

“© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

3D Ultrasound Spine Image Selection Using Convolution Learning-to-Rank Algorithm

Juan Lyu¹, Sai Ho Ling², Sunetra Banerjee², Jenny J.Y. Zheng⁴, Ka-Lee Lai³, De Yang³, Yong-Ping Zheng³, Steven Su²

Abstract—3D Ultrasound imaging has become an important means of scoliosis assessment as it is a real-time, cost-effective and radiation-free imaging technique. However, the coronal images from different depths of a 3D ultrasound image have different imaging definitions. So there is a need to select the coronal image that would give the best image definition. Also, manual selection of coronal images is time-consuming and limited to the discretion and capability of the assessor. Therefore, in this paper, we have developed a convolution learning-to-rank algorithm to select the best ultrasound images automatically using raw ultrasound images. The ranking is done based on the curve angle of the spinal cord. Firstly, we approached the image selection problem as a ranking problem; ranked based on probability of an image to be a good image. Here, we use the RankNet, a pairwise learning-to-rank method, to rank the images automatically. Secondly, we replaced the backbone of the RankNet, which is the traditional artificial neural network (ANN), with convolution neural network (CNN) to improve the feature extracting ability for the successive iterations. The experimental result shows that the proposed convolutional RankNet achieves the perfect accuracy of 100% while conventional DenseNet achieved 35% only. This proves that the convolutional RankNet is more suitable to highlight the best quality of ultrasound image from multiple mediocre ones.

I. INTRODUCTION

Scoliosis is a deformation of spinal cord to usually S or C shape and a degree of curvature occurs in the coronal plane. Since, the prime victims of this disease are adolescents, it is very important to screen schoolchildren. According to current study [1], about 0.47-5.2% of total population is affected by scoliosis. The traditional diagnose process of scoliosis is detection of spine curvature angle using Cobbs method on standing radio-graph. Traditional techniques such as MRI and CT Scan uses radiation and their frequent exposure may causes cancer.

Zheng et al. demonstrated that Scolioscan is a reliable and radiation-free system to measure spine deformity using

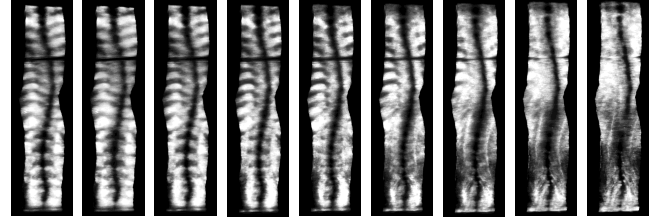


Fig. 1. The samples of nine 2D ultrasound coronal images in different depths of a 3D spine image.

3D ultrasound imaging for Adolescent Idiopathic Scoliosis (AIS) patients. This method has been proven to give superior results compared to the conventional radio-graphic Cobbs method [2]. In the recent past, the Scolioscan method has been further strengthened using an automatic measurement of spine curvature through image pre-processing and a newly developed two-fold threshold strategy [3]. However, this method needs work particularly in the area of its scalability, if it has to be applied by multiple hospitals dealing with a large group of patients. The current medical facility needs a smart system which can detect scoliosis fully automatically and quickly.

The biggest challenge in clinical diagnosis for ultrasound imaging is high inter- and intra-operator variability. By applying machine learning techniques on ultrasound image, a great success has been achieved in the diagnosis of disease with medical images [4, 5]. But low resolution and high speckle noise are some of the major problems associated with a raw ultrasound image. To overcome this, a pre-processing technique has to be applied on the raw ultrasound image. One of the pre-processing techniques used is deep learning which extracts numerical features from the ultrasound image [6]. Convolutional neural network was proposed to transform speckled, blurry ultrasound images into better quality images [7]. This our research, we aim to develop a system which is capable of automatically refining and providing a real time feedback to the user during capturing of ultrasound image.

The objective of our research is to find the best ultrasound image to detect scoliosis automatically and thereby reducing the workload of doctors. In this paper, we have used Volume Projection Image (VPI) taken from Scolioscan system. We want to select the best image from nine images of a patient as shown in Fig.1.

