



# A Deep Image Prior Learning Algorithm for Joint Selective Segmentation and Registration

Liam Burrows<sup>1</sup>, Ke Chen<sup>1()</sup>, and Francesco Torella<sup>2</sup>

<sup>1</sup> Department of Mathematical Sciences and Centre for Mathematical Imaging Techniques, University of Liverpool, Liverpool, UK  
[k.chen@liv.ac.uk](mailto:k.chen@liv.ac.uk)

<sup>2</sup> Liverpool Vascular and Endovascular Service, Liverpool University Hospitals, Liverpool L7 8XP, UK  
<https://www.liv.ac.uk/~cmchenke>

**Abstract.** Effective variational models exist for either image segmentation or image registration for a given class of problems, though robustness is a longstanding issue. This paper proposes a new and effective variational model that aims to segment a pair of images through a joint model with registration, with the advantage of only requiring geometric prior information on one image (instead of two images) and obtaining selective segmentation on both images. Moreover we develop a deep image prior based learning algorithm to achieve the same segmentation and registration results by dropping the regularisation terms from the loss function. Numerical experiments show quality results obtained from the new approach.

**Keywords:** Image segmentation · Image registration · Deep image prior

## 1 Introduction

Image registration and segmentation are two fundamental tasks in image processing. Typically they are treated as two separate processes, however some efforts have been made to combine the two into a single formulation [11, 14]. The aim of segmentation is to identify meaningful objects in an image, through their given intensity distributions. We are particularly interested in selective segmentation [1, 5, 9, 16] where a set  $\mathcal{M}$  of geometric markers is given by the user and the aim is to segment the object nearest to or containing  $\mathcal{M}$ . Hence to selectively segment two images, we need two sets of respective markers sets  $\mathcal{M}_j$ ,  $j = 1, 2$  for each image. The aim of registration is to find a transformation  $\mathbf{y}(\mathbf{x}) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  which maps a template image  $T$  to a reference image  $R$ , with  $T, R \in \Omega \subset \mathbb{R}^2$ , such that:  $T(\mathbf{y}(\mathbf{x})) \approx R(\mathbf{x})$ . Typically the transformation is written as  $\mathbf{y}(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$ , where  $\mathbf{u}(\mathbf{x})$  is the displacement vector field. Here, to focus on modeling segmentation and registration jointly, this mono-modality registration setup is sufficient for our study, assuming that images  $T, R$  are acquired

from the same source (e.g. both are digital or both are CTs). However, for multi-modal images where  $T, R$  sensing the same object come from different sources (e.g.  $T$  from infra-red and  $R$  from a digital camera) we simply employ suitably chosen transforms  $f_1, f_2$  (e.g. gradient operations) such that  $f_1(T)(\mathbf{y}(\mathbf{x})) \approx f_2(R)(\mathbf{x})$ .

For a joint model, intuitively, the segmentation (of corresponding features) may be able to guide and improve the registration process while registration can transfer an accurate segmentation result of the reference image to the template.

Our study is in the variational framework. Take the image  $T$  as an example to define our notation. We are interested in finding its feature enclosed by boundary contour  $\Gamma$ . As  $\Gamma$  is unknown, a widely used idea is to embed it in a level set function  $\phi(\mathbf{x})$  and to find this  $\phi(\mathbf{x})$  first. Then segmentation of image  $T$  amounts to partitioning the image domain  $\Omega = \Omega_1 \cup \Omega_2 \cup \Gamma$  as follows [4]:

$$\begin{cases} \Gamma = \{\mathbf{x} \in \Omega : \phi(\mathbf{x}) = 0\} \\ \Omega_1 := \text{Inside}(\Gamma) = \{\mathbf{x} \in \Omega : \phi(\mathbf{x}) > 0\} \\ \Omega_2 := \text{Outside}(\Gamma) = \{\mathbf{x} \in \Omega : \phi(\mathbf{x}) < 0\} \end{cases}$$

with the region of interest being  $\Omega_1$ .

