

A Deep Morphological Active Contour Model for Medical Image Segmentation

Ali Hatamizadeh

Computer Science Department
University of California, Los Angeles
ahatamiz@cs.ucla.edu

Xingjian Yan

Computer Science Department
University of California, Los Angeles
xjyan@g.ucla.edu

Abstract

In the proposed work, we created a fast and reliable end-to-end framework for medical image segmentation applications by employing a deep Convolutional Neural Network (CNN) and a morphological active contour model. Traditionally, active contour based models have been utilized extensively for different segmentation tasks including medical image analysis, but their accuracy is very dependent on finding the right parameters and contour initialization. Moreover, although recent advances in deep CNNs have made it possible to train end-to-end segmentation networks with good performance, such networks tend to struggle with generalization issues as they are deployed on datasets that they have not been trained on. In our proposed pipeline, We leveraged a deep CNN, namely U-Net, to generate the initial segmentation masks, and then used such masks the the morphological model for creating the final mask. The proposed method has been implemented on a unified platform and is very efficient for both training and deployment. We tested this framework the Lung Tissue Research Consortium (LTRC) data set for lung segmentation and results indicate enhanced performance in comparison to state-of-the-art segmentation methods.

1. Introduction and Related Work

With the ever-increasing computational power and readily available data, deep learning-based approaches have gained a lot of popularity for computer vision tasks such as object detection and segmentation among researchers. In the field of medical image analysis, various architectures have been proposed for segmentations, among which U-Net [17] has shown great prominence. U-Net is a 2D network with an architecture that is composed of two paths; a contraction path which captures the context and an expansion path to ensure precise localization. Based on this idea of two symmetric paths of contraction and expansion, several other networks have been proposed; 3D U-Net [4] proposes a dense volumetric segmentation network using sparse anno-

tations. V-Net [13] is a volumetric fully convolutional neural network (FCNNs), that has been shown to perform well on segmentation of MRI prostate volumes. Dense V-Net [5] proposes a similar architecture to V-Net and also takes advantage of densely connected feature stacks. While the CNN-based segmentation approaches have shown to perform well on datasets in which such networks have been trained on, they lack of robustness for cross-validation cases from other datasets. As a result, there exists a need for addition of another component that can improve the robustness and generalization capabilities of the model. We propose to use a fast and stable active contour model which can complement the output of the CNN-based models such as U-Net that has been used in this work.

Since the introduction of active contour models or snakes [8], they have been extensively used for various computer vision tasks such as segmentation [19], object boundary detection [14] and tracking [16], to mention but a few. These models leverage a variational formulation in which the image energy functional evolves the curve or surface via minimization of the functional in a steepest descent manner. A popular framework for active contour models is the level-set formulation, in which the curve of interest is evolved by propagation of the zero level-set in a dynamic partial differential equation (PDE) [15]. Solution of such a PDE can become computationally expensive and may pose stability issues. As a result, some approaches have resorted to avoid solving the PDE directly via evolving the curve by adding and removing the points that neighbor the desired interface [12]. Other approaches have proposed to substitute the mathematical terms of the PDEs that correspond to contour evolution with morphological operations with the same infinitesimal behavior. As a result, the numerical solution of the PDE can be approximated with orderly applications of such morphological operators [18]. One problem with the original model is its highly dependent behavior on proper initialization of the parameters. As a result, several models such as geodesic active contours (GAC) and Active contours without edges (ACWE) have been proposed to mitigate these issues [9, 1, 3]. GAC model computes

