Landmark Detection in Cardiac MRI Using Learned Local Image Statistics

Dwarikanath Mahapatra⋆

Department of Computer Science, ETH Zurich, Switzerland
`dwarikanath.mahapatra@inf.ethz.ch`

Abstract. We propose a supervised learning approach for detecting landmarks in cardiac images from different views. A set of candidate landmark points are obtained using morphological operations and graph cut segmentation. The final landmarks are determined using random forests (RF) classifiers which were trained on low level features derived from the neighborhood of annotated landmarks on training images. We use features like intensity, texture, shape asymmetry and context information for landmark detection. Experimental results on the STACOM LV landmark detection challenge dataset show that our approaching is promising with room for further improvement.

1 Introduction

Cardiac magnetic resonance imaging (MRI) has acquired great significance because MRI enables complete analysis of cardiac function. To overcome high inter-observer variability in interpretation of images there is a need for automatic detection of cardiac landmarks (like apex, mitral valve). Further, such automatic detection can act as a good initialization for LV segmentation.

Stralen et al. [12] proposed a method for detecting left ventricle (LV) long axis and mitral valve plane in 3D ultrasound images using circular Hough transform and dynamic programming. Fourier Mellin Transform (FMT) was used in [11] for LV localization. The template and search image are transformed into a rotation and scale invariant representation using FMT. This enables fast landmark detection using a cascade of linear detectors. In [5] a classification method to detect structures in ultrasound images is proposed. Haar features are extracted and fed to a boosted cascade of weak classifiers for detection. Karavides et al [6] detect landmarks in 3D echocardiograms using Haar features and two cascades of Adaboost classifiers. Lu et al.[10] adopt a similar approach, but with the addition of steerable features. Apart from the heart, other anatomies also appear in the same slice which can lead to ambiguities in landmark detection. To resolve such ambiguities context information was used in [9]. Context appearance is derived from its local image features like shape and appearance. The method aims to exploit the correlation of appearance and shape between different landmarks.

⋆ Corresponding author.

We propose a learning based classification approach for LV landmark detection that makes use of local appearance and context features. Local appearance information was derived from intensity, texture and shape features. Context information was included in the form of difference in curvature values over a neighborhood. These features are learned from a set of annotated training images identifying the landmark locations. The rest of the paper is organized as follows. In Section 2 we describe our feature extraction and landmark detection approach, followed by results and discussion in Section 3, and conclude with Section 4

2   Method

Our approach to detect landmarks consists of two stages. First we automatically segment the LV or RV as the case may be. All the edge points on the segmented LV/RV are examined for landmark points using random forests (RF) classifiers. This gives a set of candidate landmark points, and the two points (for mitral valve and right ventricular points) or one point (for apex and base central axis points) with the maximum votes is selected as the landmark points.

2.1   Segmenting the LV and RV

Although there are many methods in literature for LV segmentation, we desire a method that requires minimal user involvement and is fast. We use a combination of morphological operations and graph cuts to obtain the LV boundary from a test image. All the image intensities in the training and test set are normalized to lie between 0 and 1 using the $95-$th percentile intensity value such that all intensities above this value are 1. The test image (I) is first thresholded such that pixels with intensities above th are set to 1 and all other pixels are set to 0. The threshold is given by

$$th = \mu + \gamma \times \sigma^2, \tag{1}$$

where μ is the mean image intensity, σ^2 is the standard deviation and $\gamma = 2$ is a constant set empirically by testing over a large number of training images.

The morphological image opening operation (erosion followed by dilation) is applied to the thresholded image I_{th} using the *imopen* function in MATLAB to get I_{open}. For *imopen* a disk structuring element of size 3 is generated using the *strel* command in MATLAB. All the connected components in I_{open} are extracted. For each cluster of pixels we determine their intensity distributions. Previously, in the set of normalized training images we manually identify regions of the LV/RV and determine the intensity distribution over its pixels. The intensity distributions of the connected regions in I_{open} are matched with this reference distribution using the Bhattacharya metric. The region with the closest match to the reference distribution is the region of interest (ROI).

