

Active Contours for Multi-Region Segmentation with a Convolutional Neural Network Initialization

Erik Carbajal-Degante ^a, Steve Avendaño ^b, Leonardo Ledesma^a, Jimena Olveres^c, and Boris Escalante-Ramírez ^c

^aPosgrado en Ciencia e Ingeniería de la Computación, Universidad Nacional Autónoma de México, Mexico City, Mexico

^bFacultad de Ciencias, Universidad Nacional Autónoma de México, México City, México

^cDepartamento de Procesamiento de Señales, Facultad de Ingeniería, Universidad Nacional Autónoma de México, Mexico City, Mexico

ABSTRACT

Active contour-based methods are widely popular in the image segmentation field. Basically, they perform a semi-automatic region identification by partitioning the image content mainly into the foreground and background. Nevertheless, the accurate delimitation still remains as an important challenge which usually depends on how close the initial contour is placed to the object of interest (OI). Several applications of active contours require the user interaction to give prior information about the initial position as the first step, which drives the tool substantially dependent on a manual process. This paper describes how to overcome this limitation by including the expertise provided by the training stage of a Convolutional Neural Network (CNN). Despite CNN methods require a large dataset or data augmentation techniques to improve their results, the combined proposal accomplishes a pre-segmentation task with a reduced number of images to obtain the assumed locations for each OI. These results are used to initialize a multiphase active contour model that follows a level set scheme to lead a smoother multi-region segmentation with less effort. Experiments of this approach are included to compare classic techniques of contour initialization and show the benefits of our proposal.

Keywords: Active contours, convolutional neural network, initialization, multi-region segmentation

1. INTRODUCTION

The problem of image segmentation dates back to the beginning of computer vision. Segmentation basically consists of partitioning a given image into a collection of objects, or finding edges that delimit those objects. Since the image acquisition nature provides an infinite number of possible scenes, several algorithms have been proposed to improve the results in the accurate identification of an object of interest (OI). Pixel features such as intensity level, color and texture have been the most used to solve the segmentation challenge.¹

Among many other approaches, variational methods for image processing address the problem of segmentation by determining the solution of an equation corresponding to the minimization of certain functional.² An important contribution to this area includes region-based methods of active contours, which were originally conceived to separate different regions by means of evolving contours related to an specific functional.³ Several advantages can be quoted since active contours use the theory of level sets,^{4,5} adapting the problem in a discrete model so they handle complex topological changes automatically and they can be implemented by efficient stable numerical schemes. Furthermore, active contours do not require a training stage making this tool an unsupervised technique with numerous applications.⁶⁻⁸

One of the main drawbacks of active contours is the speed of convergence to a global minimum, which depends on how close the initial contour is placed to the OI.⁹ This fact makes the initialization stage to play a very important role in the final segmentation result and the number of iterations required. Recent works have

Further author information: (Send correspondence to E.C-D)
E.C-D.: E-mail: erikycd@gmail.com

successfully reformulated this problem as a convex optimization problem, allowing global minimums to be found for a non-convex functional.^{10, 11}

Another way to tackle this problem is by making use of algorithms such as Convolutional Neural Networks, which are a particular kind of Artificial Neural Networks (ANNs).¹² This type of parametric models are specific for analyzing data that has grid-like topology in account of spatial/temporal interaction of samples within a neighborhood, such as images and time series.

The principal operation employed is the convolution (eg. standard convolution, separable convolution, atrous convolution),^{13, 14} by using filters with learnable weights that pass over the image to highlight and extract the most relevant features. Other operations includes activation functions (ReLU, Softmax, Sigmoid, etc.), and non-linear filters (eg. max-pooling or average pooling).

This type of models has shown a great performance in the field of computer vision, particularly in tasks such as detection, classification and segmentation.¹⁵ Nevertheless, the major disadvantage is the great amount of training data required to achieve a good performance. Moreover, the risk of losing interpretability increases whereas more complex models are built.

