Multi-task Shape Regression for Medical Image Segmentation

Xiantong Zhen[1,2](✉), Yilong Yin[3], Mousumi Bhaduri[4], Ilanit Ben Nachum[1,2], David Laidley[4], and Shuo Li[1,2]

[1] Digital Imaging Group (DIG), London, ON, Canada
[2] The University of Western Ontario, London, ON, Canada
xzhen7@uwo.ca
[3] Shandong University, Shandong, China
[4] London Health Sciences Centre, London, ON, Canada

Abstract. In this paper, we propose a general segmentation framework of *Multi-Task Shape Regression* (MTSR) which formulates segmentation as multi-task learning to leverage its strength of jointly solving multiple tasks enhanced by capturing task correlations. The MTSR entirely estimates coordinates of all points on shape contours by multi-task regression, where estimation of each coordinate corresponds to a regression task; the MTSR can jointly handle nonlinear relationships between image appearance and shapes while capturing holistic shape information by encoding coordinate correlations, which enables estimation of highly variable shapes, even with vague edge or region inhomogeneity. The MTSR achieves a long-desired general framework without relying on any specific assumptions or initialization, which enables flexible and fully automatic segmentation of multiple objects simultaneously, for different applications irrespective of modalities. The MTSR is validated on six representative applications of diverse images, achieves consistently high performance with dice similarity coefficient (DSC) up to 0.93 and largely outperforms state of the arts in each application, which demonstrates its effectiveness and generality for medical image segmentation.

1 Introduction

Segmentation plays a fundamental role in medical image analysis, which however has long been regarded as a challenging task due to great diversity of applications in multiple modalities, huge appearance variations of images with multiple objects and high shape variabilities of objects with complex anatomical structures. However, conventional segmentation methods cannot handle all these challenges in one framework due to the lack of generality, which are usually application specific in a certain modality, designed for images with one single object and cannot segment multiple objects simultaneously. Moreover, they usually need initialization and rely on the assumption that shape contours are supported by clear edges and region homogeneity [1], which does not always hold due to overlapping of anatomical structures, complex image textures and appearances, especially with the presence of pathology.

S. Ourselin et al. (Eds.): MICCAI 2016, Part III, LNCS 9902, pp. 210–218, 2016.
DOI: 10.1007/978-3-319-46726-9_25

Shape regression has recently shown great effectiveness for medical image segmentation with significantly better performance than conventional methods [2]. The central idea is to directly estimate point coordinates on shape contours by regression, that is, to find a nonlinear regressor to associate nonrigid shape with image appearance. Compared to traditional segmentation methods, shape regression enables leveraging the advanced machine learning techniques to extract knowledge from annotated database to tackle huge shape variabilities. Moreover, shape regression removes manual interaction and initialization in conventional segmentation methods, and is computationally more efficient. However, it remains unaddressed to explicitly model coordinate correlations resulting in shapes with outlier points falling off the contour, especially with vague edge or region inhomogeneity, which leaves a theoretical deficiency to be a general model for shape regression.

A general shape regression is desired to jointly model inherent correlations among point coordinates and highly complicated relationships between image appearance and variable shapes. The points on the shape contour are spatially coherent and statistically correlated, which should be explored to capture holistic shape information to recover contours not supported by edges or region homogeneity for robust shape regression; meanwhile, the relationship between image appearance and the associated shape is highly complex and nonlinear due to great variations of image appearance and huge shape variabilities of objects, which cannot be handled by linear regression models.

In this paper, to tackle these challenges, we propose a general segmentation framework of *Multi-Task Shape Regression* (MTSR) by formulating segmentation as multi-task learning, which leverages its strength of jointly solving multiple tasks, i.e., estimating point coordinates directly and simultaneously, while enhanced by modeling their relationships, i.e., coordinate correlations. By incorporating a latent space associated with a structure matrix, the MTSR is able to simultaneously encode holistic shape information by modeling coordinate correlations via sparse learning and disentangle nonlinear relationships between image appearance and variable shapes by kernel regression.

The major contribution of this work lies in that we for the first time achieve a general segmentation framework of shape regression by multi-task learning. Compared to previous methods, the MTSR offers multiple advantages.