One limitation of thresholding is that neighboring regions of the LV are also clustered with the LV, and *imopen* is unable to remove the connections between
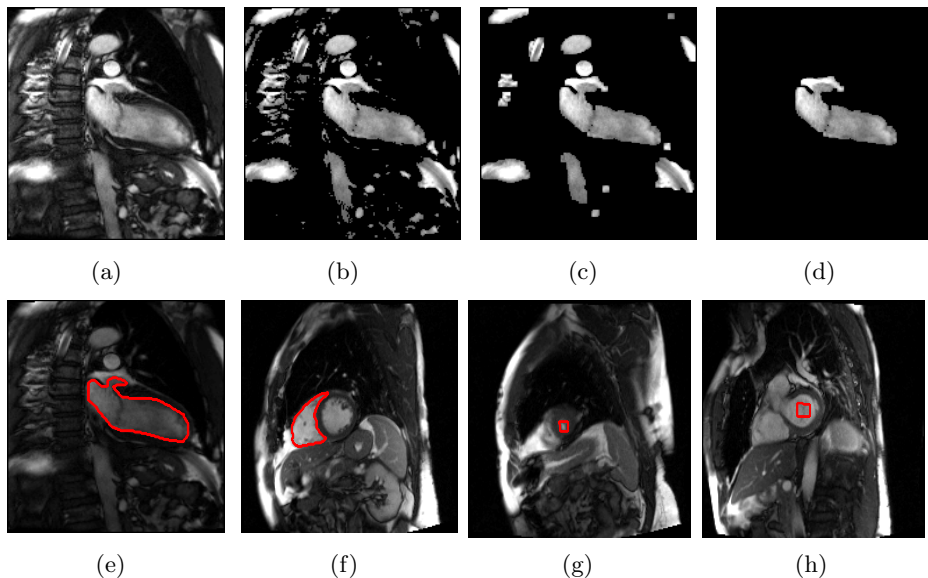
(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 1. Results of different stages of LV segmentation: (a) Original Image; (b) thresholded image; (c) morphological opening; (d) connected components; extracted edge for (e) LV and (f) RV; search areas for (g) ACA and (h) BCA

these regions. The pixels in the extracted region (from I_{open}) are again assigned labels as ROI or non-ROI using graphcuts [3]. The ROI intensity distribution previously calculated is used to determine penalties for ROI label. Similarly a reference distribution for non-ROI regions are also determined to calculate the penalty for non-ROI label. Note that only those pixels in the extracted region are labeled and not all pixels in the image. This gives a segmented ROI in which we identify the edge points. These edge points are then further classified as landmark or non-landmark using a trained RF classifier.

Figure 1 shows the results of different steps for extracting LV in a long axis (LA) image. Figure 1 (a) shows the original image followed by the thresholded image I_{th} in Fig. 1 (b). Figure 1 (c) shows the image after morphological image opening (I_{open}) and Fig. 1 (d) shows the extracted LV region using connected component analysis. As pointed out previously neighboring anatomical structures are also extracted. The result after graphcut segmentation is shown in Fig. 1 (e) with the edge points identified by the red line. This approach also works well for the RV insert points (Fig. 1 (f)). However for the apex and base central points, we do not apply graph cuts. After extracting the desired region using connected components, the center of this region is determined and a search area is defined as a bounding box of size 25×25 around this center. Figures 1 (g) and (h) show the search areas (red rectangle) for base and apex central points. The candidate search points are then classified in the next stage using RF classifiers.

2.2   Landmark Detection Using RF Classifiers

The training set provides us with annotated images that identify landmarks on
the mitral valves (MV), right ventricular inserts (RVI), centers of apex central
axis (ACA) and base central axis (BCA). For each of the annotated points we
extract features from a 35×35 neighborhood and use them to train the RF
classifier. The following features are used for training.

