COMPREHENSIVE AUTOENCODER FOR PROSTATE RECOGNITION ON MR IMAGES

Ke Yan[1], Changyang Li[1], Xiuying Wang[1], Yuchen Yuan[1], Ang Li[1], Jinman Kim[1], Biao Li[2], Dagan Feng[1]

[1]School of Information Technologies, University of Sydney
[2]Department of Nuclear Medicine, Ruijin Hospital, Shanghai Jiaotong University School of Medicine

ABSTRACT

Automatic recognition of anatomical structures is an essential prerequisite in computer aided diagnoses (CAD) such as tissue segmentation, physiological signal measurement and disease classification. However, insufficient color and speckle information in medical images pose challenges to the recognition of anatomical structures. Such challenges are evident with prostate recognition on magnetic resonance (MR) images and thus remain an open problem, although prostate cancer is an important problem that are attracting increasing interests in medical imaging. In this study, we propose an automatic approach for prostate recognition on MR images. Firstly, compared to existing works which integrate autoencoder with a specific type of classifier, we let autoencoder itself serve as a classifier and therefore lessening the impact from irregular and complex background found in prostate recognition. Secondly, an image energy minimization scheme with consideration of the coherence information from neighboring pixels is proposed to improve the recognition results with clear boundary appearance. We evaluate our method in comparison with three widely applied classifiers and the phase of atlas-based seeds-selection in prostate segmentation on a public prostate database. Our experiment results demonstrate significant superiority of our method in terms of both precision and F-measure.

Index Terms— prostate recognition, autoencoder, deep learning, classification

1. INTRODUCTION

The recognition and localization of anatomical structures are the prerequisite for many subsequent image processing procedures such as segmentation and classification [1]. It poses a challenging task because of the insufficient color information of pixels and low signal-to-noise ratio in medical images [2]. Previous works have been proposed to tackle anatomical structure recognition problems based on handcrafted features, such as steerable feature, on a wide array of imaging modalities, e.g., ileocecal valves [3], polyps [4], and livers [5] in abdominal CT, and heart chambers in ultrasound [6]. However, to our best knowledge, no work has been done on prostate recognition in MR images, although prostate cancer accounts for the second highest mortality rate among various types of cancer on males [7] and MR images prove effective for prostate diagnoses and treatments [8]. In addition to the insufficient color and speckle information, MR image artifacts, such as low contrast and blurred tissue boundary, make it even more difficult to accurately locate the prostate. Many prostate segmentation approaches, e.g., [9, 10] are often limited by the recognition techniques in prostate imaging, as accurate segmentation often requires approximate localization of the prostate object as initialization. To address this challenge, conventional prostate segmentations rely on semi-automatic methods thereby being dependent on the user [2, 11, 12]. Alternative approaches explore the use of an image atlas to define the foreground/background seeds in prostate segmentation [9, 10]. However, as noted in prior studies [8], reliance on atlas are still prone to generating errors.

While the above handcrafted-feature-based anatomical structure recognition methods [3-6] are technically sound, one major drawback is that these low-level features may fail to describe complex anatomical structure [13-16], and hence it dampens the recognition performance demonstrated by the experiment results of our work (Section 3.2) and consistent to [1, 17, 18]. Recently, autoencoder (AE) has been proven to effectively extract high-level features from input data for (patch-based) image classifications [14, 15], object detections [19], and disease diagnosis [13, 20]. It learns a set of encoding weights to construct a code vector as the feature of input data, and then learns another set of decoding weights to map the code vector into an approximate reconstruction for the input data. Although AE can learn high-level features directly from the pixel intensity values, for its application to medical images, it may be insufficient for later tissue recognition or classification from the experiment results of existing works [13, 15], especially for the low-contrast images or blurred tissue boundaries.

In this paper, to solve these issues, we propose a novel prostate recognition method on MR images by combining handcrafted features with AE. Our work has two major contributions. Firstly, different from the most works which embed a classifier on the top of the hidden layer in AE [13, 15, 19, 20], we propose a novel method to compute prostate recognition map through taking advantage of favorable capability of data reconstruction [21] from AE. Secondly, we design an image energy minimization scheme to generate a

stronger prostate recognition map with consideration of the relationship among neighboring pixels.

2. METHODS

Our method consists of two stages. In the first stage, early feature descriptors for the prostate are proposed to construct the training inputs for our proposed classifier based on the AE framework as shown in Fig. 1(a-b); the probability of a pixel belonging to the prostate can be approximated by our trained stacked autoencoder (SAE) classifier as shown in Fig. 1(c). In the second stage, we propose a new energy minimization scheme to optimize the rough probability map of prostate from the output of first stage as described in Fig. 1(d).

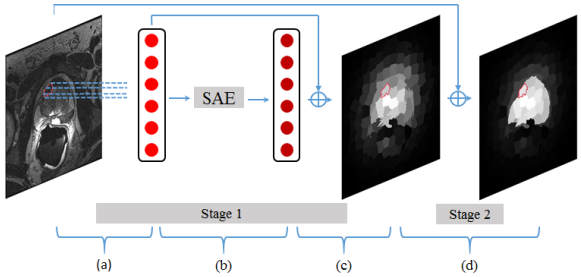

Fig. 1 Pipeline of our method. (a) Early feature extraction. (b) Superpixel reconstruction via proposed prostate AE model. (c) Superpixel classification. (d) Refinement.

2.1. Early feature descriptors

Instead of merely using the pixel intensity values, we adopt two early features, i.e. the intensity descriptor and the position descriptor, which reflects most critical information about pixel-value and spatial information. These two descriptors are the inputs for the AE. In this work, we use superpixel as the atomic homogenous region in further operations to boost performance in recognition around tissue boundaries [22]. Formally, an image $I \in R^{m\times n}$ is segmented into N superpixels via the SLIC algorithm [23]. We denote a superpixel as P. As suggested in [24], the superpixel is first whitened via zero phase component analysis (ZCA) to make the pixels less correlated with each other. An early feature vector f(P) is then derived for P as follows.

2.1.1. Intensity descriptor

For the intensity histogram IH(P) of superpixel P, the number of bins is set to 20 empirically in our experiment. Then, the intensity histogram $IH(P) \in R^{20\times 1}$ is normalized to have a uniform sum to eliminate the effect caused by the different number of pixels within different superpixels.

2.1.2. Position descriptor

Since the superpixels are of irregular shapes, we exploit bounding boxes to approximate their spatial locations. We denote the bounding box of P as $C(P) = \{c_v(\alpha_{v,1}, \alpha_{v,2}): v = 1,2\}$, where c_1 and c_2 are the top-left coordinate and bottom-right coordinate of C(P) in image $I \in R^{m\times n}$ respectively. $\alpha_{v,1}$ and $\alpha_{v,2}$ are the values of c_v corresponding to x-axis and y-axis respectively. The position descriptor $POS(P) \in R^{4\times 1}$ of superpixel P is then calculated by

$$POS(P) = \left\{\frac{\alpha_{v,u}}{(2-u)n+(u-1)m} : v = 1,2; u = 1,2\right\} \quad (1)$$

2.1.3. Early feature vector

With the early feature descriptors proposed above, a superpixel-wise feature vector f(P) with 24 dimensions is generated as

$$f(P) = \{IH(P); POS(P)\} \in R^{24\times 1} \quad (2)$$

2.2. Prostate stacked autoencoder model

After obtaining the early feature vectors of prostate superpixels, we can build a SAE [13] to extract high-level features and perform reconstruction of input for later classification. While most works [13, 15] train AE by both positive and negative samples, the training set in our work consists of only positive samples, which focuses on the prostate feature extraction and thus lessens the impacts by the irregular and complex background that may impede feature extraction.

To train a single-hidden-layered prostate AE, a training set $F = \{f(P_i): i = 1, \dots, K\}$ containing K early feature vectors of prostate superpixels are input to the AE network. The input vector $f(P_i)$ is transformed into a hidden feature representation a_i by an activation function $g(\cdot)$ with the following formula:

$$a_i(f(P_i); \theta^{(1)}) = g\left(W^{(1)} f(P_i) + b^{(1)}\right) \quad (3)$$

where $\theta^{(1)}$ is the parameter vector including weight matrix $W^{(1)}$ and bias term $b^{(1)}$; as a common practice, we use the sigmoid function $g(\phi) = 1/(1+\exp(-\phi))$ as the activation function. A decoder then maps the hidden feature representation a_i back to an approximate reconstruction $\widehat{f(P_i)} \in R^{24\times 1}$ in a similar transformation

$$\widehat{f(P_i)}(a_i; \theta^{(2)}) = g\left(W^{(2)} a_i + b^{(2)}\right) \quad (4)$$

With the training set F of K samples, the latent features of input data can be learned by minimizing the cost function

$$J(\theta) = \frac{1}{K}\sum_{i=1}^{K}\frac{1}{2}\left\|f(P_i) - \widehat{f(P_i)}\right\|^2 + \frac{\lambda}{2}\sum_{i=1}^{s^{(1)}}\sum_{j=1}^{s^{(2)}}(W_{ij}^{(1)})^2 \quad (5)$$

where the first term in $J(\theta)$ is an average sum-of-squares error term and the second term is a weight decay term that tends to decrease the magnitude of the weight and prevent

overfitting [25], with a weight decay parameter λ. $s^{(1)}$ and $s^{(2)}$ are the numbers of nodes in the input layer and hidden layer respectively. As suggested in [19], a sparsity constraint is imposed on the hidden nodes to enhance the probability of linear separability. We use gradient descent optimization algorithm to update θ in iterations and back-propagation algorithm is applied to calculate the partial derivatives in this process.

We further construct SAE [19, 21] to learn highly nonlinear and complex patterns, and perform feature presentation in the input images. As shown in Fig. 2, in a SAE structure, the early feature vector is input to the first (bottom) AE, the hidden nodes of which are concatenated as a new feature vector to the subsequent (higher-level) AE. The greedy layer-wise algorithm is adopted to obtain the corresponding parameter $\theta^{(l)}$ of the l-th layer, and back-propagation is then applied again to tune the parameters of all layers at one time. Typically in our work, we stack three AEs to construct the prostate SAE model and hence obtain a totally six layer network including three encoding layers and three decoding layers (Fig. 2).

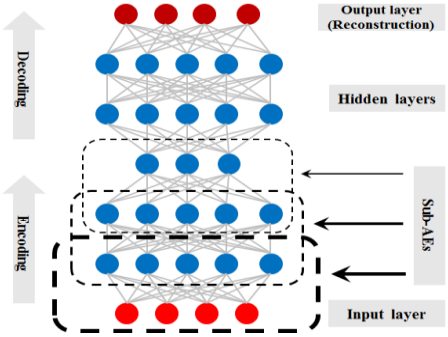

Fig. 2 Architecture of the proposed SAE. The output of each layer is the input for its subsequent layer. The output of the last layer is a reconstruction for input data.

2.3. Superpixel classification

With the trained prostate SAE model, the superpixels of the input image can thus be classified for prostate recognition. Different from other deep learning algorithms (i.e. convolution neural network), AE is capable of not only intrinsic and latent feature learning for input data (3), but also data reconstruction (4). Therefore, we can calculate the reconstruction errors defined in (7) for each superpixel in a prostate MR image through the fixed prostate SAE model. Specifically, with the all parameters $I = \{\theta^{(l)}: l = 1, \dots, L\}$ learned in SAE, for a superpixel P, set

$$f(P)^{(l+1)} = g(W^{(l)} f(P)^{(l)} + b^{(l)}) \qquad (6)$$

where l is the index of network layer. We initialize the first step of the iteration $f(P)^{(1)}$ as the early feature vector $f(P)$ of the superpixel P. Then the reconstruction error of P is calculated by $f(P)$ and $f(P)^{(L+1)}$:

$$err(P) = \sum_{\omega=1}^{24} \left\| f(P)_\omega - f(P)^{(L+1)}{}_\omega \right\|^2 \qquad (7)$$

where $f(P)_\omega$ and $f(P)^{(L+1)}{}_\omega$ are the ωth elements of $f(P)$ and $f(P)^{(L+1)}$ respectively.

As the SAE model is learned from the set of prostate superpixels (positive samples), the unlabelled superpixels belonging to prostate have low reconstruction errors, while those belonging to background have high reconstruction errors. Hence, we use the reconstruction error to measure the probability $D^{sp\_AE}(P)$ of a superpixel P being prostate tissue:

$$D^{sp\_AE}(P) = \exp(-\tau \times err(P)) \qquad (8)$$

where τ controls the distance between different superpixel's reconstruction errors within an image and is set to 100 empirically.

With calculating D^{sp_AE} of all the superpixels in an image $I \in R^{m \times n}$, we may obtain a rough prostate recognition map $D^{AE} = \{d^{AE}{}_i \in [0,1]: i = 1, \dots, m \times n\}$ after normalization of D^{AE}, as shown in the third column of Fig. 3.

2.4. Refinement

The rough prostate recognition map D^{AE} may generate wrong labels of background near the prostate, as it is a local estimation without considering the spatial and intensity coherence among superpixels. In this sub-section, a refined prostate recognition map with better suppressed background, more smooth inner region and clear boundary is generated based on D^{AE}.

Given a one-channel image I, our task in this stage is to assign a label $O_p \in \{0,1\}$ to a pixel p to measure whether p belongs to foreground or not. For the set of pixels' labelling $O = \{O_p: p \in I\}$, this can be solved by minimizing the energy function [26]

$$E(O) = \sum_{p \in I} H(O_p) + \xi \sum_{(p,q) \in Y} \frac{1}{1 + \sqrt{3(I_p - I_q)^2}} \cdot T(O_p \neq O_q) \qquad (9)$$

where Y is a set of all pairs of neighboring pixels. $H(O_p)$ is the cost for assigning a label O_p to a pixel p. We directly use local estimated recognition map D^{AE} to approximate the label-cost of pixels. Specifically, $H(O_p)$ is set to $D^{AE}(p)$ if O_p is a background label and $1 - D^{AE}(p)$ if O_p is a foreground label. The second term in (9) encourages intensity and spatial coherence by penalizing discontinuities [27] between neighboring pixels, with the parameter ξ controlling the scale of discontinuity penalty. $T(\cdot)$ is 1 if the condition inside the parentheses is true and 0 otherwise.

We adopt maximum flow algorithms [27] to minimize (9) and generate the corresponding prostate recognition map

D^{mf}. Then the final prostate recognition map $D = (D^{AE} + D^{mf})/2$ is formed to measure the probability of each pixel being prostate.

3. EXPERIMENTS

3.1. Experimental setup

The prostate MR Image Segmentation 2012 (PROMISE12) database [28] is used in this study. It contains 50 cases, with each case composed of 15 to 54 prostate transverse T2-weighted MR images. Manual segmentation are available for each case and used as the ground truth.

In the prostate SAE model, the hyperparameters of each sub-AE, i.e. the number of hidden nodes Z, and weight decay parameter λ, are derived empirically and listed in Table 1. As suggested in [26], to achieve better performances, we compute five recognition maps using five superpixel scales with $N = 200, 250, 300, 350, 400$ respectively in an image. Then, we linearly combine the five recognition maps as the final recognition result.

We perform 10-fold cross validation on the dataset. For each image, we first resize it to 320∗320 pixels, and then increase its contrast by mapping the intensity values to new values such that 1% of data is saturated at low and high intensities of the image [29]. An atlas-based seeds-selection in segmentation approach (RW) [9] and three popular classifiers, i.e. support vector machine (SVM) with radial basis function kernel, random forest (RF), and naive Bayes (NB), are chosen as comparison methods.

