

Automatic liver segmentation by integrating fully convolutional networks into active contour models

Ines VATI^{1,2}

¹École des Ponts ParisTech, Champs-sur-Marne, France

²MVA, ENS Paris-Saclay, Cachan, France

^{1,2}Email ines.vati@eleves.enpc.fr

1 Réponses aux questions

Dans cette première partie, je réponds aux questions demandé en français mais le reste du rapport sera en anglais.

1.1 Quel est le problème traité ?

Le problème traité est la segmentation automatique du foie à partir d'images médicales, qui sont notamment des CT scans. La segmentation du foie est une étape cruciale dans le diagnostic et le traitement des maladies du foie. En effet, elle permet de quantifier le volume du foie, de détecter des lésions et de suivre leur évolution. La segmentation manuelle est très fastidieuse et prend beaucoup de temps.

Les auteurs proposent une méthode de segmentation automatique en intégrant les réseaux de neurones convolutionnels (FCNs) dans les modèles de contours actifs (ACM).

Un modèle FCN-8 est entraîné à prédire une carte de distance structurée en couche. Il prédit la couche dans laquelle se trouve chaque pixel de l'image d'entrée. La sortie de ce réseau est utilisée pour définir une force externe F_{FCN} qui est intégrée dans le modèle ACM.

1.2 Quelles sont les équations et méthodes numériques utilisées. Peut-on éventuellement donner une formulation plus mathématique au problème ?

Soient C une courbe et ψ une fonction Lipschitz tel que $C = \{(x, y) | \psi(x, y) = 0\}$ est le niveau 0 de cette fonction. ψ est une fonction de distance signée telle que $\psi(x, y) > 0$ si (x, y) est à l'extérieur de l'objet délimité par C et $\psi(x, y) < 0$ si (x, y) est à l'intérieur de C .

Nous considérons une bande étroite de pixels adjacents à l'ensemble de niveau zéro L_0 . Nous définissons les voisinages de L_0 en deux types de couches : L_1, \dots, L_{N-1} , les couches qui sont à l'extérieur de la structure, et L_{-1}, \dots, L_{-N+1} , les couches qui sont à l'intérieur de la structure. Une couche est définie par son intervalle de distance à l'ensemble de niveau zéro, c'est-à-dire $L_i = \{(x, y) | (i - 0.5)\delta \leq \psi(x, y) \leq (i + 0.5)\delta\}$, pour $|i| < N - 2$.

Par exemple, pour $N = 4$, les labels de chaque couche sont $\{-3, -2, -1, 0, 1, 2, 3\}$ et

$$\begin{aligned} L_{-3} &= \{(x, y) | \infty < \psi(x, y) < -2.5\delta\} \\ L_{-2} &= \{(x, y) | -2.5\delta < \psi(x, y) \leq -1.5\delta\} \\ L_{-1} &= \{(x, y) | -1.5\delta < \psi(x, y) \leq -0.5\delta\} \\ L_0 &= \{(x, y) | -0.5\delta < \psi(x, y) \leq 0.5\delta\} \\ L_1 &= \{(x, y) | 0.5\delta < \psi(x, y) \leq 1.5\delta\} \\ L_2 &= \{(x, y) | 1.5\delta < \psi(x, y) \leq 2.5\delta\} \\ L_3 &= \{(x, y) | 2.5\delta < \psi(x, y) \leq \infty\} \end{aligned}$$

Le réseau de neurones est entraîné à prédire cette carte de label pour une image CT donnée en entrée.

Entraînement du réseau de neurones. Donnons une formulation mathématique de ce problème.

$\mathcal{X} \subset \mathbb{R}^{D \times H \times W}$ est l'espace des images CT. D , H et W sont respectivement la profondeur, la hauteur et la largeur de l'image.

Soit $f_\theta : \mathcal{X} \mapsto \mathbb{R}^{D \times H \times W \times 2N-1}$ la fonction apprise par le réseau de neurones qui prend en argument une image $I \in \mathcal{X}$. La carte des labels est obtenue en prenant la valeur maximale sur la dernière dimension de $f_\theta(I)$.

Entraîner un réseau de neurones consiste à optimiser les poids du réseaux, notés θ , de façon à minimiser une fonction de coût L qui mesure l'écart entre les prédictions du réseau et les vrais labels. Cette optimisation est généralement faite par descente stochastique du gradient.

