Deep Active Contour Network for Medical Image Segmentation

Mo Zhang[1,2,3], Bin Dong[4,5,1], Quanzheng Li[6]

[1]Center for Data Science, Peking University, Beijing 100871, China; [2]Center for Data Science in Health and Medicine, Peking University, Beijing 100871, China; [3]Laboratory for Biomedical Image Analysis, Beijing Institute of Big Data Research, Beijing 100871, China; [4]Beijing International Center for Mathematical Research (BICMR), Peking University, Beijing 100871, China; [5]Institute for Artificial Intelligence, Peking University, Beijing 100871, China; [6]Center for Advanced Medical Computing and Analysis, MGH/BWH Center for Clinical Data Science, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02115, USA.

Abstract. Image segmentation is vital to medical image analysis and clinical diagnosis. Recently, convolutional neural networks (CNNs) have achieved tremendous success in this task, however, it performs poorly at recognizing precise object boundary due to the information loss in the successive downsampling layers. To overcome this problem, we integrate an active contour model (convexified Chan-Vese model) into the CNN structure (DenseUNet), forming a new framework called deep active contour network (DACN). Instead of manual setting, DACN applies a CNN backbone to learn the initialization and parameters of active contour model (ACM) automatically. The proposed DACN leverages the advantage of ACM to detect object boundaries accurately, which can be trained in an end-to-end differential manner. The experimental results on two public datasets demonstrate the effectiveness of DACN, and the trimap experiment confirms the superior ability of DACN to obtain precise boundary delineation.

Keywords: Active contour model, Boundary delineation, Semantic segmentation

1 Introduction

Semantic segmentation is a central theme in the area of medical image processing. It is a process of splitting an image into several sub regions. In the past decades, various algorithms have been implemented for this topic ranging from thresholding [21], region growing [1], clustering [20] to active contour models (ACMs) [14, 3, 2]. Nowadays, convolutional neural networks (CNN) have significantly improved the performance of many segmentation tasks, mainly benefiting from its powerful ability to learn informative hierarchical features directly from data. However, as illustrated in [10, 25], it is rather difficult for CNN to recognize the object boundary precisely. On the other hand, active contour models

(ACMs) are a series of approaches to fit a curve for the object contour in the image, which tend to generate accurate localization of boundaries. In this circumstance, an intuitive idea is to leverage the advantage of ACM to compensate for the defects of deep learning models.

Recently, researchers have developed many techniques to combine ACM with CNN. Some works presented a novel loss function based on the energy functional of ACM, such as Mumford-Shah loss [16], AC loss [5], level set loss [17] and active contour loss [9]. At the meantime, some investigators attempted to incorporate ACM into deep learning methods in an end-to-end fashion. For instance, the work [11] utilized level set method to assist networks in detecting salient objects more precisely, however, its performance depends heavily on the fine tuning of parameters γ and λ in the level set formulation. In contrast, in order to avoid parameter tuning, DSAC [23] and DARNet [5] trained a CNN to learn ACM parameterizations automatically, nevertheless, they still lack robustness and efficiency as the result is sensitive to the manual initialization of target contours. In addition, Hatamizadeh *et al.* [10] presented a new architecture named DALS combining ACM and CNN to segment various lesions in medical imaging, however, the ACM module in DALS is not trainable only working as a post-processing step.

In this work, to overcome these limitations, we propose an end-to-end deep active contour network (DACN) for segmentation in medical imaging. In the proposed DACN, both the pixel-wise parameter maps and initial contours of ACM are learned directly from data by a CNN. It is a completely automatical and differentiable framework. More specifically, we use the classic DenseUNet as the CNN backbone, and select an improved active contour model (convexified Chan-Vese model) which reduces the original energy functional to a convex minimization. We evaluate our DACN method on two public datasets: Herlev dataset and ISIC 2017 Skin Lesion dataset, where the new framework yields superior results compared to UNet, DenseUNet as well as other competitive models. Moreover, segmentation maps obtained by DACN have more sophisticated delineation of object edges, demonstrated by the additional trimap experiment.

