



A Novel Bayesian Model Incorporating Deep Neural Network and Statistical Shape Model for Pancreas Segmentation

Jingting Ma¹(✉), Feng Lin¹, Stefan Wesarg³, and Marius Erdt^{1,2}

¹ School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore

{jma012,asflin,merdt}@ntu.edu.sg

² Fraunhofer Singapore, Nanyang Technological University, Singapore, Singapore

³ Visual Healthcare Technologies, Fraunhofer IGD, Darmstadt, Germany
stefan.wesarg@igd.fraunhofer.de

Abstract. Deep neural networks have achieved significant success in medical image segmentation in recent years. However, poor contrast to surrounding tissues and high flexibility of anatomical structure of the interest object are still challenges. On the other hand, statistical shape model based approaches have demonstrated promising performance on exploiting complex shape variabilities but they are sensitive to localization and initialization. This motivates us to leverage the rich shape priors learned from statistical shape models to improve the segmentation of deep neural networks. In this work, we propose a novel Bayesian model incorporating the segmentation results from both deep neural network and statistical shape model for segmentation. In evaluation, experiments are performed on 82 CT datasets of the challenging public NIH pancreas dataset. We report 85.32 % of the mean DSC that outperforms the state-of-the-art and approximately 12 % improvement from the predicted segment of deep neural network.

Keywords: Bayesian model · Deep neural networks
Statistical shape model · Pancreas segmentation

1 Introduction

With the rapid development of Convolutional Neural Networks (CNNs) in semantic segmentation, deep neural networks like U-Net [1], SegNet [2] have become a popular trend in medical image segmentation and achieved remarkable success in segmentation of many organs, e.g. liver, lung and spleen. However, segmentation of challenging organs such as pancreas still remains difficulties due to the relatively small region in the whole volume, highly complex anatomical structure and significantly ambiguous boundary. On the other hand, usually the amount of labeled medical image data is limited which inhibits the segmentation from achieving considerable accuracy. To tackle these challenges, we aim to

propose a robust segmentation approach for pancreas, which is one of the most challenging organs.

Numerous works focus on pancreas segmentation in literature, and the majority of them adopt deep neural networks with various refinement methods. In [3, 4], a coarse-to-fine framework is designed where the coarse network is trained to obtain the rough segment and remove the background regions, afterwards the shrunken region is passed to the fine network for precise segmentation. In [5], a Recurrent Neural Network (RNN) combining with CNN layers is employed to exploit spatial relations among successive slices. On the other hand, traditional machine learning approaches are demonstrated to be useful in segmentation framework for locally fine-tuning, e.g., random forests are utilized in feature extraction and classification following the deep neural networks in [6, 7] and Gaussian Mixture Model is employed to refine the U-Net in [8].

Considering of the ambiguities on boundary, it is well worth to leverage the 3D shape variabilities to distinguish the non-visible boundary, this motivates us to employ statistical shape models in segmentation framework. Through back projection onto the shape model, the corruptness on input shape is supposed to be corrected. Owing to the high variability of pancreas shape, we adopt the robust kernel statistical shape model presented in [9] as it has compelling advantages in handling corrupted and highly deformable training data than conventional PCA models. However, the model based approaches are sensitive to initialization, thus a deep neural network plays an important role in providing a rough segmentation for shape model initialization. With this motivation, we integrate the segmentation from deep neural network and statistical shape model within a Bayesian model for pancreas segmentation. A novel optimization principle joint with image feature and shape prior is proposed to guide segmentation. Our approach is demonstrated to be promising and efficient in terms of evaluation.

2 Method

In this section, we elaborate our segmentation approach starting with the deep neural network architecture, followed by the Bayesian model. Let us assume we have a set of 3D CT volumes $I = \{I_1, \dots, I_N\}$ and corresponding ground truth mask $Y = \{Y_1, \dots, Y_N\}$ for training. We extract shapes $S = \{S_1, \dots, S_N\}$ from the ground truth mask to train the robust kernel statistical shape model [9], defined as $RKSSM(S|\Phi; V; K)$, where Φ represents the implicit feature space, V decides the eigenvectors in kernel space, K is the robust kernel matrix with elements $K_{ij} = \kappa(S_i, S_j) = \Phi(S_i)^T \Phi(S_j)$ and κ is the kernel trick function.