Table 1: Hyperparameters in the prostate SAE model.

	sub-AE 1	sub-AE 2	sub-AE 3
Z	60	40	16
λ	8e-4	4e-4	4e-4

3.2. Evaluation

We evaluate the recognition performance using precision-recall (PR) curve and F-measure [20, 26]. Both Table 2 and Fig. 4 shows that our method outperform the four comparison methods in terms of both PR curve and F-measure. More specifically, even our unrefined results outperform the refined results of the comparison methods in precision. This is mainly attributed to the SAE for high-level feature learning and data reconstruction, while the comparison methods recognize prostate directly from the low-level early features. Fig. 3 qualitatively demonstrates that our proposed refinement significantly contributes to foreground smoothness and background suppression. The refinement poses relatively low effect around the prostate with blurred boundary as illustrated in the second row of Fig. 3. The reason is that the neighboring pixels around the boundary does not differentiate much, thus causing a large penalty in the second term of (9), which encourages to assign same labels to these pixels around the boundary of prostate. However, from Table 2, it can be seen that our proposed refinement improves all the methods in precision and F-measure.

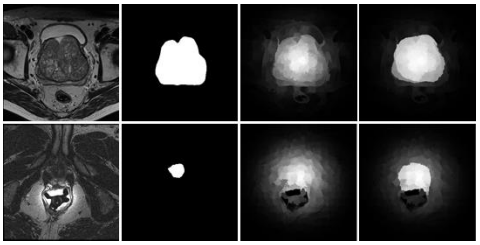

Fig. 3 Examples of prostate recognition results by our method. Left to right: raw image, ground truth, rough recognition map by the first stage, and final recognition map.

Table 2: Precision and F-measures of our method and comparison methods for prostate recognition on PROMISE12 database. The best results in each column are shown in bold.

	Precision		F-measure	
	Not refined	Refined	Not refined	Refined
OURS	**0.8518**	**0.8699**	**0.6798**	**0.6832**
RW	0.8284	0.8286	0.6617	0.6220
SVM	0.5554	0.6394	0.5415	0.6238
RF	0.4870	0.5506	0.5189	0.5766
NB	0.3539	0.4894	0.3906	0.5033

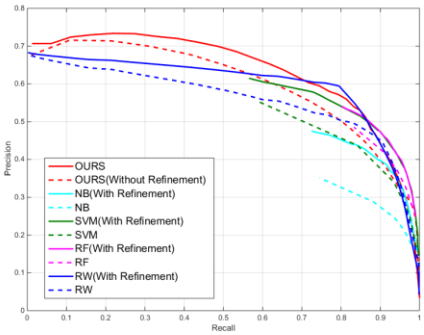

Fig. 4 PR curves of our method and comparison methods for prostate recognition on PROMISE12 database. The recognition results by comparison methods are also refined by our proposed approach for better evaluation (solid lines).

4. CONCLUSION

We propose an automatic prostate recognition method on MR images based on SAE. Compared to the most existing works with AE, we let the SAE itself serve as a classifier to focus on the prostate feature extraction. An image energy minimization scheme is then proposed to optimize the prostate recognition map constructed by SAE. Our method is compared against three benchmark classifiers and atlas-based seeds-selection approach on the PROMISE12 database, demonstrating superiority in both PR curves and F-measures.

5. REFERENCES

[1]Y. Zheng, B. Georgescu, H. Ling, S. K. Zhou, M. Scheuering, and D. Comaniciu, "Constrained marginal space learning for efficient 3D anatomical structure detection in medical images," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 194-201.

[2]M. Samiee, G. Thomas, and R. Fazel-Rezai, "Semi-automatic prostate segmentation of MR images based on flow orientation," in *Signal Processing and Information Technology, 2006 IEEE International Symposium on*, 2006, pp. 203-207.

[3]L. Lu, A. Barbu, M. Wolf, J. Liang, L. Bogoni, M. Salganicoff, *et al.*, "Simultaneous detection and registration for ileo-cecal valve detection in 3D CT colonography," in *Computer Vision–ECCV 2008*, ed: Springer, 2008, pp. 465-478.

[4]L. Lu, A. Barbu, M. Wolf, J. Liang, M. Salganicoff, and D. Comaniciu, "Accurate polyp segmentation for 3D CT colongraphy using multi-staged probabilistic binary learning and compositional model," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1-8.

[5]H. Ling, S. K. Zhou, Y. Zheng, B. Georgescu, M. Suehling, and D. Comaniciu, "Hierarchical, learning-based automatic liver segmentation," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1-8.

[6]L. Yang, B. Georgescu, Y. Zheng, P. Meer, and D. Comaniciu, "3D ultrasound tracking of the left ventricle using one-step forward prediction and data fusion of collaborative trackers," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1-8.

[7]M. M. Center, A. Jemal, J. Lortet-Tieulent, E. Ward, J. Ferlay, O. Brawley, *et al.*, "International variation in prostate cancer incidence and mortality rates," *European urology*, vol. 61, pp. 1079-1092, 2012.

[8]S. Ghose, A. Oliver, R. Martí, X. Lladó, J. C. Vilanova, J. Freixenet, *et al.*, "A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images," *Computer methods and programs in biomedicine*, vol. 108, pp. 262-287, 2012.

[9]A. Li, C. Li, X. Wang, S. Eberl, D. D. Feng, and M. Fulham, "Automated Segmentation of Prostate MR Images Using Prior Knowledge Enhanced Random Walker," in *Digital Image Computing: Techniques and Applications (DICTA), 2013 International Conference on*, 2013, pp. 1-7.

[10]S. Martin, V. Daanen, and J. Troccaz, "Atlas-based prostate segmentation using an hybrid registration," *International Journal of Computer Assisted Radiology and Surgery*, vol. 3, pp. 485-492, 2008.

[11]R. Zwiggelaar, Y. Zhu, and S. Williams, "Semi-automatic segmentation of the prostate," in *Pattern Recognition and Image Analysis*, ed: Springer, 2003, pp. 1108-1116.

[12]D. Flores-Tapia, G. Thomas, N. Venugopa, B. McCurdy, and S. Pistorius, "Semi automatic MRI prostate segmentation based on wavelet multiscale products," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, 2008, pp. 3020-3023.

[13]S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*, 2014, pp. 1015-1018.

[14]Q. Li, W. Cai, and D. D. Feng, "Lung image patch classification with automatic feature learning," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 2013, pp. 6079-6082.

[15]D. Kumar, A. Wong, and D. A. Clausi, "Lung Nodule Classification Using Deep Features in CT Images," in *Computer and Robot Vision (CRV), 2015 12th Conference on*, 2015, pp. 133-138.

[16]L. Wang, H. Lu, X. Ruan, and M.-H. Yang, "Deep Networks for Saliency Detection via Local Estimation and Global Search," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3183-3192.

[17]R. Fakoor, F. Ladhak, A. Nazi, and M. Huber, "Using deep learning to enhance cancer diagnosis and classification," in *Proceedings of the International Conference on Machine Learning*, 2013.

[18]Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu, "Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features," *Medical Imaging, IEEE Transactions on*, vol. 27, pp. 1668-1681, 2008.

[19]J. Han, D. Zhang, S. Wen, L. Guo, T. Liu, and X. Li, "Two-Stage Learning to Predict Human Eye Fixations via SDAEs," 2015.

[20]J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, *et al.*, "Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology images," 2015.

[21]Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, pp. 1798-1828, 2013.

[22]X. Ye, G. Beddoe, and G. Slabaugh, "Automatic graph cut segmentation of lesions in CT using mean shift superpixels," *Journal of Biomedical Imaging*, vol. 2010, p. 19, 2010.

[23]R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, pp. 2274-2282, 2012.

[24]N. Wang, J. Melchior, and L. Wiskott, "An analysis of Gaussian-binary restricted Boltzmann machines for natural images," in *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 2012, pp. 287-292.

[25]Y. Ouyang, W. Liu, W. Rong, and Z. Xiong, "Autoencoder-Based Collaborative Filtering," in *Neural Information Processing*, 2014, pp. 284-291.

[26]N. Tong, H. Lu, X. Ruan, and M.-H. Yang, "Salient Object Detection via Bootstrap Learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1884-1892.

[27]Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, pp. 1124-1137, 2004.

[28]G. Litjens, R. Toth, W. van de Ven, C. Hoeks, S. Kerkstra, B. van Ginneken, *et al.*, "Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge," *Medical image analysis*, vol. 18, pp. 359-373, 2014.

[29]S. Qureshi, *Embedded image processing on the TMS320C6000TM DSP: examples in code composer studioTM and MATLAB*: Springer Science & Business Media, 2005.