IMPLICIT KERNEL PRESENTATION AWARE OBJECT SEGMENTATION FRAMEWORK

Jincao Yao[1], *Huimin Yu*[1,2], *Roland Hu*[1]

[1]*Department of Information Science and Electronic Engineering, Zhejiang University, China*
[2]*State Key Laboratory of CAD & CG, Zhejiang University, China*
yaojincao@zju.edu.cn, yhm2005@zju.edu.cn, haoji_hu@zju.edu.cn

ABSTRACT

Given a set of training shapes and an input image with a shape similar to some of the elements in the training set, this paper introduces a new implicit kernel sparse model with a twofold goal. First, to obtain an implicit kernel sparse neighbor based combination that best represents the object. Second, to accurately segment the object taking into accounts both the high-level implicit kernel presentation and the low-level image information. A new energy function that combines the variational image segmentation with the implicit kernel presentation is introduced to accomplish both goals simultaneously. The experimental results on the public datasets show the superior capabilities of the proposed model.

Index Terms— Object segmentation, sparse representation, kernel method, shape prior

1. INTRODUCTION

As for an input object of which the shape has been partly contaminated or damaged, how to use the prior knowledge extracted from its shape neighbors to recover the original shape of the object is a challenging task [1-5]. Recently, several works tried to treat this problem as a sparse representation based segmentation process [6, 7], such as the sparse coding based segmentation method in [8], the Sparse Shape Composition (SSC) in [9, 10]. Our previous work also proposed a Probabilistic based Shape Sparse Representation (P-SSR) to solve this problem [11]. However, the models above either required the training set follow a certain distribution or used an explicit linear projection to build the model [8, 9, 11], and these requirements limited the shape representation ability. Considering that the kernel method has been proven to have a strong ability to deal with the complex shape statistic [1, 12], as a follow-up of [11], we proposed a new implicit kernel sparse neighbors based object segmentation models in this paper.

The contribution of this paper lies in two-fold. 1) a novel Implicit Kernel Sparse Representation (IKSR) model was proposed in this paper. We proved that the model was equivalent to a reconstruction error constrained sparse shape representation in the Hilbert space. Thus, it provided a new way to solve the problem of the sparse representation in the non-linear shape space; 2) we formulated a new segmentation energy function based on the proposed IKSR model. The energy minimization could activate the competition between the image based energy and high-level IKSR energy, and this competition eventually drove the model to segment the object with consideration of both the image information and the high-level presentation in the kernel space.

2. BACKGROUND

Suppose there is a training shape set $Q=[q_1,q_2,\cdots,q_N]\in I$, N is the number of the samples. A kernel method tries to define a nonlinear map: $\varphi: I\to F$ and a Mercer kernel function $k(\cdot,\cdot)$, such that $F: k(q_i,q_j)=(\varphi(q_i)^T\cdot\varphi(q_j))$, F is a Hilbert space [12]. There are many available kernel functions can be used, without loss of generality, we use a Gaussian kernel in this paper.

In real applications, a Principal Components Analysis (PCA) method is often combined with the kernel projection to make it more efficient. The main idea is to find the corresponding eigenvectors $V\in F\setminus\{0\}$ satisfying $\lambda V=CV$, λ is the eigenvalue, and C is the covariance matrix. In numerical calculation, a matrix K with $K_{ij}=(\varphi(q_i)^T\cdot\varphi(q_j))$ is often defined to change the eigenfunction above into $N\lambda\alpha=K\alpha$, because solving λ and α are mach easier. To represent $\varphi(\cdot)$, we first need to compute the coefficient m_k^q

$$m_k^q=\sum_{i=1}^N\alpha_i^k k(q_i,q) \quad (1)$$

Where $\alpha^k=[\alpha_1^k,\cdots,\alpha_N^k]^T$, $\alpha=[\alpha^1,\cdots,\alpha^n]$ is the reconstruct weight, and n is the number of the vector. Once we obtain the coefficient $m^q=[m_1^q,m_2^q,\cdots,m_n^q]^T$, the projected $\varphi_P(q)$ can be represented as

$$\varphi_P(q) = \sum_{i=1}^{n} m_i^q V^i \quad (2)$$

Here we assumed the projected shapes have been centralized. For the uncentralized shape, one can easily use a general K to accomplish the centralization [12].