Since the advantage of a level set functions  $\phi$  enabling to model arbitrary changes in  $\Omega_1$ 's topology is well known, their use in modeling segmentation or other interface problems is common. However, variational models involving  $\phi$  are almost always non-convex because integration over  $\Omega_1$  is replaced by integration over  $\Omega$  involving  $H(\phi)$ , where  $H$  is the Heaviside function, which is non-convex; likewise integration over  $\Omega_2$  will be by over  $\Omega$  involving  $1 - H(\phi)$ . Non-convex models suffer from the problems of non-unique solutions and strong dependence on initialisation. This apparent disadvantage may be overcome by the idea of relaxation [3] where one solves for the new variable  $\theta = H(\phi)$  instead of  $\phi$ , and then relaxes the constraint  $\theta \in \{0, 1\}$  (binary set) to  $\theta \in [0, 1]$  (continuous interval). Then once we found  $\theta$ , a threshold of it will yield the required segmentation i.e. domain  $\Omega_1$ . This is the approach that we shall adopt in this work.

Assume that two images  $T, R$  are given and a shape or feature of  $T$  is given in terms of a binary image (segmented), which can be converted to a level set function representation  $\phi_0$ . To design a joint model of registration and segmentation, we have to understand how a given segmentation of image  $T$  defined in domain  $\Omega_1$  via  $H(\phi_0(\mathbf{x}))$  interacts with the segmentation of image  $R$  of domain  $\Omega_1$  via  $H(\phi(\mathbf{x}))$ , and how registration  $T(\mathbf{x} + \mathbf{u})$  comes into the formulation. It turns out that the key step is to evolve  $H(\phi_0(\mathbf{x} + \mathbf{u}))$  to guide segmentation of  $R$ , and not to set up a new and unknown level set  $\phi(\mathbf{x})$  [11, 14, 21]. Then it becomes natural to combine registration and segmentation.

Finally there is much freedom in choosing a regulariser, either by hand-crafting or learning from data, for both types of problems. While this facilitates optimised tuning for specific applications, it is generally hard to design a robust regulariser. For instance, Chan-Vese type models (or others based on Mumford-Shah framework) minimise the length of segmented boundaries via  $\nabla H(\phi)$ , while one may regularise the Euler elastica of the boundary [26]. There are even more choices of regularisers for image registration ranging from  $H_1$  semi-norm to Euler

elastica. Our work will utilise the implicit regularisation offered by the architecture of a neural network.

This paper presents an algorithm that segments two related images and find a mapping registering these two images, by using only a single marker set  $\mathcal{M}$ . Our work differs from previous models [11, 14, 21] in three aspects: i) we employ a relaxed formulation for segmentation part; ii) we incorporate geometric markers of  $T$  only to segment both  $T$  and  $R$  selectively; iii) we propose a deep image prior based approach for an accurate solution of our joint registration and segmentation model.

## 2 A New Variational Model

In this section, we outline the proposed joint segmentation-registration problem in the variational setting. We propose to utilise the selective segmentation framework of [18] and re-purpose it for joint segmentation and registration using the ideas from [14] and [11] models.

Most variational approaches to joint registration segmentation models usually assume that the region of interest on  $T$  is already segmented and given. The task then is focused on finding the displacement map using the given segmented region of  $T$  as additional information to aid the registration process. For our proposed model, we assume no prior knowledge and aim to selectively segment an object in an image, and simultaneously register the image to another. As this section will demonstrate, models of this class involve many terms, resulting in many parameters to tune for the user. In the next section, we propose implementing our model into a deep image prior framework, in which we can remove two regularisation terms (one in the segmentation problem and one in the registration problem). The removal of such terms reduces the complexity of tuning parameters and makes a joint framework with no prior knowledge more viable. This section however introduces the joint problem with explicit regularisation.

