}^k=\frac{p_{ij}^k}{1-p_{ij}^k}$

The form of $\frac{\partial p_i}{\partial p_{ij}^k}$ is dependent on the choice of the *softmax* function, which can be deduced from Table 1 by replacing $g_l(v_l)$ with p_i and v_i with p_{ij}^k. Derivative $\frac{\partial \mathcal{L}}{\partial p_i}$ is the same as Eq. (9), and $\frac{\partial p_{ij}^k}{\partial h_{ij}^k}$ is expressed as:

$$\frac{\partial p_{ij}^k}{\partial h_{ij}^k} = 2p_{ij}^k\left(1-p_{ij}^k\right). \tag{18}$$

We further summarize the weights w_{ij}^k/w_i in Table 2. Recall that w_i is the given prior weight for the ith bag.

Note that p_i and $\mathcal{L}(h)$ depend on each h_{ij}^k. We optimize $\mathcal{L}(h^1,\ldots,h^k)$ using the coordinate descent method cycling through k, which is a non-derivative optimization algorithm (Bertsekas and Bertsekas, 1999). In each phase we add a weak classifier to h^k while keeping all other weak classifiers fixed. Details of the MCIL are demonstrated in Algorithm 3. The parameter K is the number of cancer subtypes, and the parameter T is the number of weak classifiers in Boosting. Notice that the outer loop is for each weak classifier while the inner loop is for the kth strong classifier.

In summary, the overall MCIL strategy can be described as follows. We introduce the latent variables y_{ij}^k, which denotes the instance x_{ij} belonging to the kth cluster; we encode the concept of clustering by re-weighting the instance-level weight w_{ij}^k. If cluster kth can classify an instance to be positive, thus the corresponding weights of the instance and bag for other clusters decrease in re-weighting. Thus, it forms a competition among different clusters.

Algorithm 3. MCIL-Boost

Input: Bags $\{x_1,\ldots,x_n\},\{y_1,\ldots,y_n\},K,T$
Output: $h^1,\ldots,h^K$
for $t=1\to T$ **do**
  for $k=1\to K$ **do**
    Compute weights $w_{ij}^k=-\frac{\partial \mathcal{L}}{\partial p_i}\frac{\partial p_i}{\partial p_{ij}^k}\frac{\partial p_{ij}^k}{\partial h_{ij}^k}$
    Train weak classifiers h_t^k using weights $|w_{ij}^k|$
    $h_t^k=\arg\min_h\sum_{ij}\mathbf{1}\left(h\left(x_{ij}^k\right)\neq y_i\right)|w_{ij}^k|$
    Find α_t via line search to minimize $\mathcal{L}(.,h^k,.)$
    $\alpha_t^k=\arg\min_\alpha\mathcal{L}\left(.,\mathbf{h}^k+\alpha h_t^k,.\right)$
    Update strong classifiers $h^k\leftarrow h^k+\alpha_t^k h_t^k$
  end for
end for

3.4. Contextual constraints

Most existing MIL methods are conducted under the assumption that instances within a bag are distributed independently, without considering the inter-dependences among instances; this leads to some degree of ambiguity. For example, an instance considered to be positive in a bag may be an isolated point or noise. In this situation, it will lead to incorrect recognition of cancer tissues. Rich contextual information has been proven to play a key role in fully supervised image segmentation and labeling (Tu and Bai, 2010). To further improve our algorithm, we take into consideration such contextual information to enhance the robustness of the MCIL. For convenience, this extension is called context-constrained multiple clustered instance learning (cMCIL). The key to the cMCIL is a formulation for introducing the neighborhood information as a prior for the MCIL. Note that the cMCIL is still implemented within the framework of the MCIL. The distinction between MCIL and cMCIL is illustrated in Fig. 3.

We define the new loss function in cMCIL as:

$$\mathcal{L}(h) = \mathcal{L}_A(h) + \lambda \mathcal{L}_B(h), \tag{19}$$

where $\mathcal{L}_A(h)$ is the standard MCIL loss function taking the form as Eq. (4). $\mathcal{L}_B(h)$ imposes a neighborhood constraints (in a way a smoothness prior) over the instances to reduce the ambiguity during training; it encourages the nearby image patches to be within the same cluster.

$$\mathcal{L}_B(h) = \sum_{i=1}^{n} w_i \sum_{(j,m)\in E_i} v_{jm} \|p_{ij} - p_{im}\|^2, \tag{20}$$

where λ weighs the importance of the current instance and its neighbors. 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 Euclidean spatial distance (on the image, denoted as d_{jm}) between them. Nearby instances have more contextual influence than instances that are far away from each other. In our experiment, we chose $v_{jm} = \exp(-d_{jm})$, such that higher weights will be put on closer pairs.

According to Eq. (19), we rewrite $\frac{\partial \mathcal{L}(h)}{\partial h_{ij}^k}$ as

$$\frac{\partial \mathcal{L}(h)}{\partial h_{ij}^k} = \frac{\partial \mathcal{L}_A(h)}{\partial h_{ij}^k} + \lambda \frac{\partial \mathcal{L}_B(h)}{\partial h_{ij}^k}, \tag{21}$$

and

$$\frac{\partial \mathcal{L}_B(h)}{\partial p_{ij}^k} = w_i \sum_{(j,m)\in E_i} 2 v_{jm} \left(p_{ij}^k - p_{im}^k\right). \tag{22}$$

we further rewrite the derivative of $w_{ij}^k = -\frac{\partial \mathcal{L}}{\partial h_{ij}^k}$ as:

$$w_{ij}^k = -\frac{\partial \mathcal{L}}{\partial h_{ij}^k} = -\left(\frac{\partial \mathcal{L}_A}{\partial p_i} \frac{\partial p_i}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k} + \frac{\partial \mathcal{L}_B}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k}\right). \tag{23}$$

The derivatives $\frac{\partial p_i}{\partial p_{ij}^k}$ and $\frac{\partial p_{ij}^k}{\partial h_{ij}^k}$ have been given previously (see the subsection of MCIL). $\frac{\partial \mathcal{L}_A(h)}{\partial p_i}$ takes the same form of $\frac{\partial \mathcal{L}(h)}{\partial p_i}$ in Eq. (9).

The optimization procedure for cMCIL is similar to MCIL. With the weight w_{ij}^k, we can train the weak classifier h_t^k by optimizing weighed error to obtain a strong classifier: $\mathbf{h}^k \leftarrow \mathbf{h}^k + \alpha_t^k h_t^k$. The details of cMCIL are similar to those of MCIL as demonstrated in Algorithm 3 except that the weight w_{ij}^k is replaced by Eq. (23).

4. Experiments

To illustrate the advantages of MCIL, we conduct experiments on two medical image datasets. In the first experiment, without loss of generality, we use colon tissue microarrays to perform joint classification, segmentation and clustering. For convenience, tissue microarrays are called histopathology images. In the second experiment, cytology images (Lezoray and Cardot, 2002) are used to further validate the effectiveness of MCIL. All the methods in the following experiments, unless particularly stated, are conducted under the same experimental settings and based on the same features, which are declared as follows.

