Guide-wire Detection Using Region Proposal Network for X-ray Image-guided Navigation

Li Wang[1], Xiao-Liang Xie[1], Gui-Bin Bian[1], Zeng-Guang Hou[1,2], Xiao-Ran Cheng[1], and Pusit Prasong[1]
[1] State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[2] CAS Center for Excellence in Brain Science and Intelligence Technology, Beijing 100190, China
Email: {wangli2014, xiaoliang.xie, guibin.bian, zengguang.hou, chengxiaoran2015}@ia.ac.cn, and pusitprasong@gmail.com

***Abstract*—Detection of surgical devices, in particular of guide-wire detection, is prerequisite during image-guided navigation in percutaneous coronary intervention (PCI). Guide-wire detection is a challenging task for following reasons: (i) X-ray images have a low signal-to-noise rate (SNR); (ii) there is a high similarity between guide-wires and some other adjacent anatomical skeletons' contours; (iii) guide-wires have various shapes and their motion is complex and nonlinear. Traditionally, guide-wires are detected using curve fitting method, and third-order B-spline curve model is always used to fit guide-wires, while B-spline fitting method has some obvious shortcomings such as it is a semi-automatic method which needs manual initialization, and it is not a real-time method because of high computational complexity. Recently, with the availability of large annotated datasets and the accessibility of hardware resources with GPUs, it is succeeded in detecting general objects with convolutional neural networks (ConvNet). In this paper, we present a novel image-based fully-automatic and real-time approach with ConvNet for guide-wires detection. ConvNet method is robust to guide-wires' various poses and other structures' effects. We evaluate our method on 22 different sequences of X-ray images. The detection accuracy evaluated by average precision (AP) reaches 89.2% and the detection speed achieves 40fps. Our experiment result shows a promising for accurate and real-time guide-wires detection in PCI navigation with ConvNet model.**

I. Introduction

PCI is an interventional radiology therapy implies inserting guide-wires into patient's vascular system under the monitoring of X-ray videos [1]. It has been used increasingly often in recent years. Over the past years, guide-wires detection in X-ray images has been gained interests among navigation in PCI, and wide range of applications such as 3-D guide-wires reconstruction and respiratory motion tracking, rely upon it. Guide-wire detection is a critical step during the navigation as well, which can offer image-guided visual feedback to physicians. However, guide-wires detection is a tricky task for some reasons: (i) X-ray images have a low SNR, during the PCI, X-ray contrast agent is injected in low energy for the sake of reducing X-ray radiation to patients and physicals; (ii) guide-wires have various shapes and they appear as thin, dark curves in 2-D X-ray images, so some adjacent anatomical skeletons' contours which have a high similarity with guide-wires [2] can lead to error detection; (iii) the motion of guide-wires is nonlinear and complex because of the interactions between physicians and patients, as well as patients' respiratory. These limitations make accurate and robust guide-wires detection a challenging task. Over the few years, several methods have been proposed to meet these challenges. Most of these methods aim at enhancing X-ray images and fitting guide-wires with B-spline model [3] [4]. Hessian filter [5], steerable filter and enhancing diffusion [6] are mainly used for X-ray image enhancement. The main idea of B-spline fitting method always present in two steps: (i) **local** interest pixels detection. It needs to compute a probabilistic map to search some interest points which have high probabilities of presence on guide-wires, or these key pixels need to be initialized manually; (ii) **global** curve segments detection. This step is focused on finding an efficient link between the key interest points in step (i), then these key pixels can be connected together to form guide-wire segments. This method is always aligned along a pattern of appearance, continuity and smoothness. Both steps need to take a consideration of all possible adjacent pixels. Consequently, this method has a great computational complexity, and several other technical challenges need to be overcome [7] [8] (e.g. step (i) is sensitive to key points miss-detection, when the key points are detected incorrectly, the method needs to be restarted, and smoothness of guide-wires is the most tough problem to be handled). Soon afterwards, some machine learning methods are proposed [9], these methods are mainly based on some hand-crafted features which have limited description abilities, and their detection speed is limited with hardware resources.

Considering the obvious shortcomings of above methods. We propose a novel method based on ConvNet model with GPUs for guide-wire detection, and a modified annotation method for dataset annotation, it is a fully-automatic, robust and real-time method for guide-wire detection. The ConvNet model has the following significant superiorities, first of all, it can detect guide-wires without manual intervention relative to [7] and [8]. Secondly, compared with [9], the ConvNet method can extract richer and more essential features and is robust to guide-wires' various shapes and other structures' disturbances, and the detection result of each frame is not interdependent, which can make contribution to the guide-wires detection accuracy improvement. The last but not least, the parallel computing resources via GPUs make it feasible to detect guide-wires in real-time. Our modified annotation method makes the bounding boxes are more compact to guide-wires and contain fewer backgrounds, which can do

978-1-5090-6182-2/17/$31.00 ©2017 IEEE

benefit to improve detection results. The rest of paper is organized as follows: Section II provides an introduction of our method. Section III provides our experiments, Section IV is a discussion between some additional experiments, and Section V gives some conclusions.