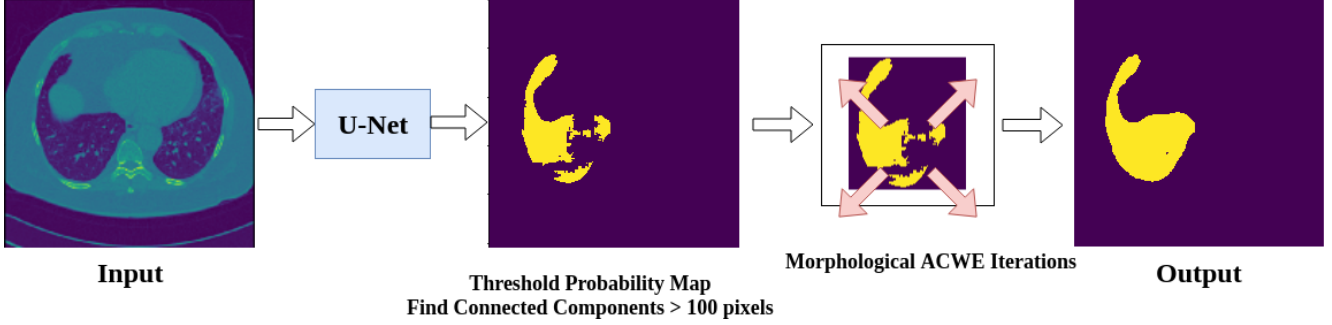


Figure 1. Deep MAC method pipeline.

the geodesics or minimal distance curves based on a Riemannian space defined by the image content. Consecutively, ACWE model, which is based on Mumford-Shah functional for segmentation and level-sets, does not require well-defined boundaries and is less dependent on prior initializations.

Several efforts have also been proposed to combine deep learning approaches with active contour models in order to address some of the aforementioned deficiencies. Hu *et al.* [7] proposed a model in which the network learns a level-set function for salient objects in order to produce more accurate boundaries. Le *et al.* [10] reformulated level-sets with deep recurrent neural networks for semantic segmentations. Hoogi *et al.* [6] proposed an adaptive estimation of active contour parameters by employing a CNN which outputs the corresponding probability of a point’s location with respect to the boundaries of the curve.

In this work, we utilize a 2D U-Net architecture for generation of input masks. Then these masks will be passed to a morphological ACWE model for further refinement and obtaining the final boundary. We utilize this framework for a lung segmentation task on LTRC dataset.

2. Method

2.1. U-Net

During the recent years, FCNNs have been leveraged successfully for semantic segmentation tasks [11]. These networks can accept an input with an arbitrary size and then produce a corresponding output with the same size. The working principle behind such networks is to construct a contracting path which generates high resolution features that will be combined with up-sampled outputs generated from supplementing the network with successive up-sampling operators instead of pooling operators. Consecutively, this combined output is fed to a final convolution layer with the aim of creation of a more precise output. The same strategy is also pursued in the U-Net architecture, except for the fact that the up-sampling path now consists of a larger number of feature channels. As a result, the network

has two symmetric contracting and up-sampling paths. In the proposed work, We utilize a 2D U-Net architecture with FCN in order to directly produce left and right lung segmentation probability masks. Figure 2 shows the outline of the architecture that we used. The input to the network is a single slice of CT image of height 512, width 512, and channel 1. The output of the FCN network, generated by the last convolution layer and sigmoid activation function, have the same height and width as the input, but with 2 channels. Each channel of the output segmentation probability map corresponds to the left and right lungs respectively. Moreover, LeakyRelu, which is a non-linear nonsaturating activation function, was used with $\alpha = 0.01$ after all of the 3×3 convolution layers except the for the last one. Batch normalization was applied before activation; dropout with *keepratio* = 0.5, and L2 regularization of convolution kernels with *weightdecay* = 5×10^{-5} was applied to prevent over-fitting the FCN model.

$$H(P, L) = - \sum_i [L_i \log(P_i) + (1 - L_i)(1 - P_i)] \quad (1)$$

For training we minimize the cross entropy loss between the output segmentation mask against the ground truth. The ground truth label for each input image is also a *Height* \times *Width* \times *Channel* = $512 \times 512 \times 2$ binary mask, where indicators of right and left lung appear on the first and second channels respectively. Equation 1 shows the idea of this simple loss, P and L are the FCN output and label. Each pixel of the FCN output is considered as independent in this equation, and the total loss is sum of losses of all pixels. A FCN, without the use of any prior knowledge, combined with this simple loss function, are expected to produce noisy and non-smooth segmentations. However, we hope to achieve, in our method, refinement of the rough FCN output by using it only as an initialization of the morphological ACWE.