In variational setting, the model which we propose is given as follows:

$$\begin{aligned} \min_{\theta, \mathbf{u}} & \int_{\Omega} g |\nabla \theta(\mathbf{x})| d\Omega + \lambda_1 \int_{\Omega} \Phi(T, a_1) \theta(\mathbf{x}) d\Omega + \lambda_2 \int_{\Omega} \Phi(R, c_1) \theta(\mathbf{x} + \mathbf{u}) d\Omega \\ & + \xi \int_{\Omega} \mathcal{D}(\mathbf{x}) \theta(\mathbf{x}) d\Omega + \frac{1}{2} \int_{\Omega} |T_{\mathbf{u}} - R(\mathbf{x})|^2 d\Omega + \frac{\alpha}{2} \sum_{\ell=1}^2 \int_{\Omega} |\nabla u_{\ell}|^2 d\Omega, \end{aligned} \quad (1)$$

where  $\theta \in [0, 1]$ ,  $T_{\mathbf{u}} = T(\mathbf{x} + \mathbf{u})$ ,  $\Phi(f, a) = \mu_1(f, a) - \mu_2(f, a)$ , and  $g = g(|\nabla T|) = \frac{1}{1+\epsilon|\nabla T|^2}$  is an edge detector. Here, as in [18], we define

$$\mu_1(f, c_1) = (f - c_1)^2, \quad \mu_2(f, c_1) = \begin{cases} 1 + \frac{f(x) - c_1}{\gamma_1}, & c_1 - \gamma_1 \leq f(x) \leq c_1, \\ 1 - \frac{f(x) - c_1}{\gamma_2}, & c_1 < f(x) \leq c_1 + \gamma_2, \\ 0, & \text{else,} \end{cases} \quad (2)$$

where  $\gamma_1$  and  $\gamma_2$  are parameters fixed as detailed in [18], and we fix  $a_1$  to be the average intensity of  $T$  inside the region defined by the user input  $\mathcal{M}$ . The

parameter  $c_1$  represents the average intensity of the region of interest on the image  $R$ , and as we only concern ourselves with mono-modal images we make the assumption that  $c_1 \approx a_1$ , and therefore fix  $c_1 = a_1$ .

In (1), the selective constraint  $\mathcal{D}$  is the geodesic distance as proposed in [17] used to put a penalty on objects outside the region defined by the user in the marker set  $\mathcal{M}$ . It is given by  $\mathcal{D} = \frac{\mathcal{D}_0}{\|\mathcal{D}_0\|_{L^\infty}}$ , where

$$\begin{cases} |\nabla \mathcal{D}_0(x)| = \varepsilon_D + \beta_D |\nabla T(x)|^2 + \xi_D \mathcal{D}_E(x), & x \in \Omega \\ \mathcal{D}_0(x) = 0, & x \in \mathcal{M}, \end{cases} \quad (3)$$

where  $\mathcal{D}_E$  is the Euclidean distance from  $\mathcal{M}$  and the parameters are fixed as in [17] to be  $\varepsilon_D = 10^{-3}$ ,  $\beta_D = 1000$  and  $\xi_D = 0.1$ . The solution to the Eikonal equation (3) can be solved quickly using fast sweeping or fast marching methods.

Once the optimisation is performed and a minimiser  $\theta^*$  is found, we define the foreground region representing the region of interest as:

$$\Sigma = \{x \in \Omega : \theta^*(x) > \gamma\},$$

where it is typical to select  $\gamma = 0.5$ .

Below we show some details of solving (1) in alternating minimization.

- Minimisation with respect to  $\mathbf{u}$  yields this Euler-Lagrange equations:

$$0 = \lambda_2 \Phi(R, c_1) \nabla_{\mathbf{u}} \theta(\mathbf{x} + \mathbf{u}) + (T(\mathbf{x} + \mathbf{u}) - R(\mathbf{x})) \nabla_{\mathbf{u}} T(\mathbf{x} + \mathbf{u}) - \alpha \Delta \mathbf{u}.$$

Implementing a gradient descent and semi-implicit finite difference scheme, utilising additive operator splitting (AOS) [24, 25], yields the following iterative scheme.