2. MATHEMATICAL PRELIMINARIES

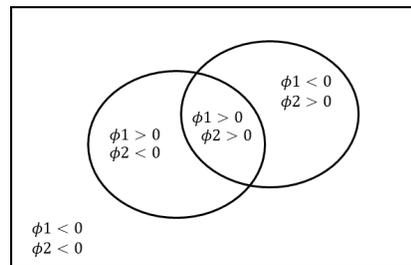
2.1 Four phase active contours based on level set

Active contour models, also called Snakes,¹⁶ have been efficient methods of unsupervised image segmentation. First, an initial curve is placed over the image. Second, by iteratively performing the minimization of a functional, shrinkage and expansion operations are carried out for capturing the region of interest within the curve. These models can be classified into two groups: **(1)** those based on edge information to guide the evolution of active contours, **(2)** and those based on regions (local and global information). A model in the latter group will be discussed in the current work.

Besides the classic methods of active contours based on regions, which use a single contour to separate up to two regions (background and foreground), the multiphase formulation arises as an extension to identify multiple regions.^{17, 18} The multiphase model requires several contours which can intersect. Each contour C_i expresses a balance of regularity of intensity levels, enclosing similar content within each curve.

The principle of multiphase formulation establishes a correspondence between $n = 2^m$ regions or phases and a partition defined by m level set functions, $C_i = \{(x, y) \in \Omega | \phi_i(x, y) = 0\}$ for $(x, y) \in \mathbb{R}^2$ in the Ω domain. A total number of phases are determined by including an array of level set functions $\Phi = \{\phi_1, \dots, \phi_m\}$ and the vector Heaviside function $H(\Phi) = \{H(\phi_1), \dots, H(\phi_m)\}$ in the classic functional $F_n(c, \Phi)$. For instance, let us consider the curve C as the union of two level set $\Phi = \{\phi_1 = 0, \phi_2 = 0\}$. Four regions divide the domain by disjoint sets as:

$$\begin{cases} R_1 = \{\phi_1 > 0, \phi_2 > 0\} \\ R_2 = \{\phi_1 > 0, \phi_2 < 0\} \\ R_3 = \{\phi_1 < 0, \phi_2 > 0\} \\ R_4 = \{\phi_1 < 0, \phi_2 < 0\} \end{cases}$$



A vector of averages $c = (c_{11}, c_{10}, c_{01}, c_{00})$ is obtained as the result of splitting the image u_0 in four regions .

We can rewrite the energy in terms of the number of classes as follows:

$$\begin{aligned}
 F_4(c, \Phi) = & \int_{\Omega} \lambda_1(u_0 - c_{11})^2 H(\phi_1) H(\phi_2) dx dy + \int_{\Omega} \lambda_2(u_0 - c_{10})^2 H(\phi_1) (1 - H(\phi_2)) dx dy \\
 & + \int_{\Omega} \lambda_3(u_0 - c_{01})^2 (1 - H(\phi_1)) H(\phi_2) dx dy + \int_{\Omega} \lambda_4(u_0 - c_{00})^2 (1 - H(\phi_1)) (1 - H(\phi_2)) dx dy \quad (1) \\
 & + \mu_1 \int_{\Omega} |\nabla H(\phi_1)| + \mu_2 \int_{\Omega} |\nabla H(\phi_2)|.
 \end{aligned}$$

The energy functional in Equation (1) is known as the 4-phase level set functional (4-PLS) and it can be solved by the gradient descend method for ϕ_1 and ϕ_2 allowing one partial differential equation (PDE) each. The terms μ_1 and μ_2 are introduced to provide flexibility to each curve and $\lambda_1, \lambda_2, \lambda_3$ and λ_4 are constant values to regulate evolution to the corresponding region.

2.2 The U-Net autoencoder

Currently, there is a wide range of CNN architectures. For instance, the U-net is a CNN with an auto-encoder like architecture for the task of semantic image segmentation.¹⁹ Due to the good performance achieved, it has been commonly used as a starting point for the development of a wide variety of models. Seeking to further increase its performance, several modifications have been proposed.²⁰⁻²³ This model is based on three main components (see Figure (1)):

- ◇ The encoder: During this process, feature maps are generated by means of applying different convolutional filters. In each stage, the spatial dimensions of the feature map are halved while the number of channels is duplicated. This drives the model to analyze context at different scales.
- ◇ The decoder: During this process, the feature map generated by the encoder goes through convolution and upsampling steps. The latter is controlled by the operation of transpose convolution. Finally, the model can learn to recover the spatial information and gradually generate the segmentation map that corresponds to the original image.
- ◇ Skip connections: These are used to connect the feature map from the encoder directly to the corresponding decoder stage so that the information at every scale handles the segmentation map generation, preventing the loss of spatial details.