La sortie du réseau de neurones est généralement normalisée par une fonction d'activation *softmax*. La prediction du réseau $\hat{y} \in \mathbb{R}^{D \times H \times W \times 2N-1}$ s'écrit alors pour tout $c \in \{1, \dots, 2N-1\}$ comme

$$\hat{y}_c = \frac{e^{[f_\theta(I)]_c}}{\sum_{i=1}^{2N-1} e^{[f_\theta(I)]_i}}$$

Un fonction de coût classique pour un problème de classification est la cross entropie. Dans le cas où il y a $2N-1$ classes, la fonction de coût est donnée par

$$L_{CE}(f_\theta(I), y) = \sum_{k=1}^D \sum_{i=1}^H \sum_{j=1}^W -\log(\hat{y}_{k,i,j,c}) \mathbb{1}_{\{y_{k,i,j}=c\}} \quad (1)$$

où $y \in \mathbb{R}^{D \times H \times W}$ est la vrai carte de labels de l'image I . D'autres fonctions de coût peuvent être utilisées.

Résolution de l'évolution des contours actifs. Les auteurs proposent une nouvelle force externe notée F_{FCN} à ajouter au modèle de contours actifs basée sur la prédiction du réseau de neurones. Plusieurs choix sont possible pour ajuster l'amplitude de cette force par rapport à la couche dans laquelle se trouve son point d'application.

Ils utilisent la méthode des ensembles de niveaux. Cela revient à résoudre l'équation suivant

$$\frac{\partial \psi}{\partial t} = (w_0 F_0 + w_1 F_{FCN} + \mu \kappa) \vec{N} \quad \text{where} \quad \kappa = \text{div}\left(\frac{\nabla \psi}{\|\nabla \psi\|}\right)$$

\vec{N} est le vecteur normal unitaire du niveau zéro C de ψ . w_i sont des poids qui permettent de régler l'importance de chaque force. μ est un paramètre de régularisation. κ est la courbure de C . F_0 correspond aux forces images conventionnelles. L'utilisateur peut choisir une fonction basée region, *region-based model* [1] ou une fonction basée contour, *edge-based model* [2] pour F_0 . Les auteurs optent pour un modèle de Chan-Vese locale basée région [3].

Pour résoudre cette équation, les auteurs utilisent la méthode *sparse field* de Whitaker [4].

1.3 Pouvez-vous situer cet article par rapport aux méthodes étudiées en cours et le comparer à des sujets proches évoqués en cours.

Cet article se situe dans le cadre de la segmentation d'images médicales, dans la partie du cours qui traite des modèles des contours actifs.

Nous avons vu dans le cours que cette méthode était très sensible à l'initialisation de la courbe donnée manuellement. En effet, un *snake* qui n'est pas suffisamment proche du bord de l'objet à segmenter n'est pas attiré par lui. Nous avons également noté qu'à cause du bruit certains points isolés, étant des maxima locaux du gradient, pouvaient bloquer la courbe. A l'instar du modèle de ballon, l'ajout de la force externe F_{FCN} rend la méthode proposée plus robuste à l'initialisation en poussant ou tirant la courbe vers le bord de l'objet à segmenter. Contrairement à la force ballon, la

courbe même si la courbe dépasse le contour d'intérêt, elle est tirée vers le contour. Dans le modèle de ballon, la courbe initiale doit être initialisée à l'intérieur de la structure. Dans le modèle proposé [5], le contour initiale peut être partiellement ou complètement à l'intérieur de la structure. Cependant, il ne peut pas être entièrement à l'extérieur à cause la présence de fragments parasite dans la carte de distance prédite par le réseau de neurones.

1.4 Quelle est l'originalité du travail (selon les auteurs)

Les auteurs proposent d'imposer une force de contrainte externe au modèle des contours actifs en utilisant la sortie d'un réseau de neurones entraîné. L'intérêt de cette nouvelle force externe est que sa direction et son intensité dépendent de la position et de la distance relative du point à la frontière de l'objet à segmenter. Selon, sa position (à l'intérieur ou l'extérieur), la force externe tire ou pousse le contour actif vers la frontière de l'objet. De plus, si le point d'application est loin de la frontière, la force est d'autant plus forte. Lorsque le point est proche de la frontière, l'amplitude de la force se réduit laissant les autres forces, comme les forces internes de régularisation et forces image, prendre le dessus et affiner précisément le bord de l'objet. Il joue en quelque sorte le rôle d'un mecanisme d'attention.

combiner des informations de haut niveau grâce à la carte de distance en couche prédite par le réseau de neurones avec des informations de bas niveau issue du modèle ACM.

L'originalité réside aussi dans le fait que le réseau de neurones n'aurait pas besoin d'être réentraîné pour chaque nouvelle image, voire pour un nouveau dataset (issu d'un autre hopital par exemple).

1.5 Quels sont les résultats nouveaux importants qui en découlent

Cette méthode est robuste à l'initialisation du contour actif. La courbe initiale peut être placée loin du bord de la structure à segmenter.

1.6 Voyez-vous des faiblesses dans l'approche présentée et avez-vous des idées pour y faire face?

Les cartes de distances prédites par le réseau de neurones présentes des artifacts.