3. THE IMPLICIT KERNEL SPARSE MODEL

The existing kernel space based methods mainly use the explicit distance $\|\varphi_P(q)-\varphi(q)\|$ to evaluate the overall error in the kernel space to guide the segmentation [1, 12]. However, it is difficult for this method to recover a cluster of $\varphi(\cdot)$ and thus form a meaningful combination to estimate the original shape. Hence, in this paper, we formulate the model in a different way.

Firstly, we observe the equation (1). Traditional methods used (1) and an input shape q to deduce the corresponding vector m^q [1, 12]. However, in this paper, we try to understand (1) in a reverse manner. We assume that the input q is unknown, and treat the vector m as the input, and then explore the property of the corresponding output. By constraining the input m in a different from, we can obtain several interesting propositions.

Proposition 1. *Assuming that* $M=[m^{q_1}, m^{q_2}, \cdots m^{q_N}]$ *with* $m^{q_i}=[m_1^{q_i},\cdots,m_n^{q_i}]^T$ *represents a set of kernel projected coefficient that is corresponding to the given training shape set* $Q=[q_1,q_2,\cdots,q_N]$. *Let* $s=[s_1,s_2,\cdots s_N]^T \in \mathbb{R}^N$ *be a sparse coefficient,* $(Ms)_i$ *represents the i-th element of sparse combination* Ms, *when set the input vector* $m=Ms$ *be an arbitrary sparse combination, and take it into the equation (2), and then we can define a shape set* ζ *as follows.*

$$\zeta = \left\{ \eta_P = \sum_{i=1}^{n} (Ms)_i V^i \mid s \in \mathbb{R}^N \right\} \quad (3)$$

This kernel space based shape set is a convex set.

Proof : given a $\gamma \in [0,1]$, for two variational shapes $\eta_P^1 = \sum_{i=1}^{n} (Ms^1)_i V^i \in \zeta$ and $\eta_P^2 = \sum_{i=1}^{n} (Ms^2)_i V^i \in \zeta$, we have

$$\gamma\eta_P^1 + (1-\gamma)\eta_P^2 = \sum_{i=1}^{n} \left(M\left(\gamma\left(s^1-s^2\right)+s^2\right)\right) V^i \quad (4)$$

Let $\hat{s}=\gamma\left(s^1-s^2\right)+s^2$, considering that $s^i \in \mathbb{R}^N, i=1,2$,then $\hat{s} \in \mathbb{R}^N$ and $\sum_{i=1}^{n}(M\hat{s})_i V^i \in \zeta$, thus ζ is a convex set . □

In Proposition 1, we replace the arbitrary coefficient m by a constrained sparse combination Ms. According to (3), each column vector in M is corresponding to a training sample. Therefore, for an input shape that is similar to some of the training samples, we can search a specific sparse combination with a corresponding projected shape element η_P in the set ζ, which can approximately recover the input. Therefore, based on the Proposition 1, we can deduce Proposition 2.

Proposition 2. *Assuming that there is an input shape q similar to some of the elements of the given training set* $Q=[q_1,q_2,\cdots,q_N]$. *Let* $k(\cdot,\cdot)$ *be a Mercer kernel function,* $\varphi(\cdot)$ *represents the kernel projection operator,* $s=[s_1,s_2,\cdots s_N]^T$ *represents the sparse coefficient, and* $k^q=[k(q_1,q),\cdots k(q_N,q)]^T$. *Then the minimization of the following energy function*

$$E_{IKSR}(s) = \left\| Ms - \alpha^T k^q \right\|_2^2 + \|s\|_1 \quad (5)$$

is equivalent to represent the projected input $\varphi(q)$ *with* $\sum_{j=1}^{N}\varphi(q_j)s_j$ *under the constraint of minimal sparse reconstruction error.*