The main contributions of this paper are listed as follows:
1) We present deep active contour network (DACN) for image segmentation, taking advantage of ACM to detect precise boundaries in CNN architecture.
2) We choose an improved Chan-Vese model as the ACM module, and utilize a auxiliary CNN loss function to enhance feature extraction.
3) DACN obtains good performance on Herlev and ISIC dataset, more generally, it can also be used in other segmentation tasks to improve boundary accuracy.

There are 3 main differences between the recent DALS and our DACN: 1) The ACM exploited in DALS is the classic Chan-Vese model [3] with local evolution [19], while the proposed DACN adopts an improved model of Chan-Vese which can find the global minimizer; 2) The ACM used in DALS is not involved into the training process, only serving as a post-processing step, however, our DACN is a fully end-to-end trainable framework; 3) In DALS, parameters of ACM are estimated from the network output by an exponential transformation, while that of DACN are learned directly via the CNN backbone.

Algorithm 1 Split Bregman for model (2)

Given an input image f, select the parameter λ and μ.
Initialize the algorithm by choosing $u^0 = 0, \vec{d}^0 = 0, \vec{b}^0 = 0, c_1^0 = 1, c_2^0 = 0$, and $k = 0$.
while stopping criteria are not met **do**
1: Define $r^k = \left(c_1^k - f\right)^2 - \left(c_2^k - f\right)^2$
2: $u^{k+1} = GS_{\mathrm{GCS}}\left(r^k, \vec{d}^k, \vec{b}^k\right)$
3: $\vec{d}^{k+1} = shrink\left(\nabla u^{k+1} + \vec{b}^k, \frac{1}{\lambda}\right)$
4: $\vec{b}^{k+1} = \vec{b}^k + \nabla u^{k+1} - \vec{d}^{k+1}$
5: Find $\Omega^k = \left\{x : u^k(x) \geq \gamma\right\}$
6: Update $c_1^{k+1} = \int_{\Omega^k} f dx$, and $c_2^{k+1} = \int_{\left(\Omega^k\right)^c} f dx$
7: k=k+1
end while

2 Method

2.1 Convexified Chan-Vese Model

Active contour models (ACMs), also referred as snakes, are firstly proposed by Kass et al. [14] to evolve the contours by solving an energy minimization problem. In contrast to parametric snakes, level set based ACMs define contours implicitly via a level set function in a higher dimension, where the object contours are denoted by the zero level set. In the past decades, a variety of ACMs have been developed to improve the performance of image segmentation, among which the region-based Chan-Vese model [3] is widely used. The energy functional of Chan-Vese model is formulated as:

$$\begin{aligned} F\left(c_1, c_2, C\right) =& \mu \cdot \mathrm{Length}(C) + \nu \cdot \mathrm{Area}(inside(C)) \\ &+ \lambda_1 \int_{inside(C)} \left|u_0(x,y) - c_1\right|^2 dxdy \\ &+ \lambda_2 \int_{\mathrm{outside}(C)} \left|u_0(x,y) - c_2\right|^2 dxdy, \end{aligned} \tag{1}$$

where $u_0(x, y)$ is the raw image, C is a closed curve, the first term $Length(C)$ represents the length of C, the second term denotes the area inside C, and μ,ν,λ_1,λ_2 are scalar parameters to be regulated. Moreover, c_1,c_2 represent the mean values of image $u_0(x, y)$ inside and outside the curve C respectively.