4.1. Experiment A: colon cancer histopathology images

Settings. For the parameter setting, we set $r = 20$, and $T = 200$. As mentioned before, the parameter r controls the sharpness and accuracy in the LSE and GM model. The parameter T decides the number of weak classifiers in boosting. The parameter K decides the number of cancer classes when performing clustering task. K is set to 4 in the colon cancer image experiment because the dataset contains four kinds of cancer types. For the value of parameter λ used in the loss function of cMCIL, 0.01 is selected according to an segmentation experimental result based on a cross validation.

We assume the initial equal weights for the positive and negative training data. Under this assumption, the initial weight w_i for the ith bag is set as uniform. In our experiments, we use the GM model as the *softmax* function, except for one classification experiment part, in which we use four models for comparison. The weak classifier we use is the Gaussian function. All the experimental results are reported with 5-fold cross validation. The number of training data and test data are always the half of the total number of all the data used in the experiment.

Features. Each instance is represented by a feature vector. In this work we focus on an integrated learning formulation rather than the feature design. Also to demonstrate the generality of our framework, we opt for general features instead of adopting or creating our own disease specific features. Specifically, we use widely adopted features including $L^*a^*b^*$ Color Histogram, Local Binary Pattern (Ojala et al., 2002; Ahonen et al., 2009), and SIFT (Lowe, 2004). Note that designing disease specific features is an interesting and challenging research topic itself due to the fact that cell appearance of different types of cancers may be very difference in terms of shape, size and so on. While using disease specific features may potentially improve the performance further, we leave it for future work.

In histopathology images, recent studies use some common and useful features from gray-level co-occurrence matrix (GLCM), Gabor filters, multiwavelet transforms, and fractal dimension

Table 3
Number of images in the subsets.

	NC	MTA	PTA	MA	SRC
Binary	30	30	0	0	0
multi1	30	15	9	0	6
multi2	30	13	9	8	0
multi3	50	28	8	8	6

Table 4
Run time in various algorithms (min).

	cMCIL	MCIL	MKL	MIL-Boost	Boosting	mi	MI
Features	90	90		90	5	90	90
Model	35	32		8	2	15	16
Total	125	122	70 h	95	7	105	106
Language	C++	C++	Matlab/C	C++	C++	JAVA	JAVA

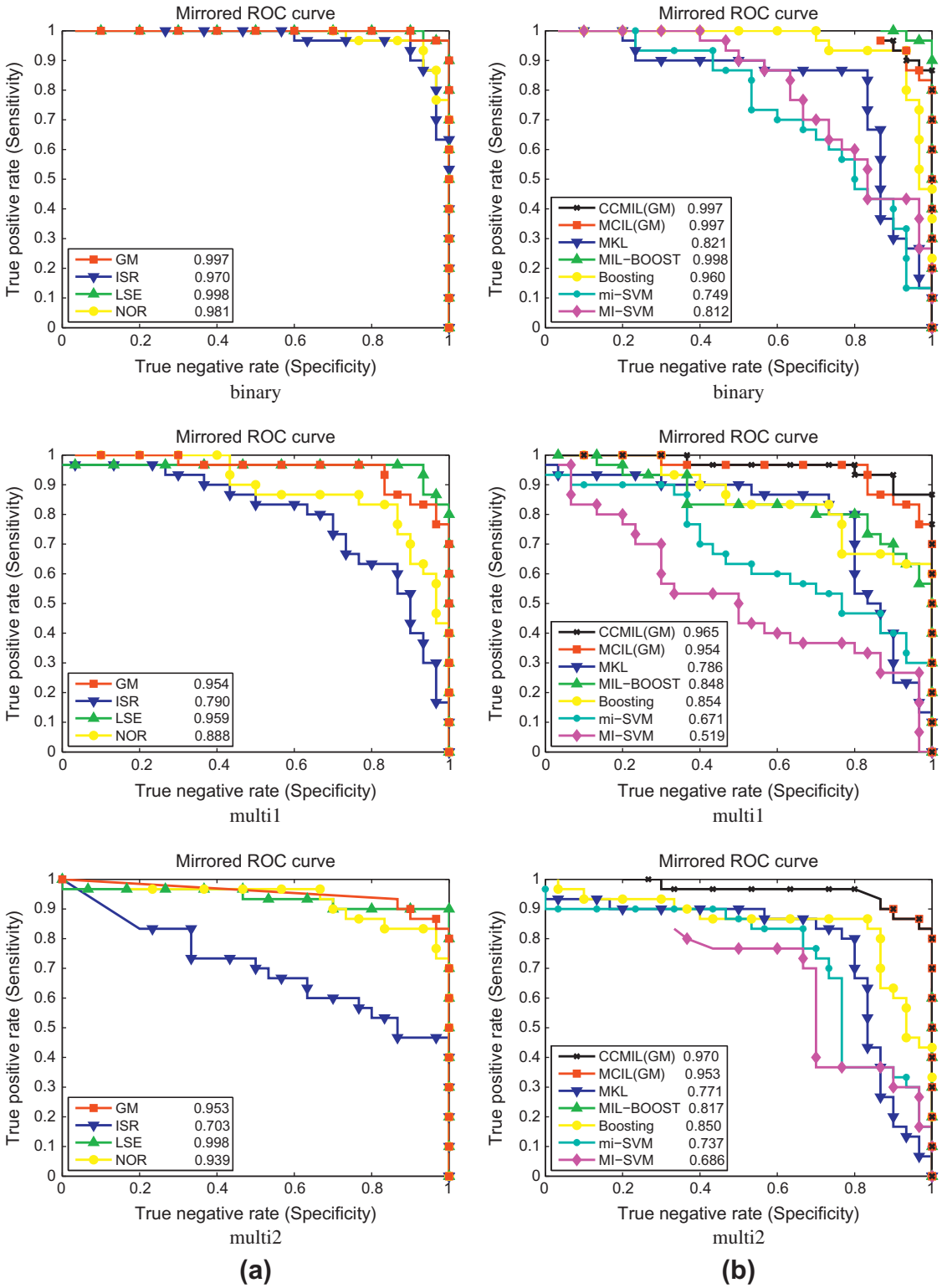


Fig. 4. ROC curves for classification in (a and b): (a) ROC curves for four *softmax* models in MCIL; LSE model and GM model fit the best for the cancer image recognition task. (b) Comparisons of image (bag)-level classification results with state-of-the-art methods on the three datasets: ROC curves for different learning methods; our proposed methods have the apparent advantages.

texture features (Huang and Lee, 2009). Therefore, we also added the similar features.

Datasets. Colon histopathology images with four cancer types are used, including Moderately or well differentiated tubular adenocarcinoma (MTA), Poorly differentiated tubular adenocarcinoma (PTA), Mucinous adenocarcinoma (MA), and Signet-ring carcinoma (SRC). These four types are the most common types in colon cancer. Combined with the Non-cancer images (NC), five classes of colon histopathology images are used in the experiments. We use the same abbreviations for each type in the following sections.

