

Chan-Vese Attention U-Net: An attention mechanism for robust segmentation.

Nicolas Makaroff¹ and Laurent D. Cohen¹

CEREMADE, UMR CNRS 7534, University Paris Dauphine, PSL Research
University, 75775 Paris, France makaroff@ceremade.dauphine.fr

Abstract. When studying the results of a segmentation algorithm using convolutional neural networks, one wonders about the reliability and consistency of the results. This leads to questioning the possibility of using such an algorithm in applications where there is little room for doubt. We propose in this paper a new attention gate based on the use of Chan-Vese energy minimization to control more precisely the segmentation masks given by a standard CNN architecture such as the U-Net model. This mechanism allows to obtain a constraint on the segmentation based on the resolution of a PDE. The study of the results allows us to observe the spatial information retained by the neural network on the region of interest and obtains competitive results on the binary segmentation. We illustrate the efficiency of this approach for medical image segmentation on a database of MRI brain images.

Keywords: Attention Mechanism · Level Set · Chan-Vese · Segmentation.

1 Introduction

Medical image segmentation is a crucial task that requires significant time and effort from medical experts. Although various solutions, including convolutional neural networks (CNNs), have been proposed to automate this process, the need for efficient and reliable methods still exists.

While CNNs have shown promising results in medical image segmentation, their lack of transparency and the sensitive nature of medical data raise concerns regarding their applicability in real-world hospital settings, especially for medical staff who are not trained in machine learning.

Convolutional neural networks have revolutionised the study of images for both classification and segmentation, the latter being the objective of interest here. Two architectures stand out among all those studied over the years, the fully convolutional neural network [6] and the famous U-Net [14]. These architectures have been tested many times on various applications such as MRI segmentation of the brain [5] or heart [13], CT scans of organs in the thoracic cavity [3]. Numerous modifications have also been made to improve the efficiency of these complex structures and in the medical field especially the U-Net has proven to be very versatile for many segmentation tasks. The rest of the paper is organised

as follows. In Section 2 we introduce our experimental method for Fast Marching Energy CNN. In Section 3 we present the main results of our experiments and provide a discussion around our work.

1.1 Related work

Several attempts have been made to integrate geometric or topological properties in the neural network to incorporate information beyond adjacent pixels for segmentation tasks. In 2019, [4] proposed a lesion segmentation method based on active contours using a U-Net-like neural network. The network predicts a segmentation mask, which is refined using a generalisation of the active contour problem from [2]. Similarly, In 2020, [15] presented a model, where the neural network predicts the parameters for initialising the active contour model and an initial contour. Learning is achieved by combining the error produced by the neural network and that produced during active contour usage. In 2021, [8] proposed a fully integrated geodesic active contour model, where the neural network learns to minimise the energy functional of the model. In this encoder-decoder network, the output is a contour map instead of a probability map for segmentation, based on the active contour method proposed by [1]. Although these methods have shown promise, there is still room for improvement in incorporating geometric and topological properties more effectively and seamlessly into deep learning-based segmentation approaches.

1.2 Contributions

The main contribution of this paper is the development of a novel hybrid segmentation method that combines deep learning with classical functional energy minimization techniques, specifically designed for medical image segmentation applications. Our method features a new attention gate, the *Chan-Vese Attention Gate*, which integrates information from the level sets method of the well-established Chan-Vese functional [2]. Unlike traditional deep learning methods that rely solely on the neural network to improve image segmentation, our approach leverages resolution information to achieve more accurate results.

To demonstrate the effectiveness of our method, we conducted comprehensive experiments on the TCGA_LGG database [10], a repository of brain images for the study of lower grade gliomas. Given the sensitive nature of medical image segmentation, it is crucial to ensure the validity of our results. Our approach achieved at least equivalent results to previous networks while remaining simple to optimise. The training time for our method is only 5% longer than the traditional approach, which is a minor increase considering the benefits. Furthermore, the difference in computation time is imperceptible during inference. Overall, our approach represents a significant advancement in medical image segmentation, offering a more accurate and efficacious solution for this critical field.

2 Methodology

The U-Net architecture The U-Net architecture is widely used for medical image segmentation. It maintains the structure of an image during transformation from an image to a vector and back to an image using features extracted during the contraction phase. The architecture takes the shape of a "U" with three parts: contraction, transition, and expansion. The contraction applies several blocks with convolution and pooling layers, doubling the feature maps at each stage. The transition uses convolution layers, while the expansion uses convolution and up-sampling layers. The information is recovered during the contraction to reconstruct the image, with the same number of expansion blocks as contraction blocks. The final output is obtained through the final convolutional layer.