Proof:

$$E(s) = \left\| Ms - \alpha^T k^q \right\|_2^2 + \|s\|_1$$

$$= \left\| \begin{bmatrix} m_1^{q_1} & \cdots & m_1^{q_N} \\ \vdots & \ddots & \vdots \\ m_n^{q_1} & \cdots & m_n^{q_N} \end{bmatrix} \begin{bmatrix} s_1 \\ \vdots \\ s_N \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^{N}\alpha_i^1 k(q_i,q) \\ \vdots \\ \sum_{i=1}^{N}\alpha_i^n k(q_i,q) \end{bmatrix} \right\|_2^2 + \|s\|_1$$

$$= \left\| \begin{bmatrix} \sum_{i=1}^{N}\alpha_i^1 k(q_i,q_1) & \cdots & \sum_{i=1}^{N}\alpha_i^1 k(q_i,q_N) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{N}\alpha_i^n k(q_i,q_1) & \cdots & \sum_{i=1}^{N}\alpha_i^n k(q_i,q_N) \end{bmatrix} \begin{bmatrix} s_1 \\ \vdots \\ s_N \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^{N}\alpha_i^1 k(q_i,q) \\ \vdots \\ \sum_{i=1}^{N}\alpha_i^n k(q_i,q) \end{bmatrix} \right\|_2^2 + \|s\|_1$$

$$= \left\| \begin{bmatrix} \sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_i^1 k(q_i,q_j)s_j \\ \vdots \\ \sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_i^n k(q_i,q_j)s_j \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^{N}\alpha_i^1 k(q_i,q) \\ \vdots \\ \sum_{i=1}^{N}\alpha_i^n k(q_i,q) \end{bmatrix} \right\|_2^2 + \|s\|_1$$

According to the Mercer Theorem[18], $k(q_i,q_j)=\varphi(q_i)^T\varphi(q_j)$

$$E(s) = \left\| \begin{bmatrix} \sum_{i=1}^{N}\alpha_i^1 \varphi(q_i)^T \left(\sum_{j=1}^{N}\varphi(q_j)s_j - \varphi(q)\right) \\ \vdots \\ \sum_{i=1}^{N}\alpha_i^n \varphi(q_i)^T \left(\sum_{j=1}^{N}\varphi(q_j)s_j - \varphi(q)\right) \end{bmatrix} \right\|_2^2 + \|s\|_1 \quad (6)$$

The first term of (6) contains two general sub-terms, namely $\alpha_i^1\varphi(q_i)^T$ to $\alpha_i^n\varphi(q_i)^T$ and $\sum\varphi(q_j)s_j - \varphi(q)$. Once the training set is given, the first sub-term is fixed. The minimization based on the second sub-term is obviously equivalent to represent the projected input $\varphi(q)$ with $\sum\varphi(q_j)s_j$ under the constraint of minimal sparse reconstruction error. □

Base on the propositions above, in this paper, we propose to use (5) as our high-level shape representation term to build the segmentation energy function.

Remark 1. *The* L_1 *norm sparse formulation is a convex formulation. Once we use Ms to regularize the input, we actually running the model on the convex set* ζ *implicitly, and that allows us to use an implicit kernel sparse*

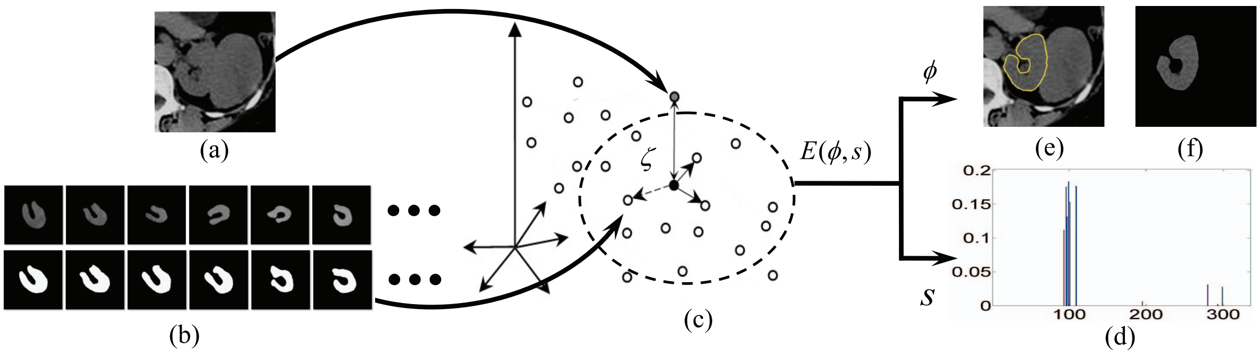


Fig. 1 (a) the input image (b) the training set, the first row is the original shapes, the second row is the normalized shapes, (c) the kernel space, the ellipse marks the convex set defined in proposition 1, (d) the recovered sparse coefficient, (e) and (f) are the extracted object.

representation to obtain a global minimization without explicitly involving the projected shape $\varphi(\cdot)$.

4 THE IMAGE BASED ENERGY

The formulations above deduce the kernel sparse shape representation. However, the input q is still unknown. To solve this problem, we rewrite q in the level-set form, and the energy becomes

$$E(s,\phi)=\left\|Ms-\alpha^{T}k^{H(\phi)}\right\|_{2}^{2}+\lambda\|s\|_{1}+\beta E_{IMAGE}(\phi) \qquad (7)$$

$H(\cdot)$ is the Heaviside function. Notice that the last term $E_{IMAGE}(\phi)$ is actually a data-driven term which causes ϕ to separate the object from the input image under the constraint of sparse representation. β is a constant parameter. There are many forms of energy that can be selected for $E_{IMAGE}(\phi)$. In this paper, we use the classical Chan and Vese (CV) model [13] as our data-driven term:

$$\begin{aligned}E_{CV}(\phi)&=\int_{\Omega}\left|I(x)-c_{+}\right|^{2}H(\phi(x))dx\\&+\int_{\Omega}\left|I(x)-c_{-}\right|^{2}\left(1-H\left(\phi(x)\right)\right)dx+\mu\int_{\Omega}\delta\left(\phi(x)\right)\left|\nabla\phi(x)\right|dx\end{aligned} \qquad (8)$$

Where, c_{+} and c_{-} are the averages of the input data inside and outside the zero level-set, respectively, and $\delta(\cdot)$ is the Dirac function. Thus the total energy includes two terms which are linearly combined with the constant β is

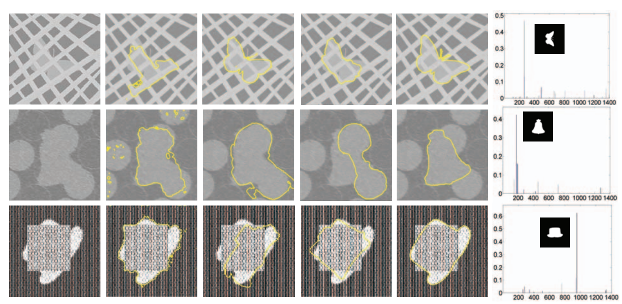
$$E(\phi,s)=E_{IKSSR}(\phi,s)+\beta E_{CV}(\phi) \qquad (9)$$

For the translation, rotation, and scaling invariance problem, we used the method in [14] to formulate ϕ as $\phi_{0}=T(\phi,\Omega)$, where $\Omega=(h,\theta,\sigma)$, h, θ and σ represent the translation vector, rotation angle and scale parameter, respectively. Minimization of s, ϕ and Ω can be easily solved by a classical alternating gradient descent scheme, considering that an alternating gradient descent scheme is the basic technique in signal processing community, we omit the details here.

Fig. 1 shows the general shape representation procedure of our method, (a) is one of the CT image from the "Open Medical Image" (OMI) dataset [15], (b) shows part of the training shapes from the OMI. Fig. 1 (c) projected the shapes into the implicit kernel space and adopted the implicit kernel neighbors to represent the object, (d) to (f) showed the recovered sparse coefficient, the curve evolution result and the extracted object, respectively.

5. EXPERIMENTAL RESULT

We tested our model on two public datasets. The experiments were running on a PC with an Intel i7 CPU. For all the experiments, we set $\beta=1$, $\lambda=1/N$ and $\mu=1.5$. First, we give some toy examples. The training set is 1400 shapes from the MPEG-7 CE-Shape-1. Fig. 2 (a) shows some synthetic shapes with different contaminations, the segmentation results of the SMs [5], the P-SSR [11] and the Explicit Shape Constrained (ESC) method in [3] are presented in (b) to (d), respectively. As is shown, once the shapes are severely damaged, the above models fail to recover the shape. Fig. 2 (e) and (f) show the segmentation



(a) (b) (c) (d) (e) (f)

Fig. 2 (a) the input images, (b) the SMs [5], (c) the P-SSR [11], (d) the ESC [3], (e) our IKSR, (f) the sparse coefficient, the embedded pictures are the original shape of the largest coefficient.

results and the recovered sparse coefficients of the proposed IKSR model. As can be seen, our model recovers the original shape even though the shape has been severely damaged. We also give the shape recovering process in the Fig. 3, and we can clearly see that the proposed function drives the evolutionary curve recovering the original shape

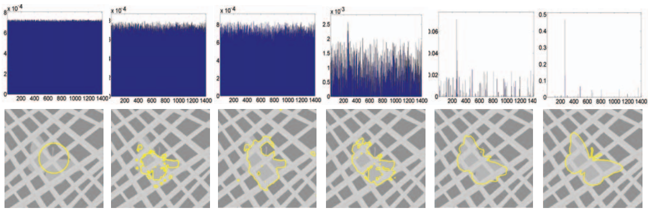
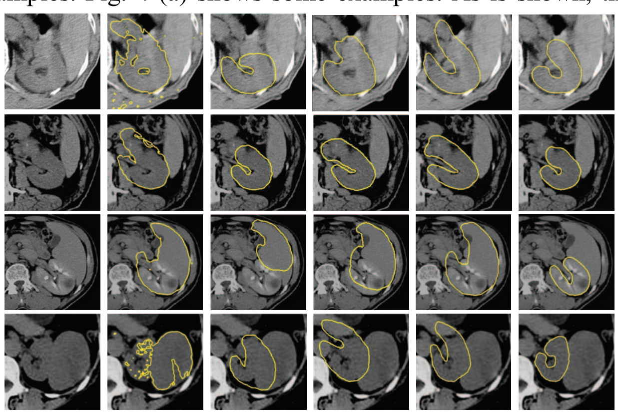


Fig. 3 The first row shows the sparse coefficients optimization process, the second row shows the curve evolution process.

of the object.

In real applications, we cannot expect the training set containing the original shape of the object, and there might be only several neighbors available. In the following experiments, the training set does not directly contain the original shape. The input images and the training set are 179 and 336 kidney CT images from the OMI dataset, receptively [15]. A k-means approach was applied to divide the training shapes into 16 clusters, and then we aligned the samples. Fig. 4 (a) shows some examples. As is shown, the noise and the connected background severely mislead the SMs [5], the P-SSR [11], the SSC [10] and the ESC [3] in (b) to (e). The models above recover some shapes that are unfaithful to the original objects (since the SSC cannot handle multiclass condition, we did not compare with it in toy example). Fig. 4 (f) is the results of our IKSR framework. As we can see, though the input shapes in the images are not directly contained in the training set, our model still successfully separates the object from the image. The original samples corresponding to the five largest coefficients were presented in the first five columns of Fig. 5. The sixth column shows recovered sparse coefficient, and the last columns is the segmentation results.



(a) (b) (c) (d) (e) (f)

Fig. 4 (a) the input images, (b) the SMs [5], (c) the P-SSR [11], (d) the SSC [10], (e) the ESC [3], (f) our IKSR.

Table 1 compares the average errors, the iteration number and the computational time of the above method. For the average errors issue, the ground truth is the manually segmentation results. The ESC method gets the least computational time and the iteration number. However, the average error of ESC method is very high. As the table shows, the segmentation error of our method is significantly better than all the other methods.

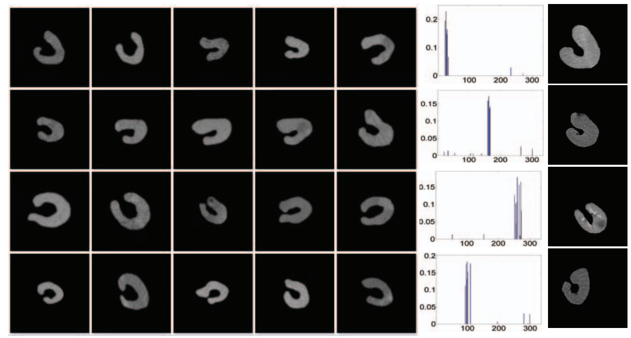


Fig. 5 The first to fifth columns are the original samples corresponding to the five largest coefficients, the sixth column is the recovered sparse coefficients, and the last column is the segmentation results of our method.

