U-ReSNet: Ultimate Coupling of Registration and Segmentation with Deep Nets

Théo Estienne[1,2]([✉]), Maria Vakalopoulou[1], Stergios Christodoulidis[3],
Enzo Battistela[1,2], Marvin Lerousseau[1,2], Alexandre Carre[2],
Guillaume Klausner[2], Roger Sun[1,2], Charlotte Robert[2],
Stavroula Mougiakakou[3], Nikos Paragios[4], and Eric Deutsch[2]

[1] Laboratoire MICS - CentraleSupélec, University Paris-Saclay,
91190 Gif-sur-Yvette, France
{theo.estienne,maria.vakalopoulou,enzo.battistela,
marvin.lerouseau}@centralesupelec.fr
[2] U1030 Molecular Radiotherapy, Paris-Sud University - Gustave Roussy - Inserm -
Paris-Saclay University, Villejuif, France
{alexandre.carre,guillaume.klausner,roger.sun,ch.robert,
eric.deutsch}@gustaveroussy.fr
[3] ARTORG Center for Biomedical Engineering Research - University of Bern,
3008 Bern, Switzerland
{stergios.christodoulidis,stavroula.mougiakakou}@artorg.unibe.ch
[4] TheraPanacea, Pépiniere Santé Cochin, Paris, France
n.paragios@therapanacea.eu

Abstract. In this study, we propose a 3D deep neural network called
U-ReSNet, a joint framework that can accurately register and segment
medical volumes. The proposed network learns to automatically gener-
ate linear and elastic deformation models, trained by minimizing the
mean square error and the local cross correlation similarity metrics. In
parallel, a coupled architecture is integrated, seeking to provide segmen-
tation maps for anatomies or tissue patterns using an additional decoder
part trained with the dice coefficient metric. U-ReSNet is trained in an
end to end fashion, while due to this joint optimization the generated
network features are more informative leading to promising results com-
pared to other deep learning-based methods existing in the literature. We
evaluated the proposed architecture using the publicly available OASIS
3 dataset, measuring the dice coefficient metric for both registration
and segmentation tasks. Our promising results indicate the potentials
of our method which is composed from a convolutional architecture that
is extremely simple and light in terms of parameters.

Keywords: Image registration · Deformable registration · Brain
tumor segmentation · 3D convolutional neural networks

1   Introduction

Registration and segmentation are two of the most well-studied problems in the medical image analysis research. Image registration seeks to find a transformation that maps two input volumes (moving and reference) to the same coordinate system, while image segmentation seeks to identify groups of voxels that belong to the same tissue structures according to their visual characteristics. For both methods a large variety of algorithms have been proposed with the deep learning based ones to achieve currently, state of the art results regarding both problems. Moreover, it is quite common, that the two methods are combined or used in the same framework. Image registration is a valuable asset for segmentation (atlas-based methods) while image segmentation provides valuable constraints to solve the registration problem since limiting registration factors such as signal variability, noise, acquisitions settings are eliminated. Therefore it is natural to consider coupling these two problems while imposing mutual consistency.

In this study, we present a framework for registration and segmentation of medical images using 3D deep convolutional neural networks. The proposed framework is inspired by the multitask learning paradigm. In particular, the proposed scheme solves and optimizes both registration and segmentation problems simultaneously. The underlying assumption is that by sharing and learning representations for both tasks, we can improve performance and generalization properties of the proposed method on unseen data sets for both tasks. This idea has conceptual similarities with [18] where the authors proposed the Atlas-Net approach in which image registration is used as data augmentation to train different networks in a step by step process.

The contributions of this paper are threefold. Firstly, we propose and present for the first time a joint framework based on deep learning for simultaneous optimisation of image segmentation and registration. The framework can produce very efficiently, due to the GPU implementation, accurate displacement and segmentation masks during a single inference. Secondly, we show that by optimising both tasks simultaneously, we can achieve similar or higher accuracy compared also to other deep learning based methods, addressing each of the tasks separately. Finally, we show that by integrating a linear transformer (affine registration) in the network, we can generate very accurate transformation maps taking advantage of the benefits of both rigid and deformable methods.