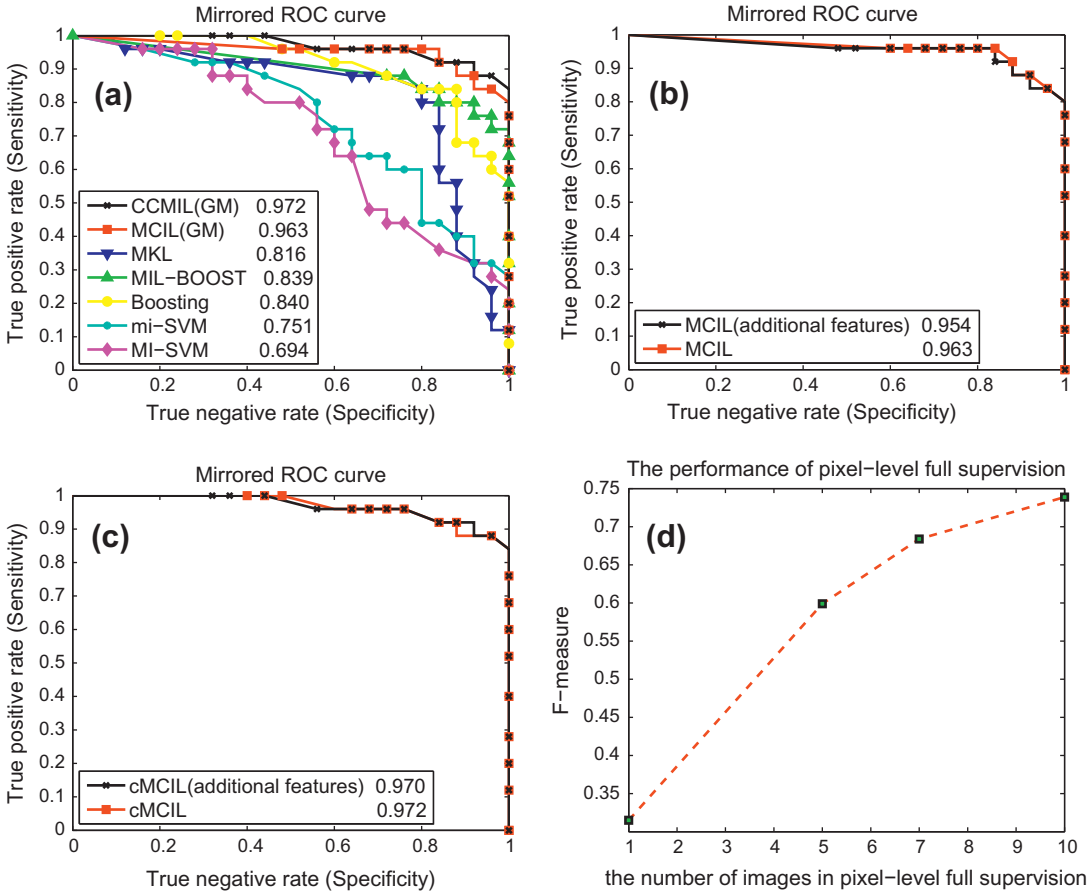


Fig. 5. ROC curves for classification on *multi*3 in (a–c). (a) Comparison with state-of-the-art methods based on the new feature set. (b and c) Comparison of MCIL/cMCIL based on two different feature set. (d) The F-measures for segmentation at varying number of images with pixel-level full supervision.

To better reflect the real world situation, we designed our dataset in an unbalanced way to match the actual distribution of the four types of cancer. According to national cancer institute (http://seer.cancer.gov/), the incidence of Moderately or well differentiated tubular adenocarcinoma accounts for 70–80%, Poorly differentiated tubular adenocarcinoma accounts for 5%, Mucinous adenocarcinoma accounts for 10%, and Signet-ring carcinoma accounts for less than 1%. The images are obtained from the NanoZoomer 2.0HT digital slice scanner produced by Hamamatsu Photonics with a magnification factor of 40. In total, we obtain 50 non-cancer (NC) images and 53 cancer images. First we down-sample the images by 5 times to reduce the computational overhead. Our segmentation therefore is conducted on the down-sampled images rather than the original images. We then densely extract patches from each image. The size of each patch is 64×64. The overlap step size is 32 pixels for training and 4 pixels for the inference. Note that each patch corresponds to an instance, which is represented by a feature vector.

We use all the images to construct four different subsets: *binary*, *multi*1, *multi*2, and *multi*3. The constituents of the four subsets are shown in Table 3. In the first three subsets, each subset contains 60 different histopathology images. *binary* refers to the subset containing only two classes: the NC class and the MTA class. It contains 30 non-cancer and 30 cancer images, and can be used to test the capability of cancer image detection. *multi*1 and *multi*2 each includes three types of cancer images and one type of non-cancer images. *multi*3 contains all four types of images. In the all four subsets, we demonstrate the advantage of the MIL formulations against the state-of-the-art supervised image categorization approaches. In *multi*2, we further show the advantage of MCIL in an integrated classification/segmentation/clustering framework.

Annotations. To ensure the quality of the ground truth annotations, images are carefully studied and labeled by well-trained experts. Specifically, each image is independently annotated by two pathologists; the third pathologist moderates their discussion until they reach the final agreement on the result. All images are labeled as cancer images or non-cancer images. Furthermore, for the cancer image, cancer tissues are annotated and their corresponding cancer subtypes are identified.

4.1.1. Image-level classification

In the experiment, we measure the image-level classification for being cancer or non-cancer images. First, the performance of the MCIL method based on different *softmax* models as mentions in Table 1 are compared.

Second, to evaluate the performance of our methods, several methods are implemented as baseline for comparison in this experiment. Since the source codes of most algorithms presented in the colon cancer image analysis literature are not always available, the image classification baseline we use here is multiple kernel learning (MKL) (Vedaldi et al., 2009) which obtains very competitive image classification results and wins the PASCAL Visual Object Classes Challenge 2009 (VOC2009) (Everingham et al., 2009). We use their implementation and the same parameters reported in their paper. For the MIL baselines, we use MI-SVM (Andrews et al., 2003), mi-SVM (Andrews et al., 2003), and

MIL-Boost (Viola et al., 2005). Moreover, we use all the instances x_{ij} to train a standard Boosting (Mason et al., 2000) by considering instance-level labels derived from bag-level labels ($y_{ij} = y_i,\ i = 1,\ldots,n,\ j = 1,\ldots,m$).

In total seven methods for colon cancer image classification are compared, including cMCIL, MCIL, MKL, MIL-BOOST, Boosting, mi-SVM and MI-SVM. Notice that MKL utilizes more discriminative features than what we use in MIL, MCIL and cMCIL, including the distribution of edges, dense and sparse visual words, and feature descriptors at different levels of spatial organization.

Moreover, to further validate the methods, special experiments on *multi*3 is conducted. In these experiments, some other features, including Hu moment and gray-level co-occurrence matrix (GLCM) (Sertel et al., 2009), are added into the original feature set to demonstrate how the feature set influences the classification result.