II. Method

Recently, the availability of large annotated datasets and the accessibility of affordable parallel computing resources via GPUs have made it feasibility to train a ConvNet model [10]. ConvNet is able to capture global structures and high-level semantic information since the network parameters are learned from the training images in a fully supervised pattern, which results in more powerful and discriminative features than the traditional hand-crafted ones such as HOG (histogram of oriented gradient) [11], SIFT (scale-invariant feature transform) [12], LBP (local binary pattern) [9] and so on. And ConvNet has been applied successfully in biomedical fields. In this paper, we apply a region proposal network to detect guide-wires efficiently and effectively. Firstly, we set up datasets comprising X-ray images and annotation bounding boxes with angles, the datasets are divided into two groups, one group for model training, and the rest group is used for testing. Then training a region proposal network model for guide-wire detection. Finally, we demonstrate the validity of our method with testing datasets, and make a comparison with traditional machine learning method with hand-crafted feature and some other modified methods.

A. Dataset Annotation

Using a ConvNet for guide-wire detection, we apply a rectangular bounding box to calibrate a guide-wire in each X-ray image, so the positions of a guide-wire are marked as (x, y, w, h), where (x, y) are the center coordinates of the bounding box, and (w, h) are the width and height of the bounding box. While during PCI navigation, guide-wires are always in various shapes, so we use a different schedule showed in the left of Fig. 1, which uses angular bounding boxes to annotate guide-wires, so the bounding boxes are denoted as (x, y, w, h, θ), and θ [1] is the angle between the rectangle long-side and the horizontal direction. Among different X-ray image sequences, the bounding boxes of guide-wires are always in different angles, while in the same sequence, allowing for the clinical experiences that contrast agent is injected in a low energy and develops in short time, during the development, the motion direction of guide-wires remains almost unchanged. As a result, we only need to compute an approximate direction in one sequence. The total datasets contain 22 different sequences, so we only need to compute 22 angles off-line and the computation complexity is free. Fig. 1 shows a comparison between our annotation method (left) and general method (right). Our bounding box is more compact and contains fewer backgrounds. What's more, the angular bounding boxes are more intuitive and natural than general vertical bounding boxes to match human visual habit [13].

[1] the angle is computed as following steps: first, the center coordinates of an angular bounding box are marked as $(x, y, 1)$, stopping rotating the bounding box around the z - axis until the bounding box to be a horizontal rectangle and the center coordinates of the angular bounding box are marked as $(x^*, y, 1)$ where $y = 0$, the rotation matrix is marked as R_z. So the θ can be computed as:

$$\begin{bmatrix} x^* \\ 0 \\ 1 \end{bmatrix} = R_z * \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, and\ R_z = \begin{bmatrix} cos(\theta) & -sin(\theta) & 0 \\ sin(\theta) & cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

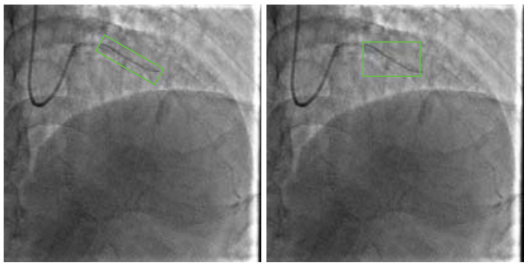

Fig. 1. A comparison between two different bounding box annotation methods. The left picture is an angular bounding box annotated with our annotation method, and the right one is the bounding box annotated with general method.

B. Region Proposal Network

Region proposal network is a fully convolutional network. It takes an image as input to extract feature and outputs a set of rectangular object proposals, each proposal with two scores ranging from 0 to 1 representing the probability that how much the proposal belongs to the object and background, and four coordinates to denote the proposal position. The region proposal network mainly comprises three parts: feature extraction, region proposals generation, and proposals regression and classification. Compared with traditional machine learning method based on a sliding window and hand-crafted features, region proposal network has advantages over the following aspects: (i) feature has stronger representation ability than hand-crafted ones; (ii) the generated proposals can be well adapted to the variability of objects due to the diverse scales and aspect ratios of proposals, while the size of detection boxes are fixed in traditional methods; (iii) there is no need to conduct time-consuming pyramid decomposition.