2.2. Morphological Operators

With the motivation of approximating the PDEs that control the curve evolution of the snake model, morphological

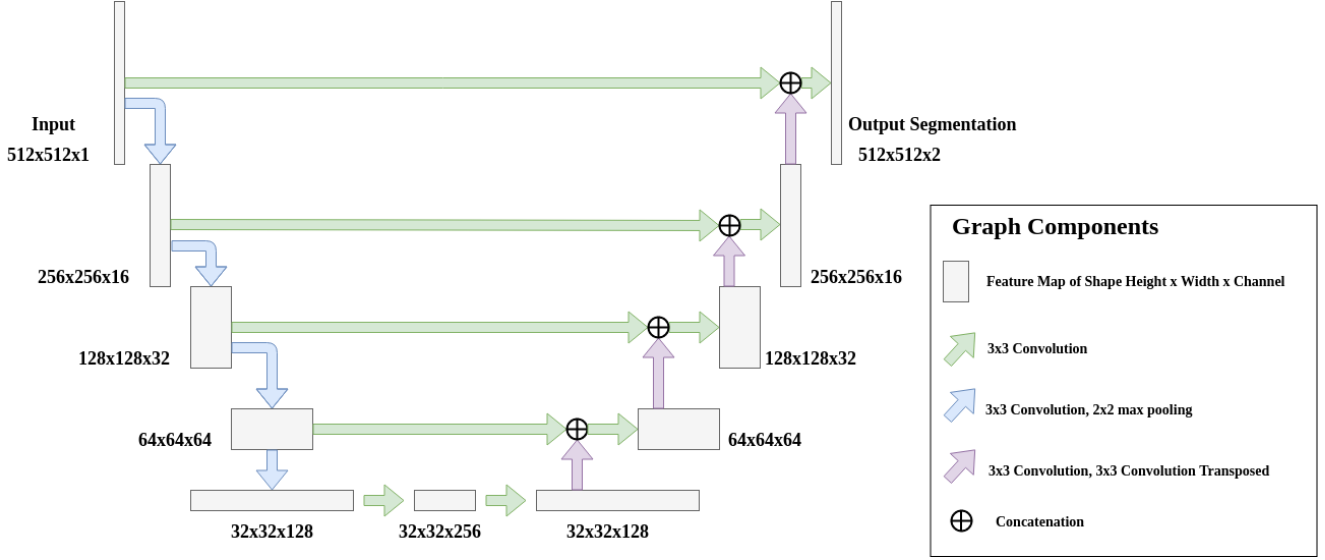


Figure 2. Proposed architecture of 2D U-Net. Size of various feature maps have been illustrated.

operators such as dilation and erosion can be utilized. The dilation operator D_h and erosion operator E_h with radius h of function u can be defined as in equation (2) [12]:

$$\begin{aligned} D_h u(x) &= \sup_{y \in hB(0,1)} u(x+y) \\ E_h u(x) &= \inf_{y \in hB(0,1)} u(x+y) \end{aligned} \quad (2)$$

In which $B(0,1)$ denotes a neighborhood of radius 1 whose center is located at 0 and hB is the scaled set of B . Due to their infinitesimal behavior, it can be shown that the successive application of morphological operators such as dilation D_h and erosion E_h with very small radius can be equivalent to the solution of equation (3) [12]:

$$\frac{\partial u}{\partial t} = \pm |\nabla u| \quad (3)$$

Moreover, if SI_h and IS_h denote the sup-inf and inf-sup of aforementioned morphological operators, then a curvature morphological operator [2, 12] can be defined as in the equation (4):

$$SI_{\sqrt{h}} \circ IS_{\sqrt{h}} u = \frac{SI_{2\sqrt{h}} u + IS_{2\sqrt{h}} u}{2} \quad (4)$$

Mrquez-Neila *et al.* [12] showed that successive of the morphological curvature operator presented in equation (4) can approximate the solution of the equation (5):

$$\frac{\partial u}{\partial t} = |\nabla u| \cdot \left(\text{div} \left(\frac{\nabla u}{|\nabla u|} \right) \right) \quad (5)$$

2.3. Morphological ACWE

The corresponding energy functional of a curve C in the ACWE model can be represented as in the equation (6) :

$$\begin{aligned} F(c_1, c_2, C) &= \mu \times \text{length}(C) + \nu \times \text{area}(\text{inside}(C)) \\ &+ \lambda_1 \int_{\text{inside}(C)} \|I(x) - c_1\| dx \\ &+ \lambda_2 \int_{\text{outside}(C)} \|I(x) - c_2\| dx \end{aligned} \quad (6)$$