The ultrasound images in different depths of a 3D image have different definitions. One of the important parameter

¹J. Lyu is with the College of Information and Communication Engineering, Harbin Engineering University, Harbin, 150001, China Juanlyu91@gmail.com

²S.H. Ling, S. Banerjee, and S.W. Su are with the School of Biomedical Engineering, University of Technology Sydney, NSW, Australia Steve.Ling@uts.edu.au, Sunetra.Banerjee@student.uts.edu.au, Steven.Su@uts.edu.au

³K.L. Lai, D. Yang, and Y.P. Zheng are with the Department of Biomedical Engineering, The Hong Kong Polytechnic University, Hong Kong kelly.lai@polyu.edu.hk, de.de.yang@connect.polyu.hk, yongping.zheng@polyu.edu.hk

⁴J.J.Y. Zheng is with the Department of Computer Science, Imperial College London jia.zheng16@imperial.ac.uk

is the the middle spine line. However, not all images can be used for the measurement of the scoliosis curve angle. To select the best image from all the depths, we propose a convolutional RankNet based on the following ideas. Firstly, as there are several similar images for each patient, it is difficult to distinguish them using classification methods. Instead, we consider it as a ranking problem. By ranking them in sequence from 1st to 9th (descending picture quality), it is easier to select the best image according to the ranking result. As the RankNet [8] is a pairwise learning-to-rank approach, we transfer our data to pairs based on their number of images before ranking. Secondly, traditional RankNet requires large computation power as its network structure used conventional ANN and our images are significantly larger, at 1790×640 . CNN can reduce the computations requirements through its weight sharing strategy and extract features more effectively than ANN. Therefore, we design a Convolution RankNet to solve the ultrasound image selection task by combining the CNN and the main principle of RankNet.

II. METHODOLOGY

In this section, we firstly explain the reason why we adopt ranking algorithm. Then, we describe the architecture of RankNet and extend it to a deep convolution structure to select the best image from the ranking result.

A. Why RankNet

For our ultrasound imaging system, it produces a 3D ultrasound image for each patient which consists of nine 2D images. The images in different depths have different imaging definitions. For our further study, our task is to select the best image from them based on a sharp, clear, continuous mid dark spine line.

However, as shown in Fig. 2, several images are of great similarities with the best image when they are adjacent, especially for the mid dark line. As a result, a deep classification model is difficult to discriminate them using the raw images. Therefore, we consider it as a ranking problem based on their probabilities of good image. The ranking algorithms, such as RankNet [8], LambdaRank [9], LambdaMART [10], and Listwise Ranking algorithms [11, 12], are widely used in web search tasks in search engine [9]. But for our problem, it becomes a simpler ranking task as we only need to rank the best image correctly, the order of rest images can be ignored. In this case, we design a convolutional RankNet which is developed based on the RankNet to solve our problem.

B. Convolutional RankNet

The traditional Ranknet [8] is a pairwise learning-to-rank algorithm, its structure is based on the traditional neural network. As convolutional neural network is much powerful to extract effective features of images, so we extend RankNet to a convolutional version.

The structure of our proposed convolutional RankNet is shown in Fig. 3. Its backbone is a 3-layer convolutional neural network. The first two layers are typical convolution, max-pooling groups. While after the third convolution layer,

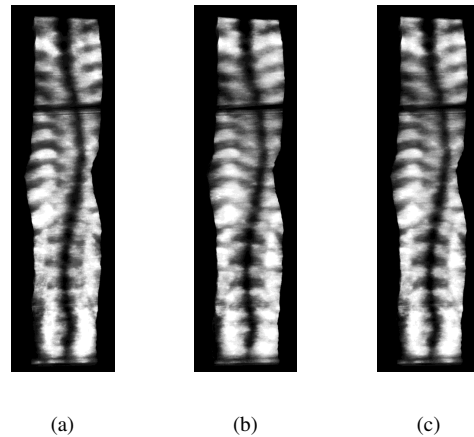


Fig. 2. The illustration of the similarity between the best image and its adjacent images. Among them, (b) is the best image, (a) and (c) are its adjacent images.

a global average pooling layer is used instead of max pooling layer. Then, a fully-connected layer is followed from it and we are able to compute the output score s . For a given input x , the network maps the input to a number $s = f(x)$. The architecture of convolutional RankNet consists of two branches, with shared weights, treated equivalent to single entity. The reason is that they use the same initialization and the same gradient during back-propagation process.