The structure of our region proposal network is showed in Fig. 2. We apply the Zeiler and Fergus model [14] (ZF) to extract guide-wires features which is mainly composed of 5 cascade convolutional layers, 2 max-pooling layers and some other implicit layers (ReLU (rectified liner units) layers and normalized layers). After the fifth convolutional filter, we can get a $W * H * 256$ feature map. Then we add an extra convolutional layer with 256 filters of kernel size of 3 * 3 on the last feature map to produce a fixed dimension feature at each spatial location which some likes a sliding window, after making a convolution with the extra convolutional filter, each sliding window corresponds to a $256-d$ feature vector. Then, the $256-d$ feature in each location is fed to two sibling convolutional layers: a bounding box regression layer (*reg*)

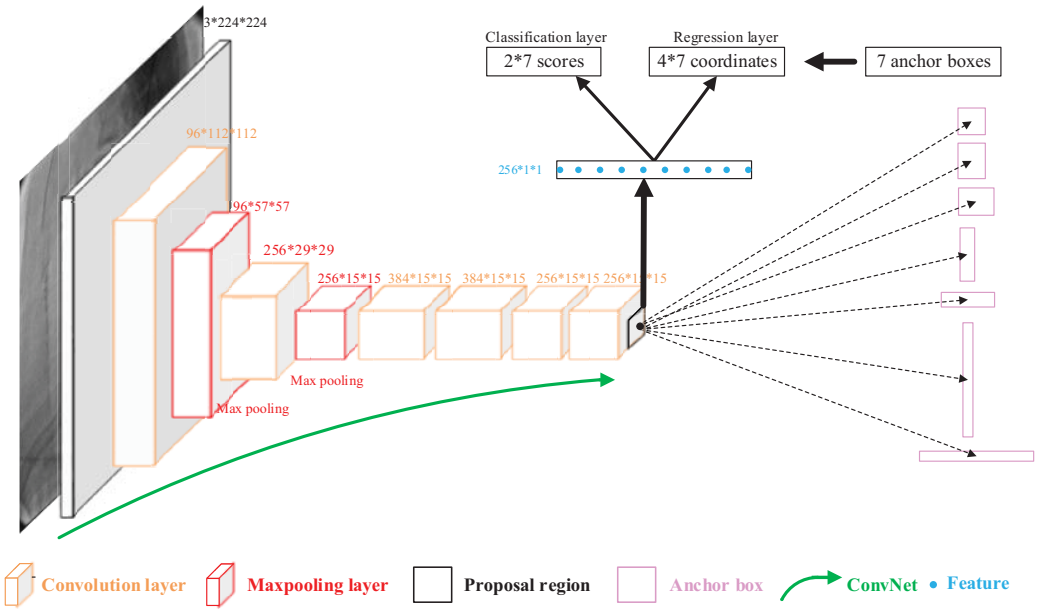

Fig. 2. Region proposal network. The first network contains 5 convolutional layers. Input is a cropped image in size of 224 * 224 (3 color channels). In first layer, input image is convolved with 96 different filters of kernel size of 7 * 7. Then the feature maps are processed in three steps: (i) passed through a ReLU (not showed); (ii) max-pooled with 3 * 3 (stride is 2); (iii) normalized (not showed). Similar operations are repeated in layers 2, 3, 4, 5. The output layer includes a classifier and a bounding box regressor [14].

and a classification layer (*cls*). The region proposal network predicts N candidate boxes at each spatial location of last convolutional feature map. The classifier is used to classify the proposals whether they are foregrounds or backgrounds based on their intersection-over-union (IoU) overlap with the ground truth boxes, and if they are considered as target, the bounding box regressor is used to modify their coordinates. Additional, $N = N_{scale} * N_{aspect}$, these proposals have N_{scale} different scales and N_{aspect} aspect ratios (h/w) and the details are discussed in part D in section II.

C. Loss Function

For training the proposal network, we use a multi-task loss for joint object classification and regression. They are defined respectively in (1) and (2), and the (3) [15] is the parameters of the regressor:

$$L_{cls}(p,u) = \sum_i L_{cls}(p_i^u) = \sum_i -logp_i^u \tag{1}$$

$$L_{reg}(t,t^*) = \sum_i^N Smooth_{L_1}(t_u^i - t_i^*), u \in \{x,y,w,h\} \tag{2}$$

$$\begin{cases} t_x = \dfrac{x-x_p}{w_p}, t_y = \dfrac{y-y_p}{h_p}, t_w = log(\dfrac{w}{w_p}), t_h = log(\dfrac{h}{h_p}) \\ t_x^* = \dfrac{x^*-x_p}{w_p}, t_y = \dfrac{y^*-y_p}{h_p}, t_w = log(\dfrac{w^*}{w_p}), t_h = log(\dfrac{h^*}{h_p}) \end{cases} \tag{3}$$