In order to solve the minimal partition problem, the level set method was introduced and the problem amounted to solving a partial differential equation (PDE) by gradient descent [3]. More recently, Chan et al. [2] proposed a convexified version of Chan-Vese model to obtain the global minimizer, which significantly improved the efficiency of computation. Following [2, 8], the original problem can be reduced to a simple convex minimization:

$$\min_{0 \leq u \leq 1} |\nabla u|_1 + \mu \langle u, r \rangle, \quad r = \left(f - c_1\right)^2 - \left(f - c_2\right)^2. \tag{2}$$

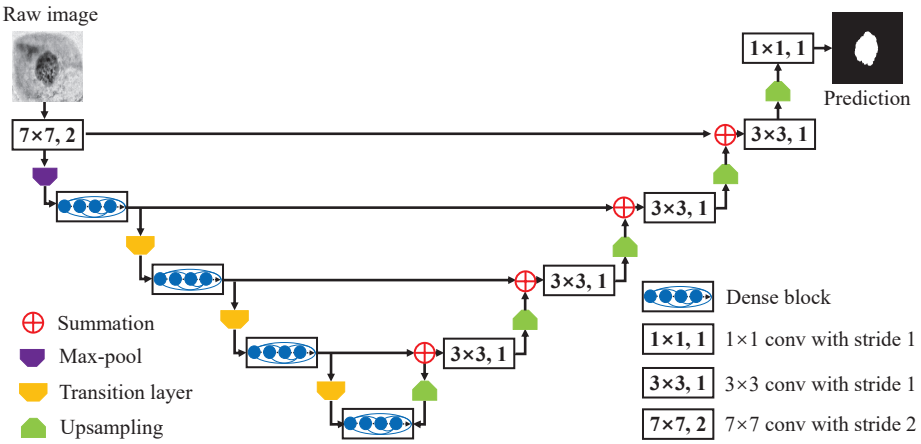

Fig. 1. The structure of DenseUNet. The raw image has a resolution of 256×256, and the setting of channels follows the original work [22].

Here f is the image to be segmented, c_1, c_2 are arbitrary fixed scalars, μ is the parameter balancing the regularization process and data term. Finally, the segmentation output can be defined as:

$$\Omega = \{x : u(x) \geq \gamma\}, \text{ for a.e. } \gamma \in [0, 1]. \tag{3}$$

The work [8] applied the split Bregman algorithm to solve the above problem, while the procedure is summarized in Algorithm 1. GS_{GCS} represents one iteration of the Gauss-Seidel method:

$$\begin{aligned} \alpha_{i,j} &= d^x_{i-1,j} - d^x_{i,j} - b^x_{i-1,j} + b^x_{i,j} + d^y_{i,j-1} - d^y_{i,j} - b^y_{i,j-1} + b^y_{i,j}, \\ \beta_{i,j} &= \frac{1}{4}\left(u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1} - \frac{\mu}{\lambda} r + \alpha_{i,j}\right), \\ u_{i,j} &= \max\left\{\min\left\{\beta_{i,j}, 1\right\}, 0\right\}, \end{aligned} \tag{4}$$

where $b^x_{i,j}$ ($d^x_{i,j}$) denotes the value of the X-direction component of $\vec{b}$ ($\vec{d}$) at pixel (i, j), $\vec{b}$ and $\vec{d}$ are the intermediate variables in Algorithm 1. In the third step, *shrink* denotes the shrinkage operator. For more details about Algorithm 1, we refer the reader to [8].

2.2 The CNN Backbone

In this paper, we choose DenseUNet [22] as our CNN backbone. As shown in Fig. 1, DenseUNet is an encoder-decoder framework which integrates the dense block [12] into the classic UNet. More specifically, in the dense block, output feature maps of each layer are transmitted to all subsequent layers as inputs, making it competent in enhancing feature propagation and alleviating gradient

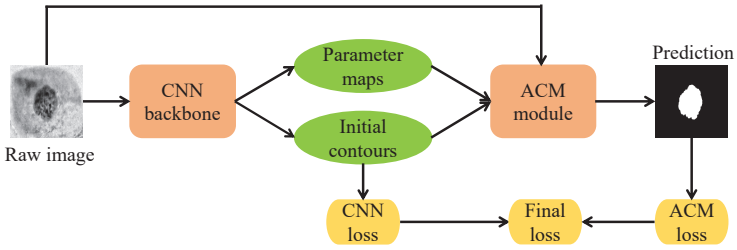

Fig. 2. The architecture of DACN. DACN is an end-to-end trainable framework with a differential ACM module. In contrast to the traditional ACM algorithm, parameters learned by the CNN backbone are pixel-wise functions $\mu(x, y), \lambda(x, y)$ rather than scalars (μ, λ in Algorithm 1).