Where parameters such as μ , ν , λ_1 and λ_2 decide the strength of each term, $I(x)$ is the image, and values of c_1 and c_2 are calculated according to equation (7) in order to minimize F [12]:

$$c_1(C) = \frac{\int_{\text{inside}(C)} I(x) dx}{\int_{\text{inside}(C)} dx}, c_2(C) = \frac{\int_{\text{outside}(C)} I(x) dx}{\int_{\text{outside}(C)} dx} \quad (7)$$

Consecutively, the Euler-Lagrange equation of the functional can be represented as in the following [12]:

$$\frac{\partial u}{\partial t} = |\nabla u| \left(\mu \text{div} \left(\frac{\nabla u}{|\nabla u|} \right) - \nu - \lambda_1 (I - c_1)^2 + \lambda_2 (I - c_2)^2 \right) \quad (8)$$

The curve of control will evolve according to equation (8) to minimize the functional F . It is noteworthy to mention that the location of point x can impact the magnitude of terms

$\lambda_1|\nabla u|(I - c_1)^2$ and $\lambda_2|\nabla u|(I - c_2)^2$ whereas if the former is bigger than the latter, then the point x is located in the exterior of the curve and vice versa. The term $\nu|\nabla u|$ denote the balloon force which determine the dilation and erosion PDEs and will be approximated by a morphological approach. In addition, the term $|\nabla u|\mu\text{div}(\frac{\nabla u}{|\nabla u|})$ represents the smoothing force which will be approximated by successive applications of the smoothing operator.

The morphological algorithm of ACWE model for a $d - 1$ hypersurface that has been defined as a level-set of $u : \mathbb{R}^d \rightarrow \mathbb{R}$ can then be represented as in equation (9) [12]:

$$u^{n+\frac{1}{3}}(x) = \begin{cases} (D_d u^n)(x) & \text{if } \nu > 0 \\ (E_d u^n)(x) & \text{if } \nu < 0 \\ u^n(x) & \text{otherwise} \end{cases}$$

$$u^{n+\frac{2}{3}}(x) = \begin{cases} 1 & \text{if } |\nabla u|^{n+\frac{1}{3}}(\lambda_1(I - c_1)^2 - \lambda_2(I - c_2)^2)(x) < 0 \\ 0 & \text{if } |\nabla u|^{n+\frac{1}{3}}(\lambda_1(I - c_1)^2 - \lambda_2(I - c_2)^2)(x) > 0 \\ u^{n+\frac{1}{3}} & \text{otherwise} \end{cases} \quad (9)$$

$$u^{n+1}(x) = (SI_{\sqrt{h}} \circ IS_{\sqrt{h}})^\mu u^{n+\frac{2}{3}}(x)$$

2.4. Combining U-Net with Morphological ACWE

As elaborated above, MACWE are designed to produce smooth and robust segmentations. For this project we test out the effectiveness of the MACWE on the multi-class left and right lung segmentation by using the FCN output as initialization. As shown in figure 1, we first binarize the 2 channel FCN probability mask output at threshold 0.5. Due to the fact that some of the binarized segmentation seems fragmented, we performed connected components analysis on the segmentation map and kept only the connected components larger than 100 pixels. We use this post processed FCN output segmentation and the definition of the initiation of the MACWE and run the MACWE over 50 iterations to get the final segmentation output.

3. Experiment and Results

3.1. Experiment Data

Data from LUNA 16 Challenge is used for our experiments. The Dataset contains a total of 888 patients, all of which have corresponding left and right lung segmentation that is generated by semi-automatic methods. Of the 888 patients we randomly selected 200 cases for training, and 50 cases for testing. For each selected case, we select the middle 50% of the CT slices and made each slice an individual training or testing sample. In total we have more than 23000 slices for training, and more than 6000 slices for testing.

3.2. Training FCN

The CT image slices have HU values ranging from -3000 to 4000 . The input images are first value clipped at -800 and 800 and then scaled to the unit range from 0 to 1 . In this way the necessary features of the lung are preserved, and also reduces risk of feature map and gradient overflowing. The FCN network was trained on a single Nvidia 1080ti GPU, which supports a maximum of 20 images per batch. Adam optimizer was used and exponential learning rate decay was applied for smoother convergence. The training loss quickly converges within 10 epochs of training, and the training binary dice coefficient reach over 0.9 after convergence.