Computational complexity. The machine (Processor: Intel (R) Core (TM)2 Quad CPU Q9400 @ 2.66 GHz 2.67 GHz; RAM: 8G; 64 Operating System) is used to evaluate the computational complexity. The data set Multi2 is used in the experiment. The feature code is C++ implementation in all these algorithms except MKL. The MKL code, including features and models, is MATLAB/C implementation from.[1] The mi-SVM and MI-SVM codes are JAVA implementation from.[2] The other codes are C++ implementation written by the authors. Table 4 shows time consuming from various algorithms. Noted that mi means mi-SVM and MI means MI-SVM. The numerical unit is minute except MKL using hour. For the computational complexity, it takes several days to train an MKL classifier for a dataset containing 60 images while it only takes about several hours using an ensemble of MIL. Compared with MIL and MCIL, because MCIL adds a loop, the training time of MCIL is more than that of MIL. The time of cMCIL is slightly more than that of MCIL due to the different loss function.

Evaluation. Receiver operating characteristic (ROC) curve is used to evaluate the performance of classification. The larger the area under the curve is, the better the corresponding classification method is.

Results. The ROC curves for four *softmax* models in MCIL are shown in Fig. 4a. According to the curves shown in the figure, it is safely to say that the LSE model and GM model fit the best for the cancer image recognition task, which is the reason why GM model is chosen in all the following experiments.

Fig. 4b shows the ROC curves for different learning methods in the three datasets. In the dataset *binary*, cMCIL, MCIL and MIL-Boost outperform well than developed MKL algorithm (Vedaldi et al., 2009) and standard Boosting (Mason et al., 2000), which shows the advantage of the MIL formulation to the cancer image classification task. cMCIL, MCIL and MIL-Boost achieve similar performance on the *binary* dataset of one class/cluster; however, when applied to the datasets *multi*1 and *multi*2, cMCIL and MCIL significantly outperform MIL-Boost, MKL, and Boosting. This reveals that the multiple clustering concept integrated in the MCIL/cMCIL framework is able to successfully deal with the complex situation in cancer image classification.

Fig. 5 further demonstrates the advantages of MCIL/cMCIL framework than other methods. Furthermore, the three results in the figure show that MCIL/cMCIL method based on new feature set can hardly outperform well than the method based on the old feature set that is very general and small. This result demonstrate that the MCIL/cMCIL method effective to detect cancer image using general feature set rather than using special medical features.

Discussion. In classification, we show the performance of both MCIL and cMCIL compared to others. Note that the performance of cMCIL (F-measure: 0.972) is almost identical to that of MCIL (F-measure: 0.963). This is expected because the contextual models mainly improve patch-level segmentation and have little effect on classification.

Different cancer types, experiment settings, benchmarks, and evaluation methods are reported in the literature. As far as we know, the code and images used in Huang and Lee (2009), Tabesh et al. (2007), and Esgiar et al. (2002) are not publicly accessible.[3] Hence, it is quite difficult to make a direct comparison between different algorithms. Below we only list their results as references. In Huang and Lee (2009), 205 pathological images of prostate cancer were chosen as evaluation which included 50 of grade 1–2, 72 of grade 3, 31 of grade 4, and 52 of grade 5. The highest correct classification rates based on Bayesian, KNN and SVM classifiers achieved 94.6%, 94.2% and 94.6% respectively. In Tabesh et al. (2007), 367 prostate images (218 cancer and 149 non-cancer) were chosen to detect cancer or non-cancer. The highest accuracy was 96.7%. 268 images were chosen to classify Gleason grading. The numbers of grades 2–5 are 21, 154, 86 and 7, respectively. The highest accuracy was 81%. In Esgiar et al. (2002), a total of 44 non-cancer images and 58 cancer images were selected to detect cancer or non-cancer. The sensitivity of 90–95% and the specificity of 86–93% were achieved according to various features.

Table 5
Colon cancer image segmentation results in F-measure of four methods. Note that standard Boosting (Mason et al., 2000) is trained under the image-level supervision.

Method	Standard boosting	MIL-Boost	MCIL	cMCIL
F-measure	0.312	0.253	0.601	0.717

4.1.2. Image segmentation

We now turn to an instance-level experiment. We report instance-level results in the dataset *multi*2 that contains 30 cancer images and 30 non-cancer images in total. Instance-level annotations for cancer images are provided by three pathologists with the procedure (two pathologists marking up and one more pathologist mediating the decision) described before.

Unsupervised segmentation techniques cannot be used as a direct comparison here since they cannot output labels for each segment. The segmentation baselines are MIL-Boost (Viola et al., 2005) and standard Boosting (Mason et al., 2000), both taking the image-level labeling as supervision. Moreover, in order to compare with the fully supervised approach with pixel-wise annotation, we provide a pixel-level full supervision method by implementing a standard Boosting method that takes the pixel-level labeling as supervision (require laborious labeling work). Experiment on varying numbers $(1, 5, 7, 10)$ of images of pixel-level full supervision are conducted.

Evaluation. For a quantitative evaluation, the F-measure is used to evaluate the segmentation result. Each approach generates a probability map P_i for each bag (image) x_i and the corresponding ground truth map is named as G_i. Then we compute F-measure as follows: $\text{Precision} = |P_i \cap G_i|/|P_i|$, $\text{Recall} = |P_i \cap G_i|/|G_i|$ and $\text{F} - \text{measure} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$.

Results and discussion. Table 5 shows the F-measure values of four methods, cMCIL, MCIL, MIL-Boost and standard Boosting. Again, standard Boosting is a supervised learning baseline that utilizes image-level supervision by treating all the pixels in the positive and negative bags as positive and negative instances respec-

[1] http://www.robots.ox.ac.uk/vgg/software/MKL/.

[2] http://weka.sourceforge.net/doc.packages/multiInstanceLearning/weka/classifiers/mi/package-summary.html.

[3] We have also tried to contact many authors working on medical segmentation related to our topic to validate our method. Unfortunately, they either did not answer our email, cannot share the data with us, or tell us that their method will fail in our task.

