A MIL-BASED INTERACTIVE APPROACH FOR HOTSPOT SEGMENTATION FROM BONE SCINTIGRAPHY

Shijie Geng   Jingyang Ma   Xiaoguang Niu   Shaoyong Jia   Yu Qiao⋆   Jie Yang

Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China.
The Key Laboratory of System Control and Information Processing, Ministry of Education, China.
{jeykigung, mary2011, 2012657, jiashaoyong, qiaoyu, jieyang}@sjtu.edu.cn

ABSTRACT

Bone scintigraphy is widely used to diagnose bone diseases. Accurate hotspot segmentation is a critical task for tumor metastasis diagnosis. In this paper, we propose an interactive approach to detect and extract hotspots in thoracic region based on a new multiple instance learning (MIL) method called EM-MILBoost. We convert the segmentation problem to a multiple instance learning task by constructing positive and negative bags according to the input bounding box. In order to be robust against noisy input, we train a region-level hotspot classifier with EM-MILBoost and develop several segmentation strategies based on it. The experimental results demonstrate that our method outperforms other methods and is robust against various noisy input.

Index Terms— Hotspot segmentation, bone scintigraphy, multiple instance learning, EM-MILBoost, region-level classifier

1. INTRODUCTION

Bone scintigraphy is very effective in diagnosing cancer and tumor metastases [1]. The abnormalities in bone scintigraphy images are called "hotspot", which generally appear to be brighter than its surroundings. Accurate hotspot segmentation is helpful for extracting information such as shape, intensity, location. However, due to poor image quality and weak boundary contrast in bone scintigraphy, accurate hotspot segmentation is a challenging task. Many computer-aided diagnosis (CAD) systems have been developed to aid clinicians in the isolation of hotspots. May Sadik et al. [2] used adaptive threshold of a specific region for hotspot segmentation. Huang et al. [3] uses linear regression model to find regional threshold to extract hotspot. Yin et al. [1] proposed a fuzzy diagnosis system which perform local-maximum-based hotspot segmentation. Experienced physicians require a convenient interactive tool for accurate hotspots segmentation. However, most of proposed methods focus on detection of
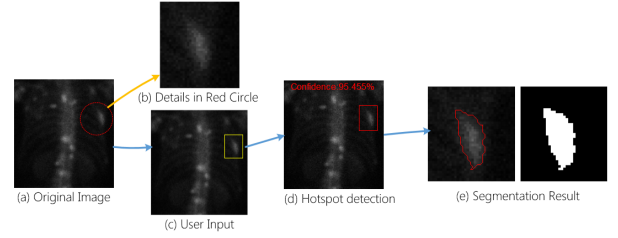

Fig. 1. The overall framework of our approach.

hotspots rather than accurate segmentation. Thresholding and region growing [4] are two widely used segmentation approaches. But none of them could achieve desired hotspot segmentation results.

MIL technique has been widely used in several scenarios, such as MILBoost [5] in object tracking , MCIL [6] and mi-Graph [7] in medical diagnosis. It can significantly reduce the efforts in human annotations and has the ability to exploit information from data automatically. Thus, we choose MIL to execute the segmentation task. MILCut [8] is a new interactive image segmentation method based on MILBoost, which achieves state-of-the-art performance in natural image datasets. It constructs negative bags from pixels outside the input bounding box. However, hotspot pixels very likely locate outside the box in bone scintigraphy and is falsely treated as normal pixels. Meanwhile, MIL-Boost may fail to distinguish hotspot from background due to weak boundary contrast. And input bounding box may not contain hotspots or not as tight as [8] says, which is called "noisy" input.

In this paper, we propose several methods to tackle these problems, Fig. 1 gives an illustration of our framework. A new bag-constructing method is first proposed to form bags from normal pixels in the belt region of the bounding box. Then a new MIL method based on EM optimization strategy, called EM-MILBoost, is developed for accurate hotspot segmentation. Finally, a region-level hotspot classifier and several strategies are developed to deal with noisy input. Experiments in Sec. 4 demonstrate the effectiveness of our framework in accurate hotspot segmentation over other methods.