Here, for the proposal region i in a batch, p_i^u is the final classier output score of proposal i for each class u (plus background), (x,y,w,h) are the regressor output of the predicted object location, (x_p,y_p,w_h,h_p) are the coordinates of each proposal, (x^*,y^*,w^*,h^*) are the ground truth, and t_i^* is the parameter of the ground truth box associated with a positive proposal and t_i is the parameter between predicted bounding box associated with a positive proposal. The entire loss is a multi-loss defined as:

$$Loss = \frac{1}{N_{cls}} L_{cls}(p,u) + \lambda [u > 0] \frac{1}{N_{reg}} L_{reg}(t,t^*)$$

Here, the loss is a weighted sum of L_{cls} and L_{reg}, the first term $L_{cls}(p,u)$ is the classification log loss for true class u of each proposal, and the second term $L_{reg}(t,t^*)$, encourages the class-specific bounding-box prediction to be as accurate as possible. If the proposal overlaps a ground truth box with IoU greater then 0.5, the truth class u is given by the class of ground truth box, otherwise $u = 0$ and the second term of loss is ignored. When $u \neq 0$, there is a balanced weight λ between two terms, and in our experiment $\lambda = 1$ [16].

D. Implementation Details

- Region Proposals
  In this paper, we generate 28 region proposals at each location on the fifth feature map with 4 scales $scales = [2,3,4,5]$ and 7 aspects $h/w = [\frac{1}{6}, \frac{1}{3}, \frac{2}{3}, \frac{1}{1}, \frac{3}{2}, \frac{3}{1}, \frac{6}{1}]$. The aspect ratios come from the statistic result of the training dataset showed in Fig. 3. So at each spatial location, *cls* layer has $2*28$ scores $[p_1, p_0]$, where p_1 is the probability of the proposals belonging to the target and p_0 is the probability of the proposals belonging to the backgrounds, and *reg* layer outputs $4*28$ coordinates $[t_x, t_y, t_w, t_h]$. For

the fifth feature map of size of $W*H$, there are totally $28*W*H$ (nearly ≈ 2000) region proposals [17].

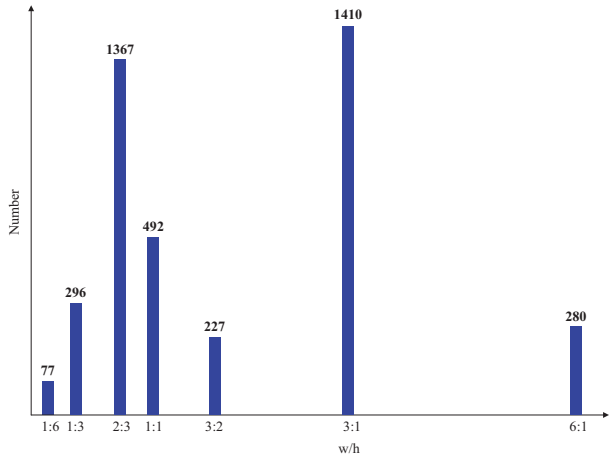

Fig. 3. Aspect ratios of region proposals. This is the aspect ration statistic result coming from the training set.

- Training
  Region proposal network is trained end-to-end based on back propagation (BP) [18] and stochastic gradient descent (SGD) method [19]. We consider 256 proposals generated from one training image as one training batch to train our model, and there are two different methods to select 256 candidate proposals. The details of batch selection principle is introduced in **Batch Selection**. The parameters of all layers are initialed by a normal distribution with zero-mean and 0.01 variance, and then they are updated with training batches.
- Batch Selection
  Each batch contains 256 proposal regions generated from each training image. There are two main methods for a batch selection.

  – Randomly Sampling
    The first method is randomly sampling. In this paper, one training image can generate $28*W*H$ (nearly ≈ 2000) proposals, then we generate a training batch with 256 proposals randomly sampled from $28*W*H$ (nearly ≈ 2000) proposals which is based on the proportion between positives and negatives is 1:1 (the positives are the proposals which have an IoU overlap with ground truth is greater than a default threshold, in our experiment, we set the threshold as 0.5, the negatives are those proposals which have an IoU overlap with ground truth is less than 0.3 and larger than 0.1), while the number of positives can not satisfy more than 128 requirements, so when positive proposal regions are fewer than 128, the batch is padded with negatives.