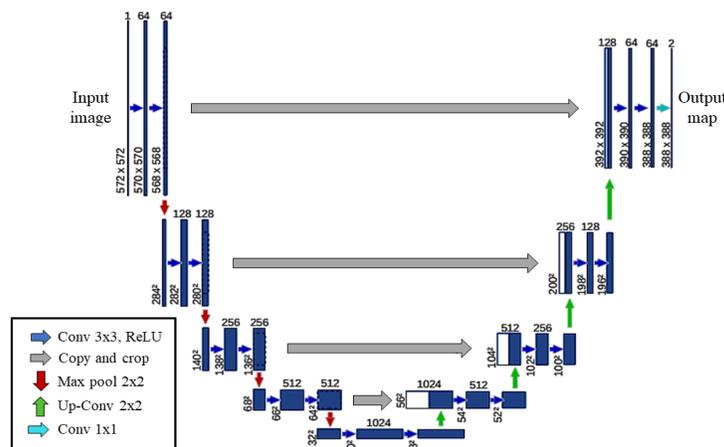


Figure 1. Scheme of the U-Net architecture and the operations performed on the input image in order to generate the segmentation map as output.

Maximum likelihood provides an strategy to estimate the model parameters which is equivalent to minimize the cross entropy between the model and the empirical distributions. In the case of categorical distribution (multiclass segmentation), this corresponds to:

$$CCE = - \sum_C \sum_i^N \hat{p}_{data}(y_i|x_i) [\log p_{model}(y_i|x_i)], \quad (2)$$

where $p_{model}(y_i|x_i)$ is the probability from the model distribution for a given pixel x_i to be assigned to the class y_i . The value $\hat{p}_{data}(y_i|x_i)$ is the probability from the empirical distribution of x_i corresponding to y_i . The optimization process is done by means of the iterative method of gradient descent.

A method to efficiently compute the modification in each parameter is the back-propagation algorithm which is used by the most part of ANNs. The error in the last layers helps to correct parameter values in the previous layers. In this sense, information about adjustments flows in the opposite direction that yields inference.

3. PROPOSED METHOD

3.1 Contributions

1. Perform a high precision automated segmentation that combines the expertise of a CNN and active contours. This hybrid model provides a solution to overcome the self limitations of each tool separately.
2. Demonstrate that the proposed model helps the network to redefine the segmentation result provided with low amount of training data.
3. Reduce the variability of convergence in the 4-PLS which is highly dependent on initialization.

3.2 Methodology

First, the input image is analyzed via the U-Net after being trained with a dataset that contains all possible classes to segment. Afterwards, the output of the U-Net yields a pre-segmentation result which provides the estimated shape and position of each OI. These candidate features are subsequently assigned to the initialization stage of the 4-PLS model as the zero level set $\Phi^{(0)} = \{\phi_1^{(0)}, \phi_2^{(0)}\}$. Parameters employed in the second stage must be defined by the user to control the evolution of each curve leading to a refined segmentation at the end of the process.

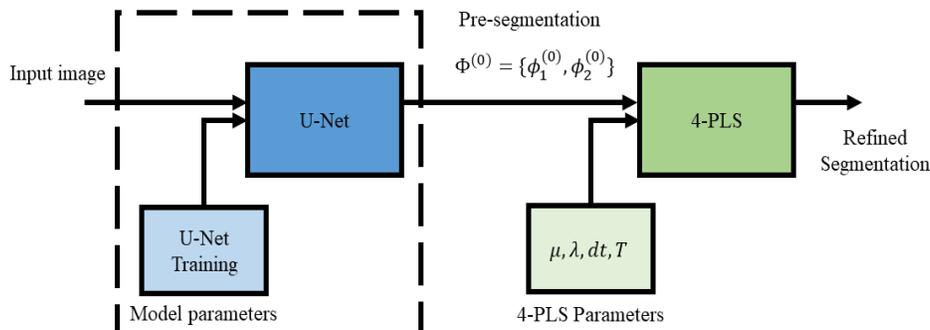


Figure 2. Block diagram of our proposal.

