COMBINING CGAN AND MIL FOR HOTSPOT SEGMENTATION IN BONE SCINTIGRAPHY

Hang Xu, Shijie Geng, Yu Qiao, Kuan Xu, Yueyang Gu*

Intelligence Learning Laboratory, Department of Automation, Shanghai Jiao Tong University, China
Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai
{sjtu.xuhang, jeykigung, qiaoyu, whxk225, guyueyang}@sjtu.edu.cn

ABSTRACT

Bone scintigraphy is widely used to diagnose bone tumor and metastasis. Accurate hotspot segmentation from bone scintigraphy is of great importance for tumor metastasis diagnosis. In this paper, we propose a new framework to detect and extract hotspots in thoracic region by integrating the techniques of both conditional generative adversarial networks (cGAN) and multiple instance learning (MIL). We first use cGAN to train a generator, which can be applied to separate input bone scan image into four anatomical regions and provide location information. A multi-dimensional feature is constructed to integrate contrast, texture and location information. We then use MIL to train a patch-level classifier with this constructed feature. In hotspot segmentation, a hotspots probability map can be estimated with the patch-level classifier. The hotspot segmentation is performed with level set method, in which the hotspot boundary is initialized based on the hotspot probability map. We evaluate the proposed framework quantitatively on the hotspot dataset, and compare it with other methods.

Index Terms— hotspot segmentation, cGAN, MIL, probability map, level set

1. INTRODUCTION

Bone scintigraphy makes it easier for doctors to diagnose cancer and tumor metastases [1]. Accurate segmentation of hotspots from bone scintigraphy serves as an important means to effectively analyze bone scintigraphy. Bone tumor and metastases show a feature of being brighter than the surrounding areas in images, so they are called hotspots. Many strategies have been developed to detect and segment hotspots. May Sadik et al. [2] used adaptive threshold in a separated region to perform hotspot segmentation. Huang et al. [3] also utilized regional threshold to extract hotspot based on linear regression model. Yin et al. [1] developed a fuzzy diagnosis system that could implement local-maximum-based segmentation of hotspot. Wang et al. [4] utilized knowledge-based method and adaptive thresholding to accomplish the segmentation of spine and ribs for convenience to complete other goals. Geng et al. also made contribution to hotspot segmentation interactively [5] and automatically [6]. They proposed a framework [6] combining multiple instance learning (MIL) and level set method.

Research on segmentation of medical images have always enjoyed great popularity. In the past, people employed traditional machine learning methods [7, 8, 9, 10]. Recently, methods based on deep network have been widely applied. Generative adversarial network (GAN) is a model for generating realistic images through adversaries between the generator and the discriminator. Souly et al. [11] improved semantic segmentation results with GAN and FCN. Hong et al. [12] utilized a GAN-based approach to address domain adaptation for the semantic segmentation of urban scenes. Singh et al. [13] used conditional generative adversarial network (cGAN) to learn the statistical invariant features and segment the images. cGAN is an extension of the GAN by introducing conditional variables into the modeling. In order to improve hotspot segmentation, we employ cGAN to support training patch-level MIL classifier.

In this paper, we present a novel framework that combines cGAN with MIL to accomplish accurate hotspot segmentation. We separate the input image into four anatomical regions by a generator trained by cGAN. In addition, we trained a patch-level MIL classifier with 38-dimension feature, including 4-dimension location feature determined by the four regions. We employ the classifier to get a probability map that shows the distribution and location of hotspots. The final segmentation is performed with level set while the initialization curve is generated by the probability map.

2. METHODOLOGY

2.1. Approach Overview

Automatic segmentation of hotspots have been investigated by Geng [6]. This paper makes some innovation and improvements on his method. We utilized more effective features in MIL, including contrast features, texture features and location features that are associated with cGAN. Figure 1 gives an illustration of our framework. Each input image will generate one new image, which has four regions with different colors

*Corresponding author: Yu Qiao (qiaoyu@sjtu.edu.cn). This research is supported by NSFC, China (No:61375048).