Attention Gate in U-Net architecture The authors of [9] presented in their paper a new attention gate especially for the case of the CNN and in particular for the U-Net. The architecture of the U-Net remains the same except for the expansion part. In this part, an attention mechanism is integrated between each block. For each block, the input and the information coming from the connections of the corresponding contraction part block pass through an attention block. The input is up-sampled in parallel and finally the two results are concatenated and sent to the convolution block.

Chan-Vese Energy Minimization Presented in [2], Chan-Vese's method is used to segment a binary image. Let I be the given grayscale image on a domain Ω to be segmented. The Chan Vese method looks for a piece-wise constant approximation of an image where there are 2 regions separated by an unknown boundary curve C . This is obtained through the minimization of the following energy depending on curve C and the constant values c_1 and c_2 inside and outside the curve:

$$E(C, c_1, c_2) = \mu \times \text{Length}(C) + \nu \times \text{Area}(\text{inside}(C)) \\ + \lambda_1 \int_{\text{inside}(C)} |I(x, y) - c_1|^2 dx dy + \lambda_2 \int_{\text{outside}(C)} |I(x, y) - c_2|^2 dx dy. \quad (1)$$

Energy minimization is simplified by replacing the curve C with a level set function ϕ . The inside region is then the set where $\phi > 0$ and the outside region the set where $\phi < 0$. With the help of the Heavyside function H , the energy becomes:

$$F(c_1, c_2, \phi) = \mu \int_{\Omega} \delta(\phi(x)) |\nabla \phi(x)| dx + \nu \int_{\Omega} H(\phi(x)) dx \\ + \lambda_1 \int_{\Omega} |I(x) - c_1|^2 H(\phi(x)) dx + \lambda_2 \int_{\Omega} |I(x) - c_2|^2 (1 - H(\phi(x))) dx, \quad (2)$$

where the term following μ represent the length of the contour, the term following ν the area inside the contour and δ the Dirac mass.

$$(P) : \arg \min_{c_1, c_2, \phi} F(c_1, c_2, \phi) \quad (3)$$

The new variable is ϕ . The energy $F(c_1, c_2, \phi)$ is minimised with respect to ϕ with a gradient descent evolution.

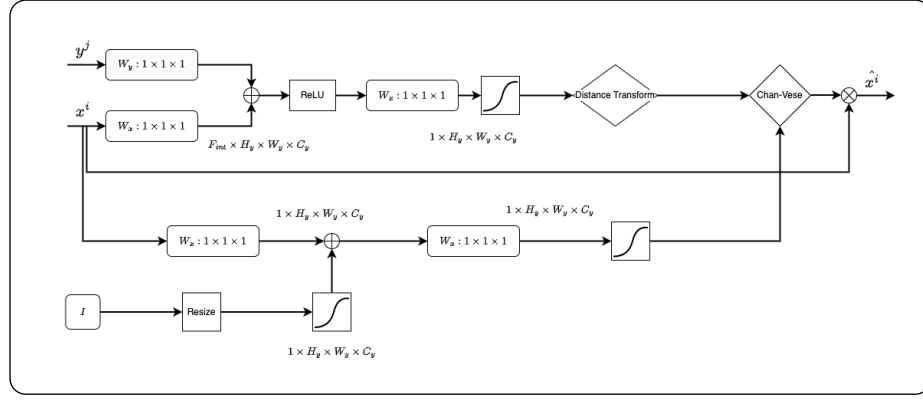


Fig. 1. Scheme of the proposed attention method. The symbols \oplus and \otimes represent respectively the addition and multiplication of the tensors.

Chan-Vese Attention in U-Net architecture In this work, we incorporate our novel attention method, the *Chan-Vese Attention Gate*, into the U-Net architecture. The attention method proposed by Oktay et al. [9] allows highlighting areas that provide more information in the skip connections. Our method expands upon this approach by incorporating information from the Chan-Vese method, which performs intermediate segmentation for each block of the expansion phase. As illustrated in Figure 1, the representation constructed by the network from the skip connection is concatenated with the information from the previous layer before undergoing several transformations. First, let us briefly review the classical attention gate proposed by Oktay et al. [9]. Given a pair of input feature maps, the attention gate computes an attention coefficient that modulates the skip connections. The attention coefficient is obtained through a gating mechanism that involves two 1×1 convolutions followed by an element-wise addition and a non-linear activation function (sigmoid). The attention gate effectively allows the network to learn which regions of the input image are more relevant to the segmentation task. Our method builds upon the classical attention gate by incorporating the Chan-Vese segmentation technique. The first key step is the distance transform applied to the classical attention representation,