Intensity and Texture Features: The mean, variance, skewness and kurtosis
of intensity values are calculated for the patch around the landmarks to get the
first set of features. Texture maps of the patch are obtained along four directions
$(0°, 45°, 90°, 135°)$ using Gabor filters. Gabor filters conform to the receptive
field properties of cortical cells, capture rich visual properties like spatial fre-
quency characteristics and orientation, and are robust to noise by incorporating
Gaussian smoothing. The Gabor filter bank is represented as

$$g_{\gamma,\omega}(x,y) = a^\gamma g \left(a^\gamma \left(x \cos\left(\omega\psi\right) + y \sin\left(\omega\psi\right) \right) a^\gamma \left(-x \sin\left(\omega\psi\right) + y \cos\left(\omega\psi\right) \right) \right) \tag{2}$$

where $\gamma = 0, \cdots, \Gamma - 1, \omega = 0, \cdots, \Omega - 1$. The mother function g is a Gaussian
defined as:

$$g(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y} \right) \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi jWx \right] \tag{3}$$

$\Gamma = 4$ is the total orientations, $\Omega = 2$ is the number of scales, the rotation factor
is $\psi = \pi/\Omega$ and the scaling factor is $a = (U_h/U_l)^{1/\Gamma-1}$. U_h and U_l determine the
frequency range of the filter bank and W is a shifting parameter in the frequency
domain.

   In [7] texture anisotropy was used as a feature to identify tumorous regions
in brain images. While they calculate anisotropy using local gradient differences
and gray level dependence histograms, we use entropy to measure anisotropy.
The texture maps are divided into 8 equal parts corresponding to 8 sectors of a
circle, and entropy determined for each sector. Figure 2 (a) shows an illustration
of the sectors in a circle that are used to compute the entropy. A higher entropy
value indicates wider distribution of texture values (hence high anisotropy), while
low entropy indicates lower anisotropy. The texture anisotropy for sector r is

$$Tex_{Anisotropy}^r = -\sum_{tex} p_{tex}^r \log p_{tex}^r. \tag{4}$$

p_{tex}^r denotes the distribution of texture values in sector r. Values of tex lie
between 0 and 1. Thus the number of texture features are 32 (8 entropy values
from each of 4 oriented filters).

Shape Asymmetry: We extend the concept of asymmetry (or anisotropy) to
shape features. Here instead of texture we operate on image curvature values.
The entropy of curvature values is determined from 8 sectors of each slice to give
us 8 values for shape asymmetry. If the curvature values have a wide distribution

it indicates greater asymmetry in shape, leading to a higher entropy value. On the other hand low entropy values indicates less shape asymmetry. The shape asymmetry measure for a sector r is given by

$$Shape^r_{Asymmetry} = -\sum_\theta p^r_\theta \log p^r_\theta. \tag{5}$$

p^r_θ denotes the probability distribution of curvature values in sector r. Values of θ lie in the range $\{-180°, 180°\}$.

Context Information. Belongie et al. [2] proposed a shape context method using distribution of relative distances for matching shapes. In [13] Tu propose "auto-context" which integrates image and contextual information from a set of trained classifiers and use it to segment brain structures from MRI. Auto context was used by Li et al. in [8] to segment the human prostate gland from computed tomography (CT) images. Object interaction priors were used in [1] for inter-vertebral disc segmentation using graph cuts. Context comes from a variety of sources. Since the human anatomy is standard, and image acquisition procedures are the same presence of one organ leads to a strong cue about the presence of another organ in medical images. We aim to capture the contextual relationship between the landmarks and its neighborhood through curvature values. Basically context information provides information of one set of objects from another set of objects. Previous works have incorporated statistical models from large training data to include context information. However, these approaches require complex modeling and inference methods. For every labeled landmark we examine a neighborhood of 35×35 and sample a few points from it. The sampling location are shown in Fig. 2 (b) (red 'X's). From the image center rays at intervals of $45°$ are drawn, and points are sampled on these rays. The maximum distance along rays at $0°, 90°, 180°, 270°$ is 17 pixels (since the neighborhood is of size 35×35) and points are sampled at distances of $2, 7, 12, 17$ pixels. The maximum distance along rays at $45°, 135°, 225°, 315°$ is $17\sqrt{2} = 24$ pixels and points are sampled at distances of $3, 8, 13, 18, 23$ pixels. This gives 36 sampled points on each slice.

At each sampled point we calculate the difference in feature values with respect to the central point (i.e., the landmark pixel) and include it in a feature vector. This gives 36 values for a single slice.

$$Cont = f(i) - f(n_i), \forall n \in N_i \tag{6}$$

where f is the feature value, n_i is the sampled neighboring points, and N_i denotes the whole in the neighborhood of the labeled voxel i.