In general, this model can be split into two main stages as shown in Figure 2: One block containing U-Net and other block displaying the 4-PLS process. Each stage performs computations separately and includes a block of parameter tuning, as explained below:

◇ **U-Net.** The dataset used by U-Net is separated into three mutually exclusive groups: Training, validation and testing. A common practice is to consider the validation stage included as part of the training stage. Therefore, validation gives us a hint of the model performance with data that has not seen before (generalization). The parameters selected are those that achieve the best result of generalization.

In order to prevent a biased estimation of the model performance on new data, a final measure is computed in the testing stage, which makes use of the so-called holdout (testing) group, that does not have influence in previous stages.

◇ **Training process.** The samples of the training sub-set are used to compute the loss (error) on the output segmentation map, this guides to modify the U-Net model parameters. Due to storage limitations, batches of data are commonly used to obtain the expected loss and update the parameters. Once all the data have passed through the model, it is said that an epoch has passed. This process can be repeated several times until error converges. Some tools used to improve model training are:

- Batch Normalization is a method that standardizes the data being processed between layers, which has proved to restrict internal covariate shift and improve the training speed of the model.^{24,25}
- Dropout is a regularization strategy that randomly selects and eliminates individual features from a particular layer. This is used to prevent coadaptation which is one of the main causes of overfitting,²⁶ forcing the model to handle variations on processed data, thus making the model more robust.

◇ **4-PSL.** According to the Equation (1), this model is able to segment up to four distinct regions. However, we propose to use two level sets without intersection ($R1 = 0$) to segment up to two OI at the same time in a given scene. This restriction allows us to have full control in the evolution of each contour individually. Final PDEs obtained after minimize the functional $F_4(c, \Phi)$ are:

$$\frac{\partial \phi_1}{\partial t} = \delta(\phi_1) \left\{ \mu_1 \operatorname{div} \left(\frac{\nabla \phi_1}{|\nabla \phi_1|} \right) + [\lambda_3(u_0 - c_{01})^2] H(\phi_2) + [\lambda_2(u_0 - c_{10})^2 - \lambda_4(u_0 - c_{00})^2] (1 - H(\phi_2)) \right\}, \quad (3)$$

$$\frac{\partial \phi_2}{\partial t} = \delta(\phi_2) \left\{ \mu_2 \operatorname{div} \left(\frac{\nabla \phi_2}{|\nabla \phi_2|} \right) + [\lambda_2(u_0 - c_{10})^2] H(\phi_1) + [\lambda_3(u_0 - c_{01})^2 - \lambda_4(u_0 - c_{00})^2] (1 - H(\phi_1)) \right\}. \quad (4)$$

◇ **Selection of parameters.** 4-PLS is a tool which depends on its own parameters to regulate the evolution. Such parameters are useful to guide the movement and velocity of each contour with prior knowledge of an expert. In the case of uncertain information about data, default values can be used, $\mu = \lambda = 1$. Moreover, the iterative process includes a temporal step dt and stopping criteria T to achieve convergence.

4. EXPERIMENTS ON MEDICAL IMAGE DATA

In this section, we carried out a comparison between our proposal and the classic methods which served as basis of our proposal (U-Net and 4-PLS).

4.1 Left and right heart ventricles segmentation in CT

Segmentation process in medical images represents a difficult task for clinicians since manual delineation of a large amount of data is highly demanding and time-consuming. For this reason, automated segmentation of an organ of interest is favorable to aid experts in their diagnosis. For instance, analysis of cardiac images has become relevant since heart failure constitutes one of the main causes of death nowadays.^{27,28}

In this experiment, the dataset used consists of 228 tomographic cardiac studies provided by the Centro Médico ABC in Mexico with a CT Siemens dual source scanner (128 channels).²⁹ These images have a resolution

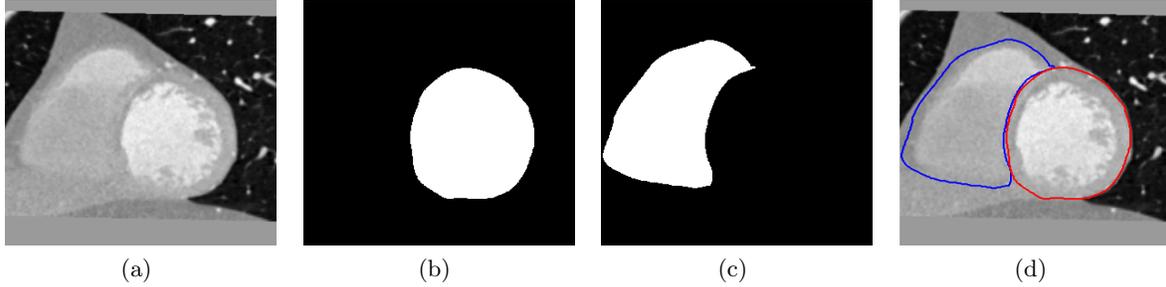


Figure 3. A CT heart image in (a), left ventricle annotation in (b), right ventricle annotation in (c) and both annotations in (d).

of 300×300 pixels with manual annotations of left and right ventricles done by a specialist. The Figure 3 displays a CT heart image with manual annotations in 3(b), 3(c) and 3(d).

First, the dataset is divided into 80% for training which leads to obtain 46 randomly selected images for testing. Second, the **U-Net** adjusts input images in size of 304×304 pixels, so the halving steps in the encoding stage can be carried out. A very common modification includes data augmentation (DA) to provide a better performance by increasing the number of samples in the dataset during training which leads to **U-Net-DA**. Figure 4 describes the validation error in terms of epochs. Note that convergence is faster in U-Net although U-Net-DA yields to a better generalization that may allow a better performance on new data.

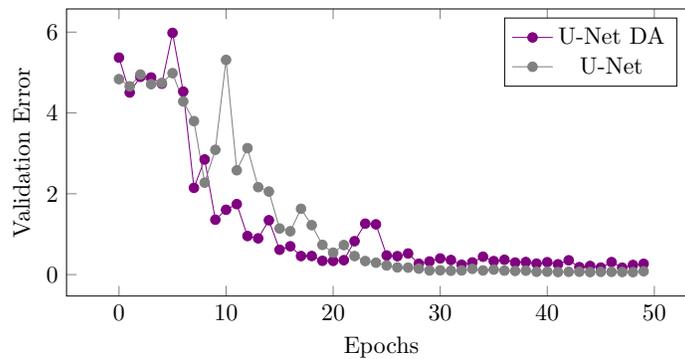


Figure 4. Curves of validation error per epoch for two U-Net versions.

On the other hand, the version called **4-PLS-UI** requires user interaction (UI) to manually adjust parameters and initialize contours. As stated in the diagram of Figure 2, the best results are then used for building the hybrid proposal which combines **U-Net-DA** and **4-PLS-UI** as consecutive main blocks. Types of parameters and values are shown in Table 1 for the studied methods.

U-Net				4-PLS	
Parameters		Data augmentation		Parameters	
Size Image	304x304	Rotation	10	λ_1	1
Batch	16	Width shift	0.1	λ_2	-2
Dropout	0.5	Height shift	0.1	μ_1	0.5
Optimization	Adam	Shear range	0.05	μ_2	0.5
Total weights	31,046,339	Zoom range	0.15	dt	0.1

Table 1. Parameters.

Finally, in order to quantify the differences among the studied methods, both Dice and Hausdorff measures were calculated according to the expert annotations, see Table 2. Dice coefficient indicates differences based on areas, in this sense, values close to 1 suggest more similar contours. Otherwise, Hausdorff distance is a standard

metric that measures separation between two contours by computing how close a point from the first boundary is from another point of the second boundary. Values close to zero indicate more alike boundaries.

Method	Left ventricle		Right ventricle	
	Dice	Hausdorff	Dice	Hausdorff
U-Net	0.911	4.78	0.866	5.14
U-Net-DA	0.931	4.42	0.883	4.99
4-PLS-UI	0.879	5.14	0.833	6.16
Proposal	0.940	4.30	0.890	4.81

Table 2. Comparison of the performance among methods.

One of the best segmentation result of our proposal and the comparison among other methods is displayed in Figure 5. The corresponding Dice and Hausdorff values for each method of left and right ventricle segmentation are shown in Table 3.