2.1 Dense-UNet Segmentation Network

DenseNet [10] has advantages in narrowing the network width, reusing features and significantly alleviating the problem of gradient vanishing. Therefore, we adopt the DenseNet in U-Net architecture by simply replacing the stacked

Conv - Relu and a following max pooling operation at each downsampling step with a 3-layer dense block with the growth rate of 4, meanwhile, keeping the upsampling path and concatenation unchanged. We use the Dice coefficient loss with a smooth value according to the most of related works that $\mathcal{L}(Z, Y) = 1 - \frac{2 \times \sum_i z_i y_i + 0.1}{\sum_i (z_i^2 + \sum_i y_i^2) + 0.1}$, where Z represents the predicted mask. Our Dense-UNet is trained with 2D slices extracted from 3D training images from Axial view, Sagittal view and Coronal view respectively, resulting in three predicted segment Z^A , Z^S and Z^C . Due to the *ReLU* activation in the output layer, the intensity range in predicted segment is in $[0, 255]$. To make use of the predicted segments in further Bayesian model, we generate probability maps $\Pi = \{\Pi_1, \dots, \Pi_N\}$ by merging the three predicted segments and feeding into a sigmoid logistic function:

$$\Pi_i = \frac{1}{1 + \exp\left(-\frac{S_i^A + S_i^Z + S_i^C}{255}\right)}, \quad (1)$$

where Π_i indicates the probability map of the i^{th} image. Using the sigmoid function to compute probability map is because (1) this is a binary segmentation task with 2 classes in total, and (2) considering the uncertain accuracy of Dense U-Net, we make the probability for each pixel in range $[0.5, 1]$ that “1” indicates the pixel has a considerable probability of being *ROI* (Region of Interest) and “0.5” indicates the pixel is unsure to be *ROI* or *NOI* (Non of Interest). Apparently, the intersection region of Z^A , Z^S and Z^C is assigned higher probabilities, and uncertain or corrupted areas receive lower probabilities.

2.2 Bayesian Model

Let the shape model *RKSSM* fed into Π for initialization (cf. Fig. 1(b)), we have an initial shape of segmentation $C = \{x_1, \dots, x_{n_P}\}$, where landmark x_i represents the i^{th} pixel in the test image. Given the test image I , probability map Π and the shape model *RKSSM*, assume the optimal shape C can be derived using Bayes’ rule as follows:

$$p(C|I, \Pi) \propto p(I, \Pi|C) p(C), \quad (2)$$

term $p(I, \Pi|C)$ is maximum likelihood estimation of C based on image and probability map and term $p(C)$ is considered as the prior distribution of the shape model. Shape C is guided towards the most probable mode by maximizing the posteriori in Eq. 2, which is equivalent to simply minimizing its negative logarithm leading to the energy function:

$$E(C) = -\log(p(I, \Pi|C)) - \log(p(C)), \quad (3)$$

the first term related to the intensity feature is solved via a Gaussian Mixture Model and the second term related to the shape prior is solved with the shape model. The optimal solution is reached by adapting the gradient descent to the energy. The overall procedure of segmentation algorithm is summarized in Algorithm 1.

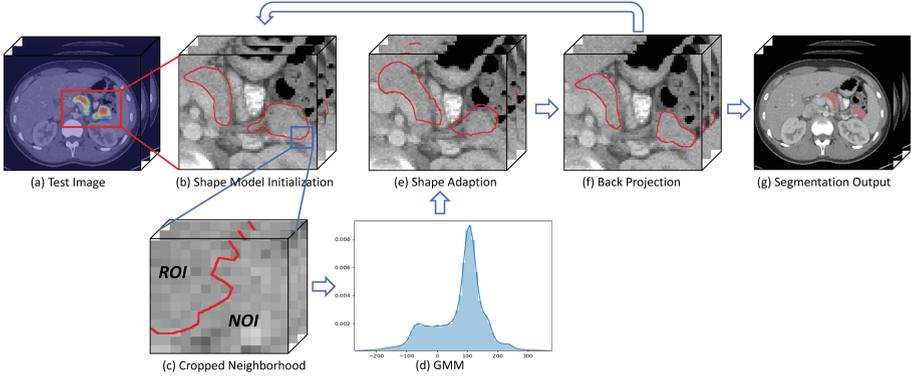


Fig. 1. This figure illustrates the pipeline of segmentation approach: given the test image with probability map (a), the shape model is initialized to fit the detected region (b); considering the neighborhood region around each landmark (c), a Gaussian Mixture Model is trained (d) to guide shape adaption (e); afterwards, project the shape onto statistical shape model (f); we obtain the segmentation output (g) when the convergence is reached.