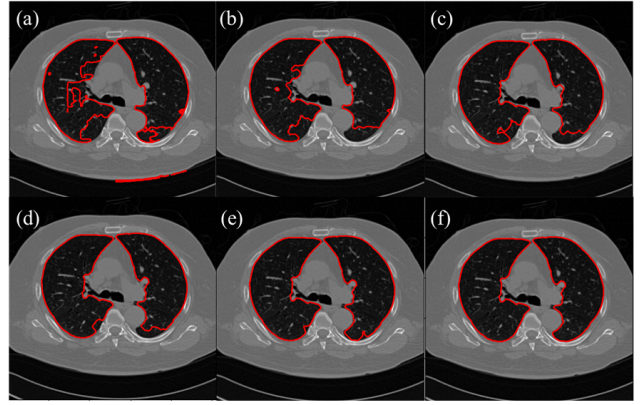


Figure 3. Evolution of the generated segmentation in the Deep MAC model. From left, the 2D U-Net outputs the segmentation masks. These masks are then used for initialization of the ACWE model as it iterates towards the final segmentation solution

3.3. MACWE refinement of FCN output

After training FCN network, we test the entire Deep MAC pipeline on the test dataset. Figure 3.2 shows an example of the MAC evolution. The red contour indicate the level set of the current iteration; from left to right figure 3.2 shows how the level set initially noisy and having holes and disconnections, but gradually evolves due to image forces and finally covers the entire lung.

The current implementation of the MAC is not well optimized, and 50 iterations takes 13 to 15 seconds depending on the CPU power. The dice score of the FCN output mask against ground truth and that of the MACWE output against ground truth are compared. Figure 4 shows the individual dice scores of 500 random samples from the test data set. The dice scores in figure 4 are sorted from highest to lowest. We can see that the top 150 dice scores of the MACEW output are higher than that of the FCN output. The average dice score of the top 200 samples of the MACEW output is 0.927 higher than that of FCN output, which is 0.915 . Figure 5 shows the moving average of all 500 random samples

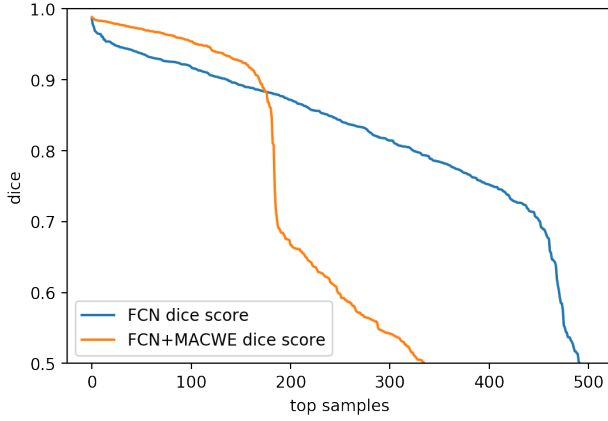


Figure 4. Dice scores of 500 random samples from test set; sorted from highest to lowest. Left and right lung are combined

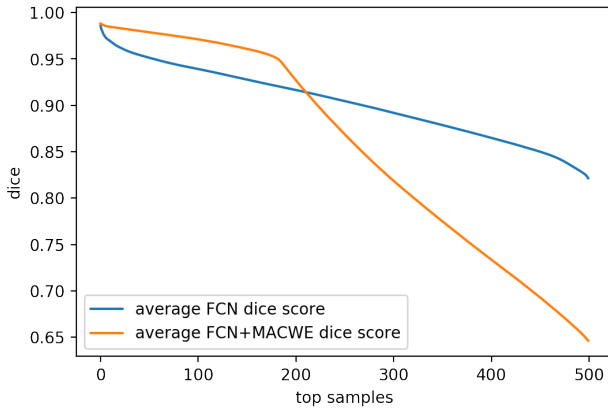


Figure 5. Moving average dice scores of 500 random samples from test set sorted from high to low. Left and right lung are combined

