
# Intervertebral Disc Labeling with Learning Shape Information, a Look once Approach

Reza Azad[1(✉)], Moein Heidari[2], Julien Cohen-Adad[3,4,5], Ehsan Adeli[6], and Dorit Merhof[1,7]

[1] Institute of Imaging and Computer Vision, RWTH Aachen University, Aachen, Germany
{azad,dorit.merhof}@lfb.rwth-aachen.de
[2] School of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran
moein_heidari@elec.iust.ac.ir
[3] Functional Neuroimaging Unit, CRIUGM, University of Montreal, Montreal, Canada
[4] NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, Canada
[5] Mila, Quebec AI Institute, Montrea, Canada
jcohen@polymtl.ca
[6] Stanford University, Stanford, USA
eadeli@stanford.edu
[7] Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany

**Abstract.** Accurate and automatic segmentation of intervertebral discs from medical images is a critical task for the assessment of spine-related diseases such as osteoporosis, vertebral fractures, and intervertebral disc herniation. To date, various approaches have been developed in the literature which routinely rely on detecting the discs as the primary step for detecting abnormality in intervertebral Discs. A disadvantage of many cohort studies is that the localization algorithm also yields to false positive detections. In this study, we aim to alleviate this problem by proposing a novel U-Net-based structure to predict a set of candidates for intervertebral disc locations. In our design, we integrate the image shape information (image gradients) to encourage the model to learn rich and generic geometrical information. This additional signal guides the model to selectively emphasize the contextual representation and to supress the less discriminative features. On the post-processing side, to further decrease the false positive rate, we propose a permutation invariant "look once" model, which accelerates the candidate recovery procedure. In comparison with previous studies, our proposed approach does not need to perform the selection in an iterative fashion. The proposed method was evaluated on the spine generic public multi-center dataset and demonstrated superior performance compared to previous work. The codes is publicly available at github.

**Keywords:** Deep learning · intervertebral disc labeling · look once · shape feature


I. Rekik et al. (Eds.): PRIME 2022, LNCS 13564, pp. 49–59, 2022.
https://doi.org/10.1007/978-3-031-16919-9_5

## 1 Introduction

The human vertebral column consists of 33 individual vertebrae stacked on top of each other and connected through the ligaments and intervertebral discs (IVDs). The vertebral column is divided into cervical, thoracic, lumbar, sacral and caudal vertebrae [3]. Each of these regions performs a vital function in the human body including, absorbing shock, load breathing, protection of the spinal cord, controlling load through the vertebral column, and so on [1]. More precisely, the IVDs act as cushions of fibrocartilage and as principal joints between vertebrae and they absorb the stress and shock the body sustains during motion and allow the spine to be flexible while preventing the vertebrae from grinding against one another. Disruption in any of the vertebral discs through aging, degeneration, or injury will result in an alteration in the corresponding disc's properties along with flaws in mechanical functionalities of adjacent tissues [19]. As a consequence, location and segmentation of intervertebral discs is a crucial task for spine disease diagnosis and provides versatile information in the quality of treatment procedure. To this end, various semi-automated and automated techniques have been proposed in the literature. These methods can be divided into two taxonomies: hand-crafted methods and deep learning-based approaches. As an example for hand-crafted dissertations, Cheng et al. [5] proposed a two-step approach where they first localize the center of each IVD by adapting a data-driven estimation framework [6] and, then, segment IVDs by classifying image pixels around each disc center as either foreground (disc) or background. Glocker et al. [11] utilized a regression forest and a probabilistic graphical model to detect and localize intervertebral discs from CT scan images. A polynomial iterative randomized Hough transform approach to segment the spine and intervertebral discs was proposed in [4]. Irrespective of the good performance of these traditional methods, in some cases they intrinsically render poor performance when compared to deep learning-based methods [2,5]. Recent advances in deep learning have facilitated investigation of robust intervertebral disc labeling [7,8,20]. In [12] the authors proposed to use a standard CNN for IVD segmentation. Dolz et al. [10] proposed an architecture called 'IVD-Net' to leverage information from multiple image modalities for inter-vertebral disc segmentation by adopting a U-Net-like architecture. In a recent article Vania et al. [20] developed a method which builds upon mask-RCNN and formulated a multi-optimization training system at a different stage to increase the computational efficiency. In another approach [21], a cross-modality method for detecting both vertebral and intervertebral discs on volumetric data has been proposed. This approach utilizes a local entropy-based texture model to localize the sacral region. Then, using three-disc entropy models, detected positions are aligned and further refined by taking into account the intensity match between regions and a spinal column template. A transfer learning-based approach is utilized by [14]. In this work, a 2D convolutional structure is exploited to detect the lumbar disc from axial images. Their proposed network uses the strength of the U-Net structure with a VGG backbone to produce a spine segmentation mask. Then, the segmented regions are used to calculate the herniation in lumbar discs. The authors of [17]

combine a fully convolutional network with inception modules to localize and label intervertebral discs. Azad et al. [3] reformulated the semantic vertebral disc labeling using pose estimation and utilized an hourglass neural network to semantically label the intervertebral discs.

The main limitation of the reviewed methods is their dependency on the regular CNN learning strategy (learning texture, shape, colour) which is not optimal for labelling anatomical structures such as intervertebral discs and usually produces both false positive (FP) and false negative (FN) detections [13]. To overcome this issue, we propose to incorporate shape information within the learning process. This additional signal guides the model to selectively emphasize the contextual representation, magnifies the structural regions and supresses the less discriminative features (e.g. color, texture).

Moreover, a principal limitation of many cohort studies is that, as they utilize the local maximum technique to locate the position of the vertebral discs in 2D space on top of the prediction masks, they encounter a substantial false positive rate. Exhaustive search tree [3], template matching [18] and point coordinate condition [17] are among the popular algorithms proposed to eliminate the FP rate. However, these approaches usually lack computational efficiency and render a poor candidate recovery. Therefore, a general method is required to handle this challenge. In this work, we propose to mitigate this limitation by bolstering the post-processing step in the intervertebral disc labeling procedure. The main idea is that, inspired by the idea of YOLO [16], we propose a permutation invariant "look once" model to increase the True Positive (TP) rate while reducing the FN detection. We re-formulate the problem by a modified version of the PointNet model [15] which is invariant to certain geometric transformations (e.g. rotation). To the best of our knowledge, this is the first post-processing algorithm that processes the whole prediction in one step without any iteration ("look once"). Our contributions are as follows:

- Adapting U-Net structure for semantic intervertebral disc labeling;
- Incorporation of shape information to further boost model performance;
- A permutation-invariant post-processing approach to reduce the FP rate;
- Publicly available implementation source code (once accepted);

## 2 Proposed Method

Our proposed method consists of two stages. In the first stage we utilize a U-Net-based structure to detect and predict semantic labeling for each intervertebral disc location. In the second stage, we propose a deep permutation invariant "look once" model to refine the prediction results and eliminate the FP candidates. In the next subsections, we will discuss each phase in more detail.

### 2.1 Semantic Intervertebral Disc Labeling

The concept of the proposed method is depicted in Fig. 1. In our novel design, we incorporate the shape information (gradient of the input image) as an additional

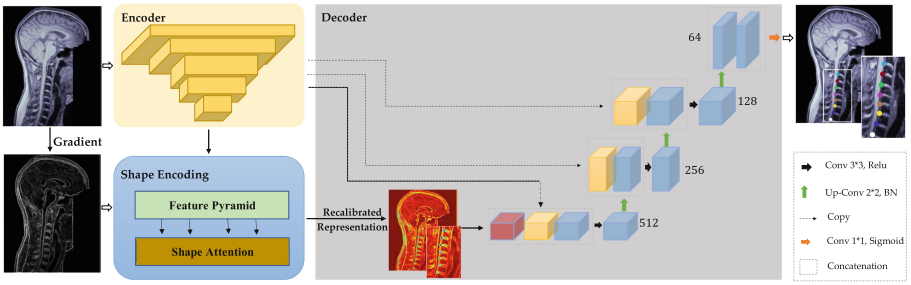


**Fig. 1.** Proposed method for intervertebral disc labeling with incorporating shape information.

signal to encourage the model to learn contextual and geometric information. To this end, we form a pyramid representation using the multi-level description resulting from each block of the encoder (U-Net encoder $E$ parametrized with $\theta$) module: $P = \{f_j = E(x, \theta), j = 0, 1, ...L\}$, where $L$ is the number of pyramid levels. Next, we propose a shape attention module. Our attention module (Fig. 2) uses the global representation of each feature map alongside the shape description to selectively emphasize the contextual representation and supress the less discriminative features. To this end, for each level of the pyramid, we learn the channel-wise recalibration parameters ($w_j^f$) and spatial recalibration parameters ($w_{sp}$) from the shape feature description ($sf$):

$$w_j^f = \sigma\left(\mathbf{W}_2\delta\left(\mathbf{W}_1 GAP_j^f\right)\right), w_{sp} = \sigma\left(\mathbf{W}_4\delta\left(\mathbf{W}_3 GAP(sf)\right)\right) \quad (1)$$

where $W_k, k \in \{1, 2, 3, 4\}$ are the learning parameters that apply to the global representation (GAP) of each pyramid level, and $\delta$ and $\sigma$ stand for the ReLU and Sigmoid activations. We form the re-calibrated description by scaling both channel and spatial dimensions: $\tilde{P}_j^f = w_{sp} \cdot (w_j^f \cdot P_j^f) + sf$. Once the re-calibration performed, we aggregate the multi-level features in a nonlinear fashion (aggregation parameter $w_{prm}$) to produce a shape-attenuating description:

$$f' = \sigma\left(\sum_{j=1}^{L} w_{prm}^j \tilde{P}_j^f\right) \quad (2)$$

Subsequently, the same decoder as in the regular U-Net, but with $V = 11$ output channels (we assume that the input image comprises, at most, 11 intervertebral discs according to [9]), is utilized to estimate the location of each intervertebral disc accordingly. Similarly, our ground truth mask consists of V channels, where in each channel the location of an intervertebral disc is labelled with a Gaussian kernel of radius 10. We employ the mean squared (MSE) loss to train the network.

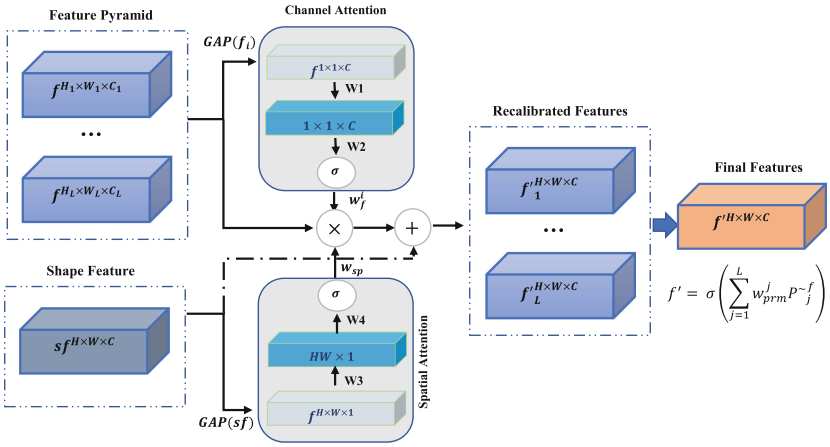


**Fig. 2.** Detailed structure of the proposed shape attention mechanism.

## 2.2 Refinement Network

Detecting intervertebral disc locations often comprises FP and FN predictions. Several post-processing approaches were proposed in the literature to overcome this problem. Rouhier et al. [17], deploys a condition-based strategy to eliminate the FP candidate generated by their countception method. In a recent article, Azad et al. [3] argues that the condition-based strategy usually fails to recover the TP candidates among the detected regions and proposes a tree-based decision space. Their approach suggests creating a search tree, where each path shows one possible combination of ordered intervertebral disc locations. Then, they calculate an error function between the general skeleton and the predicted skeleton. This iterative algorithm performs an exhaustive search and is not efficient when the number of FP is high. Template matching [18] is also another approach that seeks to reduce the FP rate by considering predefined patterns. These methods all have their assumption of particular conditions or predefined patterns in common. In addition, some of these methods perform the selection in an iterative fashion, which may not be feasible when the number of FP is high. To mitigate these issues we propose a method to 'look only once' at the noisy prediction to recover the intervertebral disc locations. To this end, we assume that, for the input image I with N intervertebral disc location, the detection model predicts a set of $M$ intervertebral disc candidates, usually $M >= N$ and $M \in R^2$ (i.e. 2D position). Taking into further consideration in a general form, we assume that the prediction model is not able to provide any semantic labelling. Thus, the objective is to recover $N$ points out $M$ which best matches the ground truth intervertebral disc locations. Since the semantic information is not provided for the predicted points, we consider it as a set of $M$ intervertebral candidates. The set is made up of unstructured data and selecting $N$ intervertebral disc location out of $M$ candidates requires the following processing permutations:

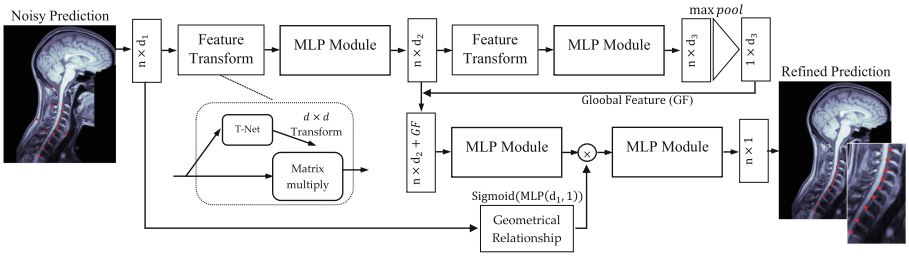


**Fig. 3.** Proposed structure for the post-processing step. The noisy prediction without a semantic label passes through the model to eliminate the FP candidates.

$$\frac{(M)!}{(N)!(M-N)!} \tag{3}$$

permutations. The processing time will dramatically increase if $M >> N$. To overcome this limitation, it is highly desirable that the post-processing algorithm processes the whole prediction at once without any iterations("look once"). Therefore, the deep model needs to be permutation invariant, i.e., any order of points should produce the same result. The proposed structure is depicted in Fig. 3. The proposed method consists of two data streams, where in the first stream (top), a series of feature transformation layers, followed by the multilayer perceptron (MLP), is designed to encode the input coordinate into a high-level representational space. The objective of this representation is to create a discriminative embedding space to characterize each point by a hidden dependency underlying the input data. Intrinsically, the transformation layer in this stream assures the robustness of the representation to the noisy samples and provides a less sensitive transformation to an affine geometrical transformation (e.g. rotation). Inspired by the permutation invariance characteristics, the MLP layer deploys a shared kernel to produce a set of representations independent of their order. Eventually, in addition to the generated feature map, a symmetric function (global pooling) is utilized to capture the shared signature among all points. We concatenate the global information with the local representation of each point to describe each intervertebral disc candidate. Details on the network structure is illustrated in Table 1. This representation more or less contains the general structure of the data, however, it still requires pair-wise relational information. To include such information, we create a geometrical representation. To this end, using the fully connected layers, we learn the embedding parameters to model the long-range geometrical dependency. The main objective of this layer is to capture the geometrical relation between points and feed it to the scaler function. We include the sigmoid function on top of the generated representation to form an attention vector. This attention vector performs the re-calibration process and adaptively scales the generated feature map. The generated final representation is then fed to the single-layer perceptron model to perform the softmax operation and to classify each candidate.

**Table 1.** Details on network architecture for the post-processing stage. We follow [15] for the structure of the Feature Transform module (including T-Net) which simply aligns the input to a feature space using an affine transformation without changing the dimension. We refer the reader to [15] for more general expositions. Note that **n** denotes the number of vertebral discs detected.

| Module | Neurons | Input-size | Output-size |
|---|---|---|---|
| MLP Module(1) | 64 | (nx3) | (nx64) |
| MLP Module(2) | 128 | (nx64) | (nx128) |
| MLP Module(3) | 512 | (nx128) | (nx512) |
| MLP Module(4) | 1024 | (nx512) | (nx1024) |

## 3 Experimental Results

In this section, we first describe the datasets and metrics used throughout our experimental evaluation. Then, we provide a deep insight into the experimental results. Our analysis was based on the publicly available Spine Generic Dataset [9]. The dataset was acquired across 42 centers (with a total of 260 participants) worldwide, accommodating both T1 and T2 MRI contrasts for each subject. Images obtained from diverse institutes, considerably varying in image quality, ages and imaging devices, render a feasibly challenging benchmark for the task of intervertebral disc labelling.

### 3.1 Metrics

To ensure the validity of the comparison of results and to draw conclusions on the applicability of our approach, we consider different comparison metrics. In the first instance, we take into account the L2 norm by calculating the distance of the vector coordinate between each predicted intervertebral disc location and the ground truth while considering the superior-inferior axis to quantify the punctuality of our proposal. In order to gain insights into the versatility of our post-processing approach, the False Positive Rate (FPR) and False Negative Rate (FNR) were selected as the primary inclusion criteria. Similar to [3], the FPR calculates the number of predictions which are at least 5 mm away from the ground truth positions. Likewise, the FNR counts the number of predictions where the ground truth has at least 5mm distance from the predicted intervertebral position.

### 3.2 Comparison of Results

We train all of our models upstream using the Adam solver with the momentum in 100 epochs with the batch size 2. In our experiments, we use an initial learning rate of 0.0001 with the decay by a factor of 0.5 at every 20th epoch, respectively. We use the same setting as explained in [17] to achieve a general

**Table 2.** Intervertebral disc labeling results on the spine generic public dataset. Note that **DTT** indicates Distance to target

| Method | T1 | | | T2 | | |
|---|---|---|---|---|---|---|
| | DTT (mm) | FNR (%) | FPR (%) | DTT (mm) | FNR (%) | FPR (%) |
| Template Matching [18] | 1.97(4.08) | 8.1 | 2.53 | 2.05(3.21) | 11.1 | 2.11 |
| Countception [17] | **1.03(2.81)** | 4.24 | 0.9 | 1.78(2.64) | 3.88 | 1.5 |
| Pose Estimation [3] | 1.32(1.33) | **0.32** | 0.0 | 1.31(2.79) | 1.2 | 0.6 |
| Baseline | 1.45(2.70) | 7.3 | 1.2 | 1.80(2.80) | 5.4 | 1.8 |
| **Proposed** | 1.2(1.90) | 0.7 | 0.0 | **1.28(2.61)** | **0.9** | **0.0** |

consensus in comparing our method with the literature and we report our findings in Table 2. Note: our baseline model uses the same structure as presented but without employing the proposed modules. The results show that our approach achieves a competitive result in T1 and T2 contrasts. Specifically, our proposed method shows superior performance in T2 contrast, where our approach prominently outperforms all other approaches in terms of FNR and distance to the target. Compared to the pose estimation approach [3], our method produces on T1 modality an average lower distance to the intervertebral locations, but there is only a small gap in distance variance. We also observe that, by removing the proposed modules the performance of the model slightly decreases, which highlights the importance of shape information in intervertebral disc labeling. Moreover, unlike the countception and template matching approaches, our method does not require a heavy preprocessing step for spinal cord region detection and outperforms these methods with both quantitative performance and inference time. In contrast to our proposal, the inference time in the two aforementioned approaches grows exponentially when the FP rates increases (see Table 3). In Fig. 4(a) we provide sample results of the proposed model on T2 modalities. It can be observed that the method precisely provides a semantic label for each IVD location without any FP predictions. It should be noted that our method requires less processing time even with large number of FP detection in opposite to the SOTA approaches (illustrated in Fig. 5).

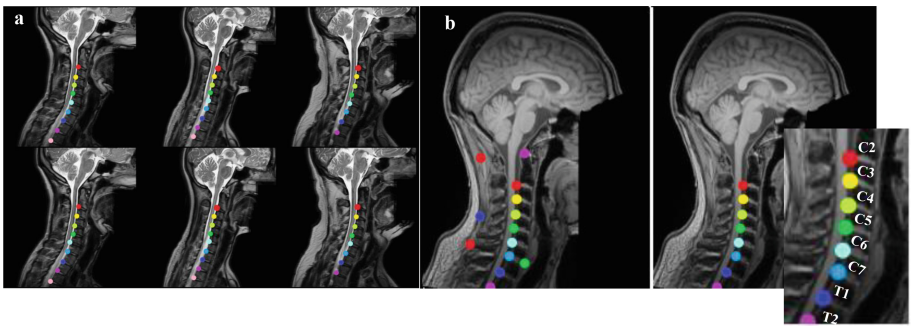


**Fig. 4.** (a): Intervertebral labeling results of three representative T2 images. upper row: ground truth, lower row: predictions. (b): Before (left) and after (right) applying look-once approach on the T1 generated noisy prediction.

**Table 3.** Performance comparison of the proposed post-processing approach vs the SOTA approach for eliminating FP detection. The experiment was done on 100 images, where for each image 20 random FP detection was added.

| Method | F1 | Accuracy | specificity | sensitivity | AUC |
|---|---|---|---|---|---|
| Template Matching [18] | 0.850 | 0.881 | 0.891 | 0.902 | 0.890 |
| Pose Estimation [3] | 0.902 | 0.921 | 0.925 | 0.914 | 0.920 |
| Proposed method (without geometrical relationship module) | 0.914 | 0.932 | 0.941 | 0.917 | 0.929 |
| **Proposed method (Only look once)** | **0.942** | **0.958** | **0.967** | **0.942** | **0.955** |

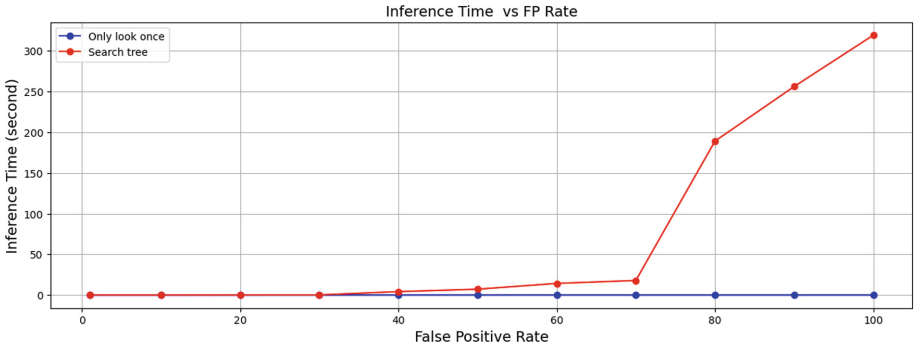


**Fig. 5.** Inference time of the proposed method vs the search-tree based approach [3]. Our method only looks once at the prediction to eliminate the FP samples while the search based approach uses an iterative algorithm.

### 3.3 Evaluation on the Noisy Prediction

To further analyze the robustness of the proposed method in the presence of noisy predictions, we attain an evaluation on the proposed "look once" post-processing method. To this end, we create a 2D Gaussian distribution around each intervertebral disc to generate new points. A sample of generated noisy image along with the model prediction is depicted in Fig. 4(b). As shown, the proposed method works well (including very fast timing) on retrieving IVD locations from the noisy prediction without relying on any predefined assumption. In addition, in our experiment (supplementary file), we observe that for the search-tree-based approach the post-processing time exponentially increased with the increase of FP rate. Similarly, the template matching method failed to recover the TP candidates in most of the cases. Whereas, our method recovered the TP samples with high precision without any iteration. Moreover, to disentangle the contribution of our proposal, we take a closer look at some additional sample detections of our method in Fig. 6 which proves its efficiency in terms of perceptual realism.

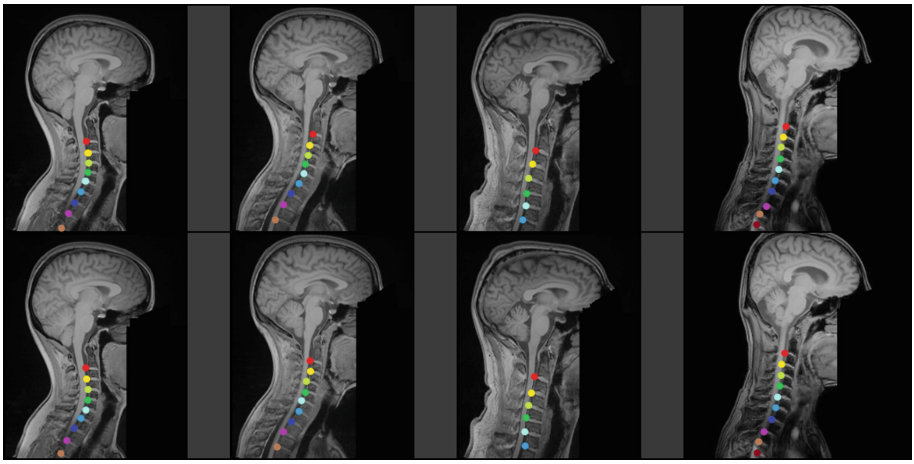

**Fig. 6.** More results of the proposed method for intervertebral disc labeling on T1w images. The first row shows the grand truth while the second row shows the predicted intervertebral disc along with the semantic labeling (color).

## 4 Conclusion

In this paper, we systematically formulate the intervertebral disc labelling problem by designing a novel method to incorporate shape information. The proposed method encourages the model to focus on learning contextual and geometrical features. Additionally, we propose a "look once" post-processing approach. Powered by this, our model alleviates the false-positive detections along with a substantial refinement in model acceleration. The results presented in this paper demonstrate the potential of our methodology across all competing methods.

## References

1. Al-kubaisi, A., Khamiss, N.N.: A transfer learning approach for lumbar spine disc state classification. Electronics **11**(1), 85 (2022)
2. Ben Ayed, I., Punithakumar, K., Garvin, G., Romano, W., Li, S.: Graph cuts with invariant object-interaction priors: application to intervertebral disc segmentation. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 221–232. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22092-0_19
3. Azad, R., Rouhier, L., Cohen-Adad, J.: Stacked hourglass network with a multi-level attention mechanism: where to look for intervertebral disc labeling. In: Lian, C., Cao, X., Rekik, I., Xu, X., Yan, P. (eds.) MLMI 2021. LNCS, vol. 12966, pp. 406–415. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87589-3_42
4. Badarneh, A., Abu-Qasmeih, I., Otoom, M., Alzubaidi, M.A.: Semi-automated spine and intervertebral disk detection and segmentation from whole spine MR images. Inform. Med. Unlocked **27**, 100810 (2021)
5. Chen, C., et al.: Localization and segmentation of 3D intervertebral discs in MR images by data driven estimation. IEEE Trans. Med. Imaging **34**(8), 1719–1729 (2015)

6. Chen, C., Xie, W., Franke, J., Grutzner, P., Nolte, L.P., Zheng, G.: Automatic x-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements. Med. Image Anal. **18**(3), 487–499 (2014)
7. Chen, J.C., Lan, T.P., Lian, Z.Y., Chuang, C.H.: A study of intervertebral disc segmentation based on deep learning. In: 2021 IEEE 4th International Conference on Knowledge Innovation and Invention (ICKII), pp. 85–87. IEEE (2021)
8. Cheng, Y.K., et al.: Automatic segmentation of specific intervertebral discs through a two-stage multiresunet model. J. Clin. Med. **10**(20), 4760 (2021)
9. Cohen-Adad, J., et al.: Open-access quantitative MRI data of the spinal cord and reproducibility across participants, sites and manufacturers. sci. data. https://doi.org/10.1038/s41596-021-00588-0
10. Dolz, J., Desrosiers, C., Ben Ayed, I.: IVD-net: intervertebral disc localization and segmentation in MRI with a multi-modal UNet. In: Zheng, G., Belavy, D., Cai, Y., Li, S. (eds.) CSI 2018. LNCS, vol. 11397, pp. 130–143. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-13736-6_11
11. 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
12. Ji, X., Zheng, G., Belavy, D., Ni, D.: Automated intervertebral disc segmentation using deep convolutional neural networks. In: Yao, J., Vrtovec, T., Zheng, G., Frangi, A., Glocker, B., Li, S. (eds.) CSI 2016. LNCS, vol. 10182, pp. 38–48. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-55050-3_4
13. Liu, L., Wolterink, J.M., Brune, C., Veldhuis, R.N.: Anatomy-aided deep learning for medical image segmentation: a review. Phy. Med. Biol. **66**, 11TR01 (2021)
14. Mbarki, W., Bouchouicha, M., Frizzi, S., Tshibasu, F., Farhat, L.B., Sayadi, M.: Lumbar spine discs classification based on deep convolutional neural networks using axial view MRI. Interdisc. Neurosurg. **22**, 100837 (2020)
15. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 652–660 (2017)
16. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
17. Rouhier, L., Romero, F.P., Cohen, J.P., Cohen-Adad, J.: Spine intervertebral disc labeling using a fully convolutional redundant counting model. arXiv preprint arXiv:2003.04387 (2020)
18. Ullmann, E., Pelletier Paquette, J.F., Thong, W.E., Cohen-Adad, J.: Automatic labeling of vertebral levels using a robust template-based approach. Int. J. Biomed. Imaging 2014 (2014)
19. Urban, J.P., Roberts, S.: Degeneration of the intervertebral disc. Arthritis Res Ther **5**(3), 1–11 (2003)
20. Vania, M., Lee, D.: Intervertebral disc instance segmentation using a multistage optimization mask-RCNN (mom-RCNN). J. Comput. Des. Eng. **8**(4), 1023–1036 (2021)
21. Wimmer, M., Major, D., Novikov, A.A., Bühler, K.: Fully automatic cross-modality localization and labeling of vertebral bodies and intervertebral discs in 3D spinal images. Int. J. Comput. Assist. Radiol. Surg. **13**(10), 1591–1603 (2018)