2. METHODOLOGY

2.1. MILboost Approach

MILBoost [5] is a boosting approach to solve MIL problem using AnyBoost framework [9]. It optimizes the cost function and trains weak classifier. In multiple instance learning, instances are not directly labeled. They are grouped in positive or negative bags. There is at least one positive instance in positive bag while all instances in negative bag are negative.

We denote a bag as $X_i = \{x_{i1}, \ldots, x_{in}\}$, in which i is the bag index and j is the instance index. The label of bag is denoted as $y_i \in \{-1, 1\}$, the hidden label of instance $y_{ij} \in \{-1, 1\}$ is latent during training. The relation between y_i and y_{ij} can be represented as: $y_i = \max_j(y_{ij})$.

According to AnyBoost , the negative log-likelihood loss function is as follows:

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

Here, p_i is the probability of bag i to be positive. p_i can be computed by generalized mean function $p_i = \frac{1}{m}(\sum_{j=1}^{m} p_{ij}^r)^{\frac{1}{r}}$ as an approximation to max function, where p_{ij} is instance-level probability to be positive. p_{ij} is define as $p_{ij} = \frac{1}{1+\exp(-2y_{ij})}$. $y_{ij} = \sum_t \alpha_t h_t(x_{ij})$ is the output of current weighted sum of weak classifiers. Then we can compute new weights as follows:

$$w_{ij} = -\frac{\partial \mathcal{L}}{\partial y_{ij}} = \begin{cases} 2\frac{p_{ij}^r - p_{ij}^{r+1}}{\sum_{j=1}^{n} p_{ij}^r} & \text{if } y_i = 1 \\ -\frac{2p_i}{1-p_i} \frac{p_{ij}^r - p_{ij}^{r+1}}{\sum_{j=1}^{n} p_{ij}^r} & \text{if } y_i = -1 \end{cases} \tag{2}$$

2.2. EM-MILboost Approach

However, MILBoost does not update the sample distribution in the process of weak classifiers training. It may affect the quality of trained classifiers. In order to overcome this problem, we apply the EM optimization strategy into MILBoost to update positive and negative distributions during weak classifier training process, which is called EM-MILBoost. This strategy gives a better approximation to real distributions of samples.

Given feature vector of all instances, we formulate positive class ω_1 and negative class ω_{-1} as two Gaussian distributions $\mathcal{N}(\mu_1, \sigma_1)$ and $\mathcal{N}(\mu_{-1}, \sigma_{-1})$. According to the sign of w_{ij}, we initialize positive distribution parameters μ_1 and σ_1 by computing mean and standard deviation from instances correspond to $w_{ij} > 0$. The negative distribution parameters μ_{-1} and σ_{-1} are initialized by computing mean and standard

deviation from instances correspond to $w_{ij} < 0$. When training weak classifier, we set equal prior for ω_1 and ω_{-1}. For each class, we calculate posterior probability of each instance based on Bayesian formula:

$$p(y_{ij} = c | x_{ij}) \propto p(x_{ij} | y_{ij} = c) = \mathcal{N}(x_{ij} | \mu_c, \sigma_c) \tag{3}$$

where $c \in \{-1, 1\}$. Then we successively apply the E-step and M-step to compute updated distributions:

E-step: compute generative probability of each class

$$p(\omega_c | x_{ij}) = \frac{\mathcal{N}(x_{ij} | \mu_c, \sigma_c)}{\sum_{k=-1,1} \mathcal{N}(x_{ij} | \mu_k, \sigma_k)} \tag{4}$$

where $c \in \{-1, 1\}$.

M-step: perform parameter estimation

$$\mu'_c = \frac{\sum_{i,j} p(\omega_c | x_{ij}) x_{ij}}{\sum_{i,j} p(\omega_c | x_{ij})} \tag{5}$$

$$\sigma'_c = \sqrt{\frac{\sum_{i,j} p(\omega_c | x_{ij})(x_{ij} - \mu'_c)^2}{\sum_{i,j} p(\omega_c | x_{ij})}} \tag{6}$$

where $c \in \{-1, 1\}$.

At last, the new mean and standard deviation of each distribution is calculated under the following update rule:

$$\mu_c \leftarrow \eta\mu_c + (1 - \eta)\mu'_c \tag{7}$$

$$\sigma_c \leftarrow \eta\sigma_c + (1 - \eta)\sigma'_c \tag{8}$$

where η is the learning rate.