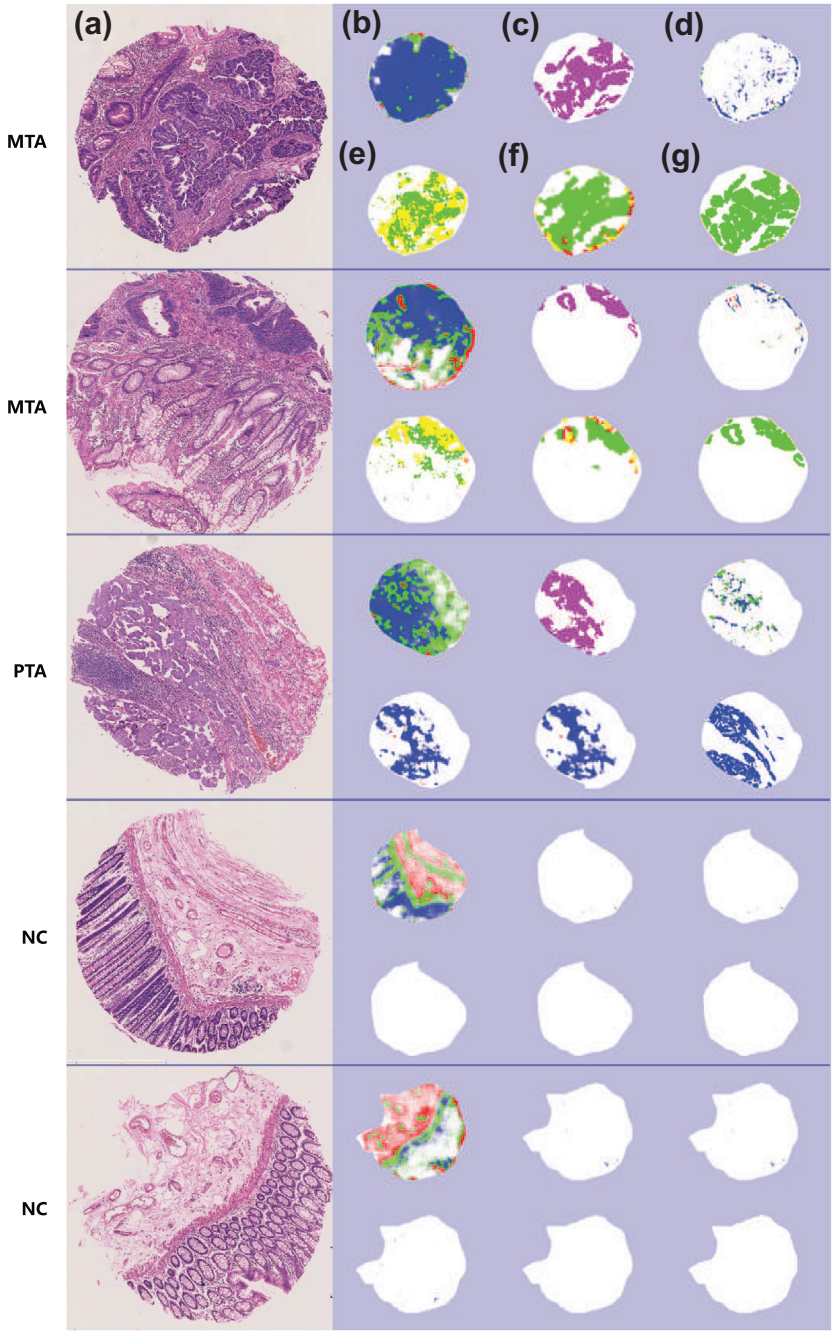


Fig. 6. Image types: from left to right. (a) The original images. (b–f) The instance-level results (pixel-level segmentation and patch-level clustering) for standard Boosting + K-means, pixel-level full supervision, MIL + K-means, MCIL and cMCIL. (g) The instance-level ground truth labeled by three pathologists. Different colors stand for different types of cancer tissues. Cancer types: from top to bottom: MTA, MTA, PTA, NC, and NC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tively. The high F-measure values of cMCIL display the great advantage of contextual constraints over previous MIL-based methods. We introduce context constraints as a prior for multiple instance learning (cMCIL), which significantly reduces the ambiguity in weak supervision (a 20% gain).

Fig. 6 shows some segmentation results of test data. According to the test results, standard Boosting with image-level supervision tends to detect non-cancer tissues as cancer tissues since it considers all the instances in positive bags as positive instances.

Since our learning process is based on image-level labels, the intrinsic label (cancer vs. non-cancer) for each patch/pixel is ambiguous. Using contextual information therefore can reduce the ambiguity on the i.i.d. (independently identically distributed) assumption. Compared with MCIL, cMCIL improves segmentation quality by reducing the intrinsic training ambiguity. Due to neighborhood constraints, cMCIL is able to reduce noises and identify small isolated areas in cancer images to achieve cleaner boundaries.

The corresponding F-measure values of the varying numbers of images of pixel-level full supervision are shown in Fig. 5d, which demonstrates that cMCIL is able to achieve comparable results (around 0.7) but without having detailed pixel-level manual annotations. Although our weakly supervised learning method requires more images (30 positive), it eases the burden of making the

pixel-wise manual annotation. In our case, it often takes 2–3 h for our expert pathologists to reach the agreement on the pixel-level ground truth while it usually costs only 1–2 min to label an image as cancerous or non-cancerous.

4.1.3. Patch-level clustering

With the same test data mentioned in segmentation, we also obtained the clustering results. For patch-level clustering, we build two baselines: MIL-Boost (Viola et al., 2005) + K-means and standard Boosting + K-means. Particularly, we first run MIL-Boost or standard Boosting to perform instance-level segmentation and then use K-means to obtain K clusters among positive instances (cancer tissues). Since we mainly focus on clustering performance here, we only include true positive instances.

Evaluation. The purity measure is used as the evaluation metric. Given a particular cluster S_r of size n_r, the purity is defined as the weighted sum of the individual cluster purities:$purity = \sum_{r=1}^{k} \frac{n_r}{n} Pu(S_r)$, where $Pu(S_r)$ is the purity of a cluster, defined as $Pu(S_r) = \frac{1}{n_r} \max_i n_r^i$. Larger purity values indicate better clustering results.

Results and discussion. The purities of cMCIL and MCIL are respectively 99.74% and 98.92%, while the purities of MIL-Boost + K-means and standard Boosting + K-means are only 86.21% and 84.37% respectively. This shows that an integrated learning framework of MCIL is better than separating the two-steps, instance-level segmentation and clustering.

We also illustrate the clustering results in Fig. 6. As shown in the figure, MCIL and cMCIL successfully discriminate cancer classes. The original MCIL method divides MTA cancer images into three clusters. Compared with MCIL, the patch-level clustering is less noisy in cMCIL. The PTA cancer tissues are mapped to blue; the MTA cancer tissues are mapped to green, yellow and red. Both MIL-Boost + K-means and standard Boosting + K-means divide one tissue class into several clusters and the results are not consistent. In the histopathology images, the purple regions around cancers are lymphocytes. For some patients, it is common that lymphocytes occur around the cancer cells and seldom appear around non-cancerous tissues although lymphocytes themselves are not considered as cancer tissues. Since a clear definition of all classes is still not available, our method shows the promising potential for automatically exploring different classes with weak supervision.

4.2. Experiment B: cytology images

Datasets. Ten cytology images together with their corresponding segmentation results (as the ground truth) are obtained from the paper (Lezoray and Cardot, 2002). We also generate additional ten background (negative) images. These images have the same background texture as the ten cytology images but without cells on them. Details of the method for texture image generation are presented in Portilla and Simoncellt (2000), in which a universal parametric model for visual texture, based on a novel set of pairwise joint statistical constraints on the coefficients of a multiscale image representation is described. For convenience, we name the cytology image as cell image (CELL) and texture image as background image (BG).