which serves as the initial contour for the Chan-Vese method. This step is made possible thanks to the use of the distance transform method proposed by [11], which renders the transformation differentiable within a classical automatic differentiation framework. A secondary branch was implemented in the proposed framework to selectively emphasise the tumorous region of the input image. This branch carefully resizes the input image to conform to the dimensions of the current layer and integrates it with the residual feature map via addition. A 1×1 convolution is then applied to the resulting map, which serves as the input for subsequent segmentation. In the second step, the modified Chan-Vese algorithm is used to iteratively segment the image a finite number of times. This segmented image is then used as a control signal to facilitate learning. The proposed approach thus provides a more refined and informative attention for the network.

Let $I \in R^{N \times H \times W \times K_I}$ represent a batch of input images, where N, H, W, K_I denote the batch size, height, width, and number of channels of the image, respectively. Let $X_i^f \in R^{F_x \times H_i \times W_i \times K}$ denote the residual feature map f at layer i and $Y_j^f \in R^{F_x \times H_i \times W_i \times K}$ the previous layer feature map at layer j and feature map f . $W_{1 \times 1}$ is a 1×1 convolution. Following the additive attention formulation:

$$q_{att}^f = \Psi^T(\sigma_1(\mathbf{W}_{1 \times 1}^T X_i^f + \mathbf{W}_{1 \times 1}^T Y_j^f + b_f) + b_\psi) \quad (4)$$

$$\alpha_i^f = \sigma_2(q_{att}^f(X_i^f; Y_j^f; \Theta_{att})) \quad (5)$$

We define D_x as the distance transform that takes a tensor of shape $H_l \times W_l \times K_l$ as input. By applying the distance transform to the former attention gate, we obtain

$$\beta_i^l = D(\alpha_i^l) = -\lambda \log(\alpha_i^l * \exp(-\frac{d(\cdot, 0)}{\lambda})) \quad (6)$$

where $d(\cdot, \cdot)$ is the Euclidean distance and $*$ is the convolution product. This information is passed as an initialization contour.

On the other side, we perform a transformation of the input image, as if using some filter (e.g., CLAHE filter [12]), as follows:

$$\gamma_i^f = \sigma_2(\mathbf{W}_{1 \times 1}^T (\mathbf{W}_{1 \times 1}^T X_i^f + \sigma_2(I)) + b_W) \quad (7)$$

Many areas of the images we wish to segment happen to have the same values as the averages of regions c_1 and c_2 as the region we ultimately want to segment. This sometimes leads the algorithm to add undesirable areas to the segmentation even though the μ and ν hyperparameters have been carefully chosen. The γ_i^f transformation solves this problem by reducing the intensity in areas far from the tumour.

Finally, the resulting mask segmentation and attention coefficient ζ_i^f is given by solving the Chan-Vese problem:

$$\zeta_i^l = CV(\gamma_i^f, \beta_i^f, \mu, \nu) \quad (8)$$

where γ_i^f is the image that supports the initial mask β_i^f . μ and ν are positive parameters.

During the backpropagation procedure, the gradients of the loss function concerning the attention gate’s parameters (marked in bold) and intermediate outputs are computed using the chain rule, which can be efficiently executed in modern deep learning frameworks. The differentiable distance transform and the Chan-Vese module enable the backpropagation to update the parameters of the attention gate and the network’s other layers. Consequently, the network can learn to focus on the most relevant regions for segmentation, improving its performance in medical image segmentation tasks.

3 Experiments

Evaluation Datasets In this study, we used the TCGA_LGG database, an openly available online repository [10] containing magnetic resonance imaging (MRI) scans of brain tumour patients. The database consists of 110 patients from The Cancer Genome Atlas (TCGA) lower-grade glioma collection, with genomic cluster data and at least one fluid-attenuated inversion recovery (FLAIR) sequence available. Table 1 provides a summary of the experimental results. The spatial resolution of the images contained in the TCGA LGG dataset is $1mm$ isotropic.

Table 1. Segmentation results (IOU) on the TCGA_LGG brain MRI database. Significant results are highlighted in bold font.