2 Methods

2.1 Turn the binary label map to layered label map

The first step is to transform a binary image mask, which serves as a segmentation result and has only two labels (0 stands for background and 1 for foreground), into a label map with multiple categories ($N > 2$). The label map contains not only spatial regional information (whether a pixel is on the foreground or background) but also boundary information (whether a pixel is on the object boundary), as well as information about the relative distance to the object boundary.

The structured label map consists of several layers, including narrow band layers near the object boundary, exterior regional layer, and interior regional layer. These layers are nonoverlapping and cover the entire image domain.

Let's consider a narrow band of grid points adjacent to the zero level set as L_0 layer. We define the neighborhoods of zero narrow band in two type of layers: L_1, \dots, L_{N-1} , the layers that are outside the structure, and L_{-1}, \dots, L_{-N+1} , the layers that are inside the structure. A layer is defined by its distance interval to the zero level set, ie $L_i = \{(x, y) | (i - 1 - 0.5)\delta \leq \psi(x, y) \leq (i + 0.5)\delta\}$

I adopt the same convention as the authors [5]. The distance for pixel outside the object (background) is positive and is negative for pixel inside the object (foreground).

2.2 Training the neural network

Mirroring the methodology used in [5], I used the same architecture for the neural network. I use the FCN-8 architecture from [6]. It uses the Caffe deep learning framework. I performed transfer learning by taking the pre-trained model from <https://github.com/wkentaro/pytorch-fcn>

2.3 Resolution of the Active contour model evolution

\vec{N} est le vecteur normal unitaire à la courbe C .

3 Implementation and results

3.1 Dataset

A new dataset that was not used by the authors and which presents new challenges. Chaos dataset <https://chaos.grand-challenge.org/Download/> (only CT scans and MRI images are available). Data corresponds to a series of DICOM images belonging to a single patient. In order to provide sufficient data that contains enough variability to be representative of the problem, the data sets in the training data are selected to represent both the difficulties that are observed on the whole database (e.g. partial volume effects for CT or bias fields for MRI) and examples of the rare but important challenges such as atypical liver shapes (Figures 4 and 5). It only contains healthy livers aligned in the same direction and patient position. However, the challenging part is the enhanced vascular structures (portal phase) due to the contrast injection.

Each axial slice has identical size of 512 9 512, the number of slices varies from [TO CHECK]

SLIVER07 dataset <https://sliver07.grand-challenge.org/Home/> (only CT scans are available). The data consists of 20 training scans and 10 test scans.

4 Conclusion

References

- [1] V. Caselles, R. Kimmel, and G. Sapiro, “Geodesic active contours,” *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, Feb. 1, 1997, ISSN: 1573-1405. DOI: [10.1023/A:1007979827043](https://doi.org/10.1023/A:1007979827043). [Online]. Available: <https://doi.org/10.1023/A:1007979827043> (visited on 03/09/2024).
- [2] T. Chan and L. Vese, “Active contours without edges,” *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, Feb. 2001, Conference Name: IEEE Transactions on Image Processing, ISSN: 1941-0042. DOI: [10.1109/83.902291](https://doi.org/10.1109/83.902291). [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/902291> (visited on 03/09/2024).
- [3] S. Lankton and A. Tannenbaum, “Localizing region-based active contours,” *IEEE Transactions on Image Processing*, vol. 17, no. 11, pp. 2029–2039, Nov. 2008, Conference Name: IEEE Transactions on Image Processing, ISSN: 1941-0042. DOI: [10.1109/TIP.2008.2004611](https://doi.org/10.1109/TIP.2008.2004611). [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/4636741> (visited on 03/09/2024).
- [4] R. T. Whitaker, “A level-set approach to 3d reconstruction from range data,” *International Journal of Computer Vision*, vol. 29, no. 3, pp. 203–231, Sep. 1, 1998, ISSN: 1573-1405. DOI: [10.1023/A:1008036829907](https://doi.org/10.1023/A:1008036829907). [Online]. Available: <https://doi.org/10.1023/A:1008036829907> (visited on 03/09/2024).
- [5] X. Guo, L. H. Schwartz, and B. Zhao, “Automatic liver segmentation by integrating fully convolutional networks into active contour models,” *Medical Physics*, vol. 46, no. 10, pp. 4455–4469, 2019, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/mp.13735>, ISSN: 2473-4209. DOI: [10.1002/mp.13735](https://doi.org/10.1002/mp.13735). [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.13735> (visited on 03/06/2024).

- [6] J. Long, E. Shelhamer, and T. Darrell, *Fully convolutional networks for semantic segmentation*, Mar. 8, 2015. arXiv: [1411.4038\[cs\]](https://arxiv.org/abs/1411.4038). [Online]. Available: <http://arxiv.org/abs/1411.4038> (visited on 03/09/2024).