  – OHEM
    The second method is based on online hard examples mining (OHEM) [20]. OHEM algorithm is a boost method of the randomly sampling. When selection the batch for training the ConvNet model, firstly, it takes all the generated nearly 2000 proposals as input and outputs the sum of classification loss and regression loss for each proposal. Then, making non-maximum suppression (NMS), which is often used as an intermediate optimization step in computer vision field to select the best one among many candidate bounding boxes, for all the proposals according to their losses. Thirdly, the proposals are sorted based on their losses according to a descending order, the top 256 samples are returned for which the current network performs worst and they are consider as the hardest examples. Finally, these 256 samples consist of a batch for network training. Therefore, OHEM is focused on the examples which are more difficult to be classified, and it makes the network to choose training samples intelligently, which will decrease the false positives and improve the detection performance.
- Classifier and regressor

III. Experiments

The testing task on the region proposal network is divided into three steps. First, we take a test image as input, using the region proposal network to extract features and then generate ≈ 256 region proposals of guide-wires. Secondly, classifying all the region proposals, and the region proposal is considered as the target if its corresponding score is larger than a threshold, and then regressing their coordinates, otherwise they are considered as the negatives. Finally, selecting the detected proposals with the highest score using NMS.

We validate our method with Caffe framework [21] build on Linux system with GPUs, all of which are available online. Our datasets come from the images from [22], which have totally 5092 frames X-ray images. They are from 22 different sequences composed of 12 different shapes of guide-wires showed in Fig. 5. We divide the 22 sequences into two sets, one is training set including 19 sequences, and the rest 3 sequences are used for testing.

Our method is based on the region proposal network in [17], while we make some changes to the network, which gives a big improvement. We apply OHEM rather than random sampling method for a batch selection used for the region proposal network training. We report an accuracy to demonstrate its validation. The accuracy is evaluated by AP which is the area under *PR* curve showed in Fig. 4, where *P* presents precision, *R* presents recall [22], the larger the area, the better the performance. From the *PR* curve in Fig. 4 and detection results in TABLE I, we can clearly proof that, compared with randomly sampling, OHEM can improve the AP by nearly 3% which demonstrates that OHEM is more effective than randomly sampling method for batches selection. Then we

make a comparison with our previous guide-wires detection results based on a cascade Adaboost classifier with LBP feature introduced in [9], whose AP is only 2.4% and the detection speed is only 8fps, which can not realize real-time detection. While the AP of our ConvNet method proposed in this paper with random sampling method for batch selection is 82.6%, and the AP value of our ConvNet method with OHEM method is 85.8%. What's more, our detection speed can achieve 40fps which is 5 times faster than [9] and can perform real-time detection. And a visualization detection result is showed in Fig. 7. Our experiment results show that the ConvNet is capable of achieving effective, efficient and real-time guide-wires detection.

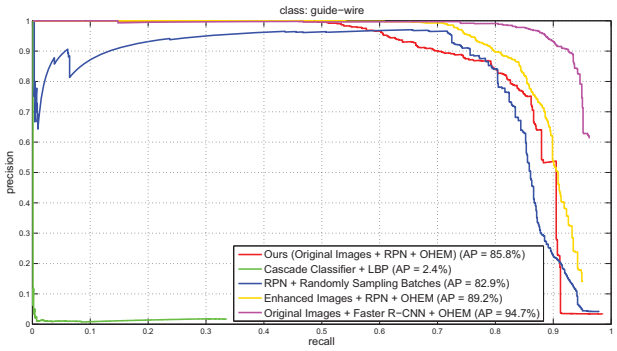

Fig. 4. Average precision (%). This a *PR* curve of all our experiments results. The red curve is the result of our proposed method with original images and OHEM method for batches selection, the green curve is the result of method in [9], the blue one is the result of our proposed method with training batches randomly sampled, the yellow one is the result of region proposal network with enhanced images, and the the pink one is the result of Faster R-CNN, a more complex Convnet, with original images.

TABLE I
DETECTION RESULTS.

Method	CPU	GPU	AP (%)	Speed (fps)
Cascade Classifier in [9]	√		2.4	8
Ours1[1]		√	**82.6**	**40**
Ours2[2]		√	**85.8**	**40**

IV. DISCUSSION

In this section, we explore changes to our architecture and justify our design choices with experiments on our X-ray image test. For a direct comparison, we also use the AP and detection speed for evaluation. We use the AP of original images with region proposal network as a baseline of all the detection methods. Our research mainly includes two aspects: does X-ray images pre-processing (it mainly refers to image enhancement) is helpful and does a more complex ConvNet model can reach better detection results.

[1]Training batches are selected with randomly sampling method.
[2]Training batches are selected with OHEM.