vanishing. The transition layer is composed of a 1×1 convolution followed by an average pooling operation with stride of 2. More details about DenseUNet can be found in the original work [22].

2.3 The DACN Architecture

In this section, we present deep active contour network (DACN) to deal with medical image segmentation. As shown in Fig. 2, the proposed DACN integrates the classic ACM (convexified Chan-Vese model) with a CNN backbone (DenseUNet) in an end-to-end differentiable manner. Firstly, the initial contours and pixel-level parameter maps ($\mu(x, y), \lambda(x, y)$) of ACM are learned from data by the CNN backbone. Secondly, these contours, maps as well as raw image are transmitted to ACM to evolve the curve iteratively via Algorithm 1. Finally, the output of ACM is compared with ground truth to produce a cross-entropy loss function (named ACM loss). In the same way, the initial contours can also yield a loss function (named CNN loss) to provide auxiliary supervision for the CNN backbone. As shown in Fig. 2, the final loss is established by the CNN loss and ACM loss as follow:

$$\mathcal{L}_{final} = \alpha \cdot \mathcal{L}_{CNN} + \mathcal{L}_{ACM}, \tag{5}$$

where α is a balancing parameter. By minimizing the final loss, the error can be back-propagated through the entire DACN architecture to guide the weights updating in the CNN backbone.

3 Experiments

We compare DACN with the widely used UNet, the baseline DenseUNet as well as recent state-of-the-art methods. To quantify the performance, we utilize several metrics including Dice, Precision, Recall, Accuracy and Hausdorff Distance. Furthermore, in order to evaluate the boundary accuracy, we also conduct the trimap experiment [4, 18]. In the trimap experiment, evaluation metrics are not

computed over the entire image, but only in the region surrounding the object boundary. Specifically, the evaluation region is generated by taking a w pixel band around the edges of objects, where w denotes the width of the target region. Trimap experiment is a more appropriate measurement for users concerned about fine contours. In this work, we calculate Dice in such edge-adjacent area with different widths to measure the model capability to delineate boundaries.

Data Description. We extensively validate DACN on two public medical image datasets: 1) Herlev dataset [13], which contains 917 images from Pap smear tests and corresponding annotations of cervical cell. The entire set is divided into three parts: 562 for training, 171 for validation and 184 for testing; 2) International Skin Imaging Collaboration (ISIC-2017) dataset [6], which aims to segment the skin lesion in dermatoscope images. The training, validation, and testing sets comprise 2000, 150, and 600 images respectively with various resolutions. As for image preprocessing, all images are resized to 256×256, followed by normalization and Contrast Limited Adaptive Histogram Equalization (CLAHE).

Implementation Details. Our implementation is based on the TensorFlow framework and all experiments are carried out on a single NVIDIA GTX1080ti GPU. We apply the Adam algorithm with default parameters to minimize the cross entropy loss function. All models are trained for 30000 epochs with batch size of 4. Each convolutional layer is followed by RELU activation and batch normalization. The hyperparameter α in the loss function is set to 0.01. Both parameter maps $(\mu(x, y), \lambda(x, y))$ of ACM are initialized to null matrix.

Table 1. Quantitative analysis of different methods on Herlev dataset.