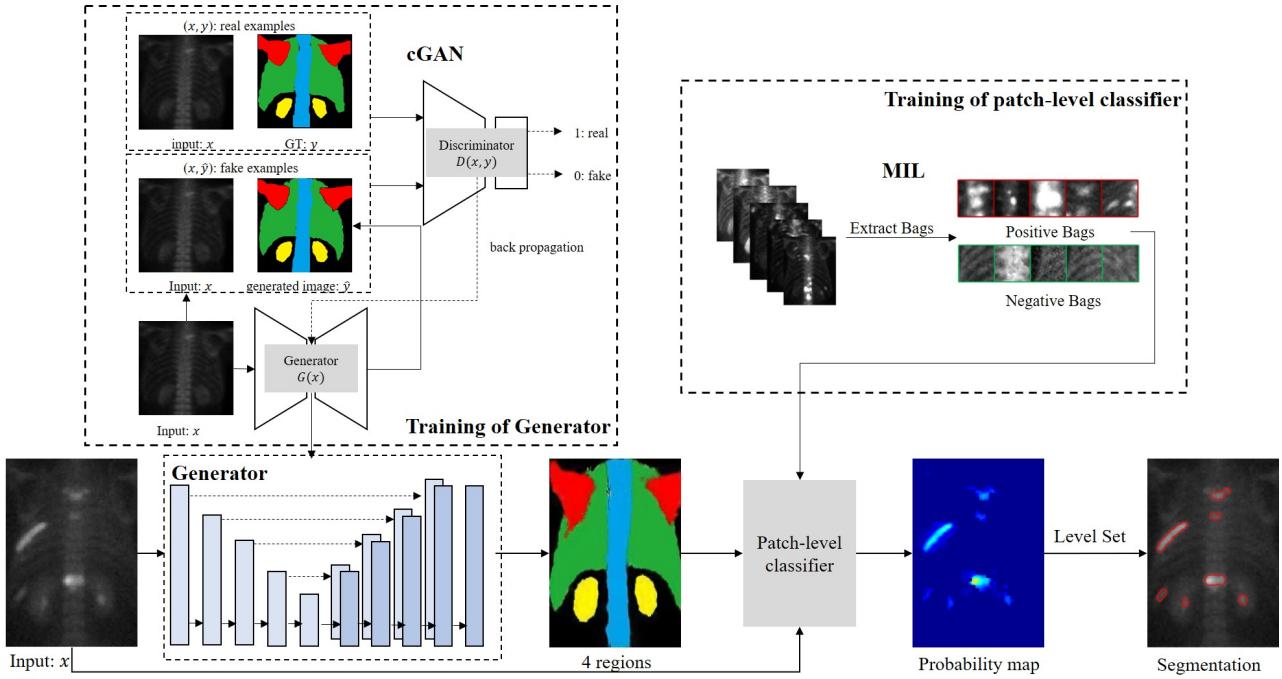

Fig. 1. Framework of our algorithm.

through the generator trained by cGAN. Then the input image will get a hotspot probability map through a patch-level classifier that is trained by MIL. Finally, we get the contour of hotspots with level set methods.

2.2. Using cGAN to Separate Scintigraphy Image

In this section, we use cGAN to separate input bone scan image into four anatomical regions. We employ pix2pix [14] model, which is based on cGAN. Unlike a conditional GAN, both the generator and discriminator observe the input image. Furthermore, pix2pix uses U-Net [15] instead of encoder-decoder for generator. The generator learns the mapping from the input to the output. In turn, the discriminator learns a loss function to train this mapping by comparing the ground-truth and the predicted output. Finally, the whole cGAN optimizes a loss function that combines a conventional binary cross-entropy loss with an adversarial term. The training of proposed cGAN is shown in Figure 1.

During the training process, pix2pix tries to minimize the following loss function:

$$\mathcal{L}_{\text{final}}(G, D) = arg \min_{G} \max_{D} \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \quad (1)$$

Here, $\mathcal{L}_{cGAN}(G, D)$ is loss function and $L1$ loss is added in order to ensure the similarity. The complete formula is shown below:

$$\begin{aligned} \mathcal{L}_{cGAN}(G, D) =& \mathbb{E}_{x,y}[\log D(x, y)] + \\ & \mathbb{E}_{x,z}[\log (1 - D(x, G(x, z)))] \end{aligned} \quad (2)$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1] \quad (3)$$

Here, x, y, z represent an original image, ground truth and random variable. $G(x, z)$ is the predicted image. The discriminator output score is $D(x, y)$ and $D(x, G(x, z))$. Besides, $L1$ loss is added in order to ensure the similarity. The $L1$ normalized distance between ground truth and the predicted image is $\|y - G(x, z)\|_1$.