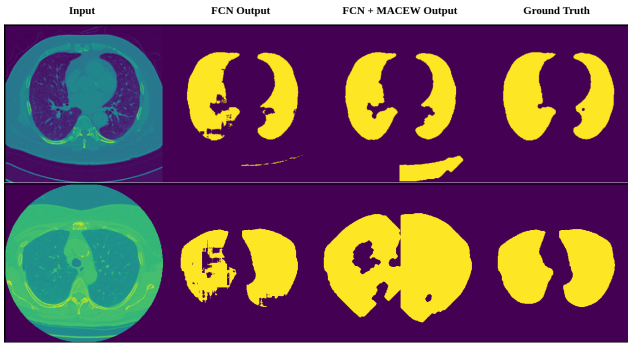


Figure 6. Remaining issues

sorted from high to low.

However, there are still cases where MACWE fails to provide a the a reasonable refinement of the FCN output. Figure 6 show two example of current issues. The first row show that MACWE though able to complete a smooth and clear lung segmentation, is rather sensitive to noises in the initialization, especially that outside of the lung. This

problem can be easily solved by doing another CCA after MACWE iterations. The second row of figure 6 show a case where the final segmentation of MACWE goes well overboard, where as the predicted segmentation of both left and right lung of FCN are reasonable. Clearly, there are much space left for improving the initial FCN output, and improving the robustness of MACWE.

4. Discussion and Future Work

References

- [1] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International journal of computer vision*, 22(1):61–79, 1997. 1
- [2] F. Catté, F. Dibos, and G. Koepfler. A morphological scheme for mean curvature motion and applications to anisotropic diffusion and motion of level sets. *SIAM Journal on Numerical Analysis*, 32(6):1895–1909, 1995. 3
- [3] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on image processing*, 10(2):266–277, 2001. 1
- [4] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432. Springer, 2016. 1
- [5] E. Gibson, F. Giganti, Y. Hu, E. Bonmati, S. Bandula, K. Gurusamy, B. Davidson, S. P. Pereira, M. J. Clarkson, and D. C. Barratt. Automatic multi-organ segmentation on abdominal ct with dense v-networks. *IEEE Transactions on Medical Imaging*, 2018. 1
- [6] A. Hoogi, A. Subramaniam, R. Veerapaneni, and D. L. Rubin. Adaptive estimation of active contour parameters using convolutional neural networks and texture analysis. *IEEE transactions on medical imaging*, 36(3):781–791, 2017. 2
- [7] P. Hu, B. Shuai, J. Liu, and G. Wang. Deep level sets for salient object detection. In *IEEE CVPR*, 2017. 2
- [8] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988. 1
- [9] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contour models. In *Computer vision, 1995. proceedings., fifth international conference on*, pages 810–815. IEEE, 1995. 1
- [10] T. H. N. Le, K. G. Quach, K. Luu, C. N. Duong, and M. Savvides. Reformulating level sets as deep recurrent neural network approach to semantic segmentation. *IEEE Transactions on Image Processing*, 27(5):2393–2407, 2018. 2
- [11] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 2
- [12] P. Marquez-Neila, L. Baumela, and L. Alvarez. A morphological approach to curvature-based evolution of curves and surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1):2–17, 2014. 1, 3, 4

- [13] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 565–571. IEEE, 2016. 1
- [14] A. K. Mishra, P. W. Fieguth, and D. A. Clausi. Decoupled active contour (dac) for boundary detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2):310–324, 2011. 1
- [15] S. Osher and R. P. Fedkiw. Level set methods: an overview and some recent results. *Journal of Computational physics*, 169(2):463–502, 2001. 1
- [16] Y. Rathi, N. Vaswani, A. Tannenbaum, and A. Yezzi. Tracking deforming objects using particle filtering for geometric active contours. *IEEE transactions on pattern analysis and machine intelligence*, 29(8):1470–1475, 2007. 1
- [17] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1
- [18] R. Van Den Boomgaard and A. Smeulders. The morphological structure of images: The differential equations of morphological scale-space. *IEEE transactions on pattern analysis and machine intelligence*, 16(11):1101–1113, 1994. 1
- [19] L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the mumford and shah model. *International journal of computer vision*, 50(3):271–293, 2002. 1