The candidate weak classifiers can be represented as:

$$h_{candidate} = sign\left\{\log\left[\frac{\mathcal{N}(x_{ij} | \mu_1, \sigma_1)}{\mathcal{N}(x_{ij} | \mu_{-1}, \sigma_{-1})}\right]\right\} \tag{9}$$

Then we find the best weak classifier subjected to following constraints from candidate weak classifiers:

$$h_t = \arg \min_{h_{candidate}} \sum_{i,j} \mathbf{1}(h_{candidate}(x_{ij}) \neq y_i) |w_{ij}| \tag{10}$$

Line search is then performed to find the weight α_t for the weak classifier which minimizes $\mathcal{L}(h)$. After training given number of weak classifiers, we get the final strong classifier :

$$h(x_{ij}) = \sum_{t=1}^{T} \alpha_t h_t(x_{ij}) \tag{11}$$

3. HOTSPOT SEGMENTATION

3.1. Pixel-level Hotspot Segmentation

Before hotspot segmentation, the bone scan image is preprocessed by applying density slicing approach to map intensity

image to RGB color space. This map aims at increasing discrimination between pixels with similar intensities. In regard to boosting method, more dimensions may yield stronger classifier, thus is helpful for accurate segmentation in bone scan image. To increase segmentation speed and achieve more local information, we apply SLIC superpixel [10] to mapped bone scan image for further preprocessing. We treat the output superpixels as instances, and construct positive bags from instance within bounding box and negative bags from the 6-pixel-width region around the bounding box (*belt region*) respectively. Compared to MILCut, our approach has an advantage that eliminate most influence of hotspot pixels outside bounding box. Fig. 2 illustrates our formulation, where negative bags are from *blue* region and positive bags from *red* region.
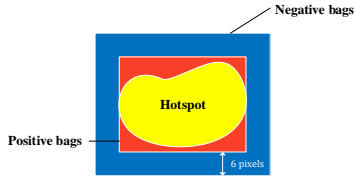

Fig. 2. Formulation to construct positive and negative bags.

After constructing positive and negative bags , we extract features for each instance. For each superpixel, we use average color of all pixels in RGB and CIELab color space to form 6 dimensional features. We then apply EM-MILBoost to train an instance-level classifier, then use the classifier to classify instances within the bounding box. According to the classification result, superpixels labeled as positive are treated as hotspots. To smooth the boundary of initial segmentation, we use the output probability to generate trimap and adopts KNN matting [11] for post-processing.

3.2. Solution to Noisy Input

3.2.1. Region-level Hotspot Classifier

To deal with noisy input, we need a classifier to determine whether there are hotspots in a certain region. We train a region-level classifier (RLC) for hotspot detection. We treat query region as a bag and densely sample 6*6 patches as instances from it. The overlap step size is 3 pixels. We then extract a 29-dimension feature vector for each instance, including 11 dimensions for intensity and 18 dimensions for texture. Intensity features consist of statistical histogram feature (mean , maximum, standard variation etc.), 4-neighborhood contrast and symmetric contrast. We use weighted difference to compute 4-neighborhood contrast:

$$contrast_{neighbor}(M, N) = \frac{1}{10} \sum_{i=1}^{n} 2^i (M_i - N_i) \quad (12)$$

where M and N are histograms of center patch and 4-neighborhood patch respectively. For symmetric contrast, we first find symmetric patch by calculating body central line as [1] does, then we compute symmetric contrast using revised chi-square distance:

$$contrast_{symmetry}(M, S) = \sum_{i=1}^{n} \frac{(M_i - S_i)^2}{M_i + S_i + 1} \quad (13)$$

where M and S are histograms of center patch and symmetric patch respectively. For texture feature, we use a variant of local binary patterns (LBP) [12], in which the threshold used is the mean intensity of the whole image. We collect 3766 instances to form 12 positive bags and 15 negative bags. Finally we apply EM-MILBoost to train the region-level classifier. Fig. 3 shows examples of classification results. We can see the classifier is both sensitive and specific.
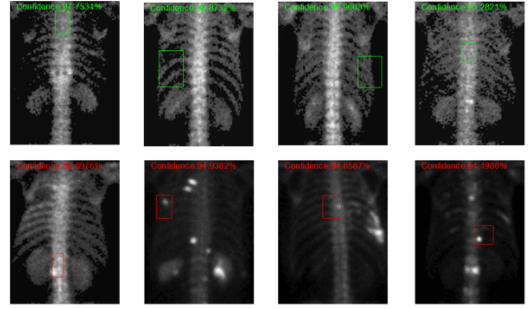

Fig. 3. Region-level classification result. Upper row: normal regions, lower row: abnormal regions.

3.2.2. Strategy to Deal with Noisy Input

The input bounding box may not satisfy the "valid" and "tight" constraints proposed by [8]. For example, if the input bounding box does not contain any hotspot, traditional methods will still execute segmentation, thus leads to significant error. To deal with "noisy" input bounding box, we put forward 4 solutions corresponding to 4 different conditions for RLC:

i) If there is no hotspot in bounding box, RLC will a make a decision not to perform segmentation.

ii) If the bounding box is not tight, false-positive bags will exist in training dataset, leads to ill-learned classifier. We use RLC to roughly locate hotspot region and generating a smaller and tighter bounding box.

iii) If there are hotspot pixels around bounding box, we apply RLC to make a selection of instances, exclude false-negative instances from negative bags.

iv) If hotspot is not completely included in bounding box, we search the 4 neighborhood strip area of the bounding box one by one. If hotspot region is found, we add in this area and construct a improved bounding box.

4. EXPERIMENT AND DISCUSSION

In this section, we apply the proposed EM-MILboost-based approach for accurate hotspot segmentation. We set parameter r to 20 in generalized mean, the update learning rate η to 0.9, the maximum number of weak classifiers T to 50. We then conduct a series of experiments to test our method compared with MCIL [6], mi-Graph [13], MIL-Boost [5], adaptive region growing [4] and thresholding method.

Bone scan images used in training and experiments are collected from Department of Nuclear Medicine, Shanghai Renji Hospital. The labels of bags used in region-level classifier training is from gold standard of radiologist diagnosis. We carefully select and construct a dataset consists of 100 representative hotspots from 46 images with different size, shape, brightness.

Fig. 4. Binary segmentation results of different methods: (a) input bounding box (*yellow*) and ground truth (*orange*); (b) our method; (c) adaptive region growing [4]; (d) MCIL [6]; (e) mi-Graph [13]; (f) MILBoost [5]; (g) thresholding method.
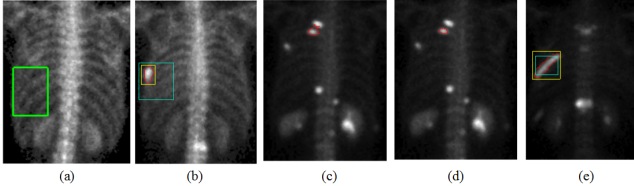

Fig. 5. The segmentation results with 4 kinds of noisy inputs: (a) no hotspot in bounding box; (b) bounding box is not tight enough; (c) hotspots exist in belt region (without instance selection); (d) hotspots exist in belt region (with instance selection); (e) bounding box is not complete enough. Meaning of colors - *green*: normal region; *red*: hotspot boundary; *cyan*: original bounding box; *yellow*: improved bounding box.

We tested these approaches on the hotspot dataset, and invited expert radiologist to annotate the ground truth of each image. Fig. 4 shows some segmentation results of these methods. Our method achieves the best segmentation results among all the methods. Fig. 4(a) gives the input bounding box and segmentation ground truth. Fig. 4(c) shows the results of adaptive region growing [4]. This method needs manually set of maximum number of growing pixels that may lead to

bigger or smaller result with different parameters. Fig. 4(d)-(f) show the results of 3 other multiple instance learning approaches. These methods do not work well and fail to exploit intrinsic information from training data. Fig. 4(g) shows the result of thresholding method. The segmentation results appear to be coarse in boundary and may have little holes inside. In contrast, Fig. 4(b) shows that our method makes accurate hotspot segmentation in cases of single, in pair or vague hotspots.