For each pair x_i, x_j , their corresponding output scores are $s_i = f(x_i)$ and $s_j = f(x_j)$, respectively. If s_i is larger than s_j , it means x_i is ranked higher than x_j . The prediction probability that x_i is ranked higher than x_j is defined as

$$P_{ij} \equiv \frac{e^{s_{ij}}}{1 + e^{s_{ij}}} \quad (1)$$

where the $s_{ij} \equiv f(x_i) - f(x_j)$, the difference of the outputs of two branches. We can see that the prediction probability is a sigmoid function of s_{ij} .

The loss function we use is cross entropy which is defined as

$$L(s_{ij}) = -\bar{P}_{ij} \log P_{ij} - (1 - \bar{P}_{ij}) \log(1 - P_{ij}) \quad (2)$$

where \bar{P}_{ij} is the target probability. In this work, since we only select the best image, $\bar{P}_{ij} \in \{0, 1\}$. If x_i ranked higher than x_j , \bar{P}_{ij} is 1, otherwise it is 0. As we know, each pair of samples are with different labels, for this task, there are total 8 pairs for each patient. Therefore, if and only if all the 8 pairs are predicted correctly, the ranking accuracy is 1; otherwise it is 0.

Above all, the convolutional RankNet can distinguish the best image from others by increasing the difference of the output scores during training process. Finally, according to the ranking result, the highest ranked image is the best image we selected from the nine ultrasound images of each patient.

III. EXPERIMENT

We first introduce the details of our data, data source, data size and data format, respectively. Then, we present how

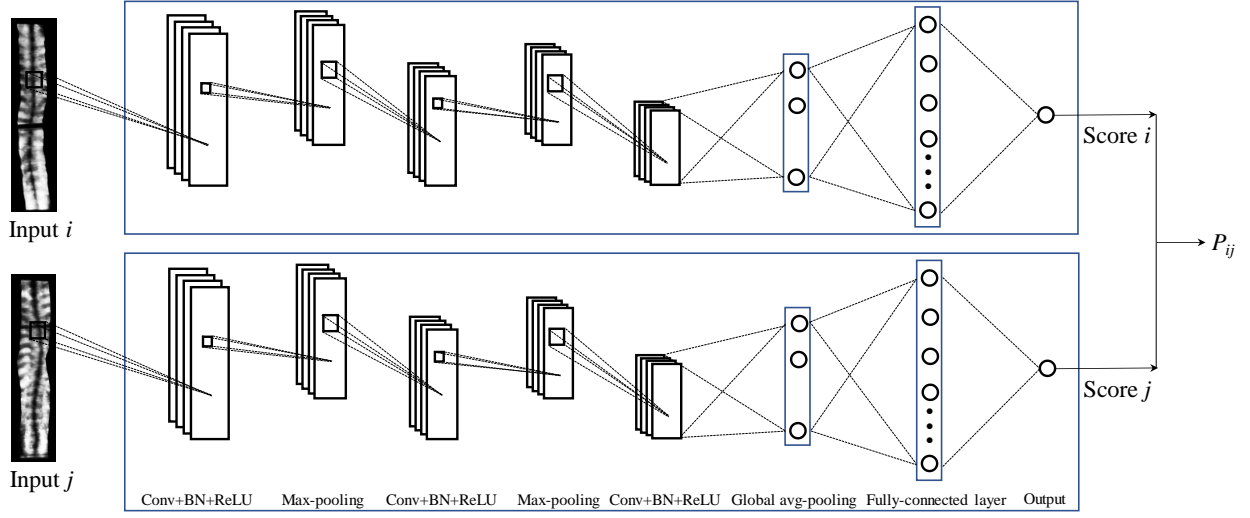


Fig. 3. The architecture of proposed convolutional RankNet.

to prepare the data before training and testing and describe the details of our proposed structure. Finally, we show the experimental result of our proposed model.