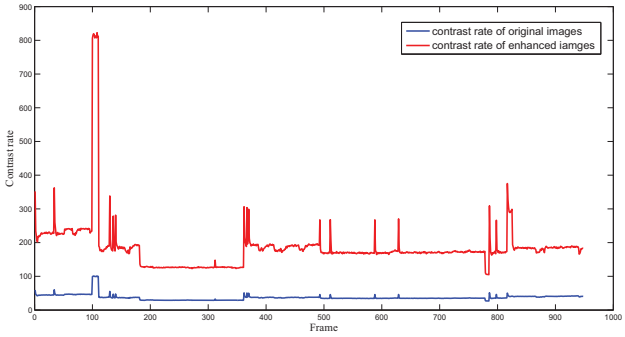

Fig. 6. A contrast rate comparison on test images between pre- and post-image enhancement. The red curve is the contrast rate of the enhanced images, and the blue one is the contrast rate of original images. The contrast rate is defined with root mean square contrast rate C_σ, where $C_\sigma = \sqrt{\frac{1}{WH} * \sum_{I_{WH}} I(x,y) - \bar{I}_{WH}}$, and $\bar{I}_{WH} = \frac{1}{WH} \sum_{I_{WH}} I(x,y)$, $\bar{I}_{WH}$ is the average pixel value of image I whose width is W and height is H, $I(x,y)$ is the pixel value at (x,y) [27].

A. *How much does image pre-processing help ?*

Top-hat enhancement algorithm is generally used to improve the medical image contrast rate. So we use a multiple Top-hat enhancement method for X-ray images pre-processing. It is mainly based on the morphological opening and closing operations [23]. The details of the multiple Top-hat enhancement algorithm are introduced in [24] [25] [26], and the ConvNet model remains region proposal network. We make a contrast rate comparison between the original images and the enhanced images on the test dataset which is showed in Fig. 6. We can clearly see that the image contrast rate improves nearly $6 \sim 7$ times. We make the guide-wire detection on original and enhanced X-ray images respectively. The detection accuracy and speed are listed in Table II. From Table II and the detection accuracy showed in Fig. 4, the AP improves nearly 3.4%, and the detection speed remains 40fps without considering the image pre-processing time which almost costs fewer time. We can conclude that the enhancement method can highlight guide-wires features which are benefit for guide-wire detection.

B. *The more complex the model, the better the results ?*

Region proposal network is a powerful method for guide-wires detection, while it is not the only one, and there are some much larger models, which are more effective at improving objects detection accuracy. So we apply Faster R-CNN [17], a most popular detection framework for our guide-wires detection task, it is a larger model than region proposal network. We fine-tune a pre-trained Faster R-CNN with our training dataset. We compare these two methods with AP on our test images, the results are showed in TABLE II. Faster R-CNN is a much larger model with more parameters than alternatives, it has two-stage classifiers and regressors which can give a better performance than region proposal network, from AP curve in Fig. 4, its detection accuracy on the test set (AP = 94.7%) outperforms much than region proposal network

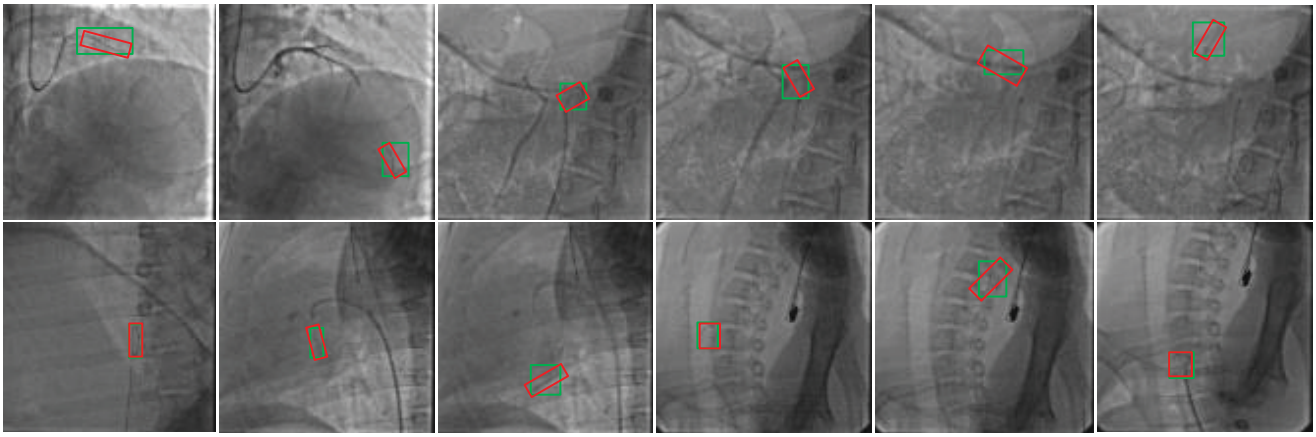

Fig. 5. X-images are sampled from each of the 12 classes. The green boxes are the ground truth which are the annotated bounding boxes with general annotating method and the red angular ones are the annotated bounding boxes with our novel annotation approach.

(AP = 85.8%), while the detection speed is nearly 2 times slower ($\approx 13 fps$). Therefore, allowing for a balance between detection accuracy and speed, we use region proposal network for guide-wire detection.

TABLE II
DETECTION RESULTS WITH THREE DIFFERENT ARCHITECTURAL CHANGES TO OUR CONVNET MODEL AND X-RAY IMAGE CONTRAST RATE.