- By formulating as a multi-task learning problem, the MTSR achieves a general shape regression framework without relying on specific assumptions and initialization, which enables segmentation of images with multiple objects from diverse applications irrespective of imaging modalities.
- By explicitly modeling correlations between coordinates via sparse learning, the MTSR can capture holistic shape information by reliably and accurately estimating each coordinate, which enables to recover contours not supported by edges and region homogeneity.
- By seamlessly working with the kernel trick, the MTSR achieves nonlinear regression, which enables disentangling highly complicated relationships between image appearance of great variations and shapes of huge variabilities.

2 Multi-task Shape Regression

The proposed MTSR formulates segmentation as multi-task regression to directly and simultaneously estimates the coordinates of the points on shape contours, where estimation of each coordinate is a regression task. By incorporating a latent space, the MTSR explicitly models coordinate correlations in a structure matrix S via the $\ell_{2,1}$-norm based sparse learning to capture the holistic information of shapes (Sect. 2.2); by kernelization, the MTSR achieves kernel regression to effectively tackle complicated nonlinear relationships between image appearance and variable shapes (Sect. 2.3).

2.1 Shape Regression by Multi-task Regression

Shape regression is to directly estimate point coordinates on the shape contour of the object by regression from input images, where the shape is represented by $\mathbf{y} = [h_1, \ldots, h_i, \ldots, h_{\mathcal{P}}, v_1, \ldots, v_i, \ldots, v_{\mathcal{P}}]^\top \in \mathbb{R}^Q$, $Q(= 2\mathcal{P})$ is the number of coordinates of the $\mathcal{P}$ points on the shape contour, h_i and v_i are the horizontal and vertical axises of the i-the point, respectively. The associated image is represented by a feature descriptor $\mathbf{x} \in \mathbb{R}^d$, e.g., the histogram of oriented gradient (HOG) [3], where d is the dimensionality.

The proposed MTSR realizes shape regression by multi-task regression to directly and simultaneously estimates coordinates while jointly capturing coordinate correlations to capture holistic shape information; and it is derived based on the widely-used fundamental multi-task regression model $\mathbf{y} = W\mathbf{x} + \mathbf{b}$, where $W = [\mathbf{w}_1, \ldots, \mathbf{w}_i, \ldots, \mathbf{w}_Q]^\top \in \mathbb{R}^{Q\times d}$ is the regression coefficient, $\mathbf{w}_i \in \mathbb{R}^d$ is the task parameter for y_i, and $\mathbf{b} \in \mathbb{R}^Q$ is the bias.

2.2 Modeling Correlation by Sparse Learning

We propose modeling the inherent correlations between point coordinates by sparse learning to learn shared features for correlated point coordinates. Since the points on the shape contour are spatially coherent and statistically correlated, encoding correlations enables to capture the holistic shape information to recover contours not supported by edges or region homogeneity.

Rather than directly imposing regularization on regression matrix W as widely explored in existing multi-task learning algorithms, we propose incorporating a latent space. On top of the latent variables, a structure matrix S is deployed to explicitly model coordinate correlations via the $\ell_{2,1}$-norm based sparse learning. Based on the least square loss function and ℓ_2 regularization, we have the following objective function w.r.t. W and S

$$\begin{aligned} &\min_{W,S} \frac{1}{\mathcal{N}} ||Y - SZ||_F^2 + \lambda ||W||_F^2 + \beta ||S^\top||_{2,1}, \\ &s.t. \;\; Z = WX \end{aligned} \tag{1}$$

where $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{\mathcal{N}}]$, $Y = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{\mathcal{N}}]$, $Z = [\mathbf{z}_1, \ldots, \mathbf{z}_i, \ldots, \mathbf{z}_{\mathcal{N}}] \in \mathbb{R}^{Q\times\mathcal{N}}$, $\mathbf{z}_i \in \mathbb{R}^Q$ is a variable in the latent space and $S \in \mathbb{R}^{Q\times Q}$ is named as the structure matrix; the $\ell_{2,1}$-norm constraint on the structure matrix S in (1) encourages to learn an S that is prone to be column sparsity [4]; the parameter β controls the column sparsity of S and a larger β induces higher sparsity; the bias is omitted since it is proven that the bias can be absorbed into the regression coefficient W [4].

Thanks to the $\ell_{2,1}$-norm based sparse learning, regressors of correlated coordinates on the shape contour are encouraged to share similar parameter sparsity patterns to capture a common set of features from the latent space, which enables to encode the inherent correlations between coordinates. Therefore, holistic shape information is effectively captured, which enables to recover the coordinates that are not supported by clear edges or region homogeneity to achieve more accurate and robust shape estimation. Moreover, the performance of all regressors can be improved by leveraging knowledge across correlated coordinates that share common features. By deploying a structure matrix S to explicitly model coordinate correlations, our MTSR can automatically learn the inherence of coordinates on the shape contour from data to cater different applications, which further improves the generality.

We highlight that due to the incorporation of the latent space associated with the structure matrix S, the proposed MTSR brings multiple attractive merits:

- The latent space decouples inputs and outputs with W and S, which enables effectively handling high image appearance variations and huge shape variability to disentangle their complex relationships.
- The structure matrix allows to explicitly encode inherent correlations between coordinates to capture the holistic shape information, which enables recovering contours not supported by clear edges or region homogeneity.

Due to the huge appearance variations of images and the high shape variabilities of objects to be segmented, the relationship between image appearance and variable shapes is complicated and highly nonlinear, which cannot be handled by linear regression and demands more powerful nonlinear regressors.

2.3 Kernelization for Nonlinear Regression

Although the objective function (1) is not guaranteed to be jointly convex with W and S, it is easy to show that kernelization can be derived with respect to W with a fixed S thanks to the incorporation of the latent space according to the Representer Theorem [5]. This enables kernel regression to handle nonlinear relationship between image appearance and shapes, while being able to encode coordinate correlations by S.

The linear representer theorem [5] is particularly useful when $\mathcal{H}$ is a reproducing kernel Hilbert space (RKHS), which simplifies the empirical risk minimization problem from an infinite dimensional to a finite dimensional optimization problem [5]. Assume that we map $\mathbf{x}_i$ to $\phi(\mathbf{x}_i)$ in some RKHS of infinite dimensionality where $\phi(\cdot)$ denotes the feature map of $\mathbf{x}_i$; the mapping serves as a nonlinear feature extraction that enables to disentangle complicated relationships between image appearance and variable shapes. The corresponding kernel

Table 1. The statistics of the six datasets.

Dataset \ Information	Task	Subjects	Images	Modalities
SKI12 [7]	Knee	20	1438	MR
CRASS12 [8]	Clavicle	20	548	CT
PROMISE12 [9]	Prostate	50	778	MR
Cardiac Bi-Ventricles (CBV) [10]	LV/RV	145	8700	MR
Cardiac 4 Chambers (C4C-MR) [6]	LV/LA/RV/RA	125	3125	MR
Cardiac 4 Chambers (C4C-CT)	LV/LA/RV/RA	101	3920	CT

function $k(\cdot,\cdot)$ satisfies $k(\mathbf{x}_i,\mathbf{x}_j)=\phi(\mathbf{x}_i)^\top\phi(\mathbf{x}_j)$. To facilitate the derivation of kernelization, we rewrite (1) in term of traces as follows:

$$\min_{W,S}\frac{1}{\mathcal{N}}tr((Y-SWX)^\top(Y-SWX))+\lambda tr(W^\top W)+\beta||S^\top||_{2,1}. \tag{2}$$

According to the linear representer theorem, we can obtain W by

$$W=\boldsymbol{\alpha}\Phi(X)^\top \tag{3}$$

where $\Phi(X)=[\phi(\mathbf{x}_1),\ldots,\phi(\mathbf{x}_i),\ldots,\phi(\mathbf{x}_\mathcal{N})]$ and $\boldsymbol{\alpha}\in\mathbb{R}^{Q\times\mathcal{N}}$. Substituting (3) into (2) gives rise to the objective function w.r.t $\boldsymbol{\alpha}$ and S:

$$\min_{\boldsymbol{\alpha},S}\frac{1}{\mathcal{N}}tr((Y-S\boldsymbol{\alpha}K)^\top(Y-S\boldsymbol{\alpha}K))+\lambda tr(\boldsymbol{\alpha}K\boldsymbol{\alpha}^\top)+\beta||S^\top||_{2,1} \tag{4}$$

where $K=\Phi(X)^\top\Phi(X)$ is the kernel matrix in the RKHS. The latent space spanned by $Z=\boldsymbol{\alpha}K$ is obtained by the linear transformation $\boldsymbol{\alpha}$ via the Representer Theorem from the KRHS induced by a nonlinear kernel K. As a result, higher-level concepts are extracted to fill the semantic gap between image representations of low-level feature descriptors and variable shapes, which enables efficient linear $\ell_{2,1}$-based sparse learning of S to explicitly model inherent correlations of coordinates to capture the holistic shape information. The objective function in (4) is non-convex jointly with $\boldsymbol{\alpha}$ and S, which fortunately can be efficiently solved by alternating optimization [6].