A. Data

Our data are collected by the Scolioscan system (Model SCN801, Telefield Medical Imaging Ltd, Hong Kong) which is developed using 3D ultrasound imaging method to obtain spine VPI for the assessment of scoliosis [2, 13-15]. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board. For each patient, we split his/her 3D spine ultrasound image into nine 2D vertebral anatomical images based on the imaging depths. The 3D ultrasound images have different sizes among the patients, so we resize them uniformly in the same size of 1790×640 , and they are in bitmap format.

B. Data Preprocessing

In this study, we have studied 3D ultrasound spine images from 150 patients. Since we only need to select the best image, we label the best image as 1 and other as 0. Our labels are marked manually by an experienced operator who based on the criteria of a clear dark line in the middle representing the spine profile, and also other spinal features as clear as possible in the image including transverse processes and ribs. Then, the pairing operation is just between the images who have different labels. Therefore, the nine images of each patient are paired into a 8-pair group. Finally, we have 150 groups data and divide them into training and testing sets by 75:75. We adopt 2-fold cross validation to get the final result.

C. Experimental Setup

For our proposed architecture, the size of convolution kernels is 7×7 with a stride of 2, and each convolutional

TABLE I
THE COMPARISON RESULT WITH DENSENET.

Number of Images	Method	Testing Accuracy (%)
Three out of Nine	DenseNet [16]	95.95
Nine	DenseNet [16]	35.00
Nine	this work	100.00

layer produce 16 features. The following max pooling layers has a kernel size of 4×4 with a stride of 4. For the fully-connected layer, there are 100 neurons. The active function of all layers is ReLU except the output layer, which is the sigmoid function. Moreover, we use Adam optimizer with a initialized learning rate of 0.001 in the training process. We train the network 20 epochs and the batch size is 1 as the size of the data is too big to train them using bigger batch size.

D. Result and Discussion

We evaluate the performance of the proposed convolutional RankNet by comparing with the classification method of DenseNet [16]. The comparison result is shown in Table 1. DenseNet, first proposed in 2017, had greatly improved the performance in classification problems. When we applied DenseNet to select the best image from three images of our data set, we got 95.95%. However, when we applied it to nine images of each of the 150 patients, the DenseNet method was unable to handle large number of images and its accuracy dropped significantly to around 35%. This showed the method is not very suitable to our purpose. To overcome this problem, we adopted RankNet. Finally, we implement the proposed convolutional RankNet on our data. After training 20 epochs and 2-fold validations, the final testing accuracy is 100%. In addition, we randomly select a pair

of images of a patient to visualize the output scores during the training process, it is shown in Fig.4. It reveals that the similar spine images have quite close output scores at the beginning, however, as the training continues, the convolutional RankNet can better distinguish them.

The experimental result indicates that the proposed convolutional RankNet is more than capable to automatically select good images from a large set of non pre-processed spine images.

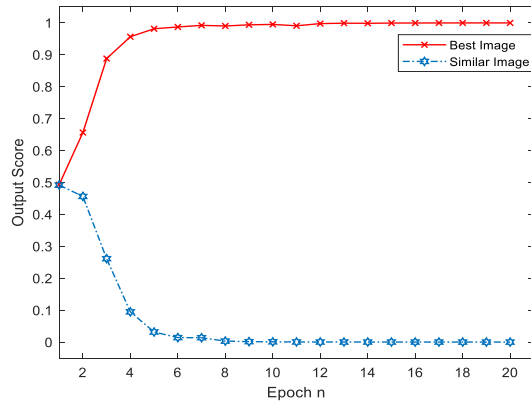


Fig. 4. The output scores of the best image and its similar image during twenty epochs.

IV. CONCLUSIONS

To select the good image from similar images using the raw data, we propose a convolutional RankNet which is a combination of traditional RankNet and CNN. RankNet is a pairwise learning-to-rank algorithm, by compare the data in pairs to find out the final ranking order. While its backbone is traditional ANN, we extend it in a convolutional version for extracting more effective features of the spine images. The result illustrate that the proposed convolutional RankNet can solve our problem perfectly.