Experiments design. To evaluate the pixel-level segmentation, we test these 20 images with 4 different methods, including pixel-level full supervision, MIL-Boost, MCIL, and cMCIL. All the four methods correctly classify the 20 images into the cell image and background image. Since all nuclei belong to the same type, the cluster concept that divides different instances into different classes is rather weak in this case. Therefore, in Experiment B we focus on the segmentation task.

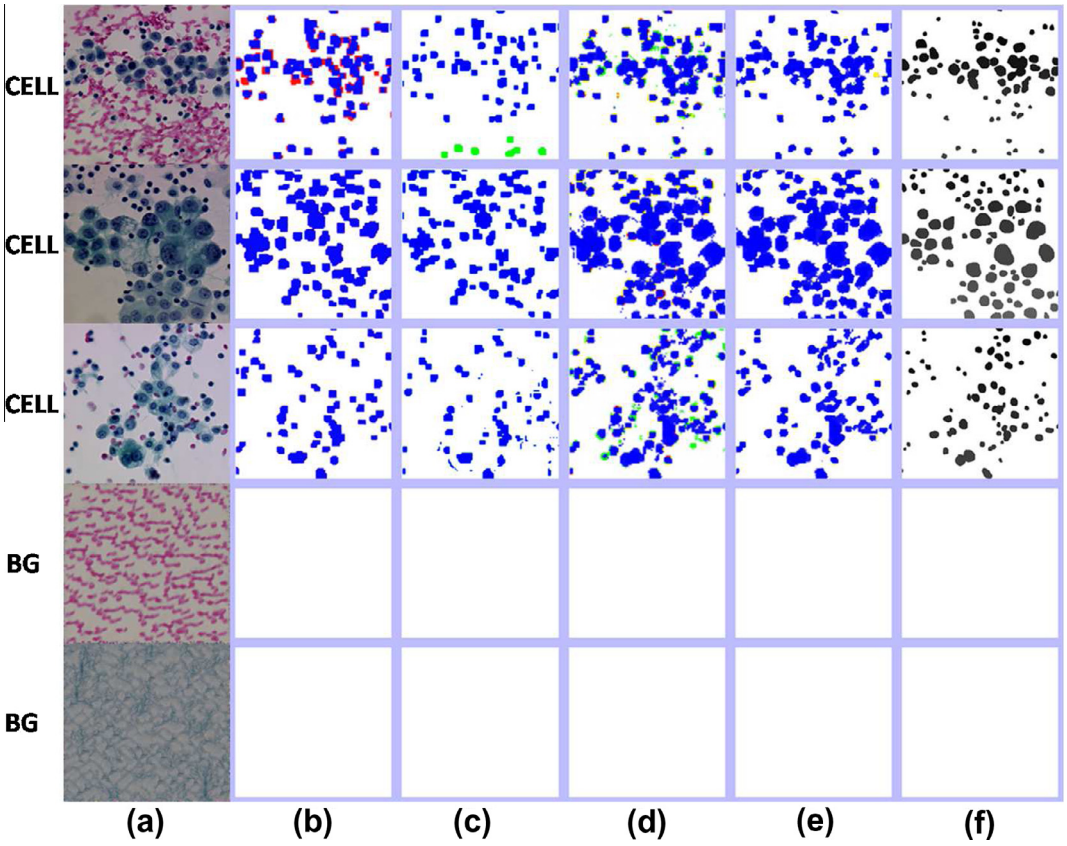


Fig. 7. Image types: from left to right. (a) The original cell images. (b–e) The segmentation results for pixel-level fully supervision, MIL-Boost, MCIL and cMCIL. (f) The ground truth images. The two bottom images are generated background images. Cytology image classes: from top to bottom: CELL, CELL, CELL, BG and BG.

Table 6
Cytology image segmentation results in F-measure of different methods.

Method	Full supervision	MIL-Boost	MCIL	cMCIL
F-measure	0.766	0.658	0.673	0.699

Results and discussion. The results are shown in Fig. 7. Same as before, supervised method with the full pixel-level supervision achieves the best performance. By comparing weakly supervised methods in Fig. 7, we observe: (1) some nuclei are missed by MIL-Boost; (2) MCIL removes some errors but also brings up noises; and (3) cMCIL further improves the results by reducing the intrinsic training ambiguity. The F-measures calculated for a quantitative evaluation are shown on Table 6, which is consistent to the qualitative illustration in Fig. 7.

The experimental results demonstrate the effectiveness of cMCIL in cytology image segmentation. MCIL significantly improves segmentation over other weakly supervised methods and it is able to achieve accuracy comparable with a fully supervised state-of-the-art method.

5. Conclusion

In this paper, we have presented an integrated formulation, multiple clustered instance learning (MCIL), for classifying, segmenting, and clustering medical images along the line of weakly supervised learning. The advantages of MCIL are evident over the state-of-the-art methods that perform the individual tasks, which include easing the burden of manual annotation in which only image-level label is required and perform image-level classification, pixel-level segmentation and patch-level clustering simultaneously.

In addition, we introduce contextual constraints as a prior for MCIL which reduces the ambiguity in MIL. MCIL and cMCIL are able to achieve comparable results in segmentation with an approach of full pixel-level supervision in our experiment. This will inspire future research in applying different families of joint instance models (conditional random fields (Lafferty et al., 2001), max-margin Markov network (Taskar et al., 2003), etc.) to the framework of MIL/MCIL, as the independence assumption might be loose.

Acknowledgments

This work was supported by Microsoft Research Asia (MSR Asia). The work was also supported by NSF CAREER award IIS-0844566 (IIS-1360568), NSF IIS-1216528 (IIS-1360566), and ONR N000140910099. It was also supported by MSRA eHealth grant, and Grant 61073077 from National Science Foundation of China and Grant SKLSDE-2011ZX-13 from State Key Laboratory of Software Development Environment in Beihang University in China. We would like to thank Department of Pathology, Zhejiang University in China for providing data and help.

Appendix A. Verification for Remark 1

We verify Remark 1 (Eq. (15)): $g_j\left(g_k\left(p_{ij}^k\right)\right) = g_{jk}\left(p_{ij}^k\right) = g_k\left(g_j\left(p_{ij}^k\right)\right)$ for each model. Given the number of clusters K and the number of instances m in each bag, we develop derivations for four models respectively:

For the NOR model:

$$g_k g_j\left(p_{ij}^k\right) = 1 - \prod_k \left(1 - \left(1 - \prod_j p_{ij}^k\right)\right) = 1 - \prod_k \left(\prod_j p_{ij}^k\right) = 1 - \prod_{j,k} p_{ij}^k = g_{jk}\left(p_{ij}^k\right) \tag{A.1}$$

For the GM model:

$$g_k g_j(p_{ij}^k) = \left(\frac{1}{K}\sum_k \left(p_i^k\right)^r\right)^{\frac{1}{r}} = \left(\frac{1}{K}\sum_k \left(\left(\frac{1}{m}\sum_j \left(p_{ij}^k\right)^r\right)^{\frac{1}{r}}\right)^r\right)^{\frac{1}{r}} = \left(\frac{1}{Km}\sum_{j,k}\left(p_{ij}^k\right)^r\right)^{\frac{1}{r}} = g_{jk}\left(p_{ij}^k\right) \tag{A.2}$$