Table 1. The comparison of the average errors, computational time and iteration number of the 179 kidney samples from OMI.

	CT image segmentation		
	AVE. Time (s)	**AVE. Error (%)**	**AVE. Iteration number**
SSC	5.13	31.09	67.54
ESC	3.01	21.27	17.59
P-SSR	9.74	19.33	21.36
SMs	19.72	46.03	86.93
Our	3.53	3.10	25.33

6. CONCLUSIONS

A novel kernel sparse neighbor based object segmentation framework, called IKSR was introduced in this paper. Two propositions were given to deduce the basic IKSR model. The energy minimization drove an evolutionary curve to segment the object taking into account both the low-level information and the high-level representation. The experimental results on two public datasets showed the satisfactory segmentation performance. In future work, we plan to consider new non-linear method to further improve the performance.

7. ACKNOWLEDGEMENTS

This work is supported by the Natural Science Foundation of China (NSFC No. 61471321 and No. 61202400) and National Key Basic Research Project of China (973 Program 2012CB316400).

8. REFERENCES

[1] S. Dambreville, Y. Rathi, A. Tannenbaum, “A frame work for image segmentation using shape models and kernel space shape priors,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1385–1399 , 2008.

[2] P. Chiwoo, J.Z. Huang, J.X. Ji, Y. Ding, “Segmentation, inference and classification of partially overlapping nanoparticles,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1–10, 2013.

[3] D. C. Chittajallu, N. Paragios, I. A. Kakadiaris, “An explicit shape-constrained MRF-Based contour evolution method for 2-D medical image segmentation,” *IEEE Journal of Biomedical and Health Informatics*, vol 18, pp. 120-129, 2014.

[4] K. Li, W. Tao, “Adaptive Optimal Shape Prior for Easy Interactive Object Segmentation,” *IEEE Transactions on Multimedia*, vol. 17, pp. 994 – 1005, 2015.

[5] F. Lecumberry, A. Pardo, G. Sapiro, “Simultaneous object classification and segmentation with high-order multiple shape models,” *IEEE Transactions on Image Processing*, vol.19, pp. 625–635, 2010.

[6] S. Berretti, A. Del Bimbo, P. Pala, “Sparse matching of salient facial curves for recognition of 3-D faces with missing parts,” *IEEE Transactions on Information Forensics and Security* vol. 8, pp. 374–389, 2013.

[7] T. Tong, W. Robin, C. Pierrick, V.H. Joseph, R. Daniel, “Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling,” J. NeuroImage, vol. 76, pp.11-23, 2013.

[8] M.W. Spratling, “Image segmentation using a sparse coding model of cortical area V1,” *IEEE Transactions on Image Processing*, vol. 22, pp. 1631–1643, 2013.

[9] S.T. Zhang, Y.Q. Zhan, M. Dewan, J.Z. Huang, D.N. Metaxas, X.S. Zhou, “Sparse shape composition: A new framework for shape prior modeling,” *IEEE International Coference on Computer Vision and Pattern Recognition*, vol. 1, pp 451-458, 2011.

[10] S.T. Zhang, Y.Q. Zhan, M. Dewan, J.Z. Huang, D.N. Metaxas, X.S. Zhou, “Towards robust and effective shape modeling: sparse shape composition,” *J. Medical Image Analysis*, vol. 16, pp. 265–277, 2012.

[11] F. Chen, H.M. Yu, R. Hu, “Shape sparse representation for joint object classification and segmentation,” *IEEE Transactions on Image Processing*, vol. 22, pp.992–1004, 2013.

[12] S. Mika, B. Scholkopf, A.J. Smola, K.R. Muller, M. Scholz, and G.Ratsch, “Kernel PCA and De-Noising in Feature Spaces,” *Advances in Neural Information Processing Systems*, vol. 11, MIT Press, 1999.

[13] T.F. Chan, L.A. Vese, “Active contours without edges,” *IEEE Trans. on Image Processing,* vol 10, pp. 266–277, 2001.

[14] T. Chan, W. Zhu, “Level set based shape prior segmentation,” *IEEE International Coference on Computer Vision and Pattern Recognition*, vol.2, pp. 1154-1170, 2005.

[15] www.openmedimg.com