

Segmentation d’images thermiques de la voûte plantaire par Deep Learning

Asma BOUGRINE, Roger LEDEE, Raphael CANALS, Rachid HARBA, Meryem JABLOUN

Univ. Orleans, PRISME, EA 4229, F45072, Orleans, France
asma.bougrine@univ-orleans.fr

Résumé – L’ulcère du pied diabétique est la principale cause d’amputation du pied. Un dépistage précoce est possible grâce à l’imagerie thermique de la voûte plantaire. Afin de développer une nouvelle technologie mobile et conviviale d’analyse thermique des surfaces plantaires du pied, nous proposons d’utiliser des images infrarouges acquises par un smartphone équipé d’une caméra thermique dédiée. Cette analyse nécessite en première phase la segmentation automatique de nos images. Pour cette tâche, nous comparons les résultats obtenus par des méthodes de segmentation sémantique de Deep Learning (FCN, Seg-Net, U-Net) avec ceux d’un récent modèle de contour actif avec *a priori* de forme. Notre base de données est composée d’images de sujets sains et pathologiques. Nous avons utilisé 80% des images de la base pour l’apprentissage des 3 réseaux et 20% pour le test. Les résultats montrent que la méthode SegNet surpasse les trois autres avec un Coefficient de Similarité de Dice (DSC) de 97,26%. Cette méthode s’est montrée également efficace et robuste pour la segmentation simultanée des deux pieds avec un DSC de 96,8%. Ainsi, SegNet est la méthode choisie pour être implémentée dans le cadre d’une application mobile d’analyse thermique de la surface plantaire du pied diabétique.

Abstract – Foot ulceration can be prevented by using thermal information of the plantar foot surface. Important indicators can be provided with a thermal infrared image that we acquired using a smartphone equipped by a dedicated thermal camera. Our aim is the segmentation of these plantar foot thermal images. To that end, we compare the previously proposed prior shape-based method and three deep learning methods namely FCN, SegNet and U-Net. Our database is composed of images coming from healthy and pathological subjects. We use 80% to train the 3 deep learning networks and 20% for test. Results show that the SegNet method outperforms the three other ones with a Dice Similarity Coefficient (DSC) equal to 97.26%. This method also shows efficiency in segmenting both feet simultaneously with a DSC equal to 96.8% for a smartphone based plantar foot thermal analysis for diabetic patients.

1 Introduction

L’ulcération du pied diabétique est un problème sévère de santé puisqu’il représente la principale raison d’amputation des membres inférieurs. L’incidence d’un ulcère du pied peut être réduite de 70% si une hyperthermie du pied est détectée. L’hyperthermie du pied est définie comme une différence de température supérieure à 2,2°C entre une région du pied et la même région sur le pied contralatéral [1]. Cette information peut être identifiée à l’aide d’une caméra infrarouge thermique, tel que rapporté dans plusieurs études [2, 3]. La plupart de ces études nécessitent un protocole d’acquisition contraignant pour cacher toutes les sources thermiques autres que celles provenant des voûtes plantaires. Notre objectif est de développer une nouvelle technologie mobile et conviviale pour analyser la température des surfaces plantaires du pied diabétique et aider à la prévention de l’ulcère. Nous proposons de nous affranchir les limitations d’acquisition et de prendre les images à main levée avec un smartphone équipé d’une caméra thermique dédiée. La segmentation automatique des images acquises dans ces conditions est difficile. En effet, les méthodes classiques de segmentation d’images échouent quand elle sont appliquées à ces images spéci-

ifiques comme nous l’avons démontré dans un travail récent [4]. Nous avons cependant présenté une méthode de contour actif basée sur le modèle de Kass *et al.* [5] en ajoutant un terme énergétique supplémentaire qui impose au contour actif de respecter la forme *a priori* du pied au cours des déformations successives. Dans un travail antérieur [6], nous avons comparé la méthode proposée à deux modèles de contour actifs basés aussi sur la forme *a priori*. Les résultats ont montré que notre méthode est plus performante que les deux autres. Mais les caractéristiques des images sont différentes d’une personne à l’autre. Prenons l’exemple d’une personne saine (ou non) dont les orteils sont toujours froids. Dans ces cas, les résultats sont loin d’être optimales

Ces dernières années, les techniques de Deep Learning ont fait des progrès spectaculaires. Initialement destinés à la classification des images, ils ont permis de résoudre une variété de problèmes de vision par ordinateur, notamment la segmentation sémantique. La segmentation sémantique exige non seulement la discrimination au niveau du pixel, mais aussi sa localisation dans l’image. Ainsi, tous les réseaux de segmentation sémantique sont construits sur deux blocs principaux. Tout d’abord, le réseau d’encodeur

(souséchantillonnage) qui s'apparente à un réseau de classification préentraîné tel que, le VGG-16 [7]. Puis vient le réseau décodeur (suréchantillonnage) qui déconvolue l'ensemble des caractéristiques en sortie du premier bloc pour fournir une carte de classification. Les approches de Deep Learning les plus populaires et appliquées dans plusieurs domaines sont le réseau entièrement convolutifs (FCN) [8], le réseau SegNet [9] et le réseau U-Net [10]. Ces méthodes exploitent les réseaux de neurones convolutifs (CNN) existants comme modèles puissants de classification. Aucune étude n'a a priori utilisé le Deep Learning pour la segmentation des images thermiques de la voûte plantaire.

Dans le présent article, nous proposons de sélectionner la méthode la plus performante pour la segmentation des images thermiques de la voûte plantaire de pied, pour une application mobile qui sera utilisée en routine clinique. Ainsi nous comparons quatre différentes méthodes ; notre méthode précédemment proposée basée sur l'*a priori* de forme et trois réseaux deep Learning, à savoir le réseau entièrement convolutifs (FCN), SegNet et U-Net. Pour entraîner ces réseaux, nous avons lancé une campagne d'acquisition de données au sein du service du pied diabétique de l'hôpital régional d'Orléans. 198 images de pieds pathologiques et 50 images provenant de groupe sain ont été acquises. Un total de 248 images constitue notre base de données.

Le reste du document est organisé comme suit. Dans la section 2, nous détaillons les méthodes de segmentation utilisées. Dans la section 3, nous décrivons notre campagne d'acquisition et le prétraitement de notre base de données. Ainsi, nous définissons les paramètres réglés et montrons les résultats qualitatifs et quantitatifs pour sélectionner, la méthode la plus appropriée de segmentation des images thermiques de la voûte plantaire. Les conclusions et les perspectives sont enfin présentées dans la dernière section.

2 Méthodes

2.1 Les méthodes de Deep Learning

Nous rappelons qu'un réseau CNN est composé par un ensemble de couches, notamment les couches convolutives permettant l'extraction des caractéristiques, les couches de pooling pour un souséchantillonnage des données, et finalement les couches entièrement connectées suivies par la fonction Softmax pour la classification. Un réseau de segmentation sémantique est composé par un réseau d'encodeur (un CNN préentraîné) puis d'un réseau de décodeur qui permet la projection de la carte de segmentation sur l'espace de l'image.

Le réseau FCN [8] : L'idée de cette approche est de transformer les modèles de classification existants et puissant en modèles entièrement convolutifs. Ceci est effectué en remplaçant les couches entièrement connectées par des couches entièrement convolutives avec des convolutions 1x1. Plutôt que produire des scores de classification, des

cartes spatiales sont générées en sortie des couches d'encodeur. En basse résolution, le processus produit une carte de présence de classe qui est suréchantillonnée à l'aide de convolutions transposées pendant la phase du décodeur. De plus, le suréchantillonnage est, à chaque étape, amélioré par la concaténation des cartes des caractéristiques prédites à partir du processus de souséchantillonnage. Enfin, pour récupérer complètement les informations spatiales perdues dans les couches de souséchantillonnage, une connexion par skip est également introduite après chaque bloc de convolution.

Le réseau SegNet [9] : Le réseau encodeur de SegNet est composé de 13 couches convolutives qui représentent les 13 premières couches convolutives du réseau VGG-16 [7]. Le processus d'apprentissage est ensuite initialisé à l'aide des poids entraînés pour la classification. Les couches entièrement connectées sont supprimées et remplacées par des cartes de caractéristiques de plus haute résolution à la plus profonde sortie de l'encodeur. Les 13 couches de l'encodeur ont leurs 13 couches correspondantes de décodeur. Le décodeur est composé d'un ensemble de couches de suréchantillonnage et de convolution. Chaque couche de suréchantillonnage du bloc décodeur correspond à une couche de max-pooling dans la partie encodeur. Ces couches suréchantillonnent les cartes de caractéristiques en utilisant les indices max-pooling correspondantes dans l'encodeur. Elles sont ensuite convoluées à l'aide d'un ensemble de filtres. Une fois que les cartes de caractéristiques sont restaurées à la résolution originale, elles sont envoyées au classificateur softmax pour générer la segmentation finale.

Le réseau U-Net [10] : Ce modèle est l'extension du réseau FCN [8] pour les images biologiques. La partie encodeur d'U-Net possède une architecture de type FCN qui extrait les caractéristiques avec des convolutions 3x3. Le bloc décodeur utilise la déconvolution qui réduit le nombre de cartes de caractéristiques tout en augmentant leurs dimensions. Ces cartes rognées à partir de l'encodeur sont copiées dans le décodeur pour éviter la perte des informations. Les cartes de caractéristiques sont transformées avec une convolution de 1x1 pour produire la carte de segmentation et finalement classer chaque pixel de l'image d'entrée avec le classificateur softmax.

2.2 Notre méthode contour actif avec *a priori* de forme[4]

Nous avons proposé de modifier la fonction de Kass *et al.* [5] en ajoutant un terme d'énergie de forme *a priori*. Ce terme calcule la différence normalisée entre la courbure de la courbe de contour actif C et la courbure de la forme *a priori* C^* lors de l'évolution du contour. L'énergie totale du modèle proposé est la somme de quatre fonctions énergétiques incluant l'énergie de forme *a priori* E_{PS} (Équation 1), l'énergie interne E_{intern} , l'énergie de l'image E_{image} et l'énergie externe E_{con} . L'énergie interne contient deux termes : longueur et courbure. L'énergie de l'image est

donnée par les informations de gradient. La contrainte externe est l'énergie du ballon.

$$E_{PS} = \gamma |C_{ss}(s) - \zeta C_{ss}^*(s)|^2, \quad (1)$$

où γ est le poids de l'énergie de forme *a priori*, $\zeta = \frac{|C_{ss}(s)|}{|C_{ss}^*(s)|}$ est le facteur de normalisation, $C_{ss}^* = \frac{\partial^2 C^*}{\partial s^2}$ et $C_{ss} = \frac{\partial^2 C}{\partial s^2}$. La méthode proposée est conçue pour être invariante par rapport à l'échelle, la rotation et la translation.

3 Base de données et résultats

3.1 Le choix de la caméra

Nous avons choisi la caméra thermique FlirOne Pro conçue pour être connectée à un smartphone. Cette caméra a une résolution de 160x120 pixels et une plage spectrale du capteur thermique de 8-14 μm . FlirOne Pro peut détecter des différences de température de 0.1°C, ce qui est suffisant pour détecter les éventuelles variations d'hyperthermie.

3.2 La base de données

Nous avons réalisé deux campagnes d'acquisition pour former une base de groupe témoin et une base de groupe pathologique; la première base est issue de 25 personnes non diabétiques qui ont participé à la première campagne d'acquisition [4]. Ce groupe est composé de 10 femmes et de 15 hommes de l'Université d'Orléans dont l'âge moyen était de 34 ans. La deuxième base est construite à partir de 36 personnes diabétiques qui ont participé à notre deuxième campagne d'acquisition au sein du Service des pieds diabétiques de l'Hôpital régional d'Orléans. Ce groupe est composé de 10 femmes et de 26 hommes dont l'âge moyen est de 69 ans. Ce groupe comprend des personnes atteintes du pied diabétique qui peuvent avoir des ulcères du pied. Nous excluons les personnes qui ont des parties amputées du pied. Les participants ont été suivis sur une période de 2 mois. Un patient peut participer aux acquisitions plusieurs fois. Au total, 99 images ont été acquises. Pour les deux campagnes nous avons respecté le protocole d'acquisition détaillé dans [4].

Nous avons utilisé 20% de la base de données correspondant à 50 images pour valider les résultats et comparer les méthodes détaillées ci-dessus. Les autres 80% de la base sont donc utilisées pour l'apprentissage des modèles de Deep Learning. Nous avons appliqué une opération d'augmentation de données à notre base d'apprentissage en appliquant des rotations et des changements de contraste. L'ensemble des données d'apprentissage utilisé comprend alors au total 1134 images.

3.3 Paramétrage et résultats

3.3.1 Le choix des paramètres

Les méthodes appliquées nécessitent un réglage de paramètres. Les paramètres choisis sont ceux qui ont donné les meilleurs résultats après plusieurs tests. Les trois méthodes de Deep Learning sont toutes basées sur le mo-

dèle de réseau préentraîné VGG-16 [7]. La profondeur de l'encodeur est choisie égale à 5 (valeur par défaut). Nous entraînons les réseaux à l'aide de la méthode de descente de gradient stochastique avec momentum (SGDM) comme optimiseur. Nous avons choisi la valeur initiale de momentum égale à 0,9 pour FCN et segNet, le momentum est fixé à 0.3 pour U-Net. Le taux d'apprentissage initial est 0,01 pour FCN et SegNet qui est la valeur par défaut pour l'optimisateur SGDM. Pour U-Net, cette valeur a été choisie égale à 0,0001. L'algorithme SGDM évalue le gradient et met à jour les paramètres à l'aide d'un sous-ensemble d'apprentissage (un mini-batch). La taille de ce mini-batch est choisie égale à 4 images pour les 3 réseaux. Lorsque l'algorithme d'apprentissage passe complètement sur l'ensemble des données d'apprentissage, une époque est exécutée. Le nombre maximum d'époques est fixé à 100. Les paramètres de notre méthode de contour actif sont α et β pour E_{intern} , W_{edge} , δ et γ qui contrôlent, respectivement, E_{image} , E_{con} et E_{PS} . L'ensemble des paramètres est : $\alpha=0.1$, $\beta=4$, $W_{edge}=20$, $\delta=0.2$ et $\gamma=35$. Le contour initial choisi est le contour de forme de référence redimensionné placé à l'intérieur de la voûte plantaire.

3.3.2 Résultats qualitatifs et quantitatifs

Nous évaluons d'abord qualitativement la performance des 4 méthodes. Nous présentons sur la Figure 1 les résultats de segmentation pour deux images sélectionnées de la base de données de test. Les résultats montrent qualitativement que les méthodes SegNet et FCN donnent de meilleurs résultats comparées à la méthode précédemment proposée basée sur la forme *a priori* et au réseau U-Net. Pour quantifier ces résultats, nous utilisons le coefficient de similarité de Dice (DSC) [11] comme une métrique d'évaluation. Ce score évalue la similarité entre la région du pied donnée par le contour de vérité terrain manuellement annoté par l'expert et les régions résultats de la segmentation. Le tableau 1 présente les scores moyens de DSC donnés par les méthodes avec leurs écarts types respectifs (STD).

TABLE 1 – Valeurs moyennes de DSC (\pm STD) données par les méthodes de segmentation.

FCN	96.16% \pm 0.85%
SegNet	97.26% \pm0.69%
U-Net	74.35% \pm 9.58%
Méthode proposée [4]	94% \pm 2%

Nous remarquons que le réseau U-Net n'a pas donné de bons résultats de segmentation. Cela peut être dû au fait que la taille de la base de données d'apprentissage n'est pas suffisante. Le meilleur résultat de segmentation est fourni par la méthode SegNet avec un DSC égal à 97,26% et un faible écart-type. Puisque notre objectif est de développer une application mobile automatique pour la visualisation et l'analyse thermique de la voûte plantaire chez les diabétiques, il est important de segmenter simultanément le

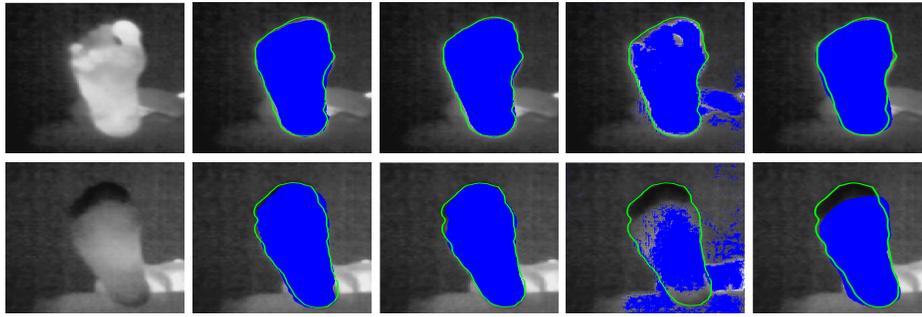


FIGURE 1 – Segmentation de deux images thermiques de la voûte plantaire (1 ère colonne). Les résultats de segmentation sont donnés par : FCN (2ème colonne), SegNet (3ème colonne), U-Net (4ème colonne) et notre méthode (5ème colonne). Les régions bleues correspondent aux résultats de segmentation tandis que la courbe verte représente la vérité de terrain.

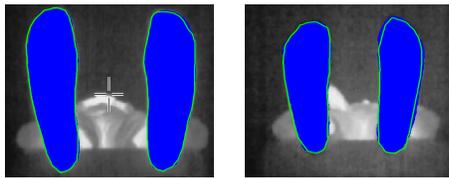


FIGURE 2 – La segmentation de 2 images thermiques contenant deux pieds en utilisant la méthode SegNet. Les régions bleues correspondent aux résultats de la segmentation tandis que la courbe verte représente les contours de la vérité terrain.

pied droit et le pied gauche. La Figure 2 montre que SegNet entraîné réussit à localiser et segmenter correctement les deux pieds dans ces deux images de la base de test. Le DSC moyen obtenu sur 25 images contenant chacune deux pieds est de $96,8\% \pm 0,81\%$.

4 Conclusions et perspectives

Dans le présent article, nous avons comparé notre méthode basée sur l'*a priori* de forme et trois méthodes de Deep Learning pour la segmentation d'images thermiques de la voûte plantaire. La comparaison effectuée sur notre base de données de test a montré la supériorité du réseau SegNet avec un DSC de $97,26\%$. Cette méthode est robuste, fournit de bons résultats et a démontré son efficacité en segmentant les deux pieds simultanément avec un DSC égal à $96,8\%$. Cela rend la méthode SegNet adaptée à une application smartphone pour l'analyse thermique de la voûte plantaire chez les patients diabétiques. En conclusion, aucun système d'isolation contraignant ne sera nécessaire, les images seront prises à main levée avec un smartphone équipé d'une caméra thermique dédiée, et le traitement sera entièrement automatique. Nous avons l'intention d'appliquer ce protocole en routine clinique pour la prévention des ulcères du pied chez les diabétiques à partir de la détection de l'hyperthermie.

Remerciements

Projet Horizon 2020 : Ce projet a été financé par le programme européen de recherche et d'innovation Horizon 2020 dans le cadre

du programme Marie Skłodowska-Curie.

Références

- [1] David G Armstrong, Katherine Holtz-Neiderer, Christopher Wendel, M Jane Mohler, Heather R Kimbriel, and Lawrence A Lavery. Skin temperature monitoring reduces the risk for diabetic foot ulceration in high-risk patients. *The American journal of medicine*, 120(12) :1042–1046, 2007.
- [2] Luis Vilcahuaman, Rachid Harba, Raphaël Canals, M Zequera, Carlos Wilches, MT Arista, L Torres, and H Arbañil. Detection of diabetic foot hyperthermia by infrared imaging. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4831–4834, 2014.
- [3] Yang Liu, Andrea Polo, M Zequera, Rachid Harba, Raphaël Canals, Luis Vilcahuaman, and Y Bello. Detection of diabetic foot hyperthermia by using a regionalization method, based on the plantar angiosomes, on infrared images. *2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, pages 1389–1392, 2016.
- [4] Asma Bougrine, Rachid Harba, Raphael Canals, Roger Ledee, and Meryem Jabloun. A joint snake and atlas-based segmentation of plantar foot thermal images. *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6, 2017.
- [5] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes : Active contour models. *International journal of computer vision*, 1(4) :321–331, 1988.
- [6] Asma Bougrine, Rachid Harba, Raphael Canals, Roger Lédée, and Meryem Jabloun. A comparison of active contour prior shape segmentation methods : application to diabetic plantar foot thermal images. *8th International Conference on Digital Image Processing and Vision*, 2019.
- [7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv :1409.1556*, 2014.
- [8] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. pages 3431–3440, 2015.
- [9] Vijay Badrinarayanan, Ankur Handa, and Roberto Cipolla. Segnet : A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. *arXiv preprint arXiv :1505.07293*, 2015.
- [10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net : Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, 2015.
- [11] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3) :297–302, 1945.