2   Related Work

Image registration methods, depending on the transformation model used, can be categorized into two distinct groups: the rigid or linear and the deformable or elastic methods. Among the different types of methods employed in medical applications, deformable registration is the most commonly used due to its richness of description [16]. The main goal of deformable registration is to determine the optimal dense transformation, per voxel, in contrast to rigid methods

where one global transformation is sought for the entire volume. Deformable registration has been addressed with a variety of methods including surface matching [12,14] or graph based approaches [7,11]. These methods have been extended to address co-registration of multiple volumes [4]. Moreover, recently a variety of deep learning based methods have been proposed, reducing significantly the computational time but maintaining the accuracy and robustness [2,3]. In particular the authors in [3] presented a deep learning framework trained for atlas-based registration of brain MR images, while in [2] the authors proposed an architecture based on dilated filters for the registration of lung MR images.

In image segmentation, deep learning based approaches are currently the state of the art. Among the variety of architectures proposed recently, fully convolutional networks that follow an autoencoder scheme, are widely used for segmentation of 2D [15] and 3D [10] medical volumes. These architectures are most of the time coupled with objective functions specific to medical imagery such as the dice coefficient [8] or introduce other constraints that can maintain the shape or other important characteristics of the organs and/ or tissues to be segmented.

There are a lot of methods in the literature proposing to address both tasks jointly. In [11] the authors present a graph-based method applied on MR brain images where the registration constraints are relaxed in the presence of brain tumors, while their formulation provides also segmentation masks of the tumor area. Using ultrasound images, a joint variational framework for the joint co-segmentation and registration has been proposed in [13], while in [5] the authors address both tasks in a joint framework, evaluating them on infant brain images. Even though there is a wide range of methods addressing image segmentation and registration jointly, showing that one solution impacts the other, according to our knowledge, this is the first time that an efficient formulation based on deep learning is presented and evaluated on brain MR images.

3   Methodology

In this study, we propose to use a single framework based on deep convolutional neural networks (CNN) for the joint registration and segmentation of medical 3D volumes. The framework consists of two branches with shared parameters, depicted in Fig. 1. During training the network uses as input a moving M and a reference R image, while its targets are their segmentation masks M_{seg} and R_{seg} respectively. In this section, we present the details for each of the important blocks of the framework as well as our final formulation for its optimization.

3.1   Linear and Elastic Registration

The network that is used for the registration is based on the one presented in [2]. In particular, the computed deformation D is a combination of a linear and elastic registration. The linear term of the deformation is generated as an intermediate output of the network and it is a 3×4 affine transformation matrix
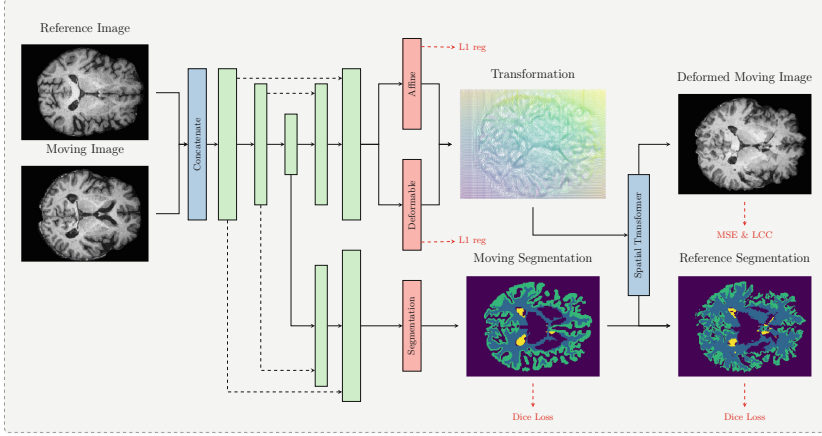
Fig. 1. The employed architecture with the registration and segmentation parts. Green layers: convolutional blocks with successive Convolution, LeakyReLu and down/upsampling operations. (Color figure online)

A, while the deformable term Φ is generated in the form of sampling coordinates for each output voxel. These sampling coordinates are generated by integrating the displacements of consecutive pixels which are also estimated as an intermediate output of the network. This approach is utilized, instead of estimating the sampling coordinates straight from the network, in order to avoid very large displacement values that are more difficult to be trained (for more details please refer to [2]). After obtaining both parts, they are both applied to the initial image jointly. The combined deformation is calculated by multiplying the affine matrix with the dense deformable sampling coordinates. More specifically, the deformed image is calculated as

$$\mathcal{W}(M, D) = \mathcal{W}\left(M, \Phi \times A\right). \tag{1}$$

where $W(\cdot)$ denotes the sampling operation which in our case consists of a backward trilinear interpolation sampling.