$$\begin{cases} u_1^{(k+1)} = \frac{1}{2} \sum_{l=1}^2 (I - 2\alpha\tau A_l)^{-1} (u_1^{(k)} - \tau g_1(u_1^{(k)}, u_2^{(k)}, \theta^{(k)})), \\ u_2^{(k+1)} = \frac{1}{2} \sum_{l=1}^2 (I - 2\alpha\tau A_l)^{-1} (u_2^{(k)} - \tau g_2(u_1^{(k)}, u_2^{(k)}, \theta^{(k)})), \end{cases} \quad (4)$$

where  $g_l(u_1, u_2, \theta) = \lambda_2 \Phi(R, c_1) \partial_{\mathbf{u}_l} \theta(\mathbf{x} + \mathbf{u}) (T(\mathbf{x} + \mathbf{u}) - R(\mathbf{x})) \partial_{\mathbf{u}_l} T(\mathbf{x} + \mathbf{u})$ ,  $l = 1, 2$ ,  $A_l$  is the discretisation of the Laplace operator  $\Delta$  along the  $l$ -coordinate direction.

- The  $\theta$  subproblem of (1), before deriving the Euler-Lagrange equation, is:

$$\begin{aligned} \min_{\theta \in [0, 1]} \mu \int_{\Omega} g |\nabla \theta(\mathbf{x})| d\Omega + \lambda_1 \int_{\Omega} \Phi(T(\mathbf{x}), a_1) \theta(\mathbf{x}) d\Omega + \lambda_2 \int_{\Omega} \Phi(R(\mathbf{x} - \mathbf{u}), c_1) \theta(\mathbf{x}) d\Omega \\ + \xi \int_{\Omega} \mathcal{D}(\mathbf{x}) \theta(\mathbf{x}) d\Omega, \quad \Rightarrow \\ 0 = \lambda_1 \Phi(T(\mathbf{x}), a_1) + \lambda_2 \Phi(R(\mathbf{x} - \mathbf{u}), c_1) + \xi \mathcal{D}(\mathbf{x}) - \mu \nabla \cdot \left( g \frac{\nabla \theta}{|\nabla \theta|} \right). \end{aligned}$$

Similar to solving the  $\mathbf{u}$  subproblem, we implement a gradient descent with semi-implicit finite difference, yielding the following iterative scheme:

$$\theta^{(k+1)} = \frac{1}{2} \sum_{l=1}^2 (I - 2\mu\tau A_l)^{-1} (\theta^{(k)} - \tau g_3(u_1^{(k)}, u_2^{(k)}, \theta^{(k)})), \quad (5)$$

where  $g_3(u_1, u_2, \theta) = \lambda_1 \Phi(T(\mathbf{x}), a_1) + \lambda_2 \Phi(R(\mathbf{x} - \mathbf{u}), c_1) + \xi \mathcal{D}(\mathbf{x})$ .

### 3 A Deep Image Prior Approach

The deep image prior (DIP) [20] approach can tackle inverse problems effectively. It is an unsupervised approach for training on a single data point, rather than having a large training set. The DIP method proposes to remove explicit regularisation from the traditional inverse problem and replace it with the implicit prior captured by the architecture of a neural network. This, combined with early stopping and varying the input to the network with noise has been shown to produce a regularisation effect on the solution, capturing generic regularity of a target image. It has been shown in [6, 15, 20, 22] that the deep image prior approach for (inverse problems of) image restoration and superresolution achieve outstanding results. Here we extend the DIP idea to image segmentation and registration, especially the joint model of these two tasks.

In our case, removing explicit regularisation and exploiting the implicit regularisation offered by the architecture of a neural network allows us to remove two terms, simplifying parameter tuning dramatically. In addition, we find the deep image prior approach outperforms the same model implemented in the classical learning setting.