Method	U-Net	Attention U-Net	Chan-Vese U-Net
Dice	0.832 ± 0.091	0.830 ± 0.023	0.824 ± 0.019
IOU	0.829 ± 0.075	0.833 ± 0.023	0.848 ± 0.021
Hausdorff	$2.390mm \pm 0.985$	$2.416mm \pm 0.775$	$2.329mm \pm 0.672$
FPR	0.010 ± 0.003	0.009 ± 0.002	0.012 ± 0.004
FNR	0.013 ± 0.004	0.015 ± 0.005	0.013 ± 0.003

Implementation Details We used a large batch of 32 for gradient update and the model parameters are optimised using an adamW optimiser [7] with learning rate 5×10^{-4} and batch normalisation. We applied standard data augmentation (resize, horizontal flip, vertical flip, random rotate, transpose, shift and scale, normalise). The Chan-Vese parameters μ and ν are set respectively to 0.1 and 1.0. The loss is computed using the addition of Dice loss and Binary Cross Entropy. The added attention layer slows down the training by an average of 1 sec out of 6 sec per batch. The code is written in Jax using Haiku framework and will soon be available.

Segmentation Results In this study, we conducted a comparative evaluation of our proposed model with the classical U-Net and the original Attention U-Net.

Table 1 provides a summary of the experimental results. Dice, IOU (Intersection Over Union), Hausdorff, FPR (False Positive Rate), and FNR (False Negative Rate) values are reported, along with their respective standard deviations (denoted as sd_i). The image resolution should be considered when interpreting the Hausdorff distance values. Our proposed model demonstrated superior IOU scores and improved false negative performance. This can be attributed to the model’s ability to focus on a smaller area of interest and the integration of the Chan-Vese method, which enables more effective capture of relevant information and reduces the risk of information loss.

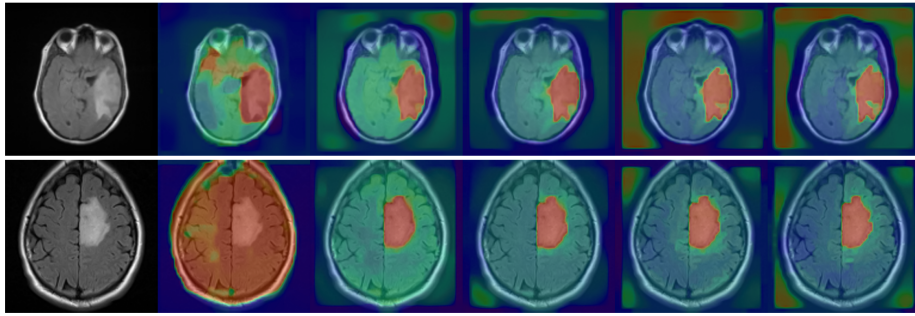


Fig. 2. Output of the Chan-Vese Attention Layer at different stage of learning (epochs:1, 50, 100, 200, 300) from left to right.

Chan-Vese Attention Masks analysis We can observe from the results of the attention layer (see Figure 2) that with the use of Chan-Vese the attention mask quickly converges to an apparently tumour-like segmentation. It takes advantage of the minimization of the Chan-Vese energy from the initialization of the mask thanks to the part inspired by the attention method of [9] but also of the use of the initial image to be segmented. Gradually the contours of the tumour become more precise and the active intensity on the tumour is the confidence on the energy to be minimised. The 0 level set was used to enable the neural network to selectively prioritise the tumour area during segmentation. The upper level set was subsequently employed to refine the tumour segmentation.

Comparison with Attention U-Net Figure 3 shows the attention output of the Chan-Vese Attention Module and the classical Attention Module. Both methods allow the neural network to focus on the tumour area. It should be noted that the method proposed by [9] obtains a finer mask on certain details of the tumour but does not manage in the framework of our study to rank the confidence of the presence of the tumour. In many places outside the tumour area we observe artefacts that do not correspond to the object of interest in the image. In contrast

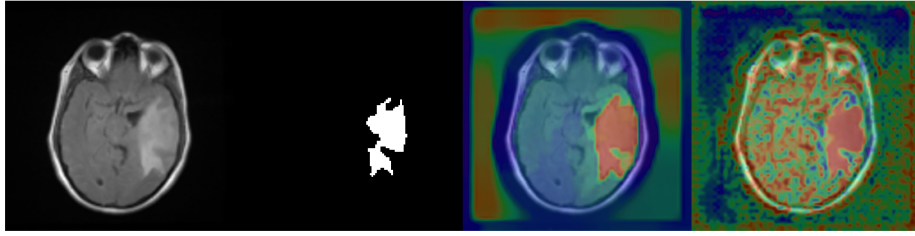


Fig. 3. Comparison of the Attention Mask between a Chan-Vese Attention and the Original Attention. (From left to right: the input MRI, the tumour to be segmented, Chan-Vese Attention mask, original Attention mask).

to these observations the method supported by Chan-Vese focuses only on the tumour area inside the skull.