Model	Dice	Precision	Recall	Accuracy	Hausdorff Distance
SP-CNN [7]	0.9000	0.8900	0.9100	-	-
PSPNet [24]	0.9070	0.9280	0.9090	-	-
DeeplabV3 [24]	0.9130	0.9170	0.9260	-	-
ASCNet [24]	0.9150	0.9100	0.9380	-	-
UNet	0.9042	0.9053	0.9308	0.9588	36.4990
DenseUNet	0.9327	0.9308	0.9459	0.9731	17.3826
DACN	**0.9454**	**0.9474**	**0.9508**	**0.9763**	**15.5814**

Table 2. Quantitative analysis of different methods on ISIC dataset.

Model	Dice	Precision	Recall	Accuracy	Hausdorff Distance
FocusNetAlpha [15]	0.8404	0.8002	0.8222	**0.9349**	-
UNet	0.7826	0.8766	0.7683	0.9135	39.1774
DenseUNet	0.8403	**0.9364**	0.8087	0.9316	26.1643
DACN	**0.8463**	0.9076	**0.8432**	0.9319	**24.4691**

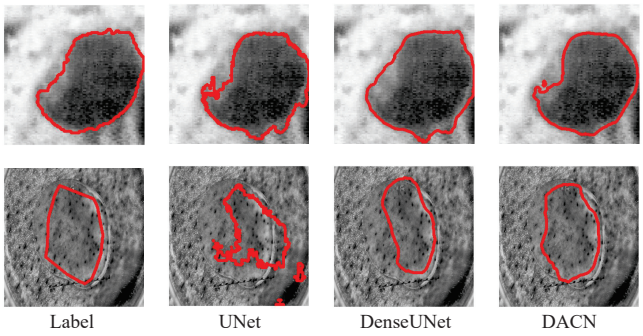

Fig. 3. Comparison of segmentation maps of different models. Top row: an example on Herlev dataset. Bottom row: an example on ISIC dataset. Compared to UNet and DenseUNet, DACN produces more accurate prediction maps.

4 Results

Evaluation of the performance of DACN. Table 1 shows the quantitative comparison of different approaches on Herlev dataset. The proposed DACN outperforms all other models including the CNN backbone DenseUNet and existing state-of-the-art frameworks on this dataset. Table 2 shows the experimental results on ISIC dataset. DACN performs best in Dice (0.8463), Recall (0.8432) and Hausdorff Distance (24.4691). Compared to the baseline DenseUNet, DACN yields a higher Recall (0.8432) with a considerable improvement (3.45%), suggesting that DACN produces less false negatives (FN). Accuracy (0.9319) of DACN is close to the optimal one (0.9349). Although Precision of DACN is lower than that of DenseUNet, it is still at a high level. Besides, it is worth mentioning that DACN gets good performance in Hausdorff Distance on both datasets. Hausdorff Distance measures the boundary distance between two surfaces, indicating that DACN tends to produce more precise localization of object boundary. Segmentation results of different methods are shown in Fig. 3. It can be observed that the target boundaries are really complex and low-contrast to background, thus resulting in the failure of UNet and DenseUNet. However, in such case, DACN still generates relatively accurate prediction maps.

Evaluation of the edge location precision. Specifically, we make trimap experiment to quantify the performance of boundary delineation. In order to compare the property of different models, we display the bar charts about related results in Fig. 4. Compared to the CNN backbone (DenseUNet), DACN gains higher Dice in all settings on both datasets. When the band widths of evaluation region are 5, 10, 15, 20 pixels, on ISIC dataset, the improvements of DACN across Dice coefficient are 4.40%, 3.59%, 3.12% and 2.45% respectively. With the increasing of band width, the gap between the CNN backbone and DACN reduces gradually, which may account for the slight improvement of Dice computed over the entire image. The above results demonstrate that DACN has