We utilise two deep image prior networks. As notation, let  $\theta(\mathbf{x}) = \theta(\Theta_1)(\mathbf{x}) = \varphi_{\Theta_1}(z_1)$  be a network parametrised by weights  $\Theta_1$  with random noise  $z_1$  as input dedicated to the segmentation task. To keep notation consistent with previous sections, we denote this as  $\theta(\mathbf{x})$ , as the first network performs the segmentation. Similarly, let  $\mathbf{u} = \mathbf{u}(\Theta_2)(\mathbf{x}) = \psi_{\Theta_2}(z_2)$  be a network parametrised by weights  $\Theta_2$ , with random noise input  $z_2$ , dedicated to the registration task. Our task is to minimise the energies in order to iteratively update the weights  $\Theta_i$ :

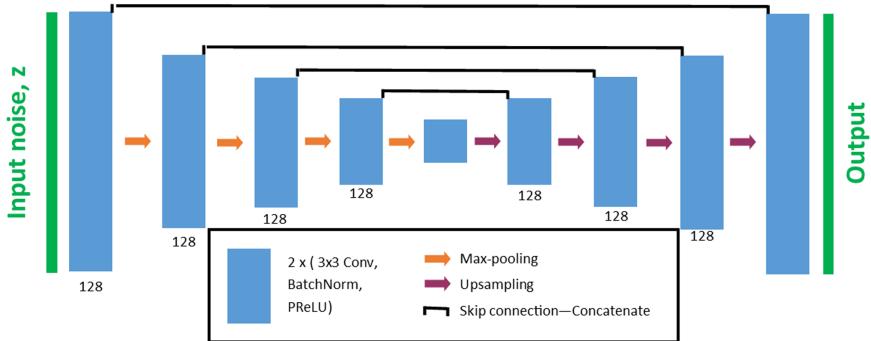
$$\Theta_1^{(k+1)} = \arg \min_{\Theta_1} \mathcal{E}_1(\Theta_1, \Theta_2^{(k)}), \quad \Theta_2^{(k+1)} = \arg \min_{\Theta_2} \mathcal{E}_2(\Theta_1^{(k)}, \Theta_2).$$

The energies  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , that serve as the loss functions for the two networks, are defined similarly as in the previous section, however without explicit regularisation:

$$\begin{aligned} L_1(\Theta_1) = \mathcal{E}_1(\Theta_1, \Theta_2^{(k)}) &= \lambda_1 \int_{\Omega} \Phi(T(\mathbf{x}), a_1) \theta(\mathbf{x}) d\Omega \\ &\quad + \lambda_2 \int_{\Omega} \Phi(R(\mathbf{x}), c_1) \theta(\mathbf{x} + \mathbf{u}) d\Omega + \xi \int_{\Omega} \mathcal{D}(\mathbf{x}) \theta(\mathbf{x}) d\Omega, \end{aligned} \tag{6}$$

$$\begin{aligned} L_2(\Theta_2) = \mathcal{E}_2(\Theta_1^{(k)}, \Theta_2) &= \lambda_2 \int_{\Omega} \Phi(R(\mathbf{x}), c_1) \theta(\mathbf{x} + \mathbf{u}) d\Omega \\ &\quad + \frac{1}{2} \int_{\Omega} (T(\mathbf{x} + \mathbf{u}) - R(\mathbf{x}))^2 d\Omega. \end{aligned} \tag{7}$$

where we recall that  $\theta(\mathbf{x})$  is the output image of the first network parametrised by weights  $\Theta_1$  and generated by input noise image  $z_1$ , and the displacement map  $\mathbf{u} = (u_1, u_2)$  is the output (in 2 images) of the second network parametrised by weights  $\Theta_2$  and generated by input noise image  $z_2$ .



**Fig. 1.** An overview of the network architecture used.

The architecture used is a typical U-Net like [19] hourglass network featuring convolutions, batch normalisations and PReLU activation functions, see Fig. 1 for the full architecture. Both networks employ this architecture, only differing in the output layer. For the output layer, the segmentation network  $\varphi_{\Theta_1}$  has a  $(1 \times 1)$  convolution followed by a Sigmoid activation in the final layer, whereas the registration network just has a  $(1 \times 1)$  convolution.