Gaussian Mixture Model Joint with Probability Map. To find the maximum likelihood of $p(I, \Pi|C)$, we train a Gaussian Mixture Model (*GMM*) based on the image intensity as the pixels are statistically independent from each other. In contrast to conventional mixture models, the probability map Π is adopted as prior weights of different components in the model. Let $X = \{x_1, \dots, x_{n_K}\}$ be a D -dimension image with n_K pixels, the probability density function of *GMM* is defined as:

$$\mathcal{P}(X|\Pi, \Theta) = \prod_{i=1}^{n_K} \{\pi_i \Psi(x_i|\Theta_R) + (1 - \pi_i) \Psi(x_i|\Theta_N)\}, \quad (4)$$

given that $\Psi(X|\Theta_R)$ follows Gaussian distribution where the parameters Θ_R consists of mean value and standard deviation of image intensity, $\Psi(X|\Theta_N)$ is defined in the same way. This *GMM* contains two independent components $\Psi(X|\Theta_R)$ and $\Psi(X|\Theta_N)$ representing *ROI* and *NOI*. As a result, the probability of pixel x_i being each component can be estimated from *GMM* in Eq. 4, we define $w_R(x_i)$ and $w_N(x_i)$ as the probability of pixel x_i being *ROI* and *NOI*:

$$\begin{aligned} w_R(x_i) &= \frac{\pi_i \Psi(x_i|\Theta_R)}{\pi_i \Psi(x_i|\Theta_R) + (1 - \pi_i) \Psi(x_i|\Theta_N)} \\ w_N(x_i) &= \frac{(1 - \pi_i) \Psi(x_i|\Theta_N)}{\pi_i \Psi(x_i|\Theta_R) + (1 - \pi_i) \Psi(x_i|\Theta_N)}. \end{aligned} \quad (5)$$

To release the non-related pixels' influence on *GMM*, only the neighborhood around each landmark is considered in training (cf. Fig. 1(c)). Let $\Omega(x_i)$ donate the cubic neighborhood around the center x_i with radius r , thus each

neighborhood contains $(2r + 1)^3$ pixels. Let $\Omega^+(x_i)$ be the region inside the shape within $\Omega(x_i)$ and $\Omega^-(x_i) = \Omega(x_i) - \Omega^+(x_i)$ be the outside region (cf. Fig. 1(c)). Therefore, the parameters Θ_R, Θ_N are trained within $\int_{x_i \in C} \Omega^+(x_i) dx$ and $\int_{x_i \in C} \Omega^-(x_i) dx$ respectively. Similarly, we obtain the mean probability μ_{wR} and μ_{wN} of being *ROI* and *NOI* by only considering the pixels in region $\int_{i=1}^{n_P} \Omega(x_i) dx$. In this way, more precise probabilities can be obtained by shrinking the region of neighborhood, leading to finer segmentation.

Theoretically, it would be ideal that the pixels inside shape C have the highest probability of being *ROI* and the pixels outside shape C have the highest probability of being *NOI*. Inspired by the popular Mumford-Shah function [11], we form the energy function term:

$$-\log(p(I, \Pi|C)) = \int_{i=1}^{n_P} \int_{j \in \Omega(x_i)} (w_R(x_j) - \mu_{wR})^2 + (w_N(x_j) - \mu_{wN})^2 + (w_R(x_j) - \mu_{wR})(w_N(x_j) - \mu_{wN}) dx, \tag{6}$$

at this stage, the landmarks are fitting to superior positions automatically in terms of the probability rules in Eq. 5. Since the pixels are statistically independent without global constraint, assume the landmark x_i will move along the outward curvature normal with direction $\vec{\mathcal{J}}(x_i)$ to reach the optimal, we compute $\frac{\partial(p(I, \Pi|C))}{\partial(C)} = 0$ to obtain the movement direction $\vec{\mathcal{J}}^*(x_i)$ for each landmark that:

$$\vec{\mathcal{J}}^*(x_i) = \frac{(w_R(x_i) - \mu_{wR})^2 - (w_N(x_i) - \mu_{wN})^2}{(w_R(x_i) - \mu_{wR})(w_N(x_i) - \mu_{wN})}, \tag{7}$$

note that for pixels $x_j \in \Omega^+(x_i)$, $\vec{\mathcal{J}}^*(x_j) < 0$, otherwise for pixels $x_j \in \Omega^-(x_i)$, $\vec{\mathcal{J}}^*(x_j) > 0$. Namely, $\vec{\mathcal{J}}^*(x_i) > 0$ indicates x_i moves along the normal to exterior and $\vec{\mathcal{J}}^*(x_i) < 0$ indicates x_i moves along the inverse direction of outward normal to interior.