In the future, we will verify it on a bigger dataset after we collect more spine images. Moreover, we would like to try to select several good images for each patient as they are quite similar, to check their performance difference in subsequent applications.

V. ACKNOWLEDGMENT

The work described in this paper was substantially supported by Research Grant Council of Hong Kong (R5017-18)

REFERENCES

- [1] M. R. Konieczny, H. Senyurt, and R. Krauspe, "Epidemiology of adolescent idiopathic scoliosis," *Journal of children's orthopaedics*, vol. 7, no. 1, pp. 3–9, 2012.
- [2] Y.-P. Zheng, T. T.-Y. Lee, K. K.-L. Lai, B. H.-K. Yip, G.-Q. Zhou, W.-W. Jiang, J. C.-W. Cheung, M.-S. Wong, B. K.-W. Ng, J. C.-Y. Cheng, et al., "A reliability and validity study for scolioscan: a radiation-free scoliosis assessment system using 3d ultrasound imaging," *Scoliosis and spinal disorders*, vol. 11, no. 1, p. 13, 2016.

- [3] G.-Q. Zhou, W.-W. Jiang, K.-L. Lai, and Y.-P. Zheng, "Automatic measurement of spine curvature on 3-d ultrasound volume projection image with phase features," *IEEE transactions on medical imaging*, vol. 36, no. 6, pp. 1250–1262, 2017.
- [4] K. Kalyan, B. Jakhia, R. D. Lele, M. Joshi, and A. Chowdhary, "Artificial neural network application in the diagnosis of disease conditions with liver ultrasound images," *Advances in Bioinformatics*, vol. 2014, no. 708279, 2014.
- [5] S. K. Veeramani and E. Muthusamy, "Detection of abnormalities in ultrasound lung image using multi-level svm classification," *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 2016, no. 29, pp. 1844–1852, 2016.
- [6] S. Joo, W. K. Moon, and H. C. Kim, "Computer-aided diagnosis of solid breast nodules on ultrasound with digital image processing and artificial neural network," in *Engineering in Medicine and Biology Society, 2004 26th Annual International Conference of the IEEE*. IEEE, 2004, pp. 1397–1400.
- [7] S. Vedula, O. Senouf, A. M. Bronstein, O. V. Michailovich, and M. Zibulevsky, "Towards ct-quality ultrasound imaging using deep learning," *CoRR*, vol. abs/1710.06304, 2017.
- [8] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender, "Learning to rank using gradient descent," in *Proceedings of the 22nd international conference on Machine learning*. ACM, 2005, pp. 89–96.
- [9] C. J. Burges, "From ranknet to lambdarank to lambdamart: An overview," *Learning*, vol. 11, no. 23-581, p. 81, 2010.
- [10] Q. Wu, C. J. Burges, K. M. Svore, and J. Gao, "Adapting boosting for information retrieval measures," *Information Retrieval*, vol. 13, no. 3, pp. 254–270, 2010.
- [11] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li, "Learning to rank: from pairwise approach to listwise approach," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 129–136.
- [12] M. Taylor, J. Guiver, S. Robertson, and T. Minka, "Softrank: optimizing non-smooth rank metrics," in *Proceedings of the 2008 International Conference on Web Search and Data Mining*. ACM, 2008, pp. 77–86.
- [13] C.-W. J. Cheung, S.-Y. Law, and Y.-P. Zheng, "Development of 3-d ultrasound system for assessment of adolescent idiopathic scoliosis (ais): and system validation," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013, pp. 6474–6477.
- [14] C.-W. J. Cheung, G.-Q. Zhou, S.-Y. Law, K.-L. Lai, W.-W. Jiang, and Y.-P. Zheng, "Freehand three-dimensional ultrasound system for assessment of scoliosis," *Journal of Orthopaedic Translation*, vol. 3, no. 3, pp. 123–133, 2015.
- [15] C.-W. J. Cheung, G.-Q. Zhou, S.-Y. Law, T.-M. Mak, K.-L. Lai, and Y.-P. Zheng, "Ultrasound volume projection imaging for assessment of scoliosis," *IEEE transactions on medical imaging*, vol. 34, no. 8, pp. 1760–1768, 2015.
- [16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.