2.3. Using MIL to Train Patch-level Classifier

In this section, we use MIL [16] method to train patch-level classifier that can produce probability map of hotspots. Based on MIL, this method can reduce the work in annotations and exploit information from data automatically. To learn detail information from input data, we need to set patches to a relatively small size. However, when the size of patches gets smaller, it will be harder for manual labeling. In multiple instance learning, instances are not directly labeled. MIL only needs bag labels. This property will significantly reduce the efforts in manual annotations. Therefore, we choose MIL to train patch-level classifier. With the region information generated by cGAN, we will estimate hotspot probability map through MIL.

2.3.1. Feature Extraction

In our approach, we select relatively large regions from bone scan images and treat them as bags. For each bag, we pick up 4×4 patches as instances. We then extract a 38-dimension feature vector for each instance, including 33-dimension texture feature, 4-dimension location feature and 1-dimension contrast feature. For symmetric contrast, we first search for symmetric patch by calculating body central line [10]. Then we compute symmetric contrast using modified chi-square distance:

$$SymmetryContrast(M,S) = \sum_{i=1}^{n} \frac{(M_i - S_i)^2}{M_i + S_i + 1} \quad (4)$$

where M and S are histograms of current patch and symmetric patch respectively. Location features are determined by which region the patch is in:

$$Loc(k) = \sum_{i=1}^{n} \varphi_k(i) \quad (5)$$

$$\varphi_k(i) = \begin{cases} 1, & pixel(i) \in region(k) \\ 0, & otherwise \end{cases} \quad (6)$$

We utilize Leung-Malik filter bank [17] to obtain the texture features. This filter is a 48 dimensional filter bank proposed by Leung and Malik. The feature vector based on the response of LM filter bank can well describe the image texture features. We extract a simplified feature vector, which consists of 24 directional filters, 6 Gaussian Lapacian filters and 3 Gaussian filters.

2.3.2. MIL Algorithm

We need to extract positive and negative bags consisting of instances to construct training dataset. Figure 1 shows how MIL extract bags. We denote a bag $X_i = \{x_{i,j}\}$, where i is the bag index and j is the instance index. During training, only bags have the label denoted as $y_i \in \{1, -1\}$. Let $P(X_i)$ is the probability of bag i, $p(x_{ij})$ is the probability of instance j in bag i. $P(X_i) = \max_j(p(x_{ij}))$. The loss function is

$$\begin{aligned} \mathcal{L}(P) = -\sum_{i=1}^{n} [\mathbf{1}(y_i = 1) \log P(X_i) + \\ \mathbf{1}(y_i = -1) \log (1 - P(X_i))] \end{aligned} \quad (7)$$

New weights can be computed: $w_{ij} = -\frac{\partial \mathcal{L}(P)}{\partial p(x_{ij})}$. These weights will be used to select best weak classifier p_k. We get the strong classifier after training weak ones: $p(x_{ij}) = \sum_{k=1}^{K} \alpha_k h_k(x_{ij})$. Here, line search is employed to find the coefficient α_k.

2.4. Using Level Set to Extract Hotspot Contour

The method of level set with local signed difference(LSD) [9] is used to refine the contour of hotspots based on the probability map of hotspots. Since LSD considers both global and local information by calculating the order of local clusters, it is able to deal with intensity inhomogeneity and weak object boundaries of bone scan images. This may lead to more robust segmentation performance.

We use C to denote the contour evolving in image domain Ω. $f_1(x)$ and $f_2(x)$ are local clusters inside and outside contour C. The LSD energy can be expressed as follows:

$$\begin{aligned} \mathcal{E}(C) = \int_{\Omega} sgn(f_2(x) - f_1(x)) |f_2(x) - f_1(x)| dx + \\ \mu Length(C) \end{aligned} \quad (8)$$

Let a Lipschitz function $\phi : \Omega \rightarrow R$ to represent LSD energy. We employ the hotspots probability map $\mathcal{P}$:

$$\phi_0 = \phi(x, t = 0) = \begin{cases} -C, & x \in x | \mathcal{P} > threshold \\ 0, & x \in x | \mathcal{P} = threshold \\ C, & otherwise \end{cases} \quad (9)$$

In order to minimize the energy, we utilize gradient descent: $\frac{\partial \phi}{\partial t} = -\frac{E(\phi)}{\partial \phi}$.