For context information we can use any or all of the previously mentioned features (intensity, texture or shape). However in practice we find that shape context features provides the maximum discrimination ability in comparison to other features. Inclusion of intensity and texture context features increases the length of the feature vector (also the computation time), but does not lead to a significant increase in accuracy of landmark detection over shape features.
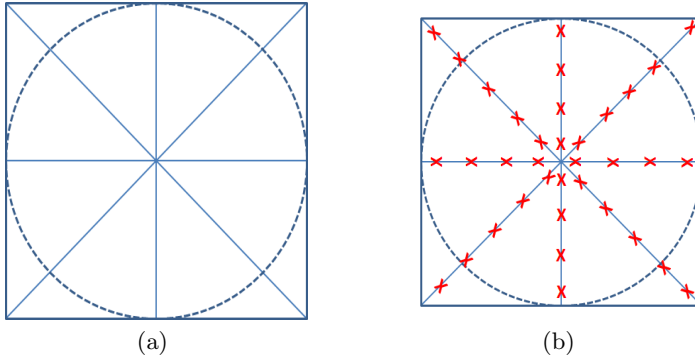
(a)   (b)

Fig. 2. (a) Illustration of sectors for entropy calculation in texture anisotropy and shape asymmetry; (b) sampling locations to derive context information

Therefore we include only the shape context features in the final feature vector. Here we clarify that to derive context information we use only shape features, while for incorporating low level information we use shape, texture and intensity features. The final landmark(s) among the candidate points are those which have the maximum number of votes by the RF classifier.

3   Experimental Results and Discussion

We use the database of STACOM LV landmark detection challenge workshop 2012 [4]. The training data consists of 100 patients with the images acquired in the LA and SA. Our classifiers were trained on features derived from all the 100 datasets. We test our method on 4 validation datasets of the STACOM landmark detection challenge - validations sets $1, 2, 3, 4$- a total of 80 patients. The extracted feature vectors for each point are 80 dimensional $(4 + 32 + 8 + 36)$. While the manual annotations gave us positive samples, we automatically extract background points in the landmarks' vicinity to get sufficient negative samples to train the RF classifier. For each landmark type (MV,RVI,ACA and BCA) we have different sets of classifiers. All the classifiers had 50 trees.

For every dataset there is only one ACA and BCA point, while there are many RVI and MV points depending upon the number of acquired images. In order to calculate the error measures the Euclidean distance between the detected landmark and the actual landmark is determined. The performance is summarized in Table 1. Figure 3 gives the error distributions for each type of landmark point. A major percentage of the errors is within 10 mm for all cases. However there are cases of high error, particularly for MV points.

Figure 4 shows the detected landmarks by our method (green 'x') and the actual landmarks as red 'x' for patient number 95. Visual results and error measures show that our method is promising and detects landmarks quite close to the actual position. We implemented the whole method in MATLAB on a Pentium Core 2 Duo, 2.66 GHz processor. The average time to detect landmarks

Table 1. Average error measures for landmark detection on 4 validation datasets comprising 80 patients. Values are given for mean±standard, minimum, maximum and median error. Values are in units of mm.

	Mean±std	min	max	median
ACA	5.66±5.39	0.60	33.48	4.16
BCA	7.01±7.99	0.25	52.28	4.51
MV	5.67±5.83	0.02	64.66	4.19
RVI	6.79±5.49	0.14	45.56	5.44

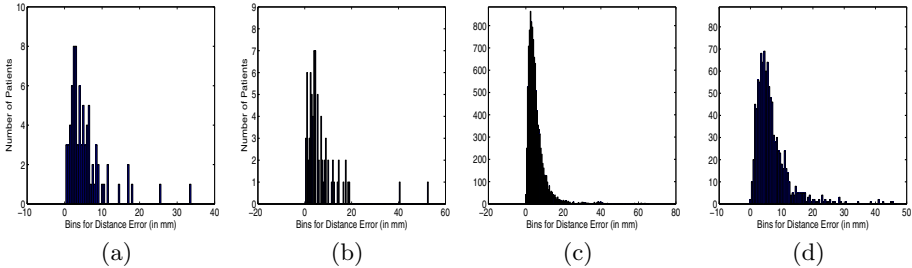(a)　　　　　　　(b)　　　　　　　(c)　　　　　　　(d)

Fig. 3. Distribution of landmark errors for: (a) ACA; (b) BCA; (c) MV; (d) RVI

in a 192×152 pixel image was 23 seconds, thus indicating scope for a faster and efficient approach. The generation of candidate landmark points took less than 2 seconds on an average.