Method	AP (%)	Speed on GPU (fps)
Original + Our ConvNet[2]	85.8	**40**
Enhanced + Our ConvNet[3]	89.2	**40**
Original + Faster R-CNN[4]	**94.7**	20

V. CONCLUSIONS

This paper introduces a novel and effective method for guide-wire detection in PCI navigation. Compared with traditional detection methodour method is based on the ConvNet which combines the recently proposed region proposal network with a different bounding boxes annotation method. Our ConvNet model can improve guide-wires' descriptors which is helpful to improve the detection accuracy, and the modified annotation method which uses angular bounding boxes to calibrate the guide-wires, facilitates the training of a more robust regressor. To justify our design choices, we conduct extensive experiments to evaluate the availability. Our experiments mainly focus on two aspects: on the one hand, we use different detection model: a traditional simple method (a cascade Adaboost classifier with LBP feature) and a more complex network model (faster R-CNN); on the other hand, we apply different kinds of detection framework on original images and enhanced ones which have higher contrast rate than original images. Finally, considering detection accuracy and speed, we employing region proposal network model with OHEM on enhanced images for guide-wire detection. The promising detection accuracy (AP = 89.2%) and outperforming detection speed (40fps) is a convincing proof of our choice. What's more, we find our proposed method is not sensitive to guide-wires' various appearances and is robust to other skeletons' interferences, so it further validate the reliability of our method. Our further work will focus on guide-wire tracking which can benefit for cardiac motion monitoring.

[2]Ours with original images and a region proposal network.

[3]The method with pre-processing images and a region proposal network.

[4]The method with original images and a Faster R-CNN model combined with OHEM.

Acknowledgements This research is supported in part by the National Natural Science Foundation (NNSF) of China (Grants 61533016, 61611130217, 61673379, U1613210 and 61421004), the Strategic Priority Research Program of the CAS (Grant XDB02080000), the Beijing science and technology project (Z161100001516004) and the Beijing Natural Science Foundation (Grant 4161001).

REFERENCES

[1] F. W. Mohr, M.-C. Morice, A. P. Kappetein, T. E. Feldman, E. Ståhle, A. Colombo, M. J. Mack, D. R. Holmes, M.-a. Morel, N. Van Dyck *et al.*, "Coronary artery bypass graft surgery versus percutaneous coronary intervention in patients with three-vessel disease and left main coronary disease: 5-year follow-up of the randomised, clinical syntax trial," *Lancet*, vol. 381, no. 9867, pp. 629–638, 2013.

[2] H. Heibel, B. Glocker, M. Groher, M. Pfister, and N. Navab, "Interventional tool tracking using discrete optimization," *IEEE Transactions on Medical Imaging*, vol. 32, no. 3, pp. 544–555, 2013.

[3] B.-J. Chen, Z. Wu, S. Sun, D. Zhang, and T. Chen, "Guidewire tracking using a novel sequential segment optimization method in interventional x-ray videos," in *IEEE Proceedings of Symposium on Biomedical Imaging (ISBI)*. IEEE, 2016, pp. 103–106.

[4] P.-L. Chang, A. Rolls, H. De Praetere, E. Vander Poorten, C. V. Riga, C. D. Bicknell, and D. Stoyanov, "Robust catheter and guidewire tracking using b-spline tube model and pixel-wise posteriors," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 303–308, 2016.

[5] H. Fazlali, N. Karimi, S. M. R. Soroushmehr, S. Sinha, S. Samavi, B. Nallamothu, and K. Najarian, "Vessel region detection in coronary x-ray angiograms," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 1493–1497.

[6] A. Hernández-Vela, C. Gatta, S. Escalera, L. Igual, V. Martin-Yuste, M. Sabate, and P. Radeva, "Accurate coronary centerline extraction, caliber estimation, and catheter detection in angiographies," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 6, pp. 1332–1340, 2012.

[7] T. H. Heibela, B. Glockera, M. Grohera, N. Paragios, N. Komodakis, and N. Navaba, "Discrete tracking of parametrized curves," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1754–1761.

[8] N. Honnorat, R. Vaillant, and N. Paragios, "Graph-based geometric-iconic guide-wire tracking," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2011, pp. 9–16.

[9] L. Wang, X.-L. Xie, Z.-J. Gao, G.-B. Bian, and Z.-G. Hou, "Guide-wire detecting using a modified cascade classifier in interventional radiology," in *IEEE Proceedings of Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 1240–1243.

[10] H. R. Roth, A. Farag, L. Lu, E. B. Turkbey, and R. M. Summers, "Deep convolutional networks for pancreas segmentation in ct imaging," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Medical Imaging*. International Society for Optics and Photonics, 2015, p. 94131G.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1. IEEE, 2005, pp. 886–893.