For the LSE model:

$$g_k g_j\left(p_{ij}^k\right) = \frac{1}{r}\ln\left(\frac{1}{K}\sum_k \exp\left(rp_i^k\right)\right) = \frac{1}{r}\ln\left(\frac{1}{K}\sum_k \exp\left(r\frac{1}{r}\ln\left(\frac{1}{m}\sum_j \exp\left(rp_{ij}^k\right)\right)\right)\right) = \frac{1}{r}\frac{1}{Km}\sum_{j,k}\exp\left(rp_{ij}^k\right) = g_{jk}\left(p_{ij}^k\right) \tag{A.3}$$

For the ISR model:

$$g_k g_j\left(p_{ij}^k\right) = \sum_k \frac{p_i^k}{1-p_i^k} \Big/ \left(1 + \sum_k \frac{p_i^k}{1-p_i^k}\right) \tag{A.4}$$

$$\sum_k \frac{p_i^k}{1-p_i^k} = \sum_k \frac{\sum_j \frac{p_{ij}^k}{1-p_{ij}^k} \Big/ \left(1 + \sum_j \frac{p_{ij}^k}{1-p_{ij}^k}\right)}{1 - \sum_j \frac{p_{ij}^k}{1-p_{ij}^k} \Big/ \left(1 + \sum_j \frac{p_{ij}^k}{1-p_{ij}^k}\right)} = \sum_{j,k} \frac{p_{ij}^k}{1-p_{ij}^k} \tag{A.5}$$

$$g_k g_j\left(p_{ij}^k\right) = \frac{\sum_k \frac{p_i^k}{1-p_i^k}}{1 + \sum_k \frac{p_i^k}{1-p_i^k}} = \frac{\sum_{j,k} \frac{p_{ij}^k}{1-p_{ij}^k}}{1 + \sum_{j,k} \frac{p_{ij}^k}{1-p_{ij}^k}} = g_{jk}\left(p_{ij}^k\right) \tag{A.6}$$

Now we show $g_{jk}\left(p_{ij}^k\right) = g_k g_j\left(p_{ij}^k\right)$ for each *softmax* models. $g_{jk}\left(p_{ij}^k\right) = g_j g_k\left(pij^k\right)$ could also be given in the same way. Thus Remark 1 (Eq. (15)) could be verified.

References

Ahonen, T., Matas, J., He, C., Pietikäinen, M., 2009. Rotation invariant image description with local binary pattern histogram fourier features. In: Scandinavian Conference on Image Analysis.

Altunbay, D., Cigir, C., Sokmensuer, C., Gunduz-Demir, C., 2010. Color graphs for automated cancer diagnosis and grading. IEEE Trans. Biomed. Eng. 57, 665–674.

Andrews, S., Tsochantaridis, I., Hofmann, T., 2003. Support vector machines for multiple-instance learning. In: Advances in Neural Information Processing Systems.

Artan, Y., Haider, M.A., Langer, D.L., van der Kwast, T.H., Evans, A.J., Yang, Y., Wernick, M.N., Trachtenberg, J., Yetik, I.S., 2010. Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields. IEEE Trans. Image Process. 19, 2444–2455.

Artan, Y., Haider, M.A., Langer, D.L., van der Kwast, T.H., Evans, A.J., Yang, Y., Wernick, M.N., Trachtenberg, J., Yetik, I.S., 2012. A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. IEEE Trans. Biomed. Eng. 59, 1205–1218.

Babenko, B., Dollár, P., Tu, Z., Belongie, S., 2008. Simultaneous learning and alignment: multi-instance and multi-pose learning. In: European Conference on Computer Vision Workshop on Faces in Real-Life Images.

Babenko, B., Yang, M.H., Belongie, S., 2011. Robust object tracking with online multiple instance learning. IEEE Trans. Pattern Anal. Mach. Intell. 33, 1619–1632.

Bertsekas, D.P., Bertsekas, D.P., 1999. Nonlinear Programming, second ed. Athena Scientific.

Boucheron, L.E., 2008. Object- and Spatial-Level Quantitative Analysis of Multispectral Histopathology Images for Detection and Characterization of Cancer. Ph.D. thesis. University of California, Santa Barbara.

Dietterich, T., Lathrop, R., Lozano-Pérez, T., 1997. Solving the multiple instance problem with axis-parallel rectangles. Artif. Intell. 89, 31–71.

Dollár, P., Babenko, B., Belongie, S., Perona, P., Tu, Z., 2008. Multiple component learning for object detection. In: European Conference on Computer Vision.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification, second ed. Wiley-Interscience.

Dundar, M., Fung, G., Krishnapuram, B., Rao, B., 2008. Multiple instance learning algorithms for computer aided diagnosis. IEEE Trans. Biomed. Eng. 55, 1005–1015.

Dundar, M., Badve, S., Raykar, V., Jain, R., Sertel, O., Gurcan, M., 2010. A multiple instance learning approach toward optimal classification of pathology slides. In: International Conference on Pattern Recognition, pp. 2732–2735.

Esgiar, A., Naguib, R., Sharif, B., Bennett, M., Murray, A., 2002. Fractal analysis in the detection of colonic cancer images. IEEE Trans. Inform. Technol. Biomed. 6, 54–58.

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2009. The PASCAL Visual Object Classes Challenge 2009 (VOC2009) Results. <http://www.pascal-network.org/challenges/VOC/voc2009/workshop/index.html>.

Felzenszwalb, P.F., Girshick, R.B., McAllester, D.A., Ramanan, D., 2010. Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell. 32, 1627–1645.

Fung, G., Dundar, M., Krishnapuram, B., Rao, B., 2006. Multiple instance algorithms for computer aided diagnosis. In: Advances in Neural Information Processing Systems 19 (NIPS 2006), Vancouver, CA, pp. 1015-1021.

Fung, G., Dundar, M., Krishnapuram, B., Rao, R., 2007. Multiple instance learning for computer aided diagnosis. In: Advances in Neural Information Processing Systems, pp. 425–432.

Galleguillos, C., Babenko, B., Rabinovich, A., Belongie, S., 2008. Weakly supervised object recognition and localization with stable segmentations. In: European Conference on Computer Vision.

Gärtner, T., Flach, P.A., Kowalczyk, A., Smola, A.J., 2002. Multi-instance kernels. In: International Conference on Machine Learning.

Huang, P.W., Lee, C.H., 2009. Automatic classification for pathological prostate images based on fractal analysis. IEEE Trans. Med. Imag. 28, 1037–1050.

Jin, R., Wang, S., Zhou, Z.H., 2009. Learning a distance metric from multi-instance multi-label data. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 896–902.

Keeler, J.D., Rumelhart, D.E., Leow, W.K., 1990. Integrated segmentation and recognition of hand-printed numerals. In: Advances in Neural Information Processing Systems, pp. 285–290.