2.4 Training and Prediction

In the training stage, the MTSR is trained on annotated data with ground truth of contours; in the prediction stage, given a new input $\mathbf{x}_t$, the point coordinates on the shape contour of objects to be segmented can efficiently be predicted by $\hat{\mathbf{y}}_t=S\boldsymbol{\alpha}K_t$, where $K_t=\Phi(X)^\top\phi(\mathbf{x}_t)$. Segmentation is then obtained based on the predicted shape contours.

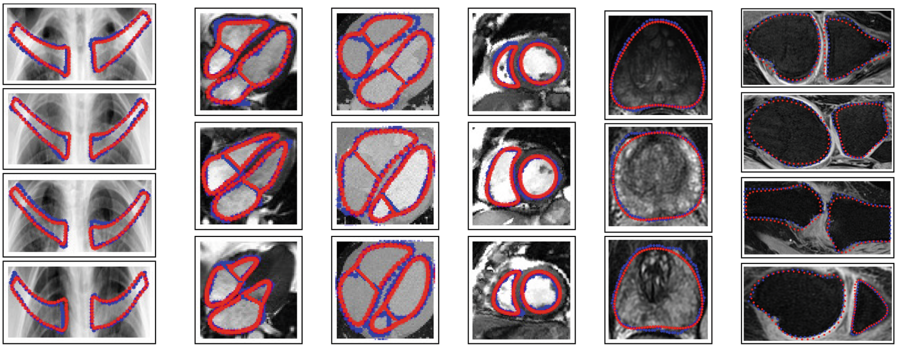

Fig. 1. Segmentation results: (from left to right) clavicle, cardiac four chambers MR/CT, bi-ventricles, prostate, knee (red: MTSR and blue: ground truth). (Color figure online)

3 Experiments

Our method has been validated by extensive experiments on six representative applications for clinical image segmentation and achieves high performance on all six applications with a dice similarity coefficient (DSC) [11] up to 0.93 and consistently outperforms state of the arts in each application.

3.1 Datasets and Implementation Details

The six datasets contain three public datasets from Grand Challenges for prostate, knee and clavicle, and three newly collected cardiac MR/CT datasets for bi-ventricles and four chambers. The statistics of the six datasets are reported in Table 1. Following [2], contours are represented by the coordinates of a set of $\mathcal{P}$ (=100) points on the shape of an object $\{p_i = (h_i, v_i)|_{i=1:\mathcal{P}}\}$ sampled evenly along the manually annotated contour from a fixed point. In cardiac four chambers, four landmark points are indicated during manual annotation to segment atrium and ventricle, respectively. The histogram of oriented gradient (HOG) descriptor [3] is used due to its computational efficiency for image representation. To benchmark with existing methods, we measure the performance using the DSC [7,9,11] obtained using leave-one-subject-out cross validation. We compare the two shape regression models fulfilled by the MSVR [2] and adaptive k-cluster regression forests (AKRF) [12]. The multi-target kernel ridge regression (mKRR) is regarded as a baseline of multi-task learning for comparison. The parameter β is obtained by cross validation on the training set.

3.2 Results

The MTSR yields high performance in comparison to ground truth on all six applications, which demonstrates its effectiveness and generality for medical

Table 2. The dice similarity coefficients (DSC) for six different applications.

Method \ Task	Clavicle	Prostate	Knee	C4C-MR	C4C-CT	CBV
MTSR	**0.857**	**0.930**	**0.894**	**0.885**	**0.886**	**0.892**
mKRR	0.829	0.906	0.861	0.826	0.824	0.843
mSVR [2]	0.851	0.924	0.889	0.849	0.836	0.868
AKRF [12]	0.842	0.921	0.879	0.847	0.841	0.869
State of Arts	0.80 [8]	0.89 [9]	0.857 [11]	-	-	-

image segmentation. We show qualitative results in Fig. 1 and the quantitative comparison to state-of-the-art algorithms are reported in Table 2.