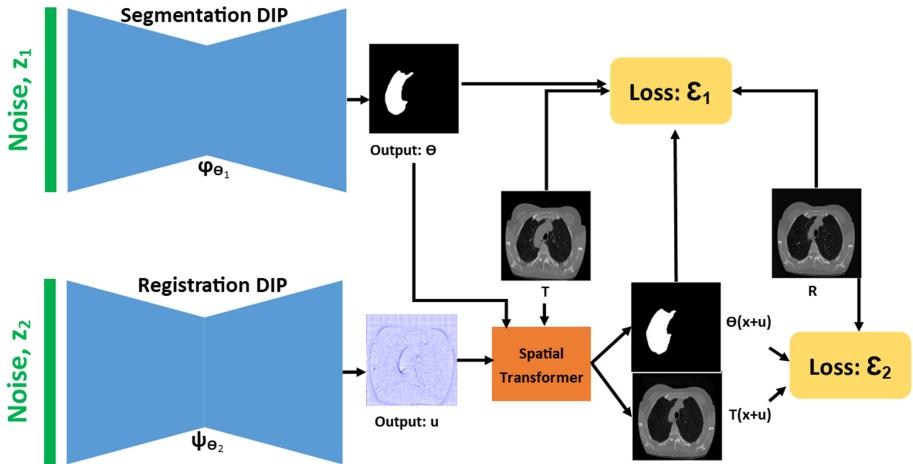
The initial inputs  $z_i^{(0)}$  to the network are each drawn randomly from a uniform distribution between 0 and 0.1, and as in [20] has the same spatial size as  $T$  and  $R$ , i.e.  $z_i^{(0)} \in \mathbb{R}^{N \times M \times C}$ , where  $N \times M$  is the image size and we typically fix  $C = 32$ . Likewise as in the original DIP approach, at each training iteration  $k$  we add additive noise  $\hat{z}_i^{(k)}$  to the input  $z_i^{(0)}$  drawn from a normal distribution, with mean 0 and standard deviations  $\frac{1}{100}$  and  $\frac{1}{30}$  for the segmentation and registration networks respectively i.e.  $z_i^{(k)} = z_i^{(0)} + \hat{z}_i^{(k)}$ .

Overall our workflow can be found in Fig. 2. Our implementation uses Keras with the Tensorflow backend, and each network uses ADAM optimiser with a learning rate of 0.001. The registration network was implemented with the aid of the voxelmorph package [2] to allow us to use a spatial transformer to interpolate the images using the output displacement map  $\mathbf{u}(\mathbf{x})$ , as shown in Fig. 2. Details of the DIP algorithm can be found in Algorithm 1.

We remark that multiple coupled DIP networks may provide a framework for a wide variety of applications [8]. Such applications include foreground/background segmentation, in which one network attempts to identify the foreground layer, another identifies the background layer, and a third identifies the binary mask. We note that this approach is similar to the approach we considered, although there is no discussion about using multiple DIP networks to tackle different inverse problems simultaneously. Moreover, Laves et al. [13] explored applying DIPs to medical image registration, exploiting the implicit regularisation that a randomly initialised CNN has to offer. Future work will look into multiple coupled DIP networks for our joint model.

**Algorithm 1.** Deep Image Prior Joint Selective Segmentation and Registration.

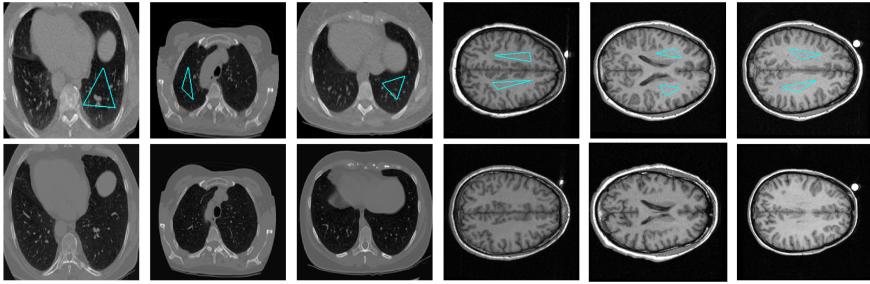
- 1: **Input:** Images  $T, R$ , user input  $\mathcal{M}$ , parameters  $\lambda_1, \lambda_2, \xi$ , learning rate (lr), max epochs.
- 2: **Initialise:**  $\Theta_i$  randomly,  $z_i \in \mathