



Automated Pathogenesis-Based Diagnosis of Lumbar Neural Foraminal Stenosis via Deep Multiscale Multitask Learning

Zhongyi Han^{1,2,3,4} · Benzheng Wei^{1,2} · Stephanie Leung⁴ · Ilanit Ben Nachum⁴ · David Laidley⁴ · Shuo Li^{3,4} 

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Pathogenesis-based diagnosis is a key step to prevent and control lumbar neural foraminal stenosis (LNFS). It conducts both early diagnosis and comprehensive assessment by drawing crucial pathological links between pathogenic factors and LNFS. Automated pathogenesis-based diagnosis would simultaneously localize and grade multiple spinal organs (neural foramina, vertebrae, intervertebral discs) to diagnose LNFS and discover pathogenic factors. The automated way facilitates planning optimal therapeutic schedules and relieving clinicians from laborious workloads. However, no successful work has been achieved yet due to its extreme challenges since 1) multiple targets: each lumbar spine has at least 17 target organs, 2) multiple scales: each type of target organ has structural complexity and various scales across subjects, and 3) multiple tasks, i.e., simultaneous localization and diagnosis of all lumbar organs, are extremely difficult than individual tasks. To address these huge challenges, we propose a deep multiscale multitask learning network (DMML-Net) integrating a multiscale multi-output learning and a multitask regression learning into a fully convolutional network. 1) DMML-Net merges semantic representations to reinforce the salience of numerous target organs. 2) DMML-Net extends multiscale convolutional layers as multiple output layers to boost the scale-invariance for various organs. 3) DMML-Net joins a multitask regression module and a multitask loss module to prompt the mutual benefit between tasks. Extensive experimental results demonstrate that DMML-Net achieves high performance (0.845 mean average precision) on T1/T2-weighted MRI scans from 200 subjects. This endows our method an efficient tool for clinical LNFS diagnosis.

Keywords Neural foraminal stenosis · Multiscale learning · Multitask learning · Deep learning

Introduction

Lumbar neural foraminal stenosis (LNFS) has attacked about 80% of the elderly population worldwide (Rajae et al. 2012). LNFS refers to a spinal nerve root of the

peripheral nervous system¹ compressed by a narrowing neural foramen. Since spinal nerve roots are responsible for transmitting sensory inputs, LNFS would cause lots of pain, muscle weakness, and even body disability (Kaneko et al. 2012). LNFS often develops with the natural aging process of people who gradually suffer various spine diseases, such as scoliosis and spondylolisthesis. However, the clinical manual diagnosis has inevitable subjectivity because experienced radiologists have different grading results even according to the same grading criterion (Lee et al. 2010). Manual diagnosis is also tedious and time-consuming (He et al. 2017b). Thus automated diagnosis system is extremely useful in providing an objective diagnosis for LNFS and relieving clinicians from laborious workloads.

✉ Benzheng Wei
wbz99@sina.com

✉ Shuo Li
slishuo@gmail.com

¹ College of Science and Technology, Shandong University of Traditional Chinese Medicine, Jinan, 250355, China

² Computational Medicine Lab, Shandong University of Traditional Chinese Medicine, Jinan, 250355, China

³ Department of Medical Imaging, Western University, London, N6A 4V2, Canada

⁴ Digital Image Group (DIG), London, ON, N6A 4V2, Canada

¹Peripheral nervous system consists of the nerves and ganglia outside the brain and spinal cord, serving as a relay between the brain and spinal cord and the rest of the body.

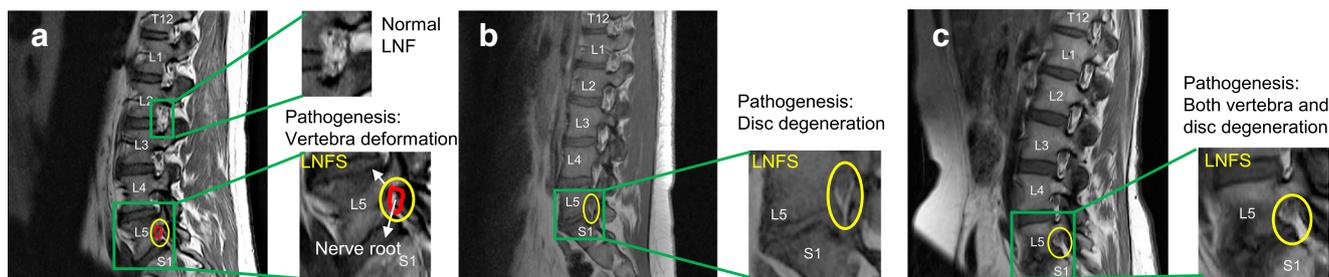


Fig. 1 Three pathological links between LNFS and its pathogenic factors. **a** An LNFS is caused by the L5 vertebral deformation, which also indicates nerve root would be compressed when LNFS does not occur (the 2nd row in Table 1). **b** An LNFS is caused by the disc degeneration between L5 and S1, which also indicates nerve root would be

compressed in a transverse direction (the 3rd row in Table 1). **c** An LNFS is caused by abnormal vertebral and disc with an osteoarthritic change in the facet joint, which also indicates nerve root would be compressed in four directions (the 4th row in Table 1)

Automated pathogenesis-based² diagnosis of LNFS on MRI efficiently assists clinicians to build optimal therapeutic schedules to prevent and control LNFS. Pathogenesis-based diagnosis of LNFS clinically does draw crucial pathological links between LNFS and its pathogenic factors. Automated pathogenesis-based diagnosis aims at drawing early and precise links for early and comprehensive diagnosis of LNFS. It would not only localize and grade neural foramina but also evaluate surrounding vertebrae and intervertebral discs, because both abnormal vertebrae and discs are important pathogenic factors as well as vital indicators of LNFS: 1) an abnormal vertebra (see Fig. 1a) has led a nerve root to be compressed by a neural foramen (Lee et al. 2010). Abnormal vertebrae are also indicating that nerve roots will be compressed in a vertical direction when LNFS does not occur (see the 2nd row in Table 1). 2) An Abnormal disc (see Fig. 1b) has led a nerve root to be compressed (Hasegawa et al. 1995). Abnormal discs are also indicating that nerve roots will be compressed in a transverse direction when LNFS does not occur (see the 3rd row in Table 1). 3) Both abnormal vertebra and disc (see Fig. 1c) not only have caused LNFS but also indicate that nerve roots will be compressed in four directions when LNFS does not occur (see the 4th row in Table 1). Therefore, automated pathogenesis-based diagnosis contributes significantly to promoting precise prediction to prevent LNFS at the earliest opportunity. It also contributes to building comprehensive pathological analysis and treatment plan to control LNFS (see the 5-8th rows in Table 1). In addition, MRI is the most appropriate test for imaging LNFS (Kim et al. 2015). Accordingly, the automated pathogenesis-based diagnosis of LNFS on MRI scans is devoted in this paper.

However, no successful work has achieved automated pathogenesis-based diagnosis of LNFS due to its extreme

three-fold challenges. 1) Multiple targets. Each lumbar spine has 17 target organs composed of six neural foramina, six intervertebral discs, and five lumbar vertebrae (see Fig. 1). 2) Multiple scales. Each type of target organ has structural complexity as well as various scales across normal and abnormal subjects. Various scales lead to the difficulty of simultaneous localization and grading. For example, the area of neural foramina accounts for 5% - 9% of lumbar MRI scans. Too small neural foramina lead to the absence of feature representation. Moreover, the shapes of organs differ from the race, sex, and age. For instance, the heights of neural foramina range from 0.61 cm to 2.27 cm while the widths range from 0.41 cm to 1.90 cm. On the contrary, normal and abnormal organs share a high similarity. 3) Multiple tasks, i.e., simultaneous localization and diagnosis of all lumbar organs, are extremely challenging than individual tasks.

To provide an efficient resolution, we propose a newly-designed deep multiscale multitask learning network (DMML-Net) integrating multiscale multi-output learning

Table 1 The automated pathogenesis-based diagnosis conducts both efficient prediction and precise diagnosis for further prevention and treatment of LNFS

Subject	LNFS	Disc	Vertebra	Prediction	P-diagnosis
#1	✓	✓	✓	✓	✓
#2	✓	✓	×	×	△
#3	✓	×	✓	×	△
#4	✓	×	×	xx	△
#5	×	✓	✓	△	×
#6	×	×	✓	△	xx
#7	×	✓	×	△	xx
#8	×	×	×	△	xxx

✓Normal; ×Abnormal; △Warning

²Pathogenesis of a disease is the biological mechanism (or mechanisms) that leads to the diseased state.

and multitask regression learning into a fully convolutional network (FCN). 1) DMML-Net is capable of robustly representing spinal structures and reinforcing the salience of target organs due to a feature boosting module merging semantic multi-level features. 2) DMML-Net is capable of the scale-invariance for various organs due to a k-means clustering method finding the rule of organs. 3) DMML-Net is capable of promoting the mutual benefit between inter-task and intra-task due to multitask regression and multitask loss modules.

Related Work

To the best of our knowledge, no one has achieved the pathogenesis-based diagnosis of LNFS. Existing works are limited in manual assessments, automated segmentation or simple grading of LNFS.

Manual assessment works demonstrate that 1) manual grading of LNFS has inevitable subjectivity and 2) pathogenic factors of LNFS do exist in vertebrae and discs. Lee et al. (2010), Park et al. (2012), and Kim et al. (2015) aimed at evaluating the reproducibility of differently manual grading criteria. Results of these studies show that experienced radiologists have different assessment results even according to the same grading criterion. Cinotti et al. (2002), Panjabi et al. (2006), and Kaneko et al. (2012) studied on the biomechanics of LNFS about how it is affected by segmental deformities and anatomical variations. Their results prove that the causes of LNFS are from multiple pathogenic factors existing in vertebrae and discs.

A few works have demonstrated the superiority of the automated way even though they are limited in the segmentation or simple grading of LNFS. He et al. (2017c) studied on unsupervised segmentation of LNFS using a multi-feature and adaptive spectral clustering method, which achieved high performance but did not realize LNFS diagnosis. He et al. (2017b) also studied on automated segmentation leveraging the combination of a support vector regression and a multiple kernel learning, which achieved accurate area estimation on multimodality images but still did not realize LNFS diagnosis. He et al. (2016) aimed at relieving the difficulty of simple grading of LNFS on MRI scans using two steps (i.e., one step for localization and another step for grading), which needed complex procedures and did not realize direct localization and grading as well as the pathogenesis-based diagnosis of LNFS.

In regard to other spinal structures' analysis, existing related works include automated detection and automated grading. Firstly, existing automated detection methods can be attributed to three types. The first type is the automated localization of one or two types of spinal structure (Alomari

et al. 2011; Corso et al. 2008; Štern et al. 2009; Zhan et al. 2012; Cai et al. 2015), which are only capable of presenting specific anatomic structure. The second type is automatic segmentation of one or two types of spinal structure (Wang et al. 2015b, b; Yao et al. 2016; Law et al. 2013a, b). The third type is simultaneous localization and segmentation (Ghosha et al. 2011; Huang et al. 2009; Klinder et al. 2008; Peng et al. 2006; Shi et al. 2007; Kelm et al. 2013). Existing detection methods have achieved accurate recognition of one or two types of spinal structure, whereas they cannot do the radiological grading of spinal structures. Secondly, a few radiological grading works include but are limited to one type of spinal structure, such as lumbar disc generation grading (He et al. 2017a; Raja'S et al. 2011), spondylolisthesis grading (Cai et al. 2017). Besides, SpineNet proposed by Jamaludin et al. (2017) has achieved multiple disc diseases grading, however it only takes the preprocessed disc's volumes as inputs so that only analyzes one type of spinal structure. In this study, not only do we analyze three types of spinal diseases at three types of spinal structures but also devote to the pathogenesis-based diagnosis of LNFS.

Methods

The newly-proposed DMML-Net (see Fig. 2) directly localizes all lumbar organs with bounding boxes, and grades all lumbar organs with crucial differential diagnoses (normal and abnormal). DMML-Net integrates a newly-designed multiscale multi-output learning and a multitask regression learning with an end-to-end framework. Multiscale multi-output learning is to predict target organs at multi-level convolutional layers of an FCN to reinforce the salience and scale-invariance. Multitask regression learning is to learn multiple complex tasks at the same time, which are expected to benefit each other to improve generalization performance. The seamlessly joint learning of multiscale and multitask ensures DMML-Net handles the structural complexity of spines, scale variability of organs, and weak relatedness of tasks. The comprehensive end-to-end workflow of DMML-Net is shown in Fig. 4. Constructed procedures of the multiscale multi-output learning architecture and formulated details of the multitask regression learning are given below.

Multiscale Multi-Output Learning Architecture Construction

Architecture construction of multiscale multi-output learning is to construct a deep FCN and select its multiscale convolutional layers extended to output layers (see Fig. 3). *First*, an FCN is constructed as the basic architecture of multiscale multi-output learning. *Second*, a k-means clustering

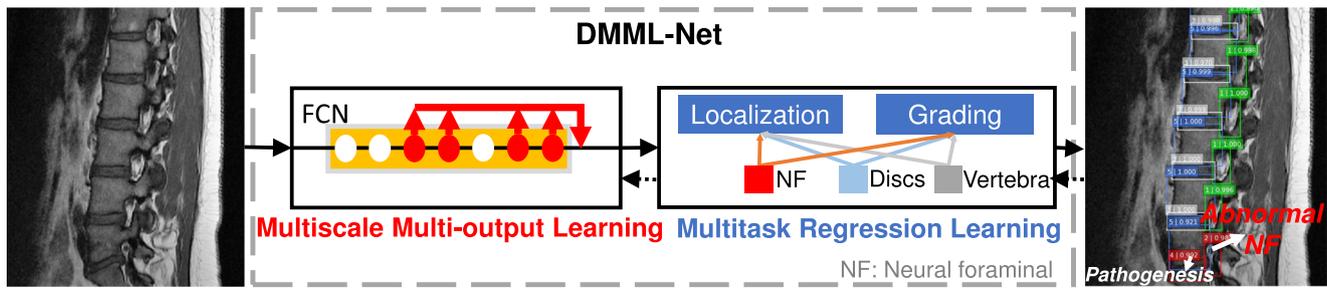


Fig. 2 DMML-Net directly localizes and grades all lumbar organs after importing clinical lumbar MRI scans. Multiscale multi-output learning predicts target organs at many convolutional layers of an FCN. Multitask regression learning links multiple tasks with parameter sharing for mutual benefit

approach is adopted to select appropriate convolutional layers as candidates of output layers. *Third*, an auxiliary feature boosting module is proposed to generate multiscale output layers by merging selected convolutional layers. These ensure multiscale multi-output learning architecture capable of semantic representations and scale-invariance for various organs.

Fully Convolutional Network

Our FCN has been constructed (see the shadow region in Fig. 4) with a deep architecture capable of extracting multi-level semantic representations of target organs. The FCN has 16 convolutional layers and five pooling layers, which are reformed from the VGG-16 network which has five convolutional blocks with five max-pooling layers and three fully connection layers (FC6, FC7, FC8) (Simonyan and Zisserman 2014). The first reformation is that the FC6, FC7, and FC8 are converted to convolutional layers (conv6, conv7, conv8). The second is that the filters of pool5 are changed from $2 \times 2 @ 2$ to $3 \times 3 @ 1$. Each of the first two convolutional blocks (conv1_x, conv2_x) has two convolutional layers, while each of the last three convolution blocks (conv3_x, conv4_x, conv5_x) has three convolutional layers. We reform the VGG-16 to the FCN rather than building a new one because the VGG-16 model trained from natural image datasets can be used for transfer learning.

Multiscale Convolutional Layers Selection by K-Means Clustering

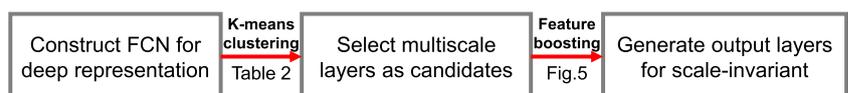
A k-means clustering is adopted to analyze the rule of target organs to select suitable multiscale convolutional layers as the candidates of output layers. Since neural foramina, discs, and vertebrae have various and small scales, only

using the last convolutional layer of the FCN has difficulty in predicting the correct locations and grades, while front layers contain more details are capable of this. The specific reason is twofold. 1) With existing convolutional neural networks going deeper; however, the feature information of their last layers are too coarse to allow correct predictions. That is because feature maps of their last layers have wider receptive fields, which results in a grosser granularity (Sun et al. 2017; Yan et al. 2014a, b, 2017a, b). Receptive field refers to the spatial extent of input volume connected to each neuron of a layer (Luo et al. 2016). An object in the feature maps of the last layer is less than one cell if the object size is less than the receptive fields of the last layer. Based on the information of less than one cell, the prediction of this object is hard to correct (Hoang Ngan Le et al. 2016). 2) On the contrary, front convolutional layers allow precise predictions because they have fine-grained information and narrower receptive fields. The practical effectiveness of the front layers has been proved by several works in computer vision domain. For instance, Hariharan et al. (2015), Liu et al. (2016), and Bell et al. (2016) have demonstrated that using multiple front layers can improve the segmentation quality and detection accuracy. However, how to choose suitable front layers based on a specific dataset is still an ongoing question. Based on this consensus, a novel strategy for the front layers selection is put forward in this study.

First, the heights, widths of target organs from the ground truth are imported as two sets of observations (X_1, X_2, \dots, X_n). Second, k-means clustering is run to split each set of observations into k clusters $S = \{S_1, S_2, \dots, S_k\}$, so as to minimize the within-cluster sum of squares:

$$\operatorname{argmin}_S \sum_{i=1}^k \sum_{X \in S_i} \|X - \mu_i\|^2, \tag{1}$$

Fig. 3 The procedures of multiscale multi-output learning architecture construction



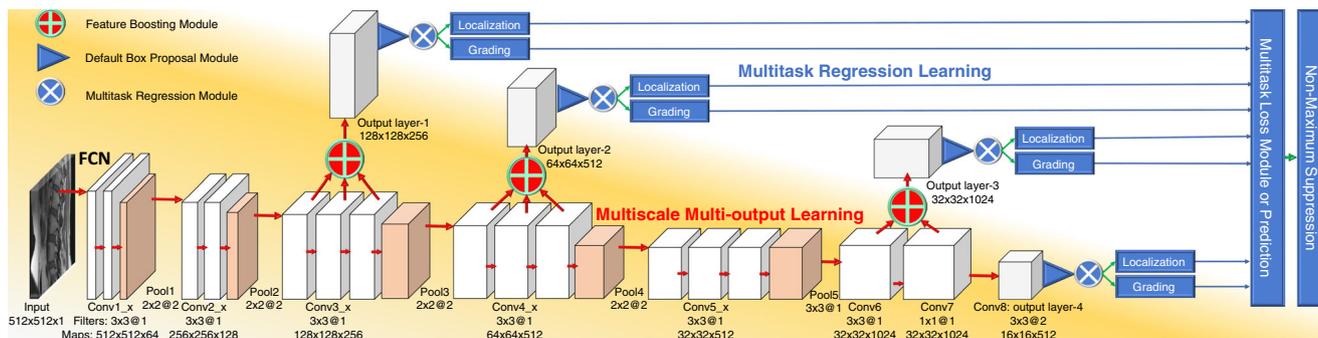


Fig. 4 The main workflow of DMML-Net. The shadow region covers the multiscale multi-output learning architecture, while the unshaded region covers the multitask regression learning

where μ_i is the centroid of a cluster S_i . k is set to 6 because each type of neural foramen, vertebra, and disc has two grades. After 1,000 iterations, optimal centroids are computed as shown in Table 2. In practice, centroids μ_1 and μ_2 are almost the size of neural foramina, centroids μ_3 and μ_4 are almost the size of intervertebral discs, μ_5 , and μ_6 are almost the size of vertebrae. Although each centroid μ_i is only corresponding to a specific scale; however, these are sufficient to choose appropriate convolutional layers. In addition, the minimum width or height of ground truth is 17 pixels while the maximum is 151 pixels, which is also a reference of choosing the first and last convolutional layers.

Based on k-means clustering results, the conv3_x, conv4_x, conv6, conv7, conv8 of the FCN have been chosen as candidates of output layers. Since receptive fields of the conv3_x block are 28 pixels (of conv3_1), 30 pixels (of conv3_2), and 32 pixels (of conv3_3), the conv3_x block has efficient receptive fields corresponding to the μ_1 and μ_2 scales of target organs. The con3_x block is chosen corresponding to μ_2 scale because the receptive field is not necessary to be equal to the scale of objects strictly, but is needed to be smaller than the objects. Similarly, the conv4_x block has been chosen to predict the μ_3 and μ_4 scales because it has effective receptive fields corresponding to μ_3 and μ_4 . Furthermore, conv6 and conv7 have been chosen to corresponding to μ_5 scales, while conv8 corresponds to μ_6 scales. The conv1_x and conv2_x blocks are not selected because they have little semantic representations although they have narrower receptive fields. The conv5_x block

is not selected because it has the similar receptive field with the conv6 layer but has less semantic representations than the conv6. The size of all neural foramina, discs, and vertebrae have the same rule on real-world lumbar MRI scans, i.e., all MRI scans can be resized to be same as our dataset, thus our method is capable of generalization in clinical practice.

Multiscale Output Layers Generation by Feature Boosting Module

The newly-proposed feature boosting module (see Fig. 5) element-wisely merges same size feature maps of selected convolutional layers in each block to generate output layers. It extracts optimal semantic representation as well as exchanges the context information between convolutional layers. The proposed feature boosting module effectively reinforces the robustness of DMML-Net at a small dataset, which is a key of successful localization and grading.

Let \mathcal{F}_j^i be feature maps of the j -th convolutional layer at the i -th block; \mathcal{B}^i be the output of feature boosting module at the i -th block. Mathematically,

$$\mathcal{B}^i = \sqrt{|\lambda_1 \mathcal{F}_1^i \star \lambda_2 \mathcal{F}_2^i \cdots \star \lambda_n \mathcal{F}_n^i| + \frac{1}{n} \sum_{\omega} \|\omega\|^2}, \quad (2)$$

where \star is an operator that depends on specific tasks. $\|\omega\|^2$ is an l_2 normalization layer to avoid the large magnitude of gradients and normalize the feature values at each point of \mathcal{F}_j^i . In this study, plus operator is used as \star with $n = 1$ because the element-wise fusion between feature maps provides more fine-grained information.

As shown in Fig. 4, conv3_x block (conv3_1, conv3_2, conv3_3), conv4_x block (conv4_1, conv4_2, conv4_3), conv6 + conv7 are applied for the feature boosting module to generate output layers. Conv8 is sole as an output layer because no layer has the same size as conv8. The constructed multiscale multi-output learning architecture of DMML-Net has the full range of receptive fields and

Table 2 Representative scales of target organs from k-means clustering analysis used to select front convolutional layers

Organ	NF		Disc		Vertebra	
	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
Width	29.55	41.72	72.41	84.82	96.68	114.57
Height	32.47	49.55	62.53	76.10	88.38	103.33

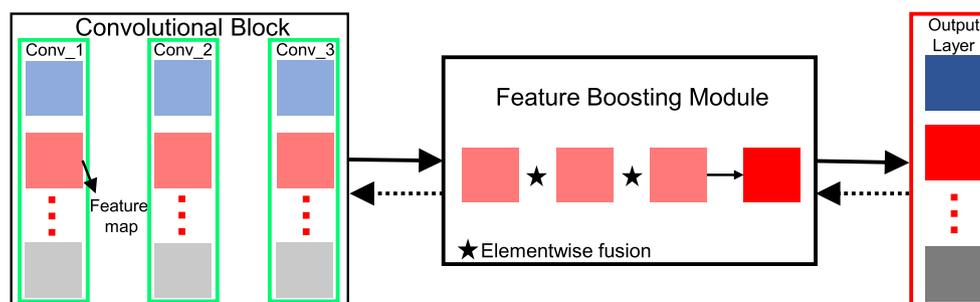


Fig. 5 Feature boosting module merges selected convolutional layers to generate output layers. It enhances the global context information and exchange information between scales, thus it is capable of the scale-invariance for various target organs

context information which are capable of covering all scales of target organs. Thus DMML-Net has scale-invariance of various organs. At the end, four output layers: conv3, conv4, conv6+7, and conv8 are in preparation for the next stage of multitask regression learning formulation.

Multitask Regression Learning Formulation

Multitask regression learning formulation of DMML-Net is to localize and grade neural foramina, discs, and vertebrae at the same time (see Fig. 6). *First*, a default box proposal module is formulated to generate default bounding boxes of target organs on multiscale output layers for further regression. *Second*, a multitask regression module is formulated to predict the refined locations and grades of default boxes. *Third*, a multitask loss module is formulated to compute both localization and grading losses. At the end, a hard negative mining approach is adopted to improve convergence of DMML-Net, while a non-maximum suppression approach is adopted to select the best bounding box for each target organ.

Organ Locations Proposal by Default Box Proposal Module

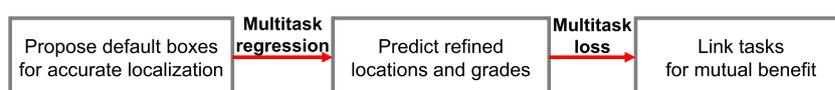
Default box proposal module is responsible for telling the DMML-Net where to look, i.e., hypothesizing locations of organs in order to obtain accurate prediction. It generates default bounding boxes of target organs on output layers before using multitask regression module to predict more refined locations and grades (see Fig. 7). So it is the foundation of correct prediction. Many works studied on the default box proposal, but the more cost-free and efficient method is Single Shot Multibox Detector (SSD, (Liu et al. 2016)). We embed k-means clustering results of target organs into the size setting of the default boxes proposal of

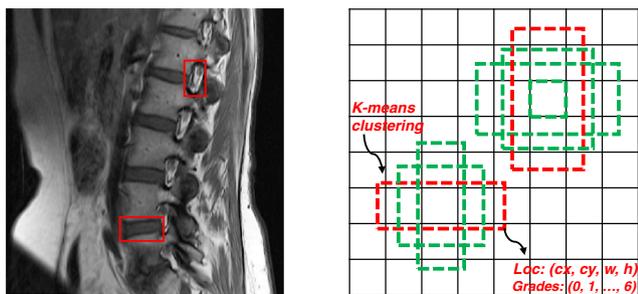
SSD. In detail, we let representative aspect ratios (Table 3) from k-means clustering approach be the aspect ratios of the default boxes, and let representative widths (Table 2) be the widths of the default boxes of the output layers. The default boxes of different output layers have the same aspect ratios but different widths. The widths of default boxes are set from large to small according to efficient receptive fields of output layers. Our default box proposal module empirically set the default box proposals according to the k-means clustering results rather than setting randomly. Therefore, our method generates the optimal default boxes for further refined localization and grading.

Simultaneous Localization and Grading by Multitask Regression Module

To achieve network parameters sharing, each multitask regression module has two convolutional filters to predict two sets of fixed vectors of locations and grades on default boxes respectively (see Figs. 4 and 7b). First, assume the feature map size of an output layer is $m \times n$ with p channels, and each feature map has k default boxes generated by the default box proposal module. Second, assume the bounding boxes of each organ are presented by 4-D vectors $\{c_x, c_y, h, w\}$ and one of the $C+1$ grades (include background), where c_x, c_y, h, w denote the center coordinate, height, width of bounding boxes respectively. Third, for each multitask regression module on one output layer, the first convolutional filter predicts the locations of a target organ by yielding $4mnkp$ outputs, while the second convolutional filter predicts the grades by yielding $(C+1)mnkp$ outputs. After running on four output layers, the multitask regression module concatenates and reshapes these outputs to be the same size as the ground truth, so as to compute the multitask loss.

Fig. 6 The main procedures of the multitask regression learning formulation





a A 512 x 512 MRI scan with examples of ground truth boxes **b** A 64 x 64 feature map with proposed default boxes

Fig. 7 Default box proposal module is responsible for hypothesizing locations of organs. **a** An input MRI scan with part ground truth boxes which are only needed in training phase. **b** Default box proposal module embeds k-means clustering results into default boxes proposal at each location of the 64 × 64 feature map. More refined locations (loc) of default boxes will be predicted by multitask regression module

Multitask Loss Module

Our multitask loss module is derived from Fast R-CNN (Girshick 2015) to compute the overall losses by summing the grading loss ($\mathcal{L}_{grading}$) and localization loss (\mathcal{L}_{loc}) via a balancing parameter λ :

$$\mathcal{L}(x_i, p^c, t_i, t_i^*) = \frac{1}{N}(\mathcal{L}_{grading}(x_i, p^c) + \lambda \mathcal{L}_{loc}(x_i, t_i, t_i^*)), \tag{3}$$

where x_i is an indicator. $x_i = 1$ if the i -th ground truth box matches the predicted bounding box, and $x_i = 0$ if the ground truth box does not match. p^c is a confidence representing the probability of a predicted grade c be a ground truth grade. N is the number of matched default boxes. If $N = 0$, the loss is set to 0. t_i is a four parameters vector representing the center coordinate of the i -th predicted box, while t_i^* is the corresponding ground truth. The weighting term λ is set to 0.5 which obtains the best performance by cross-validation. Moreover, grading loss is defined as:

$$\mathcal{L}_{grading}(x, p^c) = - \sum_{i=1}^n \sum_{j=1}^c x_i \log \frac{e^{\theta_j^T p^c}}{\sum_{j=1}^c e^{\theta_j^T p^c}}, \tag{4}$$

Table 3 Representative aspect ratios of target organs from k-means clustering analysis used to the default box proposal

Scale	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
Aspect ratio	0.39	0.83	1.08	1.48	1.83	2.32

where $\mathcal{L}_{grading}$ is a Softmax loss. x_i is an indicator. Grading error occurs when an organ is grading with a miss grade. Localization loss \mathcal{L}_{loc} is defined as:

$$\mathcal{L}_{loc}(x_i, t_i, t_i^*) = \sum_{i=1}^N \sum_{m \in \{c_x, c_y, h, w\}} x_i smooth_{L_1}(t_i^m, t_i^{m*}), \tag{5}$$

in which

$$smooth_{L_1}(z) = \begin{cases} 0.5z^2 & \text{if } |z| < 1, \\ |z| - 0.5 & \text{otherwise.} \end{cases} \tag{6}$$

L_1 smooth is used to refine bounding box coordinates capable of more robustness to outliers than L_2 smooth. The localization error occurs when an organ is detected with a misaligned bounding box, i.e., Jaccard index $J(A, B) = \frac{|A \cap B|}{|A \cup B|} < 0.5$, where A is a predicted box while B is a ground truth box.

To improve the stability and convergence of DMML-Net, a hard negative mining approach is used to select a certain percentage of negatively predicted boxes for the next iteration. After matching steps in the training phase, major predicated boxes are negatives, i.e., many predicted boxes are assigned to low confidence for an organ. Hard negative mining only preserves the negatively predicted boxes with a high confidence satisfying a constraint (Negatives: Positives = 3:1) to feed into the next iteration. Traditional approaches always use all negatives predicated boxes feeding into the next iteration, but they are time-consuming and difficult to converge. Hard negative mining has been shown as an effective strategy for reinforcing the performance of detection networks especially in computer vision domain (Wan et al. 2016; Sun et al. 2017).

To only preserve the best box, a non-maximum suppression approach is used. In the training and testing phases, one target organ has many predicted bounding boxes which have the same grade and high overlap. In practice, the satisfied result should be one organ with one best predicted box. First, the non-maximum suppression approach finds the predicted boxes based on pre-defined confidence. Second, it simultaneously sorts the predicted box which has maximum confidence and other overlap boxes by a pre-defined threshold. Finally, it suppresses unsuitable boxes in order to preserve one best box for each target organ.

The effectiveness and generalization of existing multitask learning have been demonstrated *in practice* by Evgeniou et al. (2007), Ando (2006), and Torralba et al. (2004) and *in theorem* by Baxter et al. (2000), Ando and Zhang (2005), and Ben-David and Schuller (2003). Despite the effectiveness of multitask learning in existing works, how to combine multiscale and multitask learning for an end-to-end way is still an acknowledged problem.

Ours multitask learning formulation seamlessly joins the multiscale learning with an end-to-end framework. It does reinforce the mutual benefit between tasks by sharing parameters from the multiscale multi-output learning architecture.

Experiment and Result

Dataset and Configuration

The proposed DMML-Net has been intensively evaluated on a challenging dataset which includes 200 clinical patients (Avg 60 yrs). Since these patients from multi-center are examined by multiple vendors (GE, Siemens), their MRI scans have different repetition time, echo time, magnetic field, in-plane resolution, and slice thickness. Among sequential

T1/T2-weighted MRI scans of each patient, we select one lumbar middle scan which presents neural foramina better in the sagittal direction. Therefore, the dataset has 200 lumbar scans from 200 patients such that no patient is placed in both sets of training (80%) and testing (20%). Since each lumbar scan has six neural foramina, six discs, and five vertebrae from T12 to S1, this dataset has 1200 neural foramina (518 normal, 682 abnormal), 1200 discs (627 normal, 573 abnormal), and 1000 lumbar vertebrae (690 normal, 310 abnormal). The ground truth values were annotated by one experienced physician.

Standard five-fold cross-validation is employed for performance evaluation and comparison. Data augmentation is also employed to enhance the generalization. Augmentative methods in training phase include random distortion, random flip, random brightness and contrast adjustment. The DMML-Net is implemented in Python 2.7 and Tensorflow

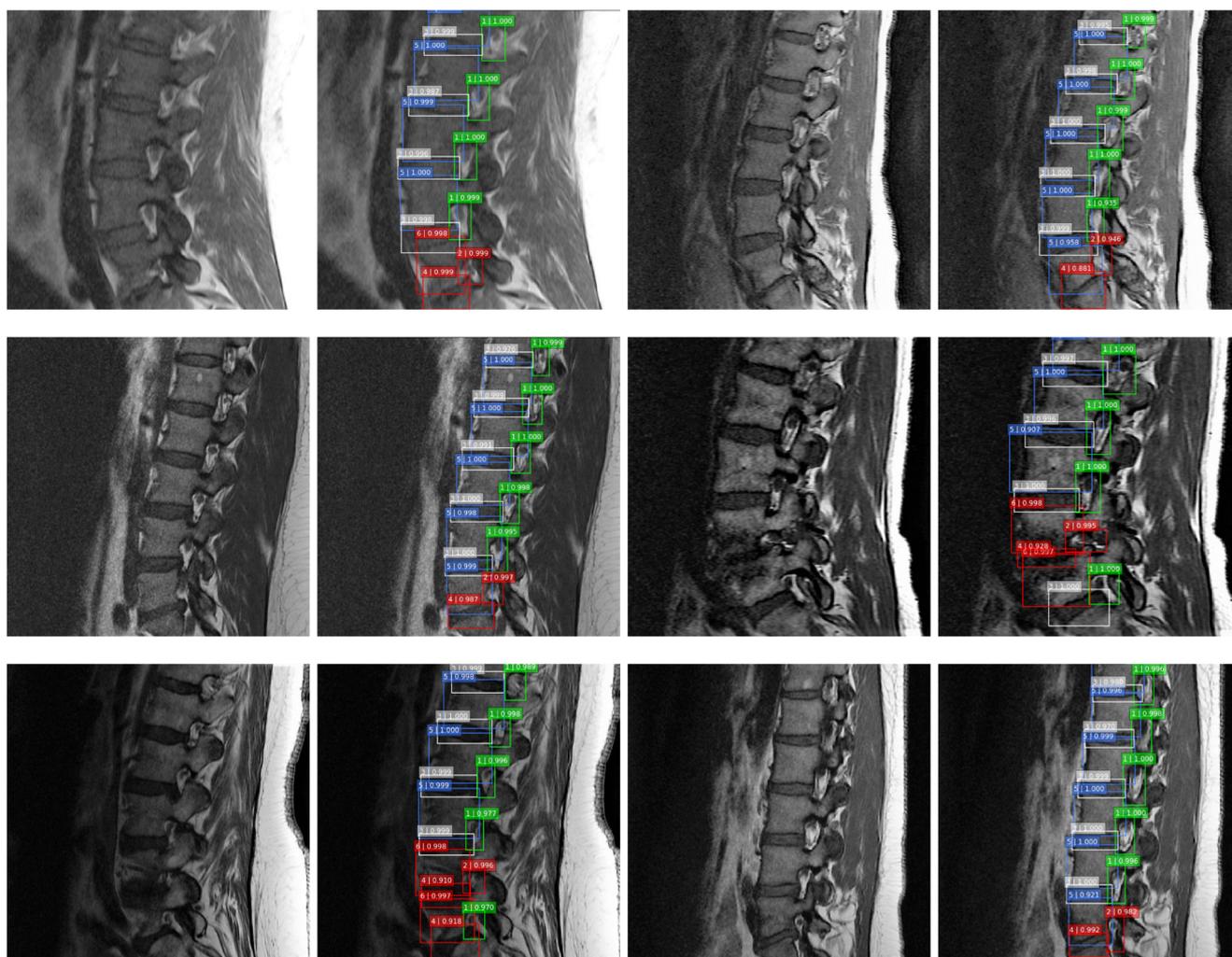


Fig. 8 DMML-Net has achieved efficient performance for pathogenesis-based diagnosis of LNFS, which endows a great tool in clinical settings. Label: 1, 3, 5 represent normal neural foramina (in

green), normal disc (in white), and normal vertebrae (in blue) respectively. Label: 2, 4, 6 represent abnormal organs (in red) respectively. The confidence of organs is next to labels

Table 4 The seamless combination of the multiscale multi-output architecture, k-means clustering method, and feature boosting module has superior effectiveness

Method	mAP_{07}/mAP_{12}	NF_{normal}	$NF_{abnormal}$	$Disc_{normal}$	$Disc_{abnormal}$	V_{normal}	$V_{abnormal}$
SSD	0.743/0.767	0.792/0.813	0.750/0.774	0.772/0.808	0.672/0.693	0.833/0.876	0.641/0.639
Faster-rcnn	0.683/0.706	0.792/0.819	0.524/0.546	0.750/0.771	0.451/0.472	0.845/0.868	0.738/0.759
-FBM	0.782/0.810	0.812/0.835	0.825/0.858	0.803/0.833	0.757/0.786	0.862/0.913	0.630/0.637
-Multiscale	0.706/0.731	0.777/0.813	0.751/0.777	0.703/0.722	0.634/0.642	0.856/0.898	0.527/0.531
-Clustering	0.718/0.731	0.616/0.633	0.558/0.555	0.866/0.891	0.718/0.744	0.871/0.876	0.679/0.684
DMML-Net	0.811/0.845	0.825/0.853	0.837/0.876	0.852/0.885	0.776/0.802	0.873/0.921	0.705/0.730

The first two rows: DMML-Net outperforms popular networks SSD and Faster-rcnn at neural foramina (NF), discs, and vertebrae (V). The last four rows: DMML-Net outperforms its ablated versions after removing feature boosting module (-FBM), multiscale output layer (-Multiscale), or k-means clustering (-Clustering)

The bold emphases represent the highest results

1.0 library (Abadi et al. 2016). Learning rate is 0.01 and optimizer is *adam*. Momentum and weight decay are 0.9 and 0.0005 respectively. Training batch size is 8 and maximum iteration is 10k using a Nvidia GPU Titan X with cuDNN v5.1 and Intel CPU Xeon(R) E5-2620@ 2.5GHz.

Evaluation Criteria

Standard precision-recall curve and mean average precision (mAP) are reported for demonstrating the sensitivity and interpretability (Everingham et al. 2010). The diagnosis of an organ is correct when both localization and grading are correct. So to evaluate localization and grading at the same time, both mAP_{07} and mAP_{12} are adopted as double assurances. Specifically, precision and recall are defined as:

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}, \tag{7}$$

where TP is true positive, FP is false positive, and FN is false negative. The mAP_{07} is used to the VOC2007 challenge (Hoiem et al. 2012). After selecting 15 recall levels $L: \{0.1, 0.2, \dots, 0.99\}$, the precision at each recall level r is interpolated by taking the maximum precision for which corresponding recall exceeds \tilde{r} :

$$P_{interp}(r) = \max_{\tilde{r}: \tilde{r} \geq r} P(\tilde{r}), \tag{8}$$

where $P(\tilde{r})$ is the measured precision at recall \tilde{r} .

$$AP_{07} = \frac{1}{15} \sum_{r \in L} P_{interp}(r), mAP_{07} = \frac{1}{C} \sum_{c_i \in C} AP_{07}(c_i). \tag{9}$$

The mAP_{12} is used to the VOC2012 challenge. It integrates the precisions at continuous recalls:

$$AP_{12} = \int_0^1 P(r)dr, mAP_{12} = \frac{1}{C} \sum_{c_i \in C} AP_{12}(c_i). \tag{10}$$

Performance Evaluation of DMML-Net

The effectiveness and advantages of DMML-Net in the pathogenesis-based diagnosis of LNFS have been demonstrated by extensive experimental results. The representative results are shown in Fig. 8. The first and third columns are raw lumbar MRI scans, while the second and fourth columns are the results of localization and grading of various organs. Our DMML-Net not only accurately diagnoses these abnormal neural foramina (Label 2 in red boxes), but also precisely discovers pathogenic factors in the surrounding abnormal intervertebral discs and vertebrae (Label 4, 6 in red boxes). Even these bounding boxes between neural foramina, discs, and vertebrae have high overlap, DMML-Net has achieved pinpoint localization and precise grading.

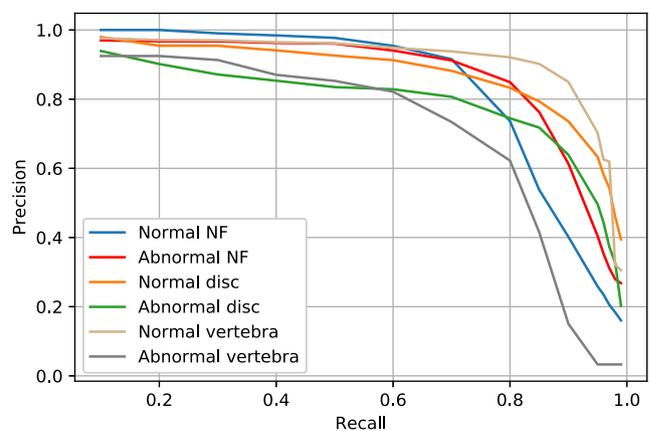


Fig. 9 The combination of multitask regression module and multitask loss module has combated the huge challenges from the localization and grading of all lumbar organs. These stable precision-recall curves also demonstrate the strengths of DMML-Net in the pathogenesis-based diagnosis of LNFS since both normal and abnormal neural foramina (NF) have high recall and precision

This demonstrates the strengths of DMML-Net in handling the structural complexity of spines, scale variability of organs, and weak relatedness of tasks.

Table 4 demonstrates the effectiveness of multitask regression learning as well as the advantage of multiscale multi-output learning. *First*, after removing the feature boosting module (-FBM), mAP_{07} and mAP_{12} decreased by 0.029, 0.035 respectively, which demonstrates the strength of feature boosting module in merging semantic

representations to reinforce the salience of numerous target organs. *Second*, after removing conv3_x and conv4_x blocks (-multiscale) and preserving two output layers (conv6+conv7 block and conv8 layer), the mAP_{07} and mAP_{12} decreased by 0.105, 0.114 respectively, which demonstrates that the multiscale multi-output learning architecture is capable of reliable scale-invariance for various organs. Moreover, only preserving the final conv8 layer as single output layer has worse results, proving the

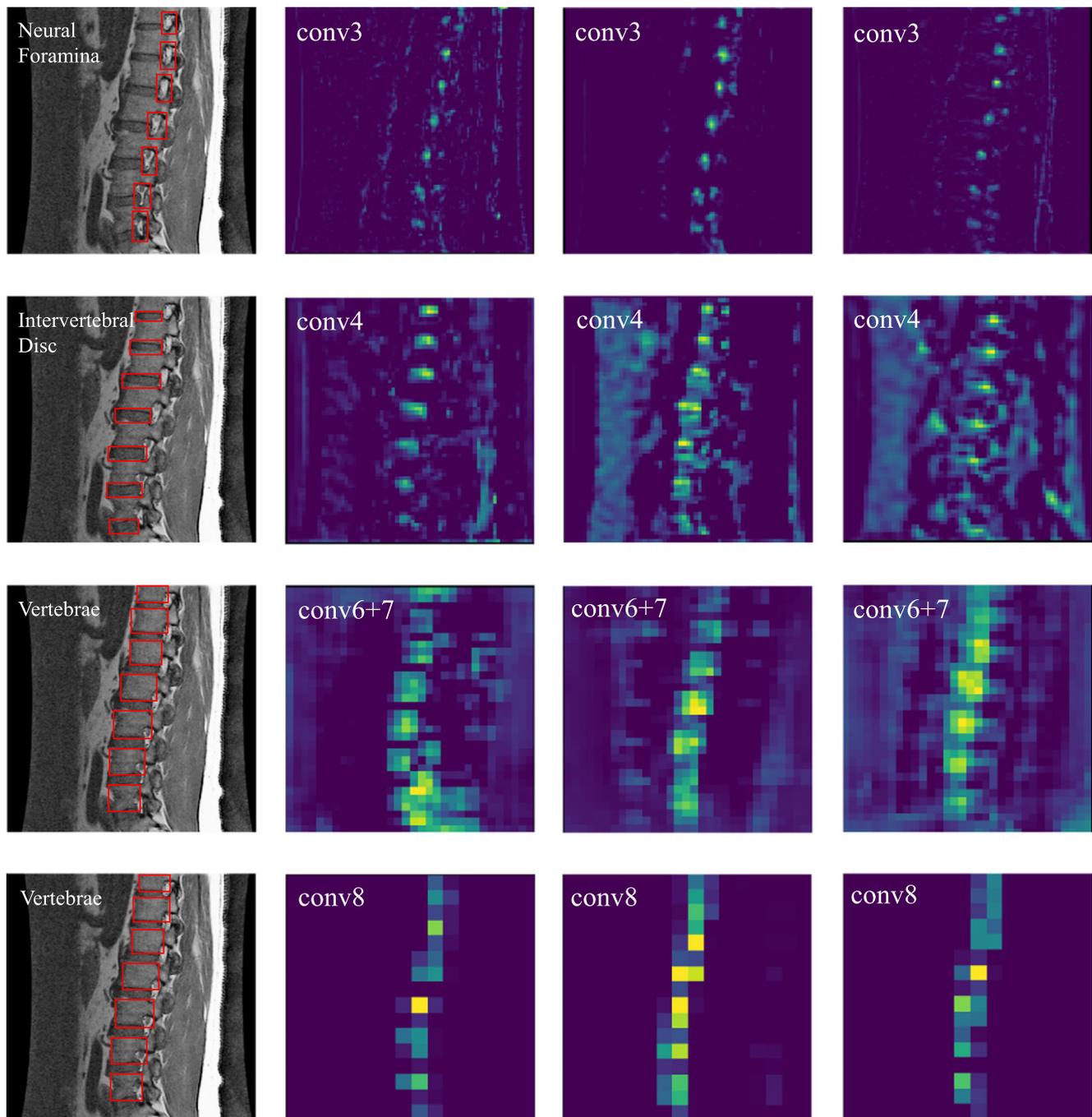


Fig. 10 Representative feature maps generated by the four output layers, for demonstrating the efficiency of DMML-Net

advantages of the selection of four output layers by a k-means clustering method. *Third*, after randomly setting the widths, aspect ratio of default boxes (-clustering) in default box proposal module, the mAP_{07} and mAP_{12} decreased by 0.093, 0.114 respectively, which demonstrates the advantage of embedding k-means clustering results into default box proposal. *Finally*, DMML-Net achieved the highest results (0.811 mAP_{07} , 0.845 mAP_{12}) in comparison with existing popular networks SSD_{512×512} (0.743, 0.767) (Liu et al. 2016) and Faster-rcnn (0.683, 0.706) (Ren et al. 2015). These demonstrate that DMML-Net is the best way for pathogenesis-based diagnosis of LNFS. The SSD model is reimplemented on Tensorflow³ which has reproduced the same performance as the original implement on Caffe. The Faster-rcnn model is also reimplemented on Tensorflow⁴ by Chen and Gupta (2017) with the same performance as the original implement on Caffe.

Figure 9 reports the precision-recall curves of multiple organs. The diagnosis of all organs achieved high precision and recall, especially the diagnosis of abnormal neural foramina achieved 0.8 recall and 0.83 precision at the same time. This demonstrates the robustness of default box proposal, multitask regression module, and multitask loss module. These stable curves also demonstrate that our multitask regression learning formulation prompted the relatedness between tasks and enhances their mutual benefit by sharing parameters. Combining Figs. 8 and 9, the sensitivity and specificity of the pathogenesis-based diagnosis of LNFS are demonstrated.

Discussion and Conclusion

Discussion According to the above comprehensive results, the proposed DMM-Net successfully resolves the three challenges: multiple targets, multiple scales, and multiple tasks, thanks to the feature boosting module with clustering, the multiscale learning, and the multitask learning. In order to better demonstrate our assumption behind clustering to select suitable output layers, we visualize the representation generated by the four output layers: conv3, conv4, conv6+7, and conv8 after the feature boosting module processing. As illustrated in Fig. 10, each row corresponds to one output layer, these generated feature maps are based on one testing MR scan. As the first row in Fig. 10, the conv3 block is focusing on neural foramina, while the conv4 block is focusing on intervertebral discs as shown in the second row. These can validate our assumption behind the multiscale learning as well as demonstrate the advantages and effectiveness of DMML-Net.

³<https://github.com/balancap/SSD-Tensorflow>.

⁴<https://github.com/endernewton/tf-faster-rcnn>.

Conclusion A deep integrated network DMML-Net has been proposed for the pathogenesis-based diagnosis of LNFS. It achieved highly accurate localization and grading of neural foramina as well as surrounding vertebra and discs when validated on a challenging dataset of 200 subjects. Joint learning of multiscale and multitask enhanced the scale-invariance of various organs and the mutual benefit between tasks. It is the first time that automated pathogenesis-based diagnosis of LNFS is investigated. The proposed method has paved a great way for other spinal disease diagnoses, such as disc degeneration, vertebral deformation, and so on.

Information Sharing Statement

Source data can be directly downloaded at https://drive.google.com/file/d/1W64dSL-DOCUt3JOEF0csn_qddmMHF4-7/view?usp=sharing; Code (python 2.7), and model (Tensorflow 1.0) developed in this manuscript are publicly available at <https://github.com/hzycn/DMML-Net/tree/master>.

Acknowledgements This work was made possible through support from Natural Science Foundation of Shandong Province in China (ZR2015FM010), Project of Shandong Province Higher Educational Science and Technology Program in China (No. J15LN20), Project of Shandong Province Traditional Chinese Medicine Technology Development Program in China (2015-026, 2017-001), and Project of Shandong Province Medical and Health Technology Development Program in China (No. 2016WS0577).

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., et al (2016). Tensorflow: large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:160304467.
- Alomari, R.S., Corso, J.J., Chaudhary, V. (2011). Labeling of lumbar discs using both pixel- and object-level features with a two-level probabilistic model. *IEEE Transactions on Medical Imaging*, 30(1), 1–10. <https://doi.org/10.1109/TMI.2010.2047403>.
- Ando, R.K. (2006). Applying alternating structure optimization to word sense disambiguation. In *Proceedings of the tenth conference on computational natural language learning, association for computational linguistics* (pp. 77–84).
- Ando, R.K., & Zhang, T. (2005). A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6, 1817–1853.
- Baxter, J. et al. (2000). A model of inductive bias learning. *Journal of Artificial Intelligence Research (JAIR)*, 12(149–198), 3.
- Bell, S., Lawrence Zitnick, C., Bala, K., Girshick, R. (2016). Inside-outside net: detecting objects in context with skip pooling and

- recurrent neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2874–2883).
- Ben-David, S., & Schuller, R. (2003). Exploiting task relatedness for multiple task learning. In *Learning theory and kernel machines* (pp. 567–580). Springer.
- Cai, Y., Osman, S., Sharma, M., Landis, M., Li, S. (2015). Multi-modality vertebra recognition in arbitrary views using 3d deformable hierarchical model. *IEEE Transactions on Medical Imaging*, 34(8), 1676–1693. <https://doi.org/10.1109/TMI.2015.2392054>.
- Cai, Y., Leungb, S., Warringtonb, J., Pandeyb, S., Shmuilovichb, O., Lib, S. (2017). Direct spondylolisthesis identification and measurement in mr/ct using detectors trained by articulated parameterized spine model. In *Proc. of SPIE* (Vol. 10133, pp. 1013,319–1).
- Chen, X., & Gupta, A. (2017). An implementation of faster rcnn with study for region sampling. arXiv preprint arXiv:170202138.
- Cinotti, G., De Santis, P., Nofroni, I., Postacchini, F. (2002). Stenosis of lumbar intervertebral foramen: anatomic study on predisposing factors. *Spine*, 27(3), 223–229.
- Corso, J.J., Raja'S, A., Chaudhary, V. (2008). Lumbar disc localization and labeling with a probabilistic model on both pixel and object features. In *International conference on medical image computing and computer-assisted intervention* (pp. 202–210). Springer.
- Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2), 303–338.
- Evgeniou, T., Pontil, M., Toubia, O. (2007). A convex optimization approach to modeling consumer heterogeneity in conjoint estimation. *Marketing Science*, 26(6), 805–818.
- Ghosh, S., Raja'S, A., Chaudharya, V., Dhillonb, G. (2011). Automatic lumbar vertebra segmentation from clinical ct for wedge compression fracture diagnosis. *Work*, 9, 11.
- Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440–1448).
- Hariharan, B., Arbeláez, P., Girshick, R., Malik, J. (2015). Hypercolumns for object segmentation and fine-grained localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 447–456).
- Hasegawa, T., An, H.S., Haughton, V.M., Nowicki, B.H. (1995). Lumbar foraminal stenosis: critical heights of the intervertebral discs and foramina. a cryomicrotome study in cadavera. *Journal of Bone and Joint Surgery (American)*, 77(1), 32–38.
- He, X., Yin, Y., Sharma, M., Brahm, G., Mercado, A., Li, S. (2016). Automated diagnosis of neural foraminal stenosis using synchronized superpixels representation. In *MICCAI (2)* (pp. 335–343). Springer.
- He, X., Landisa, M., Leunga, S., Warringtona, J., Shmuilovicha, O., Lia, S. (2017a). Automated grading of lumbar disc degeneration via supervised distance metric learning. In *Proc. of SPIE Vol* (Vol. 10134, pp. 1013,443-1).
- He, X., Lum, A., Sharma, M., Brahm, G., Mercado, A., Li, S. (2017b). Automated segmentation and area estimation of neural foramina with boundary regression model. *Pattern Recognition*, 63, 625–641.
- He, X., Zhang, H., Landis, M., Sharma, M., Warrington, J., Li, S. (2017c). Unsupervised boundary delineation of spinal neural foramina using a multi-feature and adaptive spectral segmentation. *Medical Image Analysis*, 36, 22–40.
- Hoang Ngan Le, T., Zheng, Y., Zhu, C., Luu, K., Savvides, M. (2016). Multiple scale faster-rcnn approach to driver's cell-phone usage and hands on steering wheel detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 46–53).
- Hoiem, D., Chodpathumwan, Y., Dai, Q. (2012). Diagnosing error in object detectors. *Computer Vision–ECCV, 2012*, 340–353.
- Huang, S.H., Chu, Y.H., Lai, S.H., Novak, C.L. (2009). Learning-based vertebra detection and iterative normalized-cut segmentation for spinal mri. *IEEE Transactions on Medical Imaging*, 28(10), 1595–1605.
- Jamaludin, A., Kadir, T., Zisserman, A. (2017). Spinenet: automated classification and evidence visualization in spinal mris. *Medical Image Analysis*, 41, 63–73. <https://doi.org/10.1016/j.media.2017.07.002>. <http://www.sciencedirect.com/science/article/pii/S136184151730110X>, special Issue on the 2016 Conference on Medical Image Computing and Computer Assisted Intervention (Analog to MICCAI 2015).
- Kaneko, Y., Matsumoto, M., Takaishi, H., Nishiwaki, Y., Momoshima, S., Toyama, Y. (2012). Morphometric analysis of the lumbar intervertebral foramen in patients with degenerative lumbar scoliosis by multidetector-row computed tomography. *European Spine Journal*, 21(12), 2594–2602.
- Kelm, B.M., Wels, M., Zhou, S.K., Seifert, S., Suehling, M., Zheng, Y., Comaniciu, D. (2013). Spine detection in ct and mr using iterated marginal space learning. *Medical Image Analysis*, 17(8), 1283–1292.
- Kim, S., Lee, J.W., Chai, J.W., Yoo, H.J., Kang, Y., Seo, J., Ahn, J.M., Kang, H.S. (2015). A new mri grading system for cervical foraminal stenosis based on axial t2-weighted images. *Korean Journal of Radiology*, 16(6), 1294–1302.
- Klinder, T., Wolz, R., Lorenz, C., Franz, A., Ostermann, J. (2008). Spine segmentation using articulated shape models. *Medical Image Computing and Computer-Assisted Intervention–MICCAI, 2008*, 227–234.
- Law, M.W., Tay, K., Leung, A., Garvin, G.J., Li, S. (2013a). Intervertebral disc segmentation in mr images using anisotropic oriented flux. *Medical Image Analysis*, 17(1), 43–61. <https://doi.org/10.1016/j.media.2012.06.006>. <http://www.sciencedirect.com/science/article/pii/S1361841512000874>.
- Law, M.W.K., Garvin, G.J., Tummala, S., Tay, K., Leung, A.E., Li, S. (2013b). *Gradient competition anisotropy for centerline extraction and segmentation of spinal cords* (pp. 49–61). Berlin: Springer.
- Lee, S., Lee, J.W., Yeom, J.S., Kim, K.J., Kim, H.J., Chung, S.K., Kang, H.S. (2010). A practical mri grading system for lumbar foraminal stenosis. *American Journal of Roentgenology*, 194(4), 1095–1098.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C. (2016). Ssd: single shot multibox detector. In *European conference on computer vision* (pp. 21–37). Springer.
- Luo, W., Li, Y., Urtasun, R., Zemel, R. (2016). Understanding the effective receptive field in deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 4898–4906).
- Panjabi, M.M., Maak, T.G., Ivancic, P.C., Ito, S. (2006). Dynamic intervertebral foramen narrowing during simulated rear impact. *Spine*, 31(5), E128–E134.
- Park, H.J., Kim, S., Lee, S.Y., Park, N.H., Rho, M.H., Hong, H.P., Kwag, H.J., Kook, S.H., Choi, S.H. (2012). Clinical correlation of a new mr imaging method for assessing lumbar foraminal stenosis. *American Journal of Neuroradiology*, 33(5), 818–822.
- Peng, Z., Zhong, J., Wee, W., Lee, J.H. (2006). Automated vertebra detection and segmentation from the whole spine mr images. In *27th annual international conference of the engineering in medicine and biology society, 2005. IEEE-EMBS 2005* (pp. 2527–2530). IEEE.
- Rajae, S.S., Bae, H.W., Kanim, L.E., Delamarter, R.B. (2012). Spinal fusion in the united states: analysis of trends from 1998 to 2008. *Spine*, 37(1), 67–76.

- Raja'S, A., Corso, J.J., Chaudhary, V., Dhillon, G. (2011). Toward a clinical lumbar cad: herniation diagnosis. *International Journal of Computer Assisted Radiology and Surgery*, 6(1), 119–126.
- Ren, S., He, K., Girshick, R., Sun, J. (2015). Faster r-cnn: towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91–99).
- Shi, R., Sun, D., Qiu, Z., Weiss, K.L. (2007). An efficient method for segmentation of mri spine images. In *IEEE/ICME international conference on complex medical engineering, 2007. CME 2007* (pp. 713–717). IEEE.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556.
- Štern, D., Likar, B., Pernuš, F., Vrtovec, T. (2009). Automated detection of spinal centrelines, vertebral bodies and intervertebral discs in ct and mr images of lumbar spine. *Physics in Medicine and Biology*, 55(1), 247.
- Sun, X., Wu, P., Hoi, S.C. (2017). Face detection using deep learning: an improved faster rcnn approach. arXiv preprint arXiv:170108289.
- Torralba, A., Murphy, K.P., Freeman, W.T. (2004). Sharing features: efficient boosting procedures for multiclass object detection. In *Proceedings of the 2004 IEEE computer society conference on computer vision and pattern recognition, 2004. CVPR 2004* (Vol. 2, pp. II–762–II–769). <https://doi.org/10.1109/CVPR.2004.1315241>.
- Wan, S., Chen, Z., Zhang, T., Zhang, B., Wong, K.K. (2016). Bootstrapping face detection with hard negative examples. arXiv preprint arXiv:160802236.
- Wang, Q., Lu, L., Wu, D., El-Zehiry, N., Zheng, Y., Shen, D., Zhou, K.S. (2015a). Automatic segmentation of spinal canals in ct images via iterative topology refinement. *IEEE Transactions on Medical Imaging*, 34(8), 1694–1704. <https://doi.org/10.1109/TMI.2015.2436693>.
- Wang, Z., Zhen, X., Tay, K., Osman, S., Romano, W., Li, S. (2015b). Regression segmentation for m3 spinal images. *IEEE Transactions on Medical Imaging*, 34(8), 1640–1648.
- Yan, C., Zhang, Y., Xu, J., Dai, F., Li, L., Dai, Q., Wu, F. (2014a). A highly parallel framework for hevc coding unit partitioning tree decision on many-core processors. *IEEE Signal Processing Letters*, 21(5), 573–576. <https://doi.org/10.1109/LSP.2014.2310494>.
- Yan, C., Zhang, Y., Xu, J., Dai, F., Zhang, J., Dai, Q., Wu, F. (2014b). Efficient parallel framework for hevc motion estimation on many-core processors. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(12), 2077–2089. <https://doi.org/10.1109/TCSVT.2014.2335852>.
- Yan, C., Xie, H., Liu, S., Yin, J., Zhang, Y., Dai, Q. (2017a). Effective uyghur language text detection in complex background images for traffic prompt identification. *IEEE Transactions on Intelligent Transportation Systems*, PP(99), 1–10. <https://doi.org/10.1109/TITS.2017.2749977>.
- Yan, C., Xie, H., Yang, D., Yin, J., Zhang, Y., Dai, Q. (2017b). Supervised hash coding with deep neural network for environment perception of intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, PP(99), 1–12. <https://doi.org/10.1109/TITS.2017.2749965>.
- Yao, J., Burns, J.E., Forsberg, D., Seitel, A., Rasouljan, A., Abolmaesumi, P., Hammernik, K., Urschler, M., Ibragimov, B., Korez, R., et al (2016). A multi-center milestone study of clinical vertebral ct segmentation. *Computerized Medical Imaging and Graphics*, 49, 16–28.
- Zhan, Y., Maneesh, D., Harder, M., Zhou, X.S. (2012). Robust mr spine detection using hierarchical learning and local articulated model. In *International conference on medical image computing and computer-assisted intervention* (pp. 141–148). Springer.