

Recalage métamorphique d’images par réseau de neurones résiduels

Matthis MAILLARD¹, Isabelle BLOCH^{2,1}, Pietro GORI¹

¹LTCI, Télécom Paris, Institut Polytechnique de Paris

²Sorbonne Université, CNRS, LIP6, Paris

matthis.maillard@telecom-paris.fr, isabelle.bloch@sorbonne-universite.fr,
pietro.gori@telecom-paris.fr

Résumé – En imagerie médicale, la plupart des méthodes de recalage d’images supposent implicitement une correspondance bijective entre l’image source et l’image cible. Cependant, ce n’est pas nécessairement le cas lorsqu’il s’agit d’images médicales avec des pathologies (présence d’une tumeur, d’une lésion, etc.). Pour faire face à ce problème, le modèle des *Metamorphoses* a été proposé. Il modifie à la fois la forme et l’apparence d’une image pour traiter les différences géométriques et topologiques. Cependant, le temps et la charge de calcul élevés ont entravé ses applications jusqu’à présent. Nous proposons ici une mise en œuvre de la métamorphose par apprentissage résiduel profond qui réduit considérablement le temps de calcul lors de l’inférence. De plus, nous montrons que le cadre proposé peut facilement intégrer des connaissances préalables sur la localisation des changements topologiques (par exemple les masques de segmentation) qui peuvent agir comme une régularisation spatiale pour séparer correctement les changements d’apparence et de forme.

Abstract – In medical imaging, most of the image registration methods implicitly assume a one-to-one correspondence between the source and target images. However, this is not necessarily the case when dealing with pathological medical images (presence of a tumor, lesion, etc.). To cope with this issue, the *Metamorphosis* model has been proposed. It modifies both the shape and the appearance of an image to deal with the geometrical and topological differences. However, the high computational time and load have hampered its applications so far. Here, we propose a deep residual learning implementation of Metamorphosis that drastically reduces the computational time at inference. Furthermore, we also show that the proposed framework can easily integrate prior knowledge of the localization of topological changes (*e.g.*, segmentation masks) that can act as spatial regularization to correctly disentangle appearance and shape changes.

1 Introduction

Le recalage d’images a de nombreuses applications en imagerie médicale, telles que l’analyse statistique, la fusion de modalités, la planification et le suivi chirurgical. La plupart des méthodes supposent que les deux images ont la même topologie et le même nombre de composants anatomiques, et recherchent donc un difféomorphisme [1, 17]. Cependant, dans certaines études, il est nécessaire de recalibrer des images caractérisées par des nombres différents de composants anatomiques, comme une image saine et une image avec une tumeur [5].

Une des premières stratégies pour traiter une différence de topologie entre les images source et cible est le masquage de la fonction de coût, où la lésion est exclue lors du calcul de la fonction de similarité entre les images [15]. Cependant, lorsque la lésion est trop importante, la déformation dans la zone masquée peut être très importante. Une autre stratégie, lorsqu’on veut recalibrer deux images avec et sans tumeur, consiste à modéliser d’abord la croissance de la tumeur dans l’image saine en utilisant un modèle biophysique [4, 14], puis à effectuer le recalage. Cette stratégie traite la différence topologique, puisque les images contiennent maintenant le même nombre de composants. Cependant, cette approche nécessite une initialisation par l’utilisateur, elle a un temps de calcul élevé et elle n’est pas gé-

nérique car elle n’est applicable qu’en présence de tumeurs et les modèles de croissance sont différents pour chaque type de tumeur. De manière équivalente, des techniques d’*impainting* pour effacer la tumeur ont également été proposées [18]. Cependant, elles peuvent produire des images inexactes ou irréelles lorsqu’il s’agit de tumeurs ou de lésions de grande taille.

La méthode des métamorphoses [6] est plus générique et offre des garanties théoriques pour traiter à la fois les différences de forme et d’apparence. Elle consiste à appliquer de manière répétitive (simultanément) des déformations infinitésimales de forme et d’apparence à l’image source. À notre connaissance, il n’existe aucune application clinique de cette méthode et une mise en œuvre récente [3] indique qu’elle prend beaucoup de temps.

Récemment, motivées par un temps de calcul plus rapide lors de l’inférence, plusieurs approches d’apprentissage profond ont été développées pour le recalage d’images. Les réseaux neuronaux sont entraînés sur une base de données de paires d’images et pas seulement pour une paire d’images fixe [11]. L’apprentissage est donc très long mais ce type de méthodes est beaucoup plus rapide au moment de l’inférence que les méthodes précédentes qui doivent optimiser une fonctionnelle pour chaque paire d’images. De même, l’apprentissage profond peut également être utilisé pour recalibrer des images ayant

des topologies différentes. Dans [2], les auteurs ont proposé un auto-encodeur métamorphique (MAE) pour modifier à la fois la géométrie et l'apparence d'une image. Toutefois, avec cette méthode, les changements d'intensité peuvent modifier la forme de l'image d'entrée.

Inspirés par [7], nous proposons une mise en œuvre des métamorphoses reposant sur le réseau ResNet [13] qui surmonte toutes les limitations précédentes. Ce travail constitue une extension de [16], par rapport auquel nous introduisons un nouveau schéma d'intégration pour produire des images nettes. Nos contributions sont les suivantes :

- nous utilisons un réseau résiduel [13] pour résoudre le système d'équations différentielles des métamorphoses ;
- il s'agit d'une méthode d'apprentissage où nous utilisons un ensemble d'images sources et une image cible fixée ;
- nous introduisons une régularisation locale qui consiste à limiter les changements d'intensité à une zone pré-spécifiée (un masque de segmentation). Nous montrons que cela produit une meilleure séparation entre les changements de forme et d'apparence.

2 Méthode

Formulation mathématique. Soit $\Omega \subset \mathbb{R}^d$ un domaine borné, avec $d \in \{2, 3\}$. Soit V un espace de Hilbert à noyau reproductible (RKHS) avec un noyau K de champs de vecteurs de support Ω et T fois continûment différentiables, où $T \in \mathbb{N}^*$. Soit I une image en niveaux de gris définie comme une fonction de carré intégrable et différentiable $I : \Omega \rightarrow \mathcal{R}$. Le but des métamorphoses est de modifier une image source I pour qu'elle soit parfaitement recalée avec une image cible J . Le modèle associe des déformations difféomorphiques à des changements d'intensité additifs. De façon similaire à [3, 6], l'évolution de l'image I au temps $t \in [0, 1]$ est :

$$\partial_t I_t = v_t \cdot I_t + \mu^2 z_t = -\langle \nabla I_t, v_t \rangle + \mu^2 z_t \quad \text{avec } I_0 = I \quad (1)$$

où $v_t \cdot I_t$ implique que I_t est déformé par un champ vectoriel infinitésimal $v_t \in V$ et $z_t : \Omega \rightarrow \mathbb{R}$ est la partie additive correspondant à la variation infinitésimale d'intensité (appelée image résiduelle ou momentum). L'hyperparamètre $\mu^2 \in \mathbb{R}^+$ équilibre les variations d'intensité et géométriques. Comme dans [3], nous présentons le recalage métamorphique comme un problème de correspondance inexacte minimisant la fonction de coût suivante :

$$E = \frac{1}{2} \|I_1 - J\|_2^2 + \lambda \int_0^1 \|v_t\|_V^2 + \|\mu z_t\|_2^2 dt \quad (2)$$

où le premier terme est le terme d'attache aux données classique L_2 (notons que d'autres termes pourraient également être utilisés) et le second terme, pondéré par λ , est la régularisation. Il est composé, respectivement, de l'énergie géométrique totale et de l'énergie d'intensité de la déformation. Comme indiqué

dans [3], les équations géodésiques de la métamorphose sont :

$$\begin{cases} v_t = -K * (z_t \nabla I_t) & (3a) \\ \partial_t z_t = -\nabla \cdot (z_t v_t) & (3b) \\ \partial_t I_t = -\langle \nabla I_t, v_t \rangle + \mu^2 z_t & (3c) \end{cases}$$

avec $\|v_t\|_V^2 = \langle z_t \nabla I_t, K * (z_t \nabla I_t) \rangle$, où K est un noyau gaussien.

De ce système d'équations, on peut remarquer que v_t est complètement défini par z_t et I_t , faisant ainsi de z_t la seule inconnue. Le momentum z_t a donc un double rôle. Il représente la variation d'intensité additive et il est aussi le paramètre de la déformation. Cela facilite le calcul mais rend en même temps plus difficile la séparation entre les variations de forme et d'intensité. Inspirés par [7, 10], nous proposons d'utiliser un réseau neuronal résiduel (ResNet) pour trouver la solution de ce système. Nous profitons de la similarité entre les réseaux ResNet et les solutions numériques des équations aux dérivées partielles par la méthode d'Euler pour résoudre l'équation 3b. En effet, l'intégration numérique de cette équation, en utilisant des pas de temps discrets t , s'écrit :

$$z_{t+1} = z_t - \frac{1}{T} \nabla \cdot (z_t v_t) \quad \text{pour } t = i/(T-1), i \in 0, \dots, T-1 \quad (4)$$

où T est le nombre de pas. En remplaçant la divergence par un réseau de neurones, on obtient un ResNet :

$$z_{t+1} = z_t + \frac{1}{T} f_{\theta_t}(z_t, I_t) \quad (5)$$

où f_{θ_t} est modélisé comme une couche convolutive suivie d'une fonction d'activation (*Leaky ReLU*) et de deux autres couches convolutives. Par rapport à l'équation 3b, v_t est remplacé par I_t car v_t est une fonction de I_t (et de z_t). Le réseau est construit comme une séquence de T blocs de convolution f_{θ_t} . À chaque pas de temps t , z_{t+1} est calculé à l'aide de l'équation 5. Ensuite, v_{t+1} est obtenu directement par l'équation 3a et on détermine I_{t+1} en appliquant la transformation géométrique induite par v_t et en ajoutant les résidus z_t comme dans l'équation 3c.

L'objectif est de calculer les transformations entre chaque image d'un ensemble de données d'apprentissage de N images et une image cible fixe. Par conséquent, il faut minimiser l'énergie $E_L(\theta, z_0) =$

$$\sum_{n=1}^N \left[\frac{1}{2} \|I_T^n - J\|_2^2 + \frac{\lambda}{T} \left[\sum_{t=0}^{T-1} \|v_t^n\|_V^2 + \mu^2 (\|z_0\|_2^2 + \sum_{t=1}^{T-1} \|z_t^n\|_2^2) \right] \right] \quad (6)$$

où n est l'indice de l'image I^n dans l'ensemble de données. Les paramètres à optimiser sont les poids du réseau neuronal θ et les résidus initiaux z_0 . Nous utilisons l'optimiseur Adam avec un pas de 10^{-3} . Une fois le réseau entraîné, il peut être utilisé pour recaler toute image du même type que celles de la base de donnée d'apprentissage sur l'image cible. Cela permet le recalage d'un grand nombre d'images en un temps très court (moins d'une seconde en inférence). Cela est particulièrement important, par exemple, lorsqu'il faut recalcr un grand

ensemble d'images sur un modèle commun pour faire une analyse statistique.

Régularisation locale. Le principal inconvénient des métamorphoses est qu'il est difficile de contrôler la séparation entre forme et apparence. Par exemple, une solution triviale serait de fixer la fonction de déformation géométrique globale à l'identité (pas de changement géométrique) et la carte de transformation d'apparence globale à $J - I_0$. Dans ce cas, la distance L_2 entre l'image déformée et J serait égale à 0 mais ce ne serait pas un résultat satisfaisant puisque les structures homologues doivent être appariées en utilisant uniquement les déformations géométriques alors que les changements d'apparence et topologiques (les nouvelles composantes) doivent être pris en compte par les modifications d'intensité. La séparation peut être contrôlée en réglant les hyperparamètres μ et λ . Cependant, trouver les bons paramètres est une tâche difficile et ils sont différents pour chaque couple d'images. S'ils ne sont pas correctement choisis, la carte d'apparence pourrait, par exemple, modifier la forme de l'image, faussant ainsi les résultats et leur interprétation.

Pour améliorer le contrôle de cette séparation entre forme et apparence, nous introduisons une régularisation locale qui consiste à limiter la modification en intensité à une région prédéfinie (par exemple une tumeur). Nous introduisons un masque m_0 où $m_0(x) = 1$ si le pixel x est dans la zone prédéfinie et 0 sinon. Lors de la déformation, la forme de la tumeur va être modifiée, par conséquent, le masque doit subir la même déformation. On a donc $\partial_t m_t = v_t \cdot m_t$. La transformation de l'image est alors : $\partial_t I_t = v_t \cdot I_t + \mu^2 m_t z_t$. En utilisant cette équation et $\|\sqrt{m_t} z_t\|_2^2$ comme terme de régularisation pour z_t , nous obtenons les mêmes équations géodésiques 3a et 3b.

Schéma d'intégration. La solution numérique de l'équation 3c en utilisant la méthode d'Euler est la suivante :

$$\begin{aligned} I_{t+1} &= I_t + \delta v_t \cdot I_t + \delta \mu^2 z_t \\ &= I_t \circ (Id - \delta v_t) + \delta \mu^2 z_t \end{aligned} \quad (7)$$

où le pas d'intégration est $\delta = \frac{1}{T}$. Cela indique qu'à chaque pas de temps t , I_t est déformé par le champ vectoriel v_t . Du point de vue du calcul, l'image résultante est une interpolation de I_t . Par conséquent, l'image finale I_T est obtenue après T interpolations de I_0 . L'utilisation d'un nombre élevé d'interpolations successives au plus proche voisin crée un effet de pixellisation et une interpolation bilinéaire (ou trilineaire) rend l'image floue. Pour éviter ces deux inconvénients, nous réécrivons l'équation 7 de sorte que I_{t+1} soit une déformation directe de I_0 et non de I_t . Nous remplaçons récursivement I_t par son expression en fonction de I_{t-1} dans l'équation 7 jusqu'à atteindre I_0 . On obtient :

$$I_{t+1} = I_0 \circ \phi_{0,t+1}^{-1} + \delta \mu^2 \sum_{i=0}^t z_i \circ \phi_{i+1,t+1}^{-1} \quad (8)$$

où $\phi_{i,t+1}(x)$ est la position au temps $t + 1$ de l'élément à la

position x au temps i :

$$\begin{cases} \phi_{i,t}^{-1} = Id & \text{si } i = t \\ \phi_{i,t}^{-1} = Id - \delta v_{t-1} & \text{si } i = t - 1 \\ \phi_{i,t}^{-1} = (Id - \delta v_i) \circ \dots \circ (Id - \delta v_{t-1}) & \text{sinon.} \end{cases}$$

Ici, Id désigne la fonction d'identité : $Id(x) = x$.

Avec cette expression, chaque image I_t résulte d'un seul rééchantillonnage de I_0 . Notons que même si I_{t+1} n'est pas directement calculée à partir de I_t , il est toujours nécessaire de calculer cette dernière car elle est requise pour obtenir v_t et z_t .

Comme on peut le voir sur la figure 1, ce schéma d'intégration produit des déformations beaucoup plus nettes.

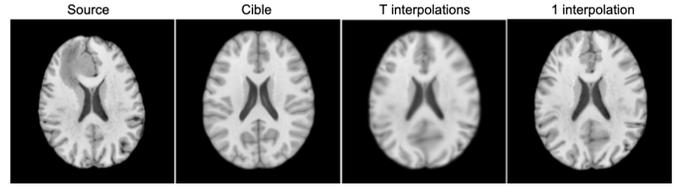


FIGURE 1 – Comparaison entre le schéma d'intégration de l'équation 8 (T interpolations) et celui de l'équation 9 (1 interpolation), en utilisant notre régularisation locale.

3 Expériences

Base de données. La base de données utilisée est BraTS 2021 [8] comprenant quatre modalités IRM et l'image de segmentation tumorale associée pour 1251 cerveaux cancéreux. Pour chaque patient, nous sélectionnons la même coupe de la modalité T1 et la sous-échantillons pour obtenir une image de taille 208×208 . Nous choisissons aléatoirement 40 images dans le jeu de données pour former un ensemble d'évaluation. L'image cible associée à cet ensemble de données est le modèle linéaire MNI152 [12]. Nous utilisons la segmentation de la tumeur comme masque pour la régularisation locale.

Résultats. Nous montrons l'intérêt de la régularisation locale sur la figure 2. Sans régularisation, la déformation géométrique de l'image source n'est pas très efficace. Les ventricules, notamment, ne sont pas bien recalés sur ceux de l'image cible. Cependant, sur la transformation totale, les ventricules sont très similaires à ceux de l'image cible. Cela signifie que la transformation en intensité a modifié la forme des ventricules. Cela n'est pas satisfaisant car nous voulons une séparation entre les transformations géométrique et d'apparence. Avec la régularisation locale, ce problème n'apparaît pas puisque la zone masquée est située en dehors des ventricules.

Pour évaluer la qualité du recalage, nous segmentons les ventricules de l'image déformée et calculons le score de Dice avec la segmentation de référence des ventricules dans l'atlas MNI (voir table 1). Comme algorithme de segmentation, nous avons entraîné un modèle U-Net sur des coupes 2D provenant du jeu de données OASIS [9]. Les segmentations de référence ont été

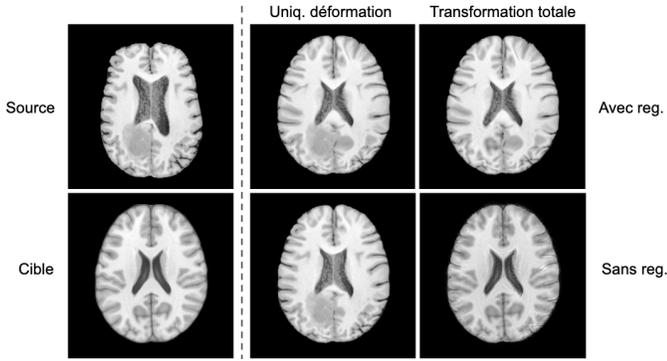


FIGURE 2 – Comparaison de notre méthode avec et sans régularisation. La colonne du milieu montre l’image source qui a subi uniquement la déformation géométrique et la dernière colonne indique la transformation totale.

obtenues à l’aide du logiciel FreeSurfer¹. Le modèle U-Net entraîné a obtenu un score de Dice moyen de 0,91 sur l’ensemble de validation.

Nous comparons notre méthode, avec et sans régularisation locale, au masquage par fonction de coût [15] et à la méthode Voxelmorph [11]. La présence de la tumeur modifie principalement la forme des ventricules (effet de masse), mais pas leur intensité. Toutes les méthodes devraient donc faire correspondre correctement les ventricules en utilisant uniquement des transformations géométriques. C’est pourquoi nous calculons le score de Dice et la distance SSD sur les images où seule la déformation géométrique a été appliquée et non sur les images avec la transformation totale. Les résultats montrent nettement l’apport de la méthode proposée avec régularisation, tant en distance SSD qu’en score de Dice, au prix d’une augmentation très limitée du temps de calcul en inférence par rapport à la méthode Voxelmorph (encore largement inférieur à une seconde).

TABLE 1 – Évaluation quantitative pour les méthodes de masquage de la fonction de coût (CFM), Voxelmorph et notre méthode avec et sans régularisation locale.

Méthode	SSD	Dice	Temps en inférence (s)
CFM	167 ± 68	69,1 ± 22,4	55 ± 1,2
Voxelmorph	187 ± 73	63,4 ± 22,6	0,01 ± 8.10⁻⁴
Sans régularisation	220 ± 20	65,4 ± 20,7	0,15 ± 0,003
Avec régularisation	147 ± 16	71,1 ± 15,2	0,16 ± 0,009

4 Conclusion

Cet article propose une méthode pour calculer les métamorphoses par réseau résiduel profond et ajoute une régularisation locale pour améliorer la séparation entre la forme et de l’apparence. Les résultats qualitatifs et quantitatifs sur le jeu de données BraTS montrent l’efficacité de la méthode. De plus, le modèle étant utilisé dans un contexte d’apprentissage, ce dernier offre un temps de calcul très rapide lors de l’inférence.

1. <https://surfer.nmr.mgh.harvard.edu/>

Références

- [1] J. Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1) :95–113, 2007.
- [2] A. Bône et al. Learning Joint Shape and Appearance Representations with Metamorphic Auto-Encoders. In *MIC-CAI*, pages 202–211, 2020.
- [3] A. François et al. Metamorphic image registration using a semi-Lagrangian scheme. In *GSI*, 2021.
- [4] A. Gooya et al. GLISTR : Glioma Image Segmentation and Registration. *IEEE TMI*, 31(10) :1941–1954, 2012.
- [5] A. Roux et al. MRI atlas of IDH wild-type supratentorial glioblastoma : Probabilistic maps of phenotype, management, and outcomes. *Radiology*, 293(3) :633–643, 2019.
- [6] A. Trouvé et al. Metamorphoses Through Lie Group Action. *Foundations of Computational Mathematics*, 5(2) :173–198, 2005.
- [7] B. Ben Amor et al. ResNet-LDDMM : Advancing the LDDMM Framework Using Deep Residual Networks. *CoRR*, 2021.
- [8] B.H. Menze et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE TMI*, 34(10) :1993–2024, 2015.
- [9] DS. Marcus et al. OASIS : Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.*, 19 :1498–1507.
- [10] F. Rousseau et al. Residual Networks as Flows of Diffeomorphisms. *JMIV*, 62(3) :365–375, 2020.
- [11] G. Balakrishnan et al. VoxelMorph : A Learning Framework for Deformable Medical Image Registration. *IEEE TMI*, 38(8) :1788–1800, 2019.
- [12] J. C. Mazziotta et al. A probabilistic atlas of the human brain : Theory and rationale for its development : ICBM. *NeuroImage*, 2(2, Part A) :89–101, 1995.
- [13] K. He et al. Deep Residual Learning for Image Recognition. In *CVPR*, pages 770–778. IEEE, 2016.
- [14] K. Scheufele et al. Coupling brain-tumor biophysical models and diffeomorphic image registration. *Comput Methods Appl Mech Eng*, 347 :533–567, 2019.
- [15] M. Brett et al. Spatial Normalization of Brain Images with Focal Lesions Using Cost Function Masking. *NeuroImage*, 14(2) :486–500, 2001.
- [16] M. Maillard et al. A deep residual learning implementation of metamorphosis. In *IEEE ISBI*, 2022.
- [17] M.F. Beg et al. Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. *IJCV*, 61(2) :139–157, 2005.
- [18] M. Sdika and D. Pelletier. Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. *Hum. Brain Mapp.*, 30(4), 2009.