Effectiveness. The effectiveness of the MTSR is shown by overcoming great image variations and huge shape variabilities caused by region inhomogeneity, vague edges, illumination change, low intensity contrast and spatial/temporal nonrigid deformations. As shown in Fig. 1, *Clavicle:* the medial part of the clavicle is heavily obscured by other anatomical structures such as the mediastinum and large vessels [8], and both femoral and tibial cartilages demonstrate high anatomical shape variations; *Knee:* The contours are discontinuous and not consistently visible due to the poor intensity contrast, and shapes vary greatly across spatial slices. *Prostate:* The contours are not supported by region intensity homogeneity because of the complex texture and illumination. *Cardiac:* Both four chambers and bi-ventricles demonstrate high variabilities caused by the large spatial and temporal deformations. However, as shown in Fig. 1 the MTSR produces contours (red) very close to ground truth (blue) on all six applications, which demonstrates its effectiveness of jointly modeling coordinate correlation and nonlinear relationship between image appearance and variable shapes.

Generality. The generality of the MTSR is shown by conquering the diversity of images with multiple objects, from a large variety of applications and in multiple imaging modalities. These images from six tasks cover a broad range of medical applications, contains varied numbers of objects from one prostate to four chambers, and are obtained in multiple modalities, i.e., MT and CT, both of which are widely used in clinical routines. However, the MTSR is able to consistently and successfully produce accurate shape contours with high performance up to 0.93 of DSC on all six applications as shown in Table 2, which validates its generality for medical image segmentation by shape regression, indicating its great potential in clinical use.

Comparison. As shown in Table 2, the MTSR achieves consistently higher performance on all applications, and substantially outperforms state-of-the-art methods with large margins up to 4.5 %. The MTSR performs much better than the MSVR/AKRF and the baseline mKRR on all six tasks, which demonstrates the strength of modeling coordinate correlations by the proposed $\ell_{2,1}$-norm based sparse learning.

4 Conclusion

In this paper, we proposed a general segmentation method, multi-target shape regression (MTSR), which formulates the segmentation of shapes as a multi-task learning problem. The MTSR is able to simultaneously capture the holistic shape information and handle highly nonlinear relationships between image appearance and variable shapes in one single framework, which enables more accurate and reliable shape regression for image segmentation with multiple varied numbers of objects, irrespective of modalities. Experiments on six diverse segmentation tasks show that the MTSR achieves consistently high performance and significantly outperforms state-of-the-art algorithms, which demonstrates its effectiveness and generality for medical image segmentation.

Acknowledgement. This work was supported by the NSFC Joint Fund with Guangdong under Key Project (Grant No. U1201258) and NSFC (Grant No. 61571147).

References

1. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. TPAMI **23**(6), 681–685 (2001)
2. Wang, Z., Zhen, X., Tay, K., Osman, S., Romano, W., Li, S.: Regression segmentation for M^3 spinal images. TMI **34**(8), 1640–1648 (2015)
3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR, vol. 1, pp. 886–893 (2005)
4. Nie, F., Huang, H., Cai, X., Ding, C.H.: Efficient, robust feature selection via joint $\ell_{2,1}$-norms minimization. In: NIPS, pp. 1813–1821 (2010)
5. Kimeldorf, G.S., Wahba, G.: A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. Ann. Math. Stat. **41**(2), 495–502 (1970)
6. Zhen, X., Islam, A., Bhaduri, M., Chan, I., Li, S.: Direct and simultaneous four-chamber volume estimation by multi-output regression. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 669–676. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24553-9_82
7. Heimann, T., Morrison, B.J., Styner, M.A., Niethammer, M., Warfield, S.: Segmentation of knee images: a grand challenge. In: Proceedings of MICCAI Workshop on Medical Image Analysis for the Clinic, pp. 207–214 (2010)
8. Hogeweg, L., Sánchez, C.I., de Jong, P.A., Maduskar, P., van Ginneken, B.: Clavicle segmentation in chest radiographs. Med. Image Anal. **16**(8), 1490–1502 (2012)
9. Litjens, G., Toth, R., van de Ven, W., Hoeks, C., Kerkstra, S., van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al.: Evaluation of prostate segmentation algorithms for MRI: the promise12 challenge. Med. Image Anal. **18**(2), 359–373 (2014)
10. Zhen, X., Wang, Z., Islam, A., Bhaduri, M., Chan, I., Li, S.: Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation. Med. Image Anal. **30**, 120–129 (2016)

11. Shan, L., Zach, C., Charles, C., Niethammer, M.: Automatic atlas-based three-label cartilage segmentation from MR knee images. Med. Image Anal. **18**(7), 1233–1246 (2014)
12. Hara, K., Chellappa, R.: Growing regression forests by classification: applications to object pose estimation. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part II. LNCS, vol. 8690, pp. 552–567. Springer, Heidelberg (2014)