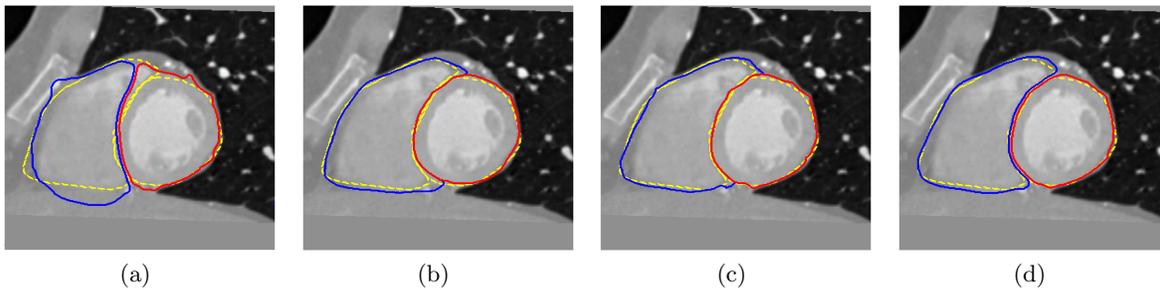


Figure 5. Graphic details. Expert annotations in yellow, segmentation result of right ventricle in blue, segmentation result of left ventricle in red for: 4-PLS-UI in (a), U-Net in (b), U-Net-DA in (c) and the proposed hybrid model in (d). See Table 3 for more details.

Method	4-PLS-UI		U-Net		U-Net-DA		Proposal	
	Left	Right	Left	Right	Left	Right	Left	Right
Dice	0.92	0.86	0.95	0.94	0.96	0.95	0.97	0.96
Hausdorff	4.58	5.38	3.74	4.24	3.01	4.79	3.01	4.20

Table 3. Segmentation results of methods displayed in Figure 5.

5. CONCLUSIONS

In this work, we presented a novel hybrid model to achieve an automated segmentation for multiple regions in an image. It's well known that the initialization stage of active contours techniques has a great influence in the evolution of each curve, for this reason, we embedded a convolutional neural network to provide an estimation of initialization. Moreover, the CNN stage of the proposed model does not require a minimum number of data for training since the output of the CNN provides the predicted shape and position of each object of interest. Finally, these features are used to achieve a refined segmentation with multiple controlled active contours that increase the accuracy in final results.

6. ACKNOWLEDGMENTS

This publication has been sponsored by the grant UNAM PAPIIT IA103119, grant SECTEI/202/2019 and Consejo Nacional de Ciencia y Tecnología (CONACYT).