Kong, J., Sertel, O., Shimada, H., Boyer, K.L., Saltz, J.H., Gurcan, M.N., 2009. Computer-aided evaluation of neuroblastoma on whole-slide histology images: classifying grade of neuroblastic differentiation. Pattern Recogn. 42, 1080–1092.

Kong, H., Gurcan, M., Belkacem-Boussaid, K., 2011. Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting. IEEE Trans. Med. Imag. 30, 1661–1677.

Lafferty, J.D., McCallum, A., Pereira, F.C.N., 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: International Conference on Machine Learning, pp. 282–292.

Lezoray, O., Cardot, H., 2002. Cooperation of color pixel classification schemes and color watershed: a study for microscopic images. IEEE Trans. Image Process. 11, 783–789.

Liang, J., Bi, J., 2007. Computer aided detection of pulmonary embolism with tobogganing and multiple instance classification in CT pulmonary angiography. In: International Conference on Information Processing in Medical Imaging, pp. 630–641.

Liu, Q., Qian, Z., Marvasty, I., Rinehart, S., Voros, S., Metaxas, D., 2010. Lesion-specific coronary artery calcium quantification for predicting cardiac event with multiple instance support vector machines. In: International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 484–492.

Loeff, N., Arora, H., Sorokin, A., Forsyth, D.A., 2005. Efficient unsupervised learning for localization and detection in object categories. In: Advances in Neural Information Processing Systems.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110.

Lu, L., Bi, J., Wolf, M., Salganicoff, M., 2011. Effective 3D object detection and regression using probabilistic segmentation features in CT images. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1049–1056.

Madabhushi, A., 2009. Digital pathology image analysis: opportunities and challenges. Imag. Med. 1, 7–10.

Maron, O., Lozano-Pérez, T., 1997. A framework for multiple-instance learning. In: Advances in Neural Information Processing Systems.

Mason, L., Baxter, J., Bartlett, P., Frean, M., 2000. Boosting algorithms as gradient descent. In: Advances in Neural Information Processing Systems.

Monaco, J.P., Tomaszewski, J.E., Feldman, M.D., Hagemann, I., Moradi, M., Mousavi, P., Boag, A., Davidson, C., Abolmaesumi, P., Madabhushi, A., 2010. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. Med. Image Anal. 14, 617–629.

Ojala, T., Pietikäinen, M., Mäenpää, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24, 971–987.

Park, S., Sargent, D., Lieberman, R., Gustafsson, U., 2011. Domain-specific image analysis for cervical neoplasia detection based on conditional random fields. IEEE Trans. Med. Imag. 30, 867–878.

Portilla, J., Simoncellt, E.P., 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. Int. J. Comput. Vis. 40, 49–71.

Ramon, J., Raedt, L.D., 2000. Multi instance neural networks. In: ICML, Workshop on Attribute-Value and Relational Learning.

Raykar, V.C., Krishnapuram, B., Bi, J., Dundar, M., Rao, R.B., 2008. Bayesian multiple instance learning: automatic feature selection and inductive transfer. In: Proceedings of the 25th International Conference on Machine Learning (ICML 2008), Helsinki, pp. 808–815.

Sertel, O., Kong, J., Shimada, H., Catalyurek, U.V., Saltz, J.H., Gurcan, M.N., 2009. Computer-aided prognosis of neuroblastoma on whole-slide images: classification of stromal development. Pattern Recogn. 42, 1093–1103.

Shotton, J., Johnson, M., Cipolla, R., 2008. Semantic texton forests for image categorization and segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8.

Soares, J.V.B., Leandro Jr., J.J.G., Cesar, R.M., Jelinek, H.F., Cree, M.J., 2006. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. IEEE Trans. Med. Imag. 25, 1214–1222.

Tabesh, A., Teverovskiy, M., Pang, H.Y., Kumar, V., Verbel, D., Kotsianti, A., Saidi, O., 2007. Multifeature prostate cancer diagnosis and Gleason grading of histological images. IEEE Trans. Med. Imag. 26, 1366–1378.

Ta, V.T., Lézoray, O., Elmoataz, A., Schüpp, S., 2009. Graph-based tools for microscopic cellular image segmentation. Pattern Recogn. 42, 1113–1125.

Taskar, B., Guestrin, C., Koller, D., 2003. Max-margin Markov networks. In: Advances in Neural Information Processing Systems.

Tu, Z., Bai, X., 2010. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 21, 1744–1757.

Tuytelaars, T., Lampert, C.H., Blaschko, M.B., Buntine, W., 2009. Unsupervised object discovery: a comparison. Int. J. Comput. Vis. 88, 284–302.

Vedaldi, A., Gulshan, V., Varma, M., Zisserman, A., 2009. Multiple kernels for object detection. In: International Conference on Computer Vision, pp. 606–613.

Vezhnevets, A., Buhmann, J.M., 2010. Towards weakly supervised semantic segmentation by means of multiple instance and multitask learning. In: IEEE Conference on Computer Vision and Pattern Recognition.

Vijayanarasimhan, S., Grauman, K., 2008. Keywords to visual categories: multiple-instance learning for weakly supervised object categorization. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8.

Viola, P.A., Jones, M.J., 2004. Robust real-time face detection. Int. J. Comput. Vis. 57, 137–154.

Viola, P.A., Platt, J., Zhang, C., 2005. Multiple instance boosting for object detection. In: Advances in Neural Information Processing Systems.

Wang, Y., Rajapakse, J.C., 2006. Contextual modeling of functional MR images with conditional random fields. IEEE Trans. Med. Imag. 25, 804–812.

Wang, J., Zucker, Jean-Daniel, 2000. Solving multiple-instance problem: a lazy learning approach. In: International Conference on Machine Learning.

Xu, Y., Zhang, J., Chang, E.I.C., Lai, M., Tu, Z., 2012a. Contexts-constrained multiple instance learning for histopathology image segmentation. In: International Conference on Medical Image Computing and Computer Assisted Intervention.

Xu, Y., Zhu, J.Y., Chang, E., Tu, Z., 2012b. Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 964–971.

Yang, L., Tuzel, O., Meer, P., Foran, D., 2008. Automatic image analysis of histopathology specimens using concave vertex graph. In: International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 833–841.

Zha, Z.J., Mei, T., Wang, J., Qi, G.J., Wang, Z., 2008. Joint multi-label multi-instance learning for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8.

Zhang, Q., Goldman, S.A., 2001. EM-DD: an improved multiple-instance learning technique. In: Advances in Neural Information Processing Systems, pp. 1–8.

Zhang, M.L., Zhou, Z.H., 2009. Multi-instance clustering with applications to multi-instance prediction. Appl. Intell. 31, 47–68.

Zhang, D., Wang, F., Si, L., Li, T., 2009. M^3IC: maximum margin multiple instance clustering. In: International Joint Conference on Artificial Intelligence.

Zhou, Z.H., Zhang, M.L., 2007. Multi-instance multilabel learning with application to scene classification. In: Advances in Neural Information Processing Systems.

Zhu, X., 2008. Semi-Supervised Learning Literature Survey. Computer Science TR 1530, University of Wisconsin-Madison.