Shape Prior. Statistical shape models are demonstrated to have a strong ability in global shape constraint. In this work, we employ the RKPCA method in [9] to train such a robust kernel model $RKSSM(S|\Phi; V; K)$. Differently, we use the model statistics to correct the erroneous modes and estimate the uncertain pieces (cf. Fig. 1(e) to (f)), which means we only focus on the back projection process. Subject to the nonlinearity of kernel space, it is sensitive to initialization of clusters. Furthermore, the shape to be projected onto the model at this stage already contains certain pieces that are supposed to be preserved. Consequently, we improve the back projection of kernel model by assigning a supervised initialization to project onto the optimal cluster. Namely, finding the j^{th} shape in training datasets S_j satisfying $\kappa(C, S_j) = \max(\kappa(C, S_i) : i = 1, \dots, N)$. Employing the shape model in Bayesian model, we consider the prior as:

$$-\log(p(C)) = \left\| \mathbb{P}_n \Phi(C) - \Phi(\hat{C}) \right\|^2 + \lambda \left\| S_j - \hat{C} \right\|^2, \tag{8}$$

the first term is the objective function employed in [9] and we add an additional term with a balance λ . $\mathbb{P}_n \Phi(x)$ denotes the projection of $\Phi(x)$ onto the principal

subspace of Φ . Afterwards, the shape projection is solved by taking gradient $\frac{\partial(-\log(p(C)))}{\partial(\hat{C})} = 0$ and the reconstructed shape vector is derived by:

$$\hat{C} = \frac{\sum_{i=1}^N \gamma_i \kappa(C, S_i) S_i - \lambda S_j}{\sum_{i=1}^N \gamma_i \kappa(C, S_i) - \lambda}, \quad \gamma_i = \sum_{k=1}^N V_i^j K_j V_i^k. \quad (9)$$

Algorithm 1. Algorithm of Segmentation with Bayesian Model

Input: a set of test images $I = \{I_1, \dots, I_{n_S}\}$, the probability maps $\Pi = \{\Pi_1, \dots, \Pi_{n_S}\}$, shape model $RKSSM$, radius $r = 2$

1. Feed shape model to the initial shape C extract from probability map
2. **while** neighborhood radius $r \geq 0$ **do**
3. Train $\mathcal{P}(X|I, \Theta)$ with current shape C in Eq. 4
4. **while** not converged **do**
5. Train GMM in Eq. 6
6. Shape Adaption in terms of Eq. 7 and obtain the new shape C^*
7. **if** $\|C^* - C\|_2 \leq \epsilon$ **break**
8. **end while**
9. Update C by back projection onto $RKSSM$ in Eq. 9
10. Shrink the neighborhood for fine tuning $r = r - 1$
11. **end while**

Output: the segment \hat{Y} from the final shape \hat{C}

3 Evaluation

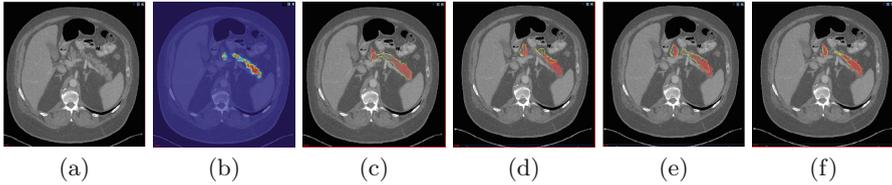
Datasets and Experiments Experiments are conducted on the public NIH pancreas datasets [12], containing 82 abdominal contrast-enhanced 3D CT volumes with size $512 \times 512 \times D$ ($D \in [181, 146]$) under 4-fold cross validation. We take the measures Dice Similarity Coefficient $DSC = 2(|Y_+ \cap \hat{Y}_+|)/(|Y_+| + |\hat{Y}_+|)$ and Jaccard Index $JI = (|Y_+ \cap \hat{Y}_+|)/(|Y_+| \cup |\hat{Y}_+|)$. For statistical shape modeling, we define the kernel trick $\kappa(x_i, x_j) = \exp(-(x_i - x_j)^2/2\sigma^2)$, where the kernel width $\sigma = 150$. In the shape projection, we set the balance term $\lambda = \frac{1}{2\sigma^2}$. We set $r = 2$ at the beginning in shape adaption with GMM . The convergence condition value for shape adaption is $\epsilon = 0.0001$.