[12] T. Lindeberg, "Scale invariant feature transform," *Scholarpedia*, vol. 7, no. 5, p. 10491, 2012.

[13] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[14] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *IEEE Conference on European Conference on Computer Vision (ECCV)*. Springer, 2014, pp. 818–833.

[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[16] R. Girshick, "Fast r-cnn," in *IEEE Proceedings of International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[18] Y. Anzai, *Pattern Recognition & Machine Learning*. Elsevier, 2012.

[19] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT*. Springer, 2010, pp. 177–186.

[20] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," *arXiv preprint arXiv:1604.03540*, 2016.

[21] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[22] Z.-Q. Feng, G.-B. Bian, X.-L. Xie, Z.-G. Hou, and J.-L. Hao, "Design and evaluation of a bio-inspired robotic hand for percutaneous coronary intervention," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 5338–5343.

[23] G. J. Banon, J. Barrera, and U. M. Braga-Neto, "Mathematical morphology and its applications to signal and image processing," in *IEEE Proceedings in Symp. Mathematical Morphology*. Springer, 2015.

[24] "Analysis of new top-hat transformation and the application for infrared dim small target detection."

[25] X. Bai, F. Zhou, and B. Xue, "Image enhancement using multi scale image features extracted by top-hat transform," *Optics & Laser Technology*, vol. 44, no. 2, pp. 328–336, 2012.

[26] L. Wang, D.-J. Li, X.-L. Xie, and Z.-G. Hou, "A vessel contour detection and estimation method for robot assisted endovascular surgery," in *Proceedings of the 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2016, pp. 5338–5343.

[27] E. Peli, "Contrast in complex images," *JOSA A*, vol. 7, no. 10, pp. 2032–2040, 1990.

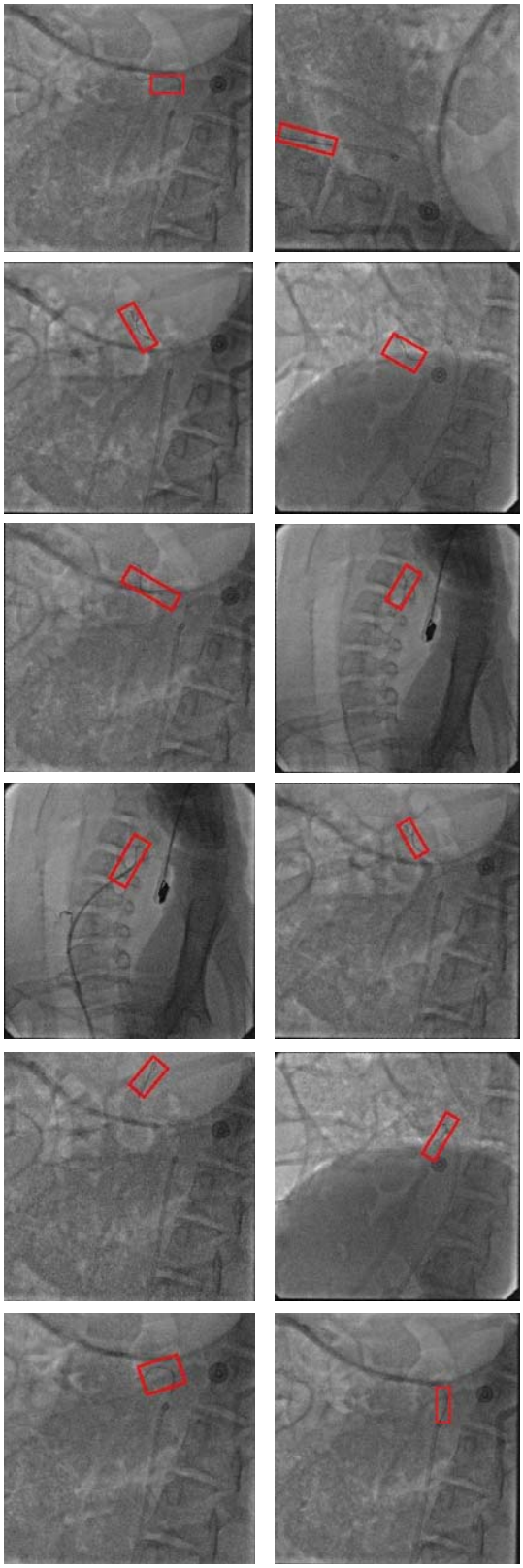

Fig. 7. Selected detection results on our X-ray image test set using the region proposal network. The model is ZF and training data is the enhanced X-ray images. Our method detects guide-wires of a wide range of aspect ratios and angles. Each output box is associated with a detection score in [0,1]. A score thresh of 0.5 is used to display these images.