4 Conclusion

In this paper, we have presented a novel segmentation approach that effectively combines classical energy minimization techniques with Deep Learning. Our proposed model, which integrates the Chan-Vese algorithm into the attention mechanism of a U-Net, demonstrates the value of incorporating non-deep learning sources of shape and structure information, particularly when dealing with sensitive medical data.

Our experimental analysis reveals that the proposed model achieves close results, with some improvements compared to both the classical U-Net and Attention U-Net in key performance metrics such as IOU scores and false negative rates. The Chan-Vese Attention Module successfully narrows down the model's focus to the tumour area, contributing to enhanced segmentation accuracy and precision. This work highlights the potential of combining diverse segmentation methods, paving the way for future research in blending various sources of shape and structure information with Deep Learning models. Ultimately, our approach aims to contribute to the development of more effective and versatile segmentation solutions.

Acknowledgements This work is in part supported by the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute).

References

1. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic Active Contours. *International Journal of Computer Vision* **22**, 61–79 (1995)

2. Chan, T., Vese, L.: Active Contours Without Edges. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society* **10** **2**, 266–77 (2001)
3. Gerard, S.E., Reinhardt, J.M.: Pulmonary Lobe Segmentation Using A Sequence of Convolutional Neural Networks For Marginal Learning. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 1207–1211 (2019). <https://doi.org/10.1109/ISBI.2019.8759212>
4. Hatamizadeh, A., Hoogi, A., Sengupta, D., Lu, W., Wilcox, B., Rubin, D., Terzopoulos, D.: Deep Active Lesion Segmentation. In: Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10. pp. 98–105. Springer (2019)
5. Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M., Biller, A.: Deep MRI brain extraction: A 3D convolutional neural network for skull stripping. *NeuroImage* **129**, 460–469 (2016)
6. Long, J., Shelhamer, E., Darrell, T.: Fully Convolutional Networks for Semantic Segmentation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2015)
7. Loshchilov, I., Hutter, F.: Fixing Weight Decay Regularization in Adam. *CoRR* **abs/1711.05101** (2017), <http://arxiv.org/abs/1711.05101>
8. Ma, J., He, J., Yang, X.: Learning Geodesic Active Contours for Embedding Object Global Information in Segmentation CNNs. *IEEE Transactions on Medical Imaging* **40**(1), 93–104 (2020)
9. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning Where to Look for the Pancreas (2018). <https://doi.org/10.48550/ARXIV.1804.03999>, <https://arxiv.org/abs/1804.03999>
10. Pedano, N., Flanders, A.E., Scarpace, L., Mikkelsen, T., Eschbacher, J.M., Hermes, B., Sisneros, V., Barnholtz-Sloan, J., Ostrom, Q.: The Cancer Genome Atlas Low Grade Glioma Collection (tcga-lgg) (version 3) [data set], The Cancer Imaging Archive (2016). <https://doi.org/10.7937/K9/TCIA.2016.L4LTD3TK>
11. Pham, D.D., Dovletov, G., Pauli, J.: A Differentiable Convolutional Distance Transform Layer for Improved Image Segmentation. In: Pattern Recognition: 42nd DAGM German Conference, DAGM GCPR 2020, Tübingen, Germany, September 28 – October 1, 2020, Proceedings. p. 432–444. Springer-Verlag, Berlin, Heidelberg (2020). https://doi.org/10.1007/978-3-030-71278-5_31
12. Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., Romeny, B.T.H., Zimmerman, J.B.: Adaptive Histogram Equalization and Its Variations. *Comput. Vision Graph. Image Process.* **39**(3), 355–368 (sep 1987). [https://doi.org/10.1016/S0734-189X\(87\)80186-X](https://doi.org/10.1016/S0734-189X(87)80186-X), [https://doi.org/10.1016/S0734-189X\(87\)80186-X](https://doi.org/10.1016/S0734-189X(87)80186-X)
13. Pop, M., Sermesant, M., Jodoin, P.M., Lalande, A., Zhuang, X., Yang, G., Young, A., Bernard, O. (eds.): Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges. Springer International Publishing (2018). <https://doi.org/10.1007/978-3-319-75541-0>, <https://doi.org/10.1007/978-3-319-75541-0>
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)

15. Zhang, M., Dong, B., Li, Q.: Deep Active Contour Network for Medical Image Segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23. pp. 321–331. Springer (2020)