Segmentation Results. We compare the segmentation results with related works using the same datasets in Table 1. In terms of the segmentation results, we report the highest 85.32% average DSC with smallest deviation 4.19, and the DSC for the worse case reaches 71.04%. That is to say, our proposed method is robust to extremely challenging cases. We can also find an improvement of JI . More importantly, we can come to the conclusion that the proposed Bayesian model is efficient and robust in terms of the significant improvement (approximately 12% in DSC) from the neural network segmentation. For an intuitive

Table 1. Pancreas segmentation results comparing with the state-of-the-art. ‘-’ indicates the item is not presented.

Method	Mean DSC	Max DSC	Min DSC	Mean JI
Ours	85.32 ± 4.19	91.47	71.04	74.61 ± 6.19
Our DenseUNet	73.39 ± 8.78	86.50	45.60	58.67 ± 10.47
Zhu et al. [4]	84.59 ± 4.86	91.45	69.92	–
Cai et al. [5]	82.40 ± 6.70	90.10	60.00	70.60 ± 9.00
Zhou et al. [3]	82.37 ± 5.68	90.85	62.43	–

view, the segmentation procedure of Bayesian model is shown in Fig. 2, where we compare the segmentation at every stage with the ground truth (in red). The *DSC* for probability map in Fig. 2(b) is 57.30%, and *DSC* for the final segmentation in Fig. 2(f) is 82.92%. Obviously, we find that the segmentation leads more precise by shrinking the radius of neighborhood.

**Fig. 2.** Figure shows the segmentation procedure of NIH case #4: (a) test image I ; (b) probability map II ; (c) initialization for shape model (ground truth mask is in red); (d)–(f) shape adaption with neighborhood radius $r = 2, 1, 0$ respectively.

4 Discussion

Motivated by tackling difficulties in challenging organ segmentation, we integrate deep neural network and statistical shape model within a Bayesian model in this work. A novel optimization principle is proposed to guide segmentation. We conduct experiments on the public NIH pancreas datasets and report the average *DSC* = 85.34% that outperforms the state-of-the-art. In future work, we will focus on more challenging segmentation tasks such as the tumor and lesion segmentation.

Acknowledgments. This research is supported by the National Research Foundation, Prime Minister’s Office, Singapore under its International Research Centres in Singapore Funding Initiative. This work is partially supported by a grant AcRF RGC 2017-T1-001-053 by Ministry of Education, Singapore.

References

1. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
2. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017)
3. Zhou, Y., Xie, L., Shen, W., Wang, Y., Fishman, E.K., Yuille, A.L.: A fixed-point model for pancreas segmentation in abdominal CT scans. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 693–701. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_79
4. Zhu, Z., Xia, Y., Shen, W., Fishman, E.K., Yuille, A.L.: A 3D coarse-to-fine framework for automatic pancreas segmentation. *arXiv preprint arXiv:1712.00201* (2017)
5. Cai, J., Lu, L., Xie, Y., Xing, F., Yang, L.: Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 674–682. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_77
6. Roth, H.R.: Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. *Med. Image Anal.* **45**, 94–107 (2018)
7. Farag, A., Lu, L., Roth, H.R., Liu, J., Turkbey, E., Summers, R.M.: A bottom-up approach for pancreas segmentation using cascaded super-pixels and (deep) image patch labeling. *IEEE Trans. Image Process.* **26**(1), 386–399 (2017)
8. Guo, Z., et al.: Deep LOGISMOS: deep learning graph-based 3D segmentation of pancreatic tumors on CT scans. *arXiv preprint arXiv:1801.08599* (2018)
9. Ma, J., Wang, A., Lin, F., Wesarg, S., Erdt, M.: Nonlinear statistical shape modeling for ankle bone segmentation using a novel kernelized robust PCA. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 136–143. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_16
10. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, p. 3 (2017)
11. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Trans. Image Process.* **10**(2), 266–277 (2001)
12. Roth, H.R., et al.: DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 556–564. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_68