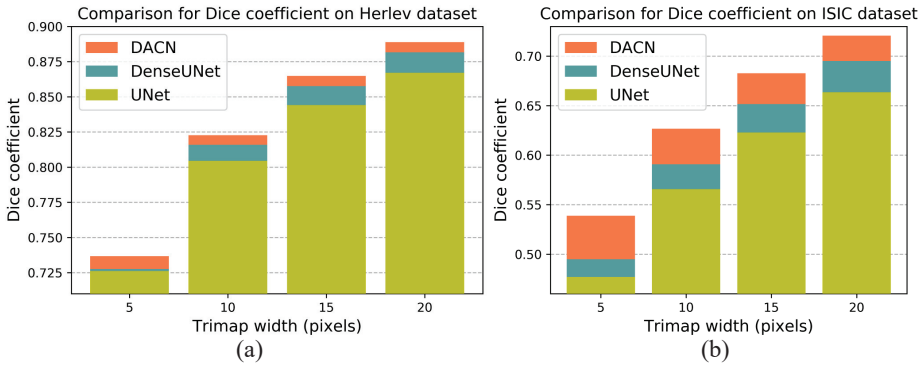

Fig. 4. The results of trimap experiment on two datasets. (a) The bar chart about trimap experiment on Herlev dataset. (b) The same bar chart on ISIC dataset. Dice is calculated in the region surrounding the object boundaries, where trimap width represents the band of this region.

strong capability to identify accurate boundaries, benefiting from the utilization of active contour models (ACMs).

For the sake of exploring the working mechanism of DACN, we show the learned initial contours and pixel-level parameter maps $\mu(x, y)$, $\lambda(x, y)$ in Fig. 5. In the initial contours, some pixels around the object boundary are difficult samples with weak confidence, making it necessary to evolve the contour more accurately by ACMs. In addition, in the $\mu(x, y)$ map, locations which have similar intensity with the target in raw image are inclined to generate higher μ values. One possible explanation is that the learned parameter maps can provide additional guidance for DACN to evolve towards the object boundary, which compensates for the faults of CNN.

5 Conclusion

In this paper, we present a novel deep active contour network (DACN) for medical image segmentation, which integrates ACM (convexified Chan-Vese model) into the DenseUNet architecture in an end-to-end differential manner. By leveraging the advantage of ACM to locate object edges, the proposed DACN tends to generate more accurate segmentation of contours. Our DACN has better performance on two public datasets compared to UNet, DenseUNet as well as several state-of-the-art models, especially for boundary delineation. Additionally, DACN can also be applied to multi-class semantic segmentation, where the issue of multi-class semantic segmentation should be decomposed into several single-class segmentations. In the future, it is worth further investigation about the working mechanism of DACN.

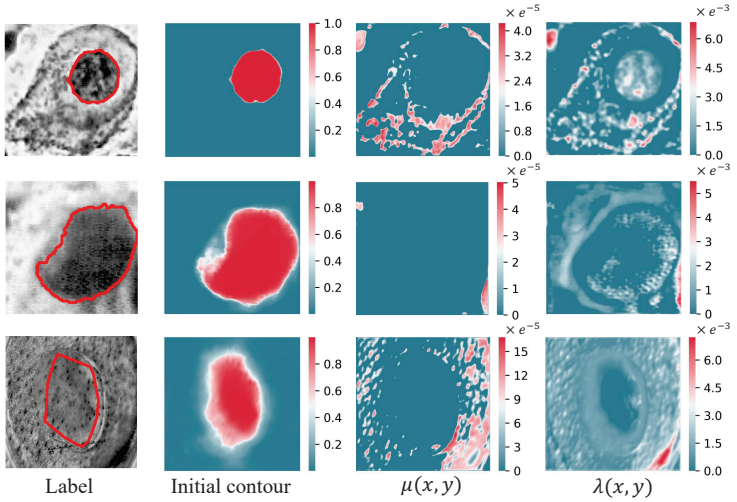

Fig. 5. Illustrations of the learned initial contours and pixel-wise parameter maps by the CNN backbone. The top two rows: examples on Herlev dataset. Bottom row: an example on ISIC dataset.