In this paper, we use Jaccard index (J) and F_1-score as metrics to measure the segmentation performance of each algorithm. Jaccard index is defined as: $J = \frac{|S \cap G|}{|S \cup G|}$, where S is segmentation result and G is ground truth. While F_1-score is given by: $F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$. Table 1 shows the quantitative evaluation of each method and demonstrates that our method outperforms other methods in both two metrics.

Algorithm	Ours	[4]	[6]	[13]	[5]	Thr
Jacc(%)	**66.0**	58.8	58.1	49.7	47.5	51.8
Prec(%)	79.6	88.2	70.1	51.9	70.1	61.4
Rec(%)	84.3	70.0	81.5	90.0	73.7	87.9
F_1-score(%)	**81.9**	78.1	75.4	65.8	71.9	72.3

Table 1. Quantitative evaluation on hotspot dataset.

Finally, we test our strategies to noisy input bounding box. Fig. 5 shows the segmentation results for 4 different types of noisy input. Fig. 5(a) shows that segmentation is not executed in case of no hotspot in bounding box. Fig. 5(b) shows the result when bounding box is not tight. Our approach shrinks the box to reasonable size and achieve a satisfactory result. Fig. 5(c) and (d) shows hotspot segmentation results without and with instance selection respectively, when hotspots exist in belt area of box. Fig. 5(c) shows that the extracted hotspot boundary is not correct without instance selection. It indicates that instance selection does make sense in this case. Fig. 5(e) shows that our strategy searches 4 neighborhood of the box and enlarges the box to reasonable size when bounding box is not complete. These results demonstrate that our strategy is effective to deal with various kinds of noisy input.

5. CONCLUSION

In this paper, we propose a novel method using multiple instance learning to segment hotspot in thoracic bone scintigraphy. EM-MILBoost algorithm is proposed for hotspot segmentation to make better approximation to distributions of positive and negative class. To deal with noisy inputs, we train region-level classifier and propose four strategies. Experiments show the effectiveness of our method.

6. REFERENCES

[1] Tang-Kai Yin and Nan-Tsing Chiu, "A computer-aided diagnosis for locating abnormalities in bone scintigraphy by a fuzzy system with a three-step minimization approach," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 5, pp. 639–654, 2004.

[2] May Sadik, Iman Hamadeh, Pierre Nordblom, Madis Suurkula, Peter Höglund, Mattias Ohlsson, and Lars Edenbrandt, "Computer-assisted interpretation of planar whole-body bone scans," *Journal of Nuclear Medicine*, vol. 49, no. 12, pp. 1958–1965, 2008.

[3] Jia-Yann Huang, Pan-Fu Kao, and Yung-Sheng Chen, "A set of image processing algorithms for computer-aided diagnosis in nuclear medicine whole body bone scan images," *Nuclear Science, IEEE Transactions on*, vol. 54, no. 3, pp. 514–522, 2007.

[4] SA Hojjatoleslami and Josef Kittler, "Region growing: a new approach," *IEEE Transactions on Image processing*, vol. 7, no. 7, pp. 1079–1084, 1998.

[5] John C. Platt Paul Viola and Cha Zhang, "Multiple instance boosting for object detection," in *Advances in neural information processing systems*, 2005, pp. 1417–1424.

[6] Yan Xu, Jun-Yan Zhu, Eric Chang, and Zhuowen Tu, "Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 964–971.

[7] Melih Kandemir, Annette Feuchtinger, Axel Walch, and Fred A Hamprecht, "Digital pathology: Multiple instance learning can detect barretts cancer," ISBI, 2014.

[8] Jiajun Wu, Yibiao Zhao, Jun-Yan Zhu, Siwei Luo, and Zhuowen Tu, "Milcut: A sweeping line multiple instance learning paradigm for interactive image segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014.

[9] Llew Mason, Jonathan Baxter, Peter Bartlett, and Marcus Frean, "Boosting algorithms as gradient descent in function space," NIPS, 1999.

[10] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 11, pp. 2274–2282, 2012.

[11] Qifeng Chen, Dingzeyu Li, and Chi-Keung Tang, "Knn matting," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 869–876.

[12] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.

[13] Zhi-Hua Zhou, Yu-Yin Sun, and Yu-Feng Li, "Multi-instance learning by treating instances as non-iid samples," in *Proceedings of the 26th annual international conference on machine learning*. ACM, 2009, pp. 1249–1256.