Importance of Different Features. Figure 5 shows error measures for different feature combinations on the training dataset. For single features Int performs worst and $Cont$ gives the best results. For two feature combination the lowest error is given by $Cont + Shape$. In the case of three features, all combinations having $Cont$ give similar results highlighting the discriminative power of context information.

Our method also has limitations. It is particularly sensitive to the accuracy of segmentation as it provides candidate points for landmark detection. Segmentation accuracy depends upon threshold value, efficacy of the image opening step and of connected component analysis. We select $\gamma = 2$ (Eqn. 1) after visual examination of the results over more than 50 images from SA and LA views. The disk size in the *imopen* operation was fixed through similar examination.

Initially we tried to localize the ROI using a machine learning approach. We divided the training images into 64 square blocks and extracted the previously mentioned features for blocks containing the annotated landmark (landmark block) as well as blocks not containing the landmark (background blocks). Context information was derived with respect to the central pixel in the block. If a test block is denoted landmark by the RF classifier, we subdivide it into 4 sub-blocks and
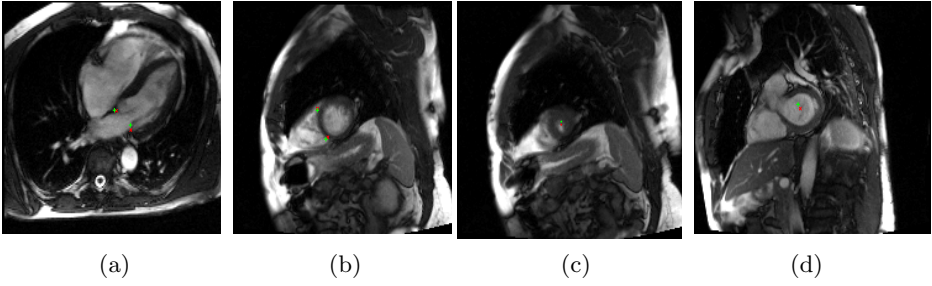
(a)   (b)   (c)   (d)

Fig. 4. Results of landmark detection: (a) MV (average error= 2.6 pixels); (b) RVI (average error= 1.9 pixels); (c) ACA (average error= 1.2 pixels); (d) BCA (average error= 2.1 pixels)
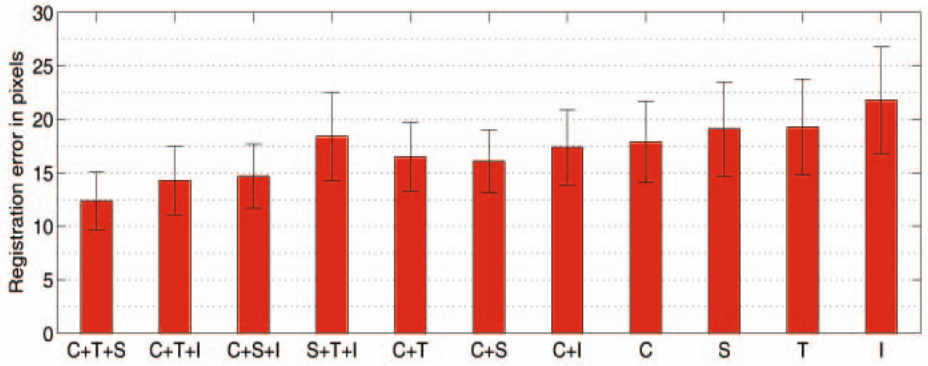

Fig. 5. Error measures(mean and standard deviation) for MV detection using different feature combinations.C-Context;S-Shape;T-Texture;I-Intensity

check for the presence of landmark from these 4 sub-blocks. Finally each pixel of the identified sub-block is analyzed for the presence of a landmark. We extract features from a 35×35 neighborhood of each pixel and train different set of classifiers for each stage. But such an approach required many computations in MATLAB. Since our features use entropy based measures sufficient samples are needed for an accurate feature value. Thus smaller blocks give erroneous values. Alternatively, larger blocks result in more computations when we analyzed each pixel's neighborhood. This approach gave fairly good results for LV and RV landmarks, but poor results for ACA and BCA points as there were not sufficient samples to train a accurate classifier that can distinguish between landmark and background blocks. In order to overcome this limitation we used morphological operations and graph cuts to find a set of candidate points for each landmark.