References

1. Adams, R., Bischof, L.: Seeded region growing. IEEE Transactions on pattern analysis and machine intelligence **16**(6), 641–647 (1994)
2. Chan, T.F., Esedoglu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. SIAM journal on applied mathematics **66**(5), 1632–1648 (2006)
3. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Transactions on image processing **10**(2), 266–277 (2001)
4. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 801–818 (2018)
5. Chen, X., Williams, B.M., Vallabhaneni, S.R., Czanner, G., Williams, R., Zheng, Y.: Learning active contour models for medical image segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 11632–11640 (2019)
6. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 168–172. IEEE (2018)
7. Gautam, S., Bhavsar, A., Sao, A.K., Harinarayan, K.: Cnn based segmentation of nuclei in pap-smear images with selective pre-processing. In: Medical Imaging 2018: Digital Pathology. vol. 10581, p. 105810X. International Society for Optics and Photonics (2018)

8. Goldstein, T., Bresson, X., Osher, S.: Geometric applications of the split bregman method: segmentation and surface reconstruction. Journal of Scientific Computing **45**(1-3), 272–293 (2010)
9. Gur, S., Wolf, L., Golgher, L., Blinder, P.: Unsupervised microvascular image segmentation using an active contours mimicking neural network. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 10722–10731 (2019)
10. Hatamizadeh, A., Hoogi, A., Sengupta, D., Lu, W., Wilcox, B., Rubin, D., Terzopoulos, D.: Deep active lesion segmentation. In: International Workshop on Machine Learning in Medical Imaging. pp. 98–105. Springer (2019)
11. Hu, P., Shuai, B., Liu, J., Wang, G.: Deep level sets for salient object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2300–2309 (2017)
12. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
13. Jantzen, J., Norup, J., Dounias, G., Bjerregaard, B.: Pap-smear benchmark data for pattern classification. Nature inspired Smart Information Systems (NiSIS 2005) pp. 1–9 (2005)
14. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International journal of computer vision **1**(4), 321–331 (1988)
15. Kaul, C., Pears, N., Manandhar, S.: Divided we stand: A novel residual group attention mechanism for medical image segmentation. arXiv preprint arXiv:1912.02079 (2019)
16. Kim, B., Ye, J.C.: Mumford–shah loss functional for image segmentation with deep learning. IEEE Transactions on Image Processing **29**, 1856–1866 (2019)
17. Kim, Y., Kim, S., Kim, T., Kim, C.: Cnn-based semantic segmentation using level set loss. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1752–1760. IEEE (2019)
18. Kohli, P., Torr, P.H., et al.: Robust higher order potentials for enforcing label consistency. International Journal of Computer Vision **82**(3), 302–324 (2009)
19. Lankton, S., Tannenbaum, A.: Localizing region-based active contours. IEEE transactions on image processing **17**(11), 2029–2039 (2008)
20. Lee, T.H., Fauzi, M.F.A., Komiya, R.: Segmentation of ct brain images using k-means and em clustering. In: 2008 Fifth International Conference on Computer Graphics, Imaging and Visualisation. pp. 339–344. IEEE (2008)
21. Li, J., Zhu, S., Bin, H.: Medical image segmentation techniques. Sheng wu yi xue gong cheng xue za zhi= Journal of biomedical engineering= Shengwu yixue gongchengxue zazhi **23**(4), 891–894 (2006)
22. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., Heng, P.A.: H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. IEEE transactions on medical imaging **37**(12), 2663–2674 (2018)
23. Marcos, D., Tuia, D., Kellenberger, B., Zhang, L., Bai, M., Liao, R., Urtasun, R.: Learning deep structured active contours end-to-end. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8877–8885 (2018)
24. Zhang, M., Zhao, J., Li, X., Zhang, L., Li, Q.: Ascnet: Adaptive-scale convolutional neural networks for multi-scale feature learning. arXiv preprint arXiv:1907.03241 (2019)
25. Zhang, Y., Chung, A.C.: Deep supervision with additional labels for retinal vessel segmentation task. In: International conference on medical image computing and computer-assisted intervention. pp. 83–91. Springer (2018)