REFERENCES

- [1] Zaitoun, N. M. and Aqel, M. J., “Survey on image segmentation techniques,” *Procedia Computer Science* **65**, 797 – 806 (2015). International Conference on Communications, management, and Information technology (ICCMIT’2015).
- [2] Mumford, D. and Shah, J., “Optimal approximations by piecewise smooth functions and associated variational problems,” *Communications on pure and applied mathematics* **42**(5), 577–685 (1989).
- [3] Chan, T. F. and Vese, L. A., “Active contours without edges,” *IEEE Transactions on Image Processing* **10**(2), 266–277 (2001).
- [4] Osher, S. and Sethian, J. A., “Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations,” *Journal of Computational Physics* **79**(1), 12 – 49 (1988).
- [5] Malladi, R., Sethian, J. A., and Vemuri, B. C., “Topology-independent shape modeling scheme,” in [*Geometric Methods in Computer Vision II*], Vemuri, B. C., ed., **2031**, 246 – 258, International Society for Optics and Photonics, SPIE (1993).
- [6] Derraz, F., Beladgham, M., and Khelif, M., “Application of active contour models in medical image segmentation,” in [*International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004.*], **2**, 675–681 Vol.2 (April 2004).
- [7] Li, C., Huang, R., Ding, Z., Gatenby, J. C., Metaxas, D. N., and Gore, J. C., “A level set method for image segmentation in the presence of intensity inhomogeneities with application to mri,” in [*IEEE Transactions on Image Processing 20 (7) (2011) 2007–2016. Sen, S.K. Pal / Information Sciences 191 (2012) 169–191*], (2011).
- [8] Olveres, J., Carbajal-Degante, E., Escalante-Ramírez, B., Vallejo, E., and García-Moreno, C. M., “Deformable models for segmentation based on local analysis,” *Accepted at Mathematical Problems in Engineering* **2017** (2017).
- [9] Chaudhury, K. N. and Ramakrishnan, K. R., “Stability and convergence of the level set method in computer vision,” *Pattern Recognition Letters* **28**, 884–893 (2007).
- [10] Spencer, J. and Chen, K., “A convex and selective variational model for image segmentation,” *Communications in Mathematical Sciences* **13**(6), 1453–1472 (2015).
- [11] Bae, E., Yuan, J., and Tai, X.-C., “Global minimization for continuous multiphase partitioning problems using a dual approach,” *International Journal of Computer Vision* **92**, 112–129 (Mar 2011).
- [12] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D., “Backpropagation applied to handwritten zip code recognition,” *Neural Computation* **1**, 541–551 (1989).
- [13] Sifre, L., *Rigid-Motion Scattering For Image Classification*, PhD thesis, Ecole Polytechnique, CMAP (2014).
- [14] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H., “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in [*ECCV*], (2018).
- [15] Ledesma, L., Olveres, J., and Escalante-Ramírez, B., “Hermite convolutional networks,” in [*Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*], Nyström, I., Hernández Heredia, Y., and Milián Núñez, V., eds., 398–407, Springer International Publishing, Cham (2019).
- [16] Kass, M., Witkin, A., and Terzopoulos, D., “Snakes: Active contour models,” *INTERNATIONAL JOURNAL OF COMPUTER VISION* **1**(4), 321–331 (1988).
- [17] Vese, L. A. and Chan, T. F., “A multiphase level set framework for image segmentation using the mumford and shah model,” *International Journal of Computer Vision* **50**(3), 271–293 (2002).
- [18] Carbajal-Degante, E., Olveres, J., and Escalante-Ramírez, B., “A multiphase active contour model based on the Hermite transform for texture segmentation,” in [*Optics, Photonics, and Digital Technologies for Imaging Applications V*], Schelkens, P., Ebrahimi, T., and Cristóbal, G., eds., **10679**, 364 – 372, International Society for Optics and Photonics, SPIE (2018).
- [19] Olaf Ronneberger, P. F. and Brox, T., “U-net: Convolutional networks for biomedical image segmentation,” *CoRR abs/1505.04597* (2015).
- [20] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J., “Unet++: A nested u-net architecture for medical image segmentation,” *CoRR abs/1807.10165* (2018).
- [21] Garcia-Uceda, A., Selvan, R., Saghir, Z., and de Bruijne, M., “A joint 3d unet-graph neural network-based method for airway segmentation from chest cts,” (2019).

- [22] Özgün Çiçek, Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O., “3d u-net: Learning dense volumetric segmentation from sparse annotation,” (2016).
- [23] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., and Rueckert, D., “Attention u-net: Learning where to look for the pancreas,” (2018).
- [24] Ioffe, S. and Szegedy, C., “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” (2015).
- [25] LeCun, Y., Bottou, L., Orr, G. B., and Müller, K.-R., “Efficient backprop,” in [*Neural Networks: Tricks of the Trade, This Book is an Outgrowth of a 1996 NIPS Workshop*], 9–50, Springer-Verlag, London, UK, UK (1998).
- [26] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research* **15**, 1929–1958 (2014).
- [27] Cheung, Y.-F., “The role of 3d wall motion tracking in heart failure,” *Nature Reviews Cardiology* **9**, 644–657 (2012).
- [28] Barba-J, L., Escalante-Ramírez, B., Vallejo Venegas, E., and Arámbula Cosío, F., “A 3d hermite-based multiscale local active contour method with elliptical shape constraints for segmentation of cardiac mr and ct volumes,” *Medical & Biological Engineering & Computing* **2017** (2017).
- [29] Olveres, J., Nava, R., Escalante-Ramírez, B., Vallejo, E., and Kybic, J., “Left ventricle hermite-based segmentation,” *Computers in Biology and Medicine* **87**, 236 – 249 (2017).