3. EXPERIMENTS AND DISCUSSION

In this section, we conduct experiments to demonstrate the effectiveness of our framework. We show that our framework achieves better segmentation results compared with Geng's approach [6] and other previous methods.

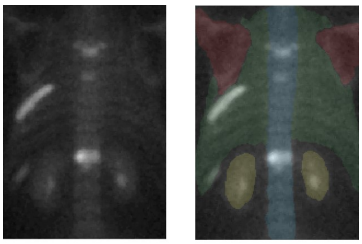

Fig. 2. Thoracic region segmentation with cGAN: (a) input image; (b) region separation result.

3.1. Implementation Details and Evaluation

The bone scan images used in training are collected from Department of Nuclear Medicine, Shanghai Renji Hospital. As the hotspots are mainly concentrated in the thoracic area, the hotspot segmentation in the thoracic area is most critical for the diagnosis of bone metastases. Therefore, this paper will focus on the thoracic bone scan images.

Evaluation. We use Jaccard index (J), Dice index (D), F_1-score (F_1) and false negative rate (FNR) as metrics to measure the hotspot 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. Dice index is defined as $D = \frac{2|S \cap G|}{|S|+|G|}$. F_1-score is given by $F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$. False negative rate is calculated by $FNR = \frac{|G - S \cup G|}{|G|}$.

We use 168 pairs of bone scan images and ground truth for training cGAN. There are also 56 pairs for testing the performance of separation. The test results is shown in Table 1. Figure 2 shows an example. The location features are reliable due to the accurate region separation with cGAN. We collect 18947 instances to form 39 positive bags and 33 negative bags for MIL training.

Method	Jacc	Dice	F_1-score	FN-rate
cGAN	0.8969	0.9453	0.9453	0.0391

Table 1. Evaluation of separation results with cGAN.

3.2. Performance Comparison

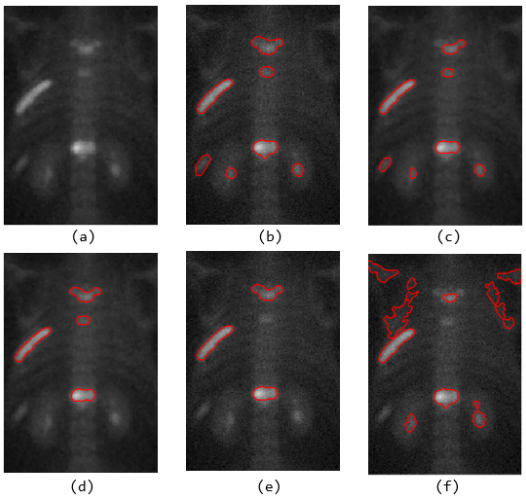

Fig. 3. Hotspot segmentation comparison among different approaches: (a) input image; (b) ground truth; (c) our method's result; (d) MIL-29's [6] result; (e) LSD's [9] result; (f) TnR's [2] result.

In the experiment, we compared our framework with four previous methods: adaptive region growing [10], region-based thresholding [2], level set evolution with local signed difference (LSD) [9] and MIL-29 [6]. We used 29 thoracic images with ground truths annotated by expert as testing dataset. Figure 3 gives an illustration of some segmentation results and ground truths. Table 2 shows the performance of all methods in terms of accurate hotspot segmentation.

Method	Jacc	Dice	F_1-score	FN-rate
RG [10]	0.5853	0.6935	0.6935	0.2881
TnR [2]	0.3729	0.5049	0.5049	0.3422
LSD [9]	0.5677	0.7029	0.7029	0.3898
MIL-29 [6]	0.6493	0.7717	0.7717	0.3122
Our method	**0.7253**	**0.8319**	**0.8319**	**0.2053**

Table 2. Performance comparison on hotspot segmentation from thoracic bone scan images.

Table 2 demonstrates that our framework implements more accurate hotspot segmentation and outperforms previous methods in all metrics. Figure 3 and 4 illustrate 2 examples, in which the hotspot contours obtained with our method have smooth boundaries and are similar to ground truths. Compared to MIL-29 [6], our method employs additional location features and select Leung-Malik filter bank instead of LBP to construct texture features. Obviously, these new features may be helpful to achieve better probability maps and thus more desirable segmentation results.