4   Conclusion

We have proposed a method to detect landmarks in cardiac MRI using a combination of graphcut segmentation, low level feature information and machine learning techniques. An initial set of candidate landmark points is obtained by thresholding, connected component analysis and graphcut segmentation. The final landmarks are identified with the help of RF classifiers trained on manual annotations of landmark points. For accurate landmark detection low level image features like intensity, texture anisotropy, shape asymmetry and context information was derived from neighborhoods of annotated landmarks. Training was performed on data from 66 patients. Results on 5 test patient databases show our approach is promising for the purpose of detecting the appropriate landmarks. But there is also scope for speeding up the detection of candidate points as well as identifying the final landmark.

References

1. Ben Ayed, I., Punithakumar, K., Garvin, G., Romano, W., Li, S.: Graph Cuts with Invariant Object-Interaction Priors: Application to Intervertebral Disc Segmentation. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 221–232. Springer, Heidelberg (2011)
2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. Patt. Anal. Mach. Intell. 24(24), 509–522 (2002)
3. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient N-D image segmentation. Intl. J. Comp. Vis. 70(2), 109–131 (2006)
4. Fonseca, C., Backhaus, M., Bluemke, D., Britten, R., Chung, J., Cowan, B., Dinov, I., Finn, J., Hunter, P., Kadish, A., Lee, D., Lima, J., Medrano-Gracia, P., Shivkumar, K., Suinesiaputra, A., Tao, W., Young, A.: The cardiac atlas project - an imaging database for computational modeling and statistical atlases of the heart. Bioinformatics 27(16), 2288–2295 (2011)
5. Georgescu, B., Zhou, X., Comaniciu, D., Gupta, A.: Database guided segmentation of anatomical structures with complex appearance. In: CVPR, pp. 429–436 (2005)
6. Karavides, T., Leung, K., Paclik, P., Hendriks, E., Bosch, J.: Database guided detection of anatomical landmark points in 3D images of the heart. In: ISBI, pp. 1089–1092 (2010)
7. Kovalev, V., Petrou, M., Bondar, Y.: Texture anisotropy in 3D images. IEEE Trans. Imag. Proc. 8(3), 346–360 (1999)
8. Li, W., Liao, S., Feng, Q., Chen, W., Shen, D.: Learning Image Context for Segmentation of Prostate in CT-Guided Radiotherapy. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 570–578. Springer, Heidelberg (2011)
9. Lu, X., Georgescu, B., Littmann, A., Mueller, E., Comaniciu, D.: Discriminative Joint Context for Automatic Landmark Set Detection from a Single Cardiac MR Long Axis Slice. In: Ayache, N., Delingette, H., Sermesant, M. (eds.) FIMH 2009. LNCS, vol. 5528, pp. 457–465. Springer, Heidelberg (2009)
10. Lu, X., Georgescu, B., Zheng, Y., Otsuki, J., Comaniciu, D.: AutoMPR: Automatic detection of standard plane in 3D echocardiography. In: ISBI, pp. 1279–1282 (2008)

11. Ma, M., Bosch, J., Reiber, J., Lelieveldt, B.: Fully automatic estimation of object pose for segmentation initialization: application to cardiac MR and echocardiography. In: SPIE Medical Imaging (2007)
12. Van Stralen, M., Leung, K., Voormolen, M., de Jong, N., van der Steen, A., Reiber, J., Bosch, J.G.: Time continuous detection of the left ventricular long axis and the mitral valve plane in 3-D echocardiography. Ultrasound in Med. Bio. 34(2), 196–207 (2008)
13. Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3D brain image segmentation. IEEE Trans. Patt. Anal. Mach. Intell. 32(10), 1744–1757 (2010)