3.2   Anatomical Segmentation

The segmentation part of the network shares the same encoder as the registration and it consists of a symmetric decoder part trained to provide segmentation masks for the moving image. Let us define ψ the function which provides the segmentation mask. The information between the registration and segmentation is shared both by using the same encoder and by comparing the predicted segmentation mask of the moving image with the segmentation mask of the reference after applying the same deformation as in Eq. 1. The final segmentation masks for both reference and moving images are defined as

$$\widehat{M}_{seg} = \psi(M, R) \tag{2}$$

$$\widehat{R}_{seg} = \mathcal{W}(\widehat{M}_{seg}, D) = \mathcal{W}\left(\widehat{M}_{seg}, \Phi \times A\right). \tag{3}$$

with $\widehat{M}_{seg}$ and $\widehat{R}_{seg}$ to indicate the predicted moving and deformed moving segmentation respectively. Here we should mention that the segmentation is trained together with the registration part in an end to end fashion.

3.3   Optimisation Strategy

To train the network, we use a combination of multiple loss functions, each of them related to a specific task. Following the studies of [2] and [3], we use two different metrics to ensure and validate the performance of the deformation, namely the mean square error and the local cross correlation (LCC):

$$\mathcal{L}_{recon} = ||R - \mathcal{W}(M, D)||^2 + LCC(R, \mathcal{W}(M, D)). \tag{4}$$

Moreover, in order to ensure that the predicted deformations are smooth for both parts we use two different regularisation terms. In particular, we regularise the predicted deformations to be close to the identity A_I and Φ_I, calculating the L_1 norm for the two obtained displacements

$$\mathcal{L}_{reg} = \alpha||A - A_I||_1 + \beta||\nabla\Phi - \nabla\Phi_I||_1. \tag{5}$$

where the regularisation parameters α and β are essential to the joint optimization. Too low values will make the network diverge generating irregular deformations, while too high values prevent the network from creating meaningful deformations by keeping them very close to the identity. For all our experiments, we set the values in an experimental manner to the following constants: $\alpha = 0.1$ and $\beta = 10^{-11}$.

For the segmentation, we calculate two different dice coefficient losses using the predicted moving segmentation $\widehat{S}_{seg}$ and its deformed form $\widehat{R}_{seg}$.

$$\mathcal{L}_{seg} = \gamma Dice(\widehat{M}_{seg}, M_{seg}) + \delta Dice(\widehat{R}_{seg}, R_{seg}). \tag{6}$$

The parameters γ and δ were set to 1 for all our experiments.

Finally, as the network is trained end to end, its final optimisation is performed by the summation of the Eqs. 4, 5 and 6, keeping the w_1 and w_2 equal to 1 for all our experiments.

$$\mathcal{L} = w_1(\mathcal{L}_{recon} + \mathcal{L}_{reg}) + w_2\mathcal{L}_{seg} \tag{7}$$

3.4   Training Scheme

Our convolutional architecture is based on the 3D autoencoder VNet [10]. As also illustrated in Fig. 1 the decoder is separated in two different parts. The encoder is composed of 2 blocks (with 8 and 16 channels respectively) of a convolution

with kernel size $2 \times 2 \times 2$ which is downsampling the size of the input by 2 and a convolution with kernel size $3 \times 3 \times 3$. The decoder follows the same structure with transposed convolutions and skip connections. Each convolution is followed by a LeakyReLu activation layer. Choosing such a small architecture, we generated a relatively small network with less than 40000 parameters, able to be trained with less amount of data.

The encoder is followed by the 3 main blocks of the architecture, the affine, deformable and segmentation blocks. Both registration and segmentation blocks are composed of 3 convolutions with kernels of size $3 \times 3 \times 3$ followed by a sigmoid activation. The affine block is composed by one global average pooling and one dense layer resulting in the affine transformation matrix.

3.5   Data and Preprocessing

For our experiments we used the T1 brain MR images from the publicly available OASIS 3 [9] dataset. All the modalities have been already resampled to a 1mm voxel grid and the skull have been removed, resulting in a volume of size $256 \times 256 \times 256$. Moreover, for these modalities, 47 different structures for left and right side of the brain are provided. Theses annotations have been automatically produced by Freesurfer [6]. For our experiments, we used 520, 67 and 156 images for respectively training, validation and test. Moreover, we performed some additional pre-processing of the images including $\mathcal{N}(0,1)$ normalization, cropping of the images to a $160 \times 176 \times 208$ and translation of the volumes such that the center of mass of the brain is moved to the center of the volume.

4   Experimental Results

We evaluated the performance of our method in both registration and segmentation tasks by calculating the Dice coefficient metric for 15 different brain structures for the registration and 3 different brain structures for the segmentation.

Our implementation is based on Keras. We use 4 Nvidia GeForce GTX 1080 GPUs for each experiment with a batch size of 4. For training we used Adam optimiser, a learning rate of 10^{-3} with a decrease of the learning by half every 15 epochs. We trained our network for approximately 80 epochs, needing approximately one day. In order to improve the convergence of the network, we initialize them with zeros weights and a bias corresponding to the identity transformation. Moreover, to prevent overfitting, we randomly shuffle the training set to generate different pairs of moving and reference images per epoch, fixing however the ones on validation and testing.
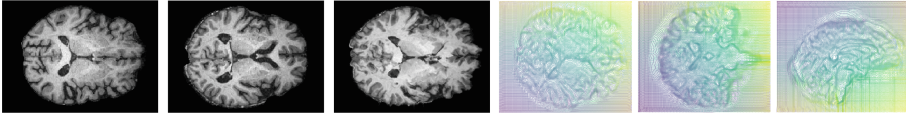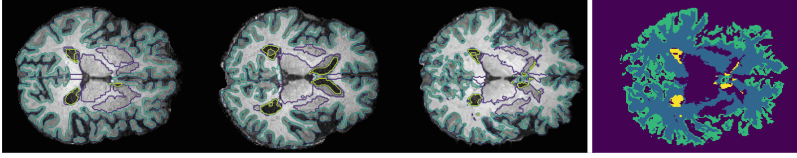
Fig. 2. A MR slice example from the evaluation of the registration task on the test set for the proposed method. From left to right: the moving, reference, deformed image and the computed displacements in axial, coronal and sagittal planes.

Table 1. The mean and standard deviation of the dice coefficient for the 15 different categories for the different evaluated methods.

	Rigid	Christodoulidis et al. [2]	UResNet
BS	0.574 ±0.17	**0.708** ± 0.146	0.706 ± 0.141
CSF	0.393 ±0.11	0.431 ± 0.111	**0.446** ± 0.12
CblmC	0.459 ±0.15	0.581 ± 0.146	**0.603** ± 0.126
CblmWM	0.504 ±0.06	0.528 ± 0.178	**0.553** ± 0.151
CeblWM	0.403 ±0.18	0.674 ± 0.0567	**0.714** ± 0.0539
Pu	0.428 ±0.16	0.464 ± 0.183	**0.479** ± 0.168
VDC	0.465 ±0.14	**0.538** ± 0.128	0.537 ± 0.132
Pa	0.329 ± 0.18	0.379 ± 0.192	**0.391** ± 0.196
Ca	0.258 ±0.21	0.331 ± 0.219	**0.435** ± 0.199
LV	0.384 ±0.15	0.598 ± 0.162	**0.704** ± 0.141
Hi	0.345 ±0.18	**0.409** ± 0.177	0.406 ± 0.171
3V	0.345 ±0.18	0.503 ± 0.173	**0.564** ± 0.173
4V	0.162 ± 0.17	0.33 ± 0.257	**0.332** ± 0.256
Am	0.238 ±0.22	0.281 ± 0.229	**0.291** ± 0.226
CeblC	0.361 ±0.04	0.489 ± 0.0571	**0.536** ± 0.0665

4.1   Results and Discussion

Evaluation of the Registration. Figure 2 shows the results of the registration for one MR slice of the moving, reference and deformed images as well as its deformation in the three different axes. Moreover, we evaluate quantitatively the performance of our method on the brain stem (BS), CSF (CSF), 4th ventricle (4V), amygdala (Am), caudate (Ca), cerebellum cortex (CblmC), cerebellum white matter (CblmWM), cerebral cortex (CeblC), cerebral white matter (CeblWM), hippocampus (Hi), lateral ventricle (LV), pallidum (Pa), putamen (Pu), ventral DC (VDC) and 3rd ventricle (3V) categories. In Table 1 the evaluation in terms of mean and standard deviation dice coefficient is presented. With rigid we indicate the dice coefficient after the translation of the volumes such that the center of the brain mass is placed in the center of the volume. Moreover, we evaluate the performance of the multi task framework reporting superior or similar dice scores with the one presented in [2], indicating that the

Fig. 3. A MR slice from the test set for the evaluation of the registration task for the proposed method. From left to right: the moving image, the reference image, the deformed image and the predicted segmentation. With blue color the CeblWM, green the CeblC and yellow the LV are indicated. (Color figure online)

segmentation part can improve more than 9% of dice score of some categories. Both evaluated methods need only a couple of seconds for the registration of a single pair due to GPU implementations. Finally, we did not provide an evaluation with the method presented in [3] as the authors provide an atlas-based approach, reporting scores in a relatively easier task than the registration to random pairs.

Evaluation of the Segmentation. We trained the network to provide segmentation maps for 3 different brain structures: the CeblWM, the CeblC and the LV, merging the annotations of both left and right regions. After evaluating the reported predictions of our proposed method, we reported dice scores 0.80 ($\pm$0.02), 0.87 ($\pm$0.04) and 0.70 ($\pm$0.02) for each class respectively. These results especially for CeblC are close to the ones reported in similar studies [1]. Moreover, Fig. 3 shows the ground truth and predicted masks superimposed with different colors on a MR slice for the moving, reference and deformed images, together with the predicted from the network mask.

5   Conclusion

In this work, we propose, U-ReSNet, a novel framework based on deep learning that couples image registration and segmentation. Our architecture is very light making it very easily trainable with less data, containing less than 40000 parameters. We provide evaluation for both registration and segmentation problems on the challenging task of brain structure detection. We report some very promising result for both problems, following the reported on literature performances of deep learning based methods addressing each problem separately. Our future steps include the extension of the proposed method, to incorporate on the segmentation component abnormal regions, such as tumor areas. Moreover, one of the limitations of our formulation is the absence of inverse and symmetry constrains. These properties are very important for the registration of medical volumes that we are currently investigating through a deformation from a mid-space paradigm that follows recent work of [17].

References

1. Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.A.: VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images. NeuroImage **170**, 446–455 (2018)

2. Stergios, C., et al.: Linear and deformable image registration with 3D convolutional neural networks. In: Stoyanov, D., et al. (eds.) RAMBO/BIA/TIA -2018. LNCS, vol. 11040, pp. 13–22. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00946-5_2

3. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 729–738. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_82

4. Dong, P., Cao, X., Zhang, J., Kim, M., Wu, G., Shen, D.: Efficient groupwise registration for brain MRI by fast initialization. In: Wang, Q., Shi, Y., Suk, H.-I., Suzuki, K. (eds.) MLMI 2017. LNCS, vol. 10541, pp. 150–158. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67389-9_18

5. Dong, P., Wang, L., Lin, W., Shen, D., Wu, G.: Scalable joint segmentation and registration framework for infant brain images. Neurocomputing **229**, 54–62 (2017). Advances in computing techniques for big medical image data

6. Fischl, B.: Freesurfer. Neuroimage **62**(2), 774–781 (2012)

7. Glocker, B., Komodakis, N., Navab, N., Tziritas, G., Paragios, N.: Dense registration with deformation priors. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 540–551. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02498-6_45

8. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS -2017. LNCS, vol. 10553, pp. 240–248. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_28

9. Marcus, D.S., Fotenos, A.F., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. J. Cogn. Neurosci. **22**(12), 2677–2684 (2010)

10. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)

11. Parisot, S., Duffau, H., Chemouny, S., Paragios, N.: Joint tumor segmentation and dense deformable registration of brain MR images. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7511, pp. 651–658. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33418-4_80

12. Postelnicu, G., Zollei, L., Fischl, B.: Combined volumetric and surface registration. IEEE Trans. Med. Imaging **28**(4), 508–522 (2009)

13. Prevost, R., Cuingnet, R., Mory, B., Correas, J.-M., Cohen, L.D., Ardon, R.: Joint co-segmentation and registration of 3D ultrasound images. In: Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (eds.) IPMI 2013. LNCS, vol. 7917, pp. 268–279. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38868-2_23

14. Robinson, E.C., et al.: Multimodal surface matching with higher-order smoothness constraints. NeuroImage **167**, 453–465 (2018)

15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
16. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: a survey. IEEE Trans. Med. Imaging **32**(7), 1153–1190 (2013)
17. Thewlis, J., Bilen, H., Vedaldi, A.: Modelling and unsupervised learning of symmetric deformable object categories. In: Advances in Neural Information Processing Systems 31. Curran Associates, Inc. (2018)
18. Vakalopoulou, M., et al.: AtlasNet: multi-atlas non-linear deep networks for medical image segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 658–666. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_75