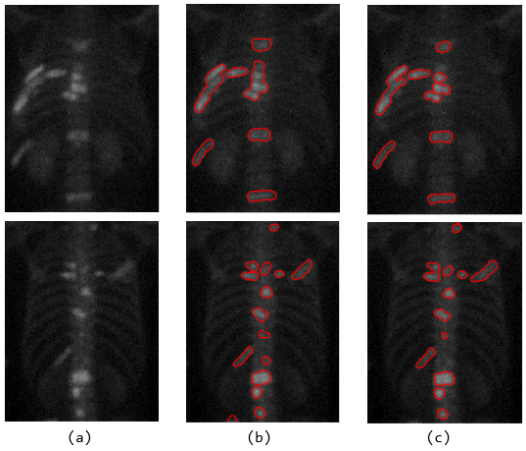

Fig. 4. Hotspot segmentation samples of our approach: (a) input image; (b) ground truth; (c) hotspot segmentation results of our method.

4. CONCLUSION

In this paper, we present a framework combining conditional generative adversarial network with multiple instance learning to accomplish accurate hotspot segmentation. In addition to contrast and texture features, we take location information into consideration, which is obtained with cGAN. Experimental results show that our framework outperform other methods and the feature of location are helpful for accurate hotspot segmentation.

5. 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," *IEEE Transactions on Medical Imaging*, 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," *IEEE Transactions on Nuclear Science*, vol. 54, no. 3, pp. 514–522, 2007.

[4] Qiang Wang, Qingqing Chang, Yu Qiao, Yuyuan Zhu, Gang Huang, and Jie Yang, "Knowledge-based segmentation of spine and ribs from bone scintigraphy," in *International Conference on Neural Information Processing*. Springer, 2011, pp. 241–248.

[5] Shijie Geng, Jingyang Ma, Xiaoguang Niu, Shaoyong Jia, Yu Qiao, and Jie Yang, "A mil-based interactive approach for hotspot segmentation from bone scintigraphy," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 942–946.

[6] Shijie Geng, Shaoyong Jia, Yu Qiao, Jie Yang, and Zhenhong Jia, "Combining cnn and mil to assist hotspot segmentation in bone scintigraphy," in *International Conference on Neural Information Processing*. Springer, 2015, pp. 445–452.

[7] Yu Qiao and Jie Yang, "Adaptive region growing based on boundary measures," in *International Conference on Neural Information Processing*. Springer, 2011, pp. 249–256.

[8] Yu Qiao, Qingmao Hu, Guoyu Qian, Suhuai Luo, and Wieslaw L Nowinski, "Thresholding based on variance and intensity contrast," *Pattern Recognition*, vol. 40, no. 2, pp. 596–608, 2007.

[9] Lingfeng Wang, Huaiyu Wu, and Chunhong Pan, "Region-based image segmentation with local signed difference energy," *Pattern Recognition Letters*, vol. 34, no. 6, pp. 637–645, 2013.

[10] Qingqing Chang, Qiang Wang, Yu Qiao, Yuyuan Zhu, Gang Huang, and Jie Yang, "Adaptive detection of hotspots in thoracic spine from bone scintigraphy," in *International Conference on Neural Information Processing*. Springer, 2011, pp. 257–264.

[11] Nasim Souly, Concetto Spampinato, and Mubarak Shah, "Semi supervised semantic segmentation using generative adversarial network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 5688–5696.

[12] Weixiang Hong, Zhenzhen Wang, Ming Yang, and Junsong Yuan, "Conditional generative adversarial network for structured domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1335–1344.

[13] Vivek Kumar Singh, Hatem A Rashwan, Farhan Akram, Nidhi Pandey, Md Mostafa Kamal Sarker, Adel Saleh, Saddam Abdulwahab, Najlaa Maaroof, Jordina Torrents-Barrena, Santiago Romani, et al., "Retinal optic disc segmentation using conditional generative adversarial network.," in *CCIA*, 2018, pp. 373–380.

[14] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[15] Olaf Ronneberger, "Invited talk: U-net convolutional networks for biomedical image segmentation," in *Bildverarbeitung für die Medizin 2017*, pp. 3–3. Springer, 2017.

[16] Yan Xu, Jun-Yan Zhu, I Eric, Chao Chang, Maode Lai, and Zhuowen Tu, "Weakly supervised histopathology cancer image segmentation and classification," *Medical image analysis*, vol. 18, no. 3, pp. 591–604, 2014.

[17] Thomas Leung and Jitendra Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International journal of computer vision*, vol. 43, no. 1, pp. 29–44, 2001.