Localization and Identification of Lumbar Intervertebral Discs on Spine MR Images with Faster RCNN Based Shortest Path Algorithm

Merve Zeybel[1,2(✉)] and Yusuf Sinan Akgul[2]

[1] TUBITAK BILGEM, Gebze, Kocaeli, Turkey
kaya.merve@tubitak.gov.tr
[2] Department of Computer Engineering, Gebze Technical University, Gebze, Kocaeli, Turkey
akgul@gtu.edu.tr

Abstract. Automatic detection and identification of the intervertebral discs on the spine MR images is a challenging task due to similarity of the discs on the same image, size and shape differences between subjects, and poor resolution. Many deep learning-based methods have been proposed recently to achieve automated detection and identification of human intervertebral discs. However, since there is usually only a small amount of labeled vertebral images available, employing an end-to-end deep learning system is not easily achievable. In this paper, we use a multi-stage deep learning system to detect and identify human lumbar discs from MRI data. We first use a Faster Region based Convolutional Neural Network (FRCNN) method to detect candidate disc positions. Each candidate from the FRCNN becomes a node in a weighted graph structure. The edge weights between the nodes are calculated using the FRCNN scores and the scores from a Binary Classifier Network (BCN) that tests compatibility of the nodes of the edge. A novel application of Dijkstra's shortest path algorithm in this network produces both localizations and identifications of the lumbar discs in a globally optimal manner. Experiments on our dataset of 80 MRI scans from 80 patients achieved very promising results as they exceeded the state of the art alternatives on similar datasets.

Keywords: Spine Lumbar MRI · Detection of intervertebral disc · Faster RCNN · Dijkstra's Algorithm

1 Introduction

Detecting the anatomical structure of the human vertebrae is crucial for many applications such as diagnosis of degenerative discs, finding herniated and slipping discs [16] and detecting abnormality in the spine. With the current medical

practice, this operation is mostly performed manually which makes it subjective, prone to human errors, time-consuming, and expensive. As a result, many computerized methods have been proposed to detect the anatomical structure of the human vertebrae. For example, Oktay and Akgul [11] introduce a model-based Markov-chain-like graphical model. Lootus *et al.* [10] present a method that employs a graphical model combined with a Deformable Part Model. These classical methods generally consider the problem as many sub-localization problems (one for each intervertebral disc or vertebra) and then combine the results of these sub-localizations in a graphical method that models the whole vertebra. The sub-localization stages usually employ hand-crafted feature extraction based machine learning techniques.

Recently, deep learning-based methods have exceeded state of the art results for the detection of human vertebral structure. Forsberg *et al.* [3] use two Convolutional Neural Network (CNN) models to assign scores to given image patches and then combine the results of these networks under a graphical model to enforce the whole vertebrae constraints. Chen *et al.* [2] also employ a random forest classifier to get coarse localization of the vertebra efficiently. These coarse positions are passed to a joint CNN model to enforce both local and pairwise constraints of the vertebrae. Still on the same topic, Wang *et al.* [14] proposes a multi-stage system that learns vertebrae specific deep features using auto-encoders and then enforces anatomical context-related constraints later.

End-to-end deep learning systems are known to produce better results compared to sequential multi-stage (or pipe-lined) systems partly because every aspect of end-to-end systems is directed towards the final goal [6]. With the end-to-end approach, since there are no intermediate stages, there are no stage combination or fusion decisions which makes the overall system more robust. However, end-to-end systems need large amounts of training data if the task at hand is not trivial. Since the labeled data for the vertebral MR images is scarce, deep learning methods usually avoid building an end-to-end system that takes the whole MR image of the vertebrae as an input and produces the final positions and labels of the individual discs. All the cited methods above propose systems with stages that employ one or more deep networks whose results are combined later, which makes it possible to train networks on small patches of the whole MR images. Although multi-stage approaches are more convenient for the network training, fusion of the resulting data in the subsequent stages should be done in a robust way to eliminate errors caused by this process. We argue that there should be guarantees of optimality for the data fusion or combination steps, which should address the robustness problems of stage based methods.

In this paper, we follow the same multi-stage deep learning approach to automatically localize and identify the InterVertebral Discs (IVD) of human vertebra from MR images. Different from the other systems, we propose a method that optimally combines the data produced by the system stages. Our method consists of two stages. In the first stage, we use Faster RCNN (FRCNN) network [12] to learn every single lumbar IVD individually. In the second stage, we use a Binary Classifier Network (BCN) to learn about two neighboring discs to use

more global context information about the candidate disc positions produced from FRCNN. Our main contribution in this paper is the fusion of prediction scores of the FRCNN's and the confidence scores of BCN's in the shortest path setting to make a globally optimal disc localization and identification decision. We build a graph whose nodes represent the candidate positions produced by the FRCNN. The edge that connects two nodes in this graph is assigned a weight produced by our BCN about these two candidate positions. The shortest path through this network makes a globally optimal disc localization and identification decision, which can be achieved in polynomial time using Dijkstra's shortest path algorithm.

Although the proposed method is not an end-to-end trainable deep network system, the results of our stages are brought together in a globally optimal manner which partially addresses the missing feedback loop problem of multi-stage systems. The proposed system is original in terms of using two very popular deep networks in a shortest path environment. The result shows that our detection accuracy is 96.25% and localization error is 1.08 mm which is comparable with the state of the art methods. The proposed system is very modular and easily implementable because it uses very well known FRCNN and BCN methods. It is also very fast because both FRCNN and BCN stages are designed to be fast. Polynomial time Dijkstra's shortest path algorithm is also very efficient.

The rest of this paper is organized as follows: Sect. 2 explains Lumbar MRI Data and the proposed method. Section 3 describes the dataset, experiment, and results and Sect. 4 presents our conclusions.

2   Lumbar MRI Data and the Proposed Method

2.1   Lumbar MRI Data

The human spine consists of 33 vertebra connected with IVDs. There are five types of vertebrae: cervical vertebrae, thoracic vertebrae, lumbar vertebrae, sacrum vertebrae, and coccyx vertebrae. Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) are the two most widely used methods to scan the vertebrae. In this study, we focus on IVDs in the lumbar vertebra region on MR images. These IVDs are L1-L2, L2-L3, L3-L4, L4-L5 and L5-S1. Figure 1 shows an example MR image with labeled lumbar IVDs from our dataset.

2.2   First Training Stage: Pre-trained Faster RCNN

The identification and localization of discs on an MR image can be considered as an object detection problem, which is usually harder than classification problems. Classical CNN models are not directly applicable for the object detection tasks. To address this issue, Girshick *et al.* [5] present a model which combines region proposals with CNN's named Region-Based CNN (RCNN). RCNN is good at object detection accuracy but slow at training and testing time. It also needs a large amount of memory. Because of these drawbacks, Girshick *et al.* [4] proposes
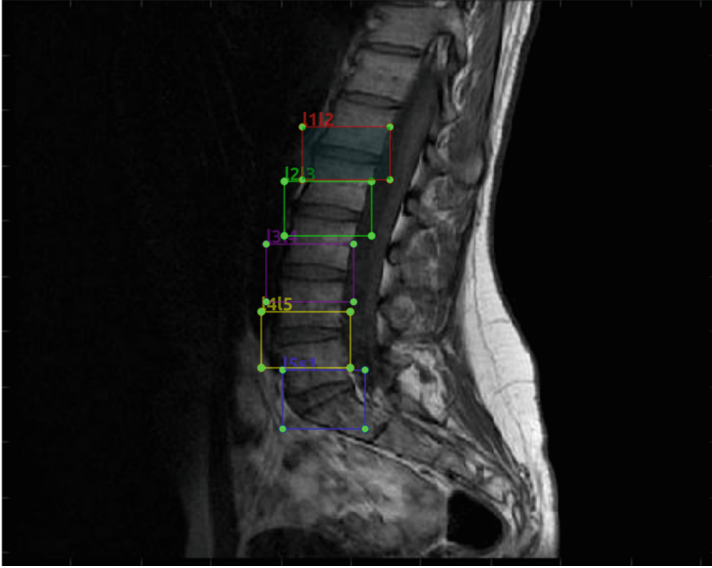
Fig. 1. Example of labeled L1-L2, L2-L3, L3-L4, L4-L5, L5-S1 discs on MR image from our dataset

a new model that uses a region of interest pooling method named Fast RCNN, which is 9 times faster than RCNN and also has better accuracy. Later, Ren *et al.* [12] used a new region proposal method to achieve higher accuracy and lower execution time than Fast RCNN. This method is called Faster RCNN (FRCNN). Due to its high accuracy and near real-time execution, we use FRCNN in our model to produce lumbar disc candidate positions for each IVD in the lumbar region.

In this stage, firstly we prepare our training data. We extract x and y coordinates, width and height information of each lumbar IVDs and label them with their disc names. To determine the beginning and the end of the lumbar spine, we also extract positions of top (T12) and bottom (S1) lumbar vertebra, which makes the system more robust. At the end, we have 7 classes (5 IVDs, T12 and S1 vertebra). We give this training data directly to FRCNN for training. Since our data is scarce and classes are very similar to each other, we do transfer learning and use the pre-trained FRCNN Inception V2 model provided by Tensorflow trained on COCO dataset. After the training, at the testing time, the model produces bounding boxes for every disc with a score. Lets denote the candidate disc positions as $c_j \in R^2$ and every lumbar disc label as $d_i \in \{L1 - L2, L2 - L3, L3 - L4, L4 - L5, L5 - S1, T12, S1\}$. We also define probability $P_E(d_i|c_j)$, which defines the likelihood of having a disc d_i at position c_j. The trained FRCNN model produces candidate discs coordinates, their labels, and scores.

FRCNN methods are very favorable in terms of their localization error, i.e., the position error of the localized discs. FRCNN can also achieve very good detection accuracy rates if there is a good amount of training data. However, the labeled IVD data is very limited for our application. Furthermore, since the appearances of the five discs in the lumbar region are very similar to each other, FRCNN methods produce many false candidates for each disc type with high prediction accuracy scores. As a result, an FRCNN cannot be used on its own for the localization and identification of all the lumbar discs. In order to address this problem, we propose to use more global context information about the candidate disc positions.

2.3 Second Training Stage: Binary Classification Network

In the previous step, the trained FRCNN model learns every IVD as an independent class and it has no idea about sequencing between them. In this stage, we create a CNN model that takes candidate image locations of two neighboring discs, such as L1-L2 and L2-L3. This network makes a binary decision that shows if these two discs are really neighbors. This way, we can lower the number of false positives produced by the FRCNN method. To prepare the training data, image patches that contain two consecutive neighboring discs are cropped (For example, one of the image patches includes L1-L2 and the other includes L2-L3.) We follow the method introduced by Karakoc *et al.* [9] for image patches cropping. Since consecutive discs are locally very similar, four patches extracted from the same center point with different scales to give more information to the model. Every patch is resized to 64×64 and combined into a 128×128 image. We designed a network model that consists of two 2D convolutional layers, 2D max-pooling layer, dropout layer, flatten layer, dense layer, dropout layer, and dense layer. Softmax is used as an activation function, stochastic gradient descent as the optimizer and categorical cross-entropy as a loss function. In the testing phase, the BCN model is given two disc patches, c_i and c_{i-1}, and it produces the probability of having these two patches as neighbors on an MRI image as $P_T(c_i, c_{i-1})$.

Although the decision produced by the BCN is more globally informed than the FRCNN model, it still uses information about only two neighboring discs. In order to get a globally optimal localization and identification results, we use the prediction scores of the FRCNN's and the confidence scores of BCN's in a graph shortest path setting.

2.4 Graphical Model for Optimal Disc Center Localization and Identification

We propose a graphical model that combines the results of these two networks for the final localization and identification of the disc centers. In the first stage, the FRCNN model produces scores and positions for every candidate individual disc. We take a maximum of five predicted candidate positions for every disc. We build a graph whose nodes represent these candidate positions. The edge

that connects two nodes is given a weight produced by our FRCNN and BCN (Fig. 2). Our edges connect two sequential candidate disc positions. To calculate the edge costs between nodes i and j, we use

$$\mathcal{W} = \begin{cases} 1 - P_E\left(d_i|c_j\right) & \text{if } i = 0 \text{ or } i = n \\ 1 - P_E\left(d_i|c_j\right) + 1 - P_T\left(c_i, c_{i-1}\right) & \text{otherwise} \end{cases} \quad (1)$$

where n is the number of discs and vertebra. We try to minimize our cost function to find the most probable disc sequence. The shortest path through this network makes a globally optimal disc localization and identification decision, which can be achieved in polynomial time using Dijkstra's shortest path algorithm. Figure 2 shows a visualization of Dijkstra's algorithm on our system.
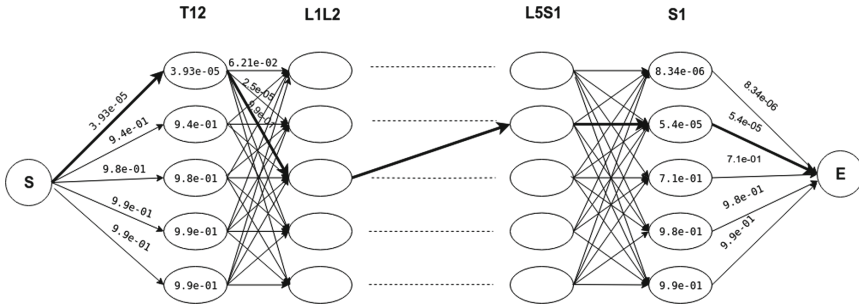

Fig. 2. Visualization of Dijkstra's Algorithm on our system. T12 and S1 nodes represent the first and last lumbar vertebra of interest. We take these vertebra to determine the starting and ending points of the lumbar spine. L1-L2, L2-L3,...L5-S1 nodes represent IVDs. The top five prediction coordinates generated from FRCNN are produced for each disc. Every edge weight is calculated based on BCN and FRCNN scores with Eq. 1.

2.5   Overall System

To sum up the overall system, in the first stage, lumbar discs and their coordinates are given to the pre-trained FRCNN model. While FRCNN model is trained on individual IVDs, in the second stage, BCN is trained to learn the relations between two consecutive discs. Figure 3 shows our visualization of the training procedures.

At the testing stage, given an image, the FRCNN model produces candidate discs. Every combination of two candidate discs are given to BCN and BCN produces a score about the probability of these two discs being neighbors. According to FRCNN and BCN results, all of the path costs starting from the first lumbar disc to the last lumbar disc are calculated. Finally, the most probable sequential disc path that has the minimal cost is found by Dijkstra's algorithm. Figure 4 shows visualization of our testing system.
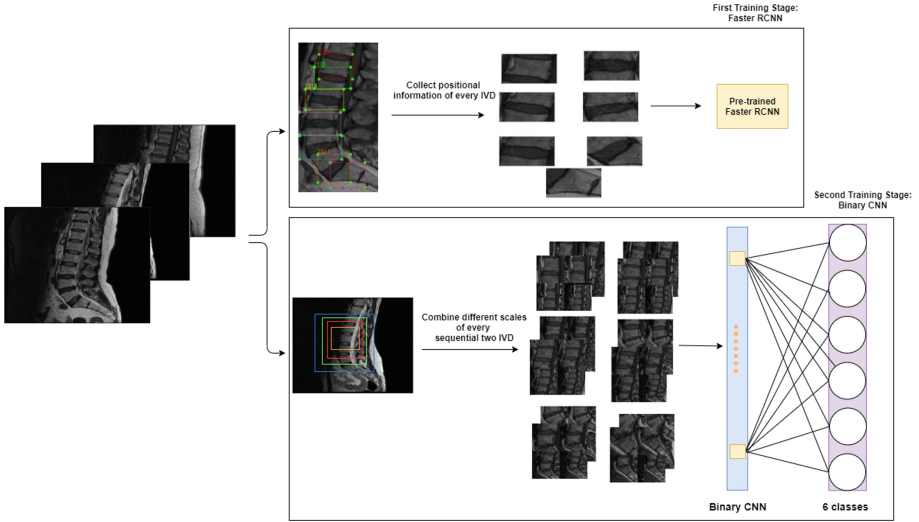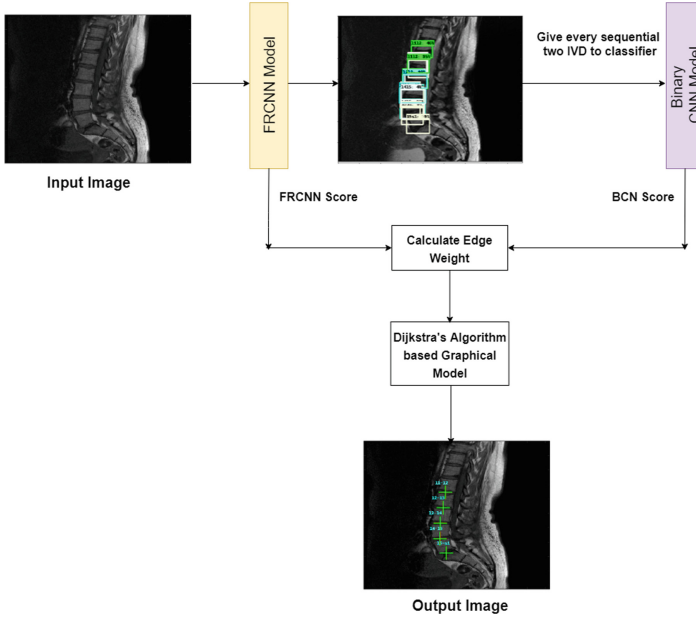
Fig. 3. Visualization of our training system. In the first training stage lumbar discs, first and last lumbar vertebra are trained with FRCNN. In the second training stage, every two consecutive disc is trained with BCN to learn the sequence of discs.

3 Experiments

In first stage of training, $\{x_{min}, x_{max}, y_{min}, y_{max}, \text{width, height, class name}\}$ features extracted for every lumbar disc by a volunteer radiologist. To extract these features, we use a tool named LabelImg[1]. With this tool, the radiologist takes discs in rectangles and labels them. The tool generates xml files for every MR image containing positional information of every disc. We have 80 MR lumbar images in our dataset. Since our data is very small for deep learning, we augment the data by resizing, rotating, scaling, shearing, and translating the images by interval values. For resizing $(450, 600)$, for rotating $(-6, +6)$, for translating $(-0.15, +0.15)$, for scaling $(-0.25, +0.25)$ and for shearing $(-0.1, +0.1)$ limit values are used as intervals. One hundred newly augmented images are generated for every image via randomly selected augmentation parameters between intervals. Since we use 10 fold cross validation as the evaluation method, at the end of the augmentation, we have 7200 training images and 800 test images for each fold. To extract positional information of these augmented pictures, we use a public augmentation library.[2] Then all of these images with their label data are used to train the FRCNN. The initial learning rate of the model is 0.0002 and the activation function is softmax. We train our model with 57 000 epoch which takes about 2 h. At testing time, FRCNN gives predicted bounding boxes and their probabilities on average 0.48 s for a single lumbar MR image.

---

[1] https://github.com/tzutalin/labelImg.
[2] https://github.com/Paperspace/DataAugmentationForObjectDetection.

Fig. 4. Visualization of our testing system. Firstly, test image is given to FRCNN which produces candidate discs. Every combination of two candidate discs are given to BCN to get probability about they come in sequence. These two scores are given to the Eq. 1 and the weight of the edges are calculated. Finally, Dijkstra's Algorithm is run to find the optimal path in polynomial time.

For the BCN model, we use the same training data with FRCNN. Firstly we combine two consecutive discs and crop the image patches from the center of this consecutive disc. Since the appearance of the combination of two consecutive discs is similar, we take four image patches with different scales from the same disc center. These four image patches resized to 64×64 and combined in an image with 128×128 size. The main aim of this process is to obtain more information about two sequential disc centers. We have 7200 training data and 800 test data like the first training stage in a single cross-validation fold. Softmax is used as an activation function, stochastic gradient descent as an optimizer and categorical cross-entropy as a loss function. The learning rate is 0.0001. The result of BCN measures the possibility of two discs coming consecutively. The model is tested with 10 fold cross-validation. The average BCN accuracy is 92%, which is not very good but we do not use BCN results by themselves. They are used in the shortest path setting along with the FRCNN outputs.

Finally, we create a weighted graph with FRCNN and BCN results. The result of the shortest path algorithm produces smallest cost disc path on average 1.1 seconds. The accuracy of 10 fold cross-validation for the overall system is 96.25%. We calculate localization errors both only on true positives and on all dataset separately. Also we find accuracy, localization error and standard deviation of

FRCNN model by itself, which makes a good baseline. To find localization error between detected and ground truth disc centers, Euclidean distance formula is used. Each pixel is 0.625×0.625 mm given by the MRI data. Figure 5 shows the box plot of the localization errors for every lumbar disc.
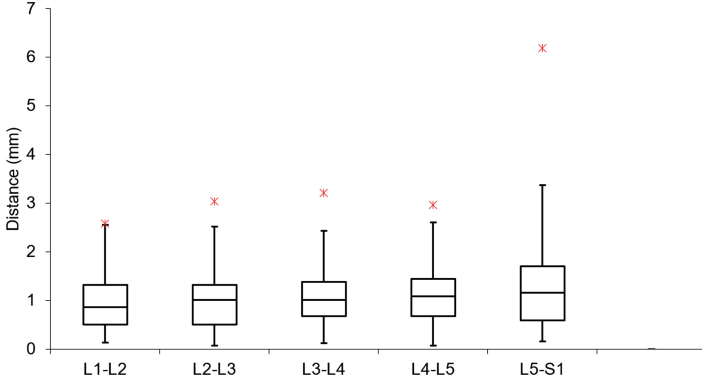
Fig. 5. The box plot shows the mean localization error for every disc. Localization error is calculated with Euclidean distance between the detected bounding box center and the ground truth disc center. In the box plot, the horizontal line in each box shows the median error, the bottom and top of each box shows the 25th and 75th percentile errors, respectively. '*' signs represent the statistical outlier errors.

Table 1 shows accuracy, localization error mean and standard deviation for FRCNN and our system. The results show that our system is quite fast and reliable.

Table 1. Values shows accuracy, localization error mean, standard deviation for FRCNN and our system with and without false positives.

Dataset	Accuracy	Mean error (mm)	Standard deviation
Only FRCNN (True positives)	85%	1.04	0.64
Only FRCNN (True positives + False positives)	85%	2.47	5.60
FRCNN + BCN + Dijkstra's Algorithm (True positives)	96.25%	1.08	0.68
FRCNN + BCN + Dijkstra's Algorithm (True positives + False positives)	96.25%	1.98	4.63

Examples of our results on MR images is shown in Fig. 6. Green plus signs are marked by the expert and red ones are the output of our system.

We use the same dataset with Oktay and Akgul [11]. Their localization error is 3.25 mm for discs and their accuracy is 97.82%. Our localization error is 1.08 mm and our accuracy is 96.25%. The execution time of our system is 1.1 s.
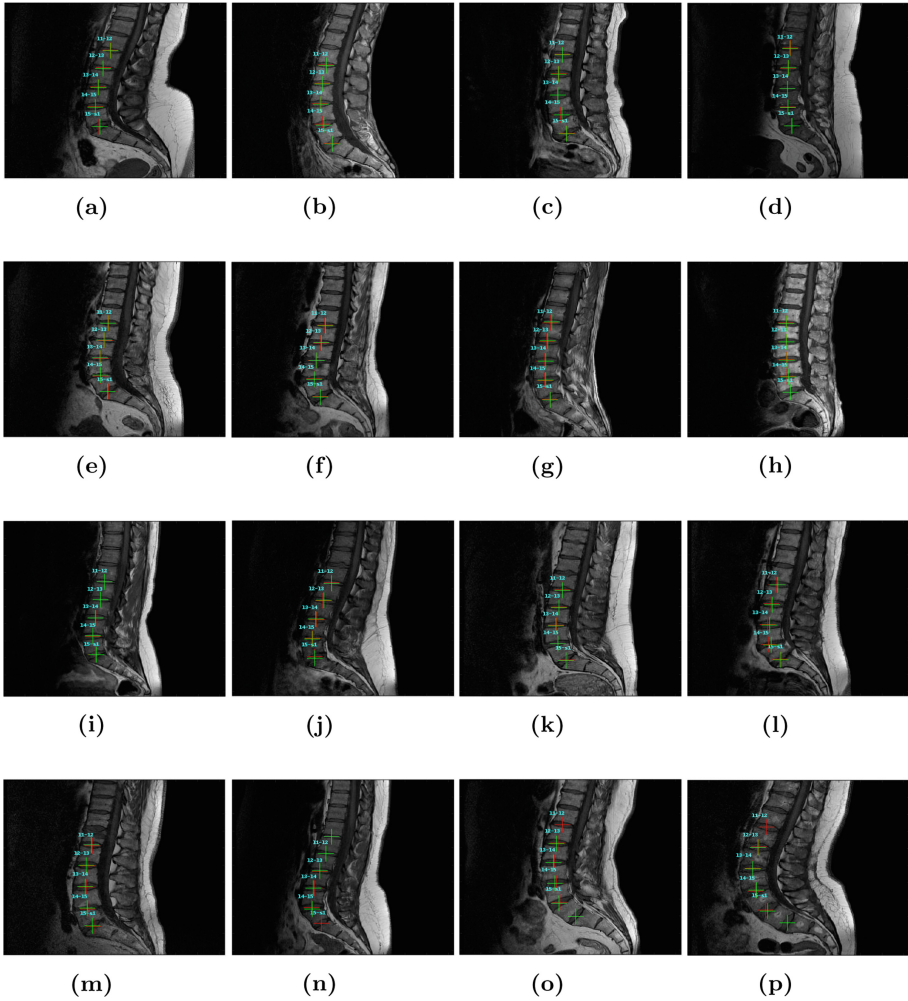
Fig. 6. Results of our system. Green plus signs are marked by the expert and the red one's are the outputs of our system. For the images from (a) to (m), the system can find all discs correctly, on the last 3 images from (n) to (p) are miss-classified images where the system is confused about all discs due to a single shift. (Color figure online)

This shows that our system can identify and localize disc in seconds with comparable performance with the state of the art.

We also compare our results with other studies in this area. Table 2 shows a comparison of our method with other studies.

Although (except Oktay and Akgul [11]) accurate comparison cannot be made because the datasets are different from other studies, we can make inferences that using two networks in a shortest path environment gives high accuracy values because of both using local and global context information. Also, our

Table 2. Comparison with the other methods.

Method	Accuracy	Dice score	Mean error (mm)	Standard deviation	Execution time
Our method	96.25%		1.08	0.68	1.1 s
Jamaludin *et al.* [8]		84	1.1	0.6	
Glocker *et al.* [7]	81%		6.10	5.5	Less than 2 min
Oktay *et al.* [11]	97%		3.1		
Suzani *et al.* [13]	95%		18.2	11.4	Less than 3 s
Chen *et al.* [2]	84%		8.82	13.04	
Lootus *et al.* [10]	84%		3.3	3.2	Less than 1 min
Cai *et al.* [1]	98.1%		3.23	2.09	
Yang *et al.* [15]	90.0%		6.4	5.9	3 s

study shows that using pre-trained FRCNN in spine MR images makes the localization error (mean error) better and can reach state of the art performance as shown by our experiments. The near real-time execution performance feature of FRCNN and Dijkstra's algorithm makes the mean execution time of our system 1.1 s per image, which is quite fast compared to other studies.

4   Conclusions

In this paper, we described our method for the automatic detection and identification of lumbar discs from the MRI data, which is very important for several applications. Although the proposed system is not trained in an end-to-end fashion, our novel employment of Dijkstra's shortest path algorithm makes the final results optimal given the available outputs of FRCNN and BCN components. Our system obtains the candidate lumbar disc positions from an FRCNN module, which are known as fast and accurate object detectors. Many false positives produced by the FRCNN module are eliminated by the shortest path algorithm that uses a second Binary Classification Network to calculate edge weights. The final localization and identification results are comparable with the state of the art methods. The run time of the system is very favorable because the main system components (FRCNN, BCN, shortest path) are known to be very efficient. This system is also easily applicable to different sequential multi-stage deep learning systems. For the future work, we plan to apply this method to 2D and 3D CT and MR images of the whole human vertebrae and discs.

References

1. Cai, Y., Landis, M., Laidley, D.T., Kornecki, A., Lum, A., Li, S.: Multi-modal vertebrae recognition using transformed deep convolution network. Comput. Med. Imag. Graph. **51**, 11–19 (2016)

2. Chen, H., et al.: Automatic localization and identification of vertebrae in spine CT via a joint learning model with deep neural networks. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 515–522. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_63

3. Forsberg, D., Sjöblom, E., Sunshine, J.L.: Detection and labeling of vertebrae in mr images using deep learning with clinical annotations as training data. J. Dig. Imag. **30**(4), 406–412 (2017)

4. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)

5. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)

6. Glasmachers, T.: Limits of end-to-end learning. In: Asian Conference on Machine Learning, pp. 17–32 (2017)

7. Glocker, B., Feulner, J., Criminisi, A., Haynor, D.R., Konukoglu, E.: Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7512, pp. 590–598. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33454-2_73

8. Jamaludin, A., Lootus, M., Kadir, T., Zisserman, A.: Automatic intervertebral discs localization and segmentation: a vertebral approach. In: Vrtovec, T., et al. (eds.) CSI 2015. LNCS, vol. 9402, pp. 97–103. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41827-8_9

9. Karakoç, N.S., Karahan, Ş., Akgül, Y.S.: Deep learning based estimation of the eye pupil center by using image patch classification. In: 2017 25th Signal Processing and Communications Applications Conference (SIU), pp. 1–4. IEEE (2017)

10. Lootus, M., Kadir, T., Zisserman, A.: Vertebrae detection and labelling in lumbar MR images. In: Yao, J., Klinder, T., Li, S. (eds.) Computational Methods and Clinical Applications for Spine Imaging. LNCVB, vol. 17, pp. 219–230. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07269-2_19

11. Oktay, A.B., Akgul, Y.S.: Simultaneous localization of lumbar vertebrae and intervertebral discs with svm-based mrf. IEEE Trans. Biomed. Eng. **60**(9), 2375–2383 (2013)

12. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)

13. Suzani, A., Seitel, A., Liu, Y., Fels, S., Rohling, R.N., Abolmaesumi, P.: Fast automatic vertebrae detection and localization in pathological CT Scans - a deep learning approach. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 678–686. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_81

14. Wang, X., Zhai, S., Niu, Y.: Automatic vertebrae localization and identification by combining deep SSAE contextual features and structured regression forest. J. Dig. Imag. **32**, 1–13 (2019). https://doi.org/10.1007/s10278-018-0140-5

15. Yang, D., et al.: Automatic vertebra labeling in large-scale 3D CT using deep image-to-image network with message passing and sparsity regularization. In: Niethammer, M., et al. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 633–644. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_50

16. Zukić, D., Vlasák, A., Egger, J., Hořínek, D., Nimsky, C., Kolb, A.: Robust detection and segmentation for diagnosis of vertebral diseases using routine MR images. In: Computer Graphics Forum, vol. 33, pp. 190–204. Wiley Online Library (2014)