Automated segmentation and area estimation of neural foramina with boundary regression model

Xiaoxu He¹,²,*, Andrea Lum¹,², Manas Sharma¹,², Gary Brahmb, Ashley Mercada,², Shuo Lia,²,⁎

¹ Digital Imaging Group (DIG), London, ON, Canada N6A 4V2
² Department of Medical Imaging, University of Western Ontario, London, ON, Canada N6A 3K7

ARTICLE INFO

Keywords:
Automated segmentation
Area estimation
Neural foramina stenosis
Boundary regression model
Multiple output support vector regression
Multiple kernel learning

ABSTRACT

Accurate segmentation and area estimation of neural foramina from both CT and MR images are essential to clinical diagnosis of neural foramina stenosis. Existing clinical routine, relying on physician's purely manual segmentation, becomes very tedious, laborious, and inefficient. Automated segmentation is highly desirable but faces big challenges from diverse boundary, local weak/no boundary, and intra-inter-modality intensity inhomogeneity. In this paper, a novel boundary regression segmentation framework is proposed for fully automated and multi-modal segmentation of neural foramina. It creatively formulates the segmentation task as a boundary regression problem which models a highly nonlinear mapping function from substantially diverse neural foramina images directly to desired object boundaries. By leveraging a seamless combination of multiple output support vector regression (MSVR) and multiple kernel learning (MKL), the proposed framework enables the domain knowledge learning in a holistic fashion which successfully handles the extreme diversity posing a tremendous challenge to conventional segmentation methods. The performance evaluation was conducted on a dataset including 912 MR images and 306 CT images collected from 152 subjects. Experimental results show that the proposed automated segmentation framework is highly consistent with physician with average DSI (dice similarity index) as high as 0.9005 (CT), 0.8984 (MR), 0.8935 (MR+CT) and BD (boundary distance) as low as 0.6393 mm (CT), 0.6586 mm (MR), 0.6881 mm (MR+CT). Based on this accurate automated segmentation, the estimated areas, highly correlated to their independent ground truth, have been achieved that the proposed automated segmentation framework is highly consistent with physician with average DSI (dice similarity index) as high as 0.9005 (CT), 0.8984 (MR), 0.8935 (MR+CT) and BD (boundary distance) as low as 0.6393 mm (CT), 0.6586 mm (MR), 0.6881 mm (MR+CT). Based on this accurate automated segmentation, the estimated areas, highly correlated to their independent ground truth, have been achieved with correlation coefficient: 0.9154 (CT) and 0.8789 (MR). Hence, the proposed approach enables an efficient, accurate and convenient tool for clinical diagnosis of neural foramina stenosis.

1. Introduction

Neural foramina stenosis (NFS), clinically defined as the narrowing of the bony exit (see Fig. 1(a)) of the spinal nerve root, is caused by abnormalities in vertebral and intervertebral disc, such as a decrease in the height of an intervertebral disc, or osteoarthritic changes in the facet joints [1,2]. Symptoms of NFS are very common, affecting up to 80% of the population worldwide, and may cause pain, disability and economic loss [3–5]. For example, each year more than 400,000 Americans suffer from lower back or leg pain [6,7]. Diagnosis and treatment of NFS, often require segmentation of neural foramina images from multiple imaging modalities for estimating its area as quantitative analysis [1,8–10]. Here, MR and CT imaging are often simultaneously required as MR is better to display the stenosis caused by disc abnormality and CT is better to display the stenosis caused by vertebra abnormality (as shown in Fig. 1(b)). For efficient diagnosis and timely treatment of NFS, manual segmentation by physician is bound to be infeasible for neural foramina images because of its known tediousness, inefficiency, and inconsistency [8,10].

Computer processing methods are highly desirable, but face big challenges due to the following complexities in segmentation of neural foramina (as shown in Fig. 1(c)):

1. Complex appearance inhomogeneity: Two types of appearance inhomogeneity are included:
   (1) Inter-modality intensity difference: In different modalities, the intensity profile of neural foramina is completely different [10].
   (2) Intra-modality appearance variation: Even for one specific modality, the structures passing neural foramina are inhomogeneous and this inhomogeneity varies with different subjects, positions, and spine abnormalities [3].

2. Great boundary variations: Two types of boundary variations are included:
   (1) Diverse boundary shape variation: The boundary shape of
neural foramina varies with different positions in spine, subjects, and spine abnormalities [11,12]. For example, the shape of boundary in normal and stenosed state is obviously different [5,12].

(2) **Local weak/no boundary:** Low intensity contrast around some parts of boundary leads to local weak/no boundary problem. For example, in CT imaging, local weak/no boundary problem commonly appears (marked in green dotted line) as neural foramina has highly similar intensity profile with the surrounding intervertebral disc.

All these mentioned complexities bring great challenges (as shown in Fig. 2) to conventional computer processing methods:

- **Infeasibility caused by great variations in intensity profiles (Fig. 2(a)):** Intensity-based methods are very sensitive to inhomogeneous intensity profiles which commonly appear in MR imaging; even worse, the intensity profiles in MR and CT are completely different. So intensity-based methods can easily be confused when segmenting images from different modalities in a single framework [13–15].

- **Infeasibility caused by inhomogeneous intensity distribution (Fig. 2(b)):** Region-based segmentation methods are very sensitive to inhomogeneous intensity distribution inside neural foramina, which brings a noise disturbance when seeking the optimal region partition. So region-based segmentation methods [15,16] stops at a false region.

- **Infeasibility caused by the combination of great boundary diversity and local weak/no boundary (Fig. 2(c)):** Semi-automated methods, based on the evolution of an initialized boundary, are sensitive to the combination of great boundary diversity and local weak/no boundary as it breaks the required assumption which considers boundaries as small deformations of an initialized boundary, and the intensity contrast around a boundary is homogenous and strong [17,18]. So semi-automated methods fail to evolve the true boundary.

- **Infeasibility caused by the weak edge, no edge, and noisy edge (Fig. 2(d)):** Edge-based methods [19] are sensitive to local weak/no edge around the desired boundary, and are easily be disturbed by the noisy edge, which is strong but does not lie in the desired boundary (indicated by line box). So edge-based methods leak the true edge of neural foramina.

Even worse, in clinical practice, these mentioned difficulties always simultaneously appear so that a more bigger challenge is posed to conventional computer processing methods. Due to the discussed enormity of the challenges, as far as we know, there is still no automated segmentation method proposed for neural foramina.

In this paper, we propose a novel regression segmentation framework for automated segmentation and area estimation of neural foramina from MR and CT images, for the first time. It creatively formulates the segmentation task as a boundary regression problem to associate extremely diverse images directly with desired boundaries. By leveraging the strength of sparse kernel machines, a highly nonlinear boundary regression model is learnt in holistic regression fashion. Such holistic regression fashion enables the learnt model with an accurate, efficient, and robust boundary prediction: (1) the holistic regression outputs where the locations of all the boundary points are regressed simultaneously instead of separately, therefore, the local weak part of the regressed boundaries are guided by the global shape prior learned from the training data; (2) the holistic regression input where each
boundary point is regressed using the full image as a signal, therefore the boundary regression model is able to capture the full context of a boundary point instead of local information, which is sensitive to local noise disturbance. Hence, our boundary regression model provides the accurate and efficient segmentation of neural foramina for images from MR and CT modalities. This study is of significant clinical importance, and its clinical validation has been accepted in Radiological Society of North America (RSNA) 2015 as an abstract.

In summary, contributions of our paper include the following three aspects:

1. **Application**: It achieved a fully automated and multi-modal segmentation tool for clinical practice to greatly reduce the heavy burden for physicians and improve the efficiency in diagnosis of neural foramina stenosis.

2. **Approach**: It creatively formulated the segmentation task as a boundary regression problem to fully leverage the advancement of machine learning in a holistic fashion for seeking the optimal segmentation which simultaneously preserves accuracy, robustness, and efficiency.

3. **Methodology**: It creatively combined holistic boundary regression and multi-feature optimal fusion together in one unified framework for capturing the highly nonlinear relationship from diverse images to desired boundaries. Here, multi-feature optimal fusion enables a new sparse discriminating image feature as the regression input to eliminate irrelevant feature disturbance; holistic boundary regression enables the regression output which regression with the guidance from the input image and the learnt global shape prior. Hence, the combination of them enables accurate boundary regression output.

This paper is structured as follows. In Section 2, the overview of the proposed segmentation framework will be described. Then the key component of this framework, boundary regression model, will be detailed in Section 3. Next, experiment evaluations are presented in Section 4 and Section 5 concludes this paper.

### 2. Method overview

The overview of our regression segmentation framework is mainly composed of two processes (as shown in Fig. 3): (1) the training process learns a highly nonlinear boundary regression model $Y = F(X)$ from the training image set $D = \{(X_i, Y_i) \mid i = 1, \ldots, l\}$ including $l$ images contoured by physicians; (2) the testing process utilizes the learnt model to directly predict the locations of neural foramina's boundary for the input image. The notation used in this paper is summarized in Table 1.

Here, boundary $Y$ as a holistic regression output is represented by the coordinates of a set of points (as described in Section 2.1). With this flexible representation, the accurate and efficient segmentation can be directly obtained from the simultaneous regression of locations for all boundary points with the guidance from the input image and the learnt global shape prior. Image $X$ as a holistic regression input is represented by a sparse discriminating image feature extracted from multiple image
With this discriminative image feature, the robust segmentation can be directly obtained from the reliable regression of each boundary point using the most relevant image information as a signal. The highly nonlinear boundary regression model \( Y = F(X) \) is built by the seamless combination of multiple output support vector regression (MSVR) and multiple kernel learning (MKL) in an unified framework (as described in Section 3). Such seamless combination generates an optimally combined kernel for accurately capturing the fully reliable context of each boundary point, and provides the simultaneous guidance from the input image and the learnt global shape prior for holistic boundary regression.

The novelty of our regression segmentation method is the seamless combination of boundary holistic regression and multiple image features fusion, which brings the great flexibility and robustness in diverse neural foramina images. Hence, our framework has the following advantages in segmentation of neural foramina: (1) accurate segmentation is obtained from the preserved boundary shape diversity; (2) robust segmentation is obtained from the solved local weak/no boundary problem guided by the learnt global shape prior from the training set; (3) efficient segmentation is obtained from the direct boundary regression, without any initialization or alignment intermediate operations.

### 2.1. Boundary holistic regression output

Our framework utilizes the strength of multi-output regression [21–23] to simultaneously predict the locations of all boundary points for the input image. This holistic regression output fashion enables the simultaneous guidance from the input image and the learned global shape prior for regression. The guidance from the input image plays an important part in regression of each boundary point and serves as a fine-tuning adjustment for the output boundary. The guidance from the learnt global shape prior provides important input for the holistic boundary regression and serves as a natural connector for the local weak/no boundary. Hence, this holistic regression output fashion simultaneously exhibits the global and local variations (as shown in Fig. 5), so that an accurate segmentation of neural foramina can be naturally obtained.

#### 2.1.1. Flexible boundary points' representation

To capture the great diversity in boundaries [5,12], flexible boundary representation (as shown in Fig. 6(a)) based on the locations of \( q \) boundary points is employed:

<table>
<thead>
<tr>
<th>Notations</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>Number of inputs</td>
</tr>
<tr>
<td>( q )</td>
<td>Number of boundary points</td>
</tr>
<tr>
<td>( Q (Q=2q) )</td>
<td>Number of outputs</td>
</tr>
<tr>
<td>( M )</td>
<td>Number of kernels/features</td>
</tr>
<tr>
<td>( l )</td>
<td>Number of training samples</td>
</tr>
</tbody>
</table>
where $(locx_t, locy_t)$ is the $t$-th boundary point's coordinates. For each boundary, the boundary points are sampled in a consistent way: the first point (yellow star in Fig. 7) is fixed at the location where the boundary is crossed by the horizontal center line, and the rest $q - 1$ points are evenly sampled along the boundary in the clockwise direction.

Using locations of boundary points as boundary representation brings the following advantages: (1) depict the local shape difference (as shown in Fig. 6(b)): the locations of the $t$-th boundary point $y_t = (locx_t, locy_t)$ can be adjusted to depict boundary with local shape difference; (2) reserve the global shape information: $t = 1, \ldots, q$ records the connectivity defining how the landmarks are joined to form the boundaries in the image; (3) represent multiple-level information: $q$ is the number of boundary point, and its value can determine different resolution of the boundary shape. In our experiments, to smoothly approximate complicated shapes, $q=100$ points were used empirically for representing neural foramina boundary.

2.1.2. Holistic regression output by MSVR

Given the training data set $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^l$ of $l$ neural foramina images. Each training image includes two parts: a feature vector of $N$ descriptive variables $X_i \in \mathbb{R}^N$ and boundary points' locations vector with $Q$ target variables $Y_i = (y^{(1)}_i, \ldots, y^{(Q)}_i), Q = 2q$. The task of boundary regression is to learn a regression model $F$ which assigns to each image, given by the vector $X$, a vector $Y$ of $Q$ target values:

$$ y^{(t)} = F(X), \quad t = 1, \ldots, Q. $$

(2)

To provide holistic regression of all boundary points, MSVR is used to minimize the following objective function:

$$ \min_{w^{(1)}, \ldots, w^{(Q)}, b^{(1)}, \ldots, b^{(Q)}} \sum_{i=1}^l \left\{ \frac{1}{2} \sum_{q=1}^Q \| w^{(q)} \|_2^2 + C \sum_{t=1}^l \mathcal{L}(y^{(t)}_i - (\phi(X_i)w^{(q)} + b^{(q)})) \right\} $$

(3)

where $\phi: \mathbb{R}^N \rightarrow \mathbb{R}^{n_h}$ is a mapping to some higher dimensional Hilbert space $\mathcal{H}$ with $n_h$ dimensions. The loss term $\mathcal{L}(\cdot)$ is defined as the $\epsilon$-insensitive loss function [21–23]. $W = (w^{(1)}, \ldots, w^{(Q)})$ is a $N \times Q$ weight matrix and $b = (b^{(1)}, \ldots, b^{(Q)})^T$ is the bias parameters.

Compared with the multiple use of single-output SVR, MSVR is faster as the $Q$ dimensional outputs share the same kernel, more accurate as the coupling among the outputs are fully considered, and more robust as the local weak/no boundary regression are guided by the global shape prior learned from the training data.

Advantages: The use of MSVR greatly increases the efficiency and accuracy of our regression model due to the following three advantages:

- Its capability of modeling a highly nonlinear mapping function from diverse images to desired boundaries enables direct segmentation, without any intermediate steps such as shape alignment or mean shape initialization.
- Its capability of predicting all the dimensions in the output vector simultaneously and dependently exactly fits the boundary prediction
problem, because the boundary spatial coherence can be fully exploited to achieve a more accurate and efficient prediction.

- Its capability of a sparse kernel machine naturally eliminates image features irrelevant to all boundary points’ regression tasks, and identifies discriminative image features for each of the tasks.

2.2. Discriminative image feature regression input

Our framework utilizes a sparse discriminative image feature as regression input to capture the fully relevant context of each boundary point. As shown in Fig. 8, this discriminative image feature is achieved by MKL which exploits optimally combined kernels that naturally correspond to multiple image features to improve the discrimination ability of the proposed framework. With MKL, the obtained image feature representation removes irrelevant and redundant information by assigning smaller kernel weights, and strengthens relevant and discriminative features with bigger weights. With this new image feature as holistic regression input, the complex relationship from the whole image to each boundary point can be accurately captured. Here, inspired by [24] which introduces MKL into SVR, we combine MKL in MSVR. This extension is the first and not investigated in the original multi-output support vector regressor (MSVR) work [25,21–23].

2.2.1. Multiple image feature extraction

To capture the full and reliable context of a boundary point from the input image, a comprehensive set of multiple image features from different levels and aspects is employed. This complete feature set consists of the following ones due to their capabilities in capturing the image context from different views and their mutual complementarity.

- **Intensity**: Intensity carries fundamental information of the input image as the intensity of neural foramina is obviously different from its surroundings.
- **Texture**: WI-SIFT [26,27] captures the texture information in the input image by computing a histogram of oriented gradients in image blocks.
- **Shape**: HOG [28–30] captures the main shape of the anatomic structure in an image based on the occurrences of gradient orientation in localized partitions of the image. Here, it describes the shape information of neural foramina, which upholds invariance to geometric and photometric transformations by operating on localized cells.
- **Semantic context**: GIST [31] summarizes an image’s semantic context by computing its response to a Gabor filter bank.

Consequently, the whole image feature set extracted from an input image \( I \) is represented as

\[
X = f(I) = [X^{\text{Intensity}}, X^{\text{WI-SIFT}}, X^{\text{HOG}}, X^{\text{GIST}}]
\]  

2.2.2. Supervised sparse discriminative image feature learning by MKL

A novel sparse discriminative image features is learnt by the seamless integration of MKL into MSVR, which achieves an optimal weight combination of four kernels corresponding to the extracted four
image features in Eq. (2). This adaptively combined kernel, in the form of the product of RBF (radial basis function) kernels, assigns smaller weight values to eliminate irrelevant feature and bigger weight values to capture the fully relevant context of a boundary point. To be specific, given a pool of RBF kernels, each individual RBF kernel $K_{d_m}$ is constructed by a different image feature $X_m, m=1,\ldots,M$ out of the four ($M=4$) employed features, so the final kernel is learned by the weighted linear combination of the kernels from this pool:

$$K_d(X, X') = \prod_{m=1}^{M} \exp(-d_m \parallel X^m - (X')^m \parallel^2)$$  

(5)

where, $\phi_d(X) : R^d \rightarrow R^H$ in Eq. (5) is a nonlinear transformation to a higher $H$-dimensional space:

$$\phi_d(X) = \left[ \sqrt{d_1} \phi_{d_1}(X) \ldots \sqrt{d_M} \phi_{d_M}(X) \right]$$  

(6)

Here, Eq. (6) has been parameterized by multiple kernel weights $d = \{d_1, \ldots, d_M\}$.

**Advantages:** The use of MKL increases the discriminant power of our regression model with the following three advantages:

- **It fuses different types of image features $\{X_m\}_{m=1}, \ldots,M\}$.**
- **It removes the irrelevant feature and strengthens the relevant feature via different kernels weights $d$.**
- **It achieves an optimal combination of different types of image features via supervised learning from the training set.**
3. Boundary regression model learning

Our boundary regression model is learned by a seamless combination of MSVR and MKL together into a unified regressor called multi-kernel multi-output support vector regressor (MKMOSVR) (as shown in Fig. 8). Such combination of MSVR and MKL boosts each other’s strength, therefore, MKMOSVR is able to learn various appearances and geometric characteristics of neural foramina into one single model.

To be specific, MKMOSVR provides the following merits to the learnt boundary regression model:

- **A highly nonlinear mapping function**: It utilizes the strength of sparse kernel machines [24,32] to learn a highly nonlinear model for associating extremely diverse images directly with desired boundaries.

- **Flexible regression output**: It utilizes the strength of multi-output regression [21–23] to simultaneously regress the locations of all boundary points for providing the flexible boundary representation, in addition, this holistic regression fashion successfully solves the weak/no boundary problem using the guidance of the global shape prior learned from the training data.

- **Discriminative regression input**: It utilizes the strength of multi-kernel learning [24,33] to optimally fuse different types of image features together for generating a new discriminant feature to capture the context of each boundary point, and this new regression input greatly increases the discriminant power of the regressor.

By leveraging these merits of MKMOSVR, the learnt boundary regression model accurately regresses desired boundaries for the input images with great variations in shape, appearance, and the disturbance from weak/no boundary.

Next, we first present the mathematic formulation of MKMOSVR, and then detail the implementation.

3.1. Mathematic formulation

The mathematic formulation of MKMOSVR is defined as follows:

$$Y = W \cdot \phi_d(X) + b$$  \hspace{1cm} (7)

Eq. (7) denotes for every $N$ dimensional image feature input $X$ composed of $M$ individual features, a $Q$ dimensional vector boundary output $Y$ is simultaneously regressed by weight vector $W = (w^{(1)}, \ldots, w^{(Q)})$, the bias parameters $b = (b^{(1)}, \ldots, b^{(Q)})$, and feature kernel weights $d = (d_1, \ldots, d_q)$.

Compared with the multiple use of single-output SVR, MKMOSVR is faster as the $Q$ dimensional outputs simultaneously regress, and is more accurate as the coupling among the outputs are fully considered. Meanwhile, compared with the conventional single kernel learning, MKMOSVR has a better learning performance as the enhanced discriminant ability successfully handles the extreme diversity in the images by fusing different kernel weights $d$ to suppress irrelevant feature to all boundary points’ regression and strengthen relevant feature to each of the boundary point regression.

Fig. 8. Supervised sparse discriminative image feature learning by MKL.
Algorithm 1. MKMOSVR Learning

\[ i \leftarrow 0; \]
\[ \text{Initialize } d^0 \text{ randomly}; \]
repeat \( k \)
\[ \begin{align*}
  & K \leftarrow K_{d^k}; \\
  & \text{Use the MSVR solver in Eq. (8) to solve the single kernel problem with fixed kernel } K \text{ and obtain } \beta; \\
  & d^{i+1} = d^i - s \left( \frac{\partial}{\partial d} - \frac{1}{2} \sum_{j=1}^{Q} (\beta^T K_j d^i) \right); \\
  & i \leftarrow i + 1;
\end{align*} \]
until converged

3.2. The implementation of MKMOSVR

The implementation contains two main phases: (1) the training phase learns a highly nonlinear mapping function from training set delineated by physicians; (2) the test phase directly regresses the holistic boundary for the input image via the learned boundary regression model.

Training phase: The regression problem in Eq. (7) is performed by optimizing the following two parts: (1) the regression hyperplane \( F_{w,b}(X) \); (2) the kernel weight vector \( d = (d_1, \ldots, d_Q) \). So the objective of the MKMOSVR is to learn the optimal parameters of \( W, b, \) and \( d \) from the training set \( D = \{(X_i, Y_i) | i = 1, \ldots, l \} \). It can be solved by a nested two step optimization procedure [24] shown in Algorithm 1: the inner loop holds the fixed kernel and learns the value of \( W, b \) that optimizes the MSVR parameters; while the outer loop optimizes the kernel parameter \( d \) using a gradient descent technique.

1. The inner loop stage (fix \( d \), compute \( W, b \)): In the inner loop stage, the kernel weight vector \( d \) is fixed, i.e., \( K_d(X, X') = \phi_d(X)\phi_d(X') \) is known, the objective of the inner loop is to solve the single kernel problem with kernel \( \phi = \phi_d \) using MSVR solver, which learns the optimal parameters of \( W = (w^{(1)}, \ldots, w^{(Q)}), b = (b^{(1)}, \ldots, b^{(Q)})^T \) by solving the optimization problem [25,22,23]:

\[
\begin{align*}
  & \min_{W,b} L(W, b, d) = \min_{W,b} \frac{1}{2} \|W - (Y - \langle W, \phi(X) \rangle) \|_2 + C \sum_{i=1}^{l} (\xi_i + \hat{\xi}_i) \text{subject to } \langle W, \phi(X_i) \rangle + b - Y_i \leq \epsilon + \xi_i, \quad Y_i - \langle W, \phi(X_i) \rangle + b \leq \epsilon + \hat{\xi}_i, \quad \xi_i, \hat{\xi}_i \geq 0, \quad i = 1, \ldots, l.
\end{align*}
\]
where \( l \) is the number of training patterns, \( \epsilon \neq 0 \) is the allowed error, \( C > 0 \) is the penalty parameter to control the penalty degree to the samples exceeding the prediction error \( \epsilon, \xi_i = (\xi_i^{(1)}, \ldots, \xi_i^{(Q)}) \) and \( \hat{\xi}_i = (\hat{\xi}_i^{(1)}, \ldots, \hat{\xi}_i^{(Q)}) \) are slack variables for exceeding the target value by more than \( \epsilon \) and for being below the target value by more than \( \epsilon \). Also, \( \langle , \rangle \) indicates the inner product of the involved arguments. Eq. (8) takes into account the prediction errors of all dimensional outputs and therefore incorporates the boundary spatial coherence.

The best solution of \( W \) can be expressed as a linear combination of the training samples in the transformed feature space [25,23]:

\[
  w^{(t)} = \sum_{i=1}^{l} \phi(X_i)\beta_i^{(t)}, \quad t = 1, \ldots, Q
\]
Accordingly, the best solution of \( b \) is

\[
  b^{(t)} = Y_i^{(t)} - C \sum_{i=1}^{l} (\phi(X_i)\beta_i^{(t)}) - \langle W, \phi(X) \rangle, \quad t = 1, \ldots, Q
\]

2. The outer loop stage (fix \( W, b \), compute \( d \)): Since \( W, b \) is known in the outer loop stage, the objective function becomes:
where \( r \) is a differential regularizer function of \( d \).

Note that \( J(d) \) only depends on \( d \). To obtain optimal kernel weights \( d \), a gradient projection is utilized in the outer loop of Algorithm 1: (1) the descent step size is chosen based on the Armijo rule to guarantee convergence [24]; (2) the descent direction is extended from the single dimensional output format in [24] to the multi-dimensional output format:

\[
\frac{\partial J}{\partial d_m} = \frac{\partial r}{\partial d_m} - \frac{1}{2Q} \sum_{j=1}^{Q} \left( H(\beta) \right)^{T} \frac{\partial K}{\partial d_m} \beta^j
\]

(12)

**Testing phase:** Once obtaining the optimal parameters of \( W, b, \) and \( d \), the \( Q \)-dimensional output \( Y \) (i.e., the locations of all boundary points) for each new input \( X \) can be computed using Eq. (7). Based on the regressed boundary, segmentation of neural foramina can be directly obtained.

### 4. Experiments and results

The superiority of the proposed approaches has been intensively evaluated by a cross-modality MR + CT dataset. The results in segmentation demonstrated a high consistency with an experienced physician, with average dice similarity index (DSI) as high as 0.8935 and an average boundary distance (BD) as low as 0.6881 mm in CT and MR images.

#### 4.1. Experiment setting

The experimental dataset includes 912 ROIs in MR modality and 306 ROIs in CT modality collected on 152 subjects. Manual segmentation and area estimation of these ROIs, used as the benchmark, were performed by one experienced physician. All ROIs are tested and trained using a leave-one-subject-out cross-validation procedure in which ROIs from the same subject is iteratively left out as the test data and the training is conducted on the ROIs from all the other subjects.

**Dataset:** These experimental ROIs are from two image sets collected on 152 subjects: (1) the first image set includes T1 MR scans collected from 110 subjects (41 men, 69 women, avg 60 ± 16 yrs) to provide 912 ROIs in MR modality. These collected MR scans, covering the entire lumbar spine, are scanned using a sagittal T1 weight MRI with repetition time (TR) of 533 ms and echo time (TE) of 17 ms under a magnetic field of 1.5 T, and the in-plane resolution is 0.5 mm × 0.5 mm with slice thickness 3.3 mm; (2) the second image set consists of spinal CT scans collected from 42 subjects (21 men, 21 women, avg 60 ± 16 yrs) to provide 306 ROIs in CT modality, these CT scans covering lumbar spine have the in-plane resolution 0.4 mm × 0.4 mm and the slice thickness 0.6 mm.

**ROI selection:** These ROIs are automatically extracted based on two selected landmarks. Firstly, choose the center of the upper pedicle and lower pedicle as two landmarks (blue stars in Fig. 9) [34]. Then a square ROI is automatically cropped with its center (red dot in Fig. 9) at the middle point of the landmarks [3,9]. Here, to simulate the inter- and intra- users’ variation in clinical practice, we let the size of ROIs change from 1.5 to 2 times the landmarks’ distance.

**Parameter setup:** The proposed framework are compared with the following publicly available out-of-state segmentation programs: for graph cuts see References [13,35], for pixel classification see Reference [14], for active contours see References [18,19], and for Mumford–Shah model see References [36,37]. The values for the input parameters of each algorithm were determined experimentally to achieve the best performance for a fair comparison. Any internal parameters were kept fixed. For active contour method, the required initial shape is the mean shape of neural foramina (as shown in Fig. 4(a)), and the resulted shape is evolved for exactly 1000 iterations; for Mumford–Shah model method, two different initial shapes inside and outside of neural foramina (as shown in Fig. 15(a) and (c)) are considered, and the resulted shape is evolved for exactly 1000 iterations.

#### 4.2. Evaluation metric

Two kinds of metric are used for evaluation:

**Metric in segmentation:** To evaluate the performance in segmentation, DSI [38] and BD [16] are used in this paper. DSI measures the closeness of boundaries between manually segmented (M) and automatically segmented (A) neural foramina:

\[
DSI = \frac{2S_{MA}}{S_{A} + S_{M}}
\]

(13)

where \( S_{A} \) is the area of the automated segmentation, \( S_{M} \) is the area of the manual segmentation by the physician, and \( S_{MA} \) is the area of the overlap region between \( M \) and \( A \). The range of DSI is \([0, 1]\), and the change from 0 to 1 represents the worst match to the best match. While BD assesses the difference in boundary shape:

\[
BD = \frac{\sum_{j=1}^{Q} D_{j}}{2d_{x}} + \frac{\sum_{j=1}^{Q} D_{j}}{2d_{y}}
\]

(14)

where \( D_{j} \) is the distance of the \( j \)-th boundary point on the boundary of \( A \) to the closest one on the boundary of \( M \), \( q_{A} \) is the total number of boundary points on the boundary of \( A \), while \( D_{j} \) is the distance of the \( j \)-th boundary point on the boundary of \( M \) to the closest one on the boundary of \( A \), and \( q_{MA} \) is the total number of boundary points on the boundary of \( M \). The range is \([0, \infty]\) for BD, and indicates from the best match to the worst match.

**Metric in area estimation:** To evaluate the performance in area estimation, Pearson’s correlation coefficient \( Corr \) [39] is used to quantitatively evaluate correlations between the automated area estimation \( S_{A} \) and the manual area estimation \( S_{M} \) by physician. If the correlation coefficient \( Corr(S_{A}, S_{M}) \) is close to 1, it indicates that \( S_{A} \) and \( S_{M} \) are positively linearly related and the scatter plot falls almost along a straight line with a positive slope.

#### 4.3. Performance analysis in segmentation

The superiority of the proposed segmentation framework is validated by: (1) qualitative analysis, which intuitively demonstrates the accurate and robust segmentations have been achieved in images with diverse appearance, weak/no boundary, and the variability in boundaries; (2) quantitative analysis, which numerically illustrates the effectiveness of the proposed framework has been achieved in three common cases in clinical practice: MR, CT, and cross-modalities.
Fig. 10. Visualization of delineation results in MR and CT modality. Each automatic segmentation is represented as a red solid contour, and its corresponding ground truth manually obtained by spine physician is represented as a blue dashed contour. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)
Fig. 11. The robustness of the proposed segmentation in images with weak/no boundary, diverse appearance and shape variation. Each automated segmentation is represented as a red solid contour. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)
The advantages from the creative combination of holistic boundary regression (MSVR) and multiple image optimal fusion (MKL) in our method were validated by two groups of experiments: (1) the higher accuracy and less computation time of holistic boundary regression was validated by the comparison between MSVR and single output support vector regression (SSVR); (2) the better discriminative ability of multiple image optimal fusion was validated by the comparison between MKL and individual kernel learning.

The advantage of holistic boundary regression: Table 4 demonstrates the superiority of MSVR over SSVR. It can be found that MSVR is superior to SSVR due to the greatly improved prediction accuracy (1.1 mm BD decrease) and the less computational time (200 times faster). The reason for this greatly improved accuracy is that the coupling between all boundary points is considered by MSVR, and provides a guidance from the learnt global shape prior in regression of boundary points on local weak/no boundary. While this coupling is ignored by SSVR which merely predicts each boundary independently. The greatly improved computation efficiency comes from the fact that the computation of MSVR is only done once as the holistic outputs share the same objective function. In comparison, SSVR needs 2q times (q is the number of boundary points, 2 represents two coordinates).

The advantage of multi-feature optimal fusion: A better learning performance of multi-kernel MSVR than any other MSVRs with a single kernel corresponding to one specific image feature has been demonstrated in Fig. 13. To be specific, a better discriminant ability of multi-kernel learning is reported in Fig. 13(a), and an improved mean accuracy and reduced standard variation of multi-kernel MSVR is reported in Fig. 13(b).

For the evaluated discriminant power in multi-kernel MSVR and single-kernel MSVR, the observed indicator is the change of the accuracy (measured by Dice) in a series of experiments where the number of training samples was increased gradually (i.e., from 110 to 1010) to simulate the added image appearance diversity in the training set. Fig. 13(a) demonstrates the corresponding results of all the compared methods. It can be found that the accuracy (measured by Dice) of multi-kernel MSVR maintains growth with the increasing of training samples while the accuracy (measured by Dice) of other MSVRs with a single kernel becomes small and unstable. This is because the adaptively combined kernel of multi-kernel learning focuses the relevant features and is immune to disturbance from irrelevant features by automatically adjusting the weight of kernels so that the stable and increasing accuracy is obtained. While the single feature fails to map non-linear separable appearance to the appropriate feature space, and is very sensitive to the noise feature disturbance, causing the accuracy very unstable and even worse with the increasing complex appearance disturbance.

With this enhanced discrimination from MKL, our regression segmentation framework has achieved better accuracy with higher mean value and smaller variance, as shown in Fig. 13(b) where a box and whisker plot of dice is used to show the distribution of segmentation accuracy in 152 clinical subjects (1218 images in total).

4.6. Method comparison

Table 5 demonstrates great improvement in accuracy of segmentation (at least 0.25 Dice increase) has been produced by the proposed

<table>
<thead>
<tr>
<th>Image modality</th>
<th>No. of experimental images</th>
<th>DSI</th>
<th>BD (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT</td>
<td>306</td>
<td>0.9005 ±</td>
<td>0.6393 ±</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0156</td>
<td>0.1308</td>
</tr>
<tr>
<td>MR</td>
<td>912</td>
<td>0.8984 ±</td>
<td>0.6586 ±</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0167</td>
<td>0.1804</td>
</tr>
<tr>
<td>MR+CT</td>
<td>1218</td>
<td>0.8935 ±</td>
<td>0.6881 ±</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0201</td>
<td>0.1849</td>
</tr>
</tbody>
</table>
approach compared to conventional segmentation methods including graph cuts [13,35], pixel classification [14], active contours [18,19], and Mumford–Shah model [36,37]. These conventional methods have produced great success in many applications. However, the high complexities in neural foramina images bring formidable challenges to these existing methods so their segmentation results are very unsatisfying with the best Dice no more than 0.76.

Fig. 14 gives a detailed illustration to demonstrate the influence of these high complexities on the performance of segmentation methods. The representative 40 images (as shown in Fig. 14(a)) are chosen to fully embody the challenges in automated segmentation of neural foramina: (1) noise intensity disturbance (i.e., neural foramina and its surroundings may share the similar intensity profile); (2) low intensity contrast (i.e., incomplete boundaries); (3) intensity inhomogeneity; (4) great diversity in boundary shape (as shown in Fig. 14(b)). These challenges lead to the infeasibility of conventional computer processing methods in our problem (as shown in Figs. 14(d)–(f) and 15).

For graph cuts, it failed (as shown in Fig. 14(d)) because the internal regions of neural foramina are highly inhomogeneous, and the unclear boundaries contain many gaps. These strong but noisy local minima severely disturb the optimal solution of graph cut.

For pixel classification method, it failed (as shown in Fig. 14(e)) due to the serious noise problem misclassified pixels, such as the similar intensities with the neighboring structures, and the highly inhomogeneous intensities for the internal regions.

For active contours, it failed (as shown in Fig. 14(f)) due to neural foramina with great shape variations and various missing boundary's parts. Specifically, the great variations in neural foramina's boundary shape bring topological changes of shape which are not allowed by active contour model, meanwhile, some missing parts of boundary leads to serious boundary leakage problem.

For Mumford–Shah model, even starting from two different
initializations (as shown in Fig. 15(a), (c)), it still failed (as shown in Fig. 15(b), (d)). This is because many parts of the desired boundary are missing and the image has noise, such as the similar intensity with the neighboring structures and the highly inhomogeneous internal region.

In comparison, our proposed regression segmentation is far superior to these existing methods (see Fig. 14(c)), and its success is derived from the following three aspects: (1) it preserves the great diversity in boundaries by introducing MSVR which simultaneously regresses the locations of all boundary points for the input image; (2) it overcomes the low intensity contrast disturbance by statically learning the relationship from the whole image to the location of each boundary point; (3) it handles the noise intensity disturbance and inter-modality intensity difference by introducing MKL which learnt a sparse discriminative image feature from $m$ multiple image features for capturing the full context of a boundary point.

5. Conclusion

Automated segmentation and area estimation of neural foramina are necessary steps in clinical diagnosis of neural foramina stenosis. Existing clinical methods, relying on the physicians’ purely manual segmentation, suffer from the unbearable tediousness, laboriousness, and inefficiency. Automated segmentation is the desirable solution but faces big challenges from diverse boundary, local weak/no boundary, intensity inhomogeneity, and completely different inter-modality intensity profiles. This paper presented a novel automated segmentation framework based on a new boundary regression model. It provides a novel way to solve the segmentation problem from a totally new perspective, and is therefore able to tackle the large diversity among spinal images. This novel framework is able to segment neural foramina images with different modalities, appearance, and boundary shapes with a powerful boundary regression model learnt by a newly proposed multiple kernel multiple output support vector regressor. By leveraging the strength of multiple kernel learning, the high complexity in images are successfully overcome; by leveraging the strength of multiple output support vector regression, the desired segmentation is directly obtained; by leveraging the strength of regression learning, the gap in weak/no boundary is connected.

The experimental results demonstrated that the proposed segmentation framework enables an accurate, robust, and automated segmen-

<table>
<thead>
<tr>
<th>Method</th>
<th>MR</th>
<th>CT</th>
<th>MR+CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph cuts</td>
<td>0.5666±</td>
<td>0.5738±</td>
<td>0.4468±</td>
</tr>
<tr>
<td></td>
<td>0.1252</td>
<td>0.0794</td>
<td>0.2972</td>
</tr>
<tr>
<td>Pixel classification</td>
<td>0.4082±</td>
<td>0.3631±</td>
<td>0.3877±</td>
</tr>
<tr>
<td></td>
<td>0.1511</td>
<td>0.2055</td>
<td>0.1708</td>
</tr>
<tr>
<td>Active contours</td>
<td>0.6462±</td>
<td>0.7513±</td>
<td>0.6726±</td>
</tr>
<tr>
<td></td>
<td>0.1706</td>
<td>0.0682</td>
<td>0.1582</td>
</tr>
<tr>
<td>Mumford–Shah model</td>
<td>0.5552±</td>
<td>0.7459±</td>
<td>0.6031±</td>
</tr>
<tr>
<td>(initialization1)</td>
<td>0.1774</td>
<td>0.0953</td>
<td>0.1807</td>
</tr>
<tr>
<td>Mumford–Shah model</td>
<td>0.6654±</td>
<td>0.7505±</td>
<td>0.6870±</td>
</tr>
<tr>
<td>(initialization2)</td>
<td>0.1293</td>
<td>0.0697</td>
<td>0.1227</td>
</tr>
<tr>
<td>Regression segmentation</td>
<td>0.8984±</td>
<td>0.9005±</td>
<td>0.8935±</td>
</tr>
<tr>
<td></td>
<td>0.0167</td>
<td>0.0156</td>
<td>0.0201</td>
</tr>
</tbody>
</table>

Fig. 14. Illustration of the difference in segmentation performance from the capability of handling the large complexities in neural foramina images. (a) 40 images with noise intensity disturbance, local weak/no boundary (indicated by yellow dot line), different inter-modality intensity profiles; (b) great diversity in their boundaries; (c) results of our regression segmentation; (d) results of graph cut; (e) results of classification based method; (f) results of active contour. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)
tation of neural foramina. Hence, with our framework, an efficient clinical tool is provided in clinical practice for freeing physicians from the tediousness and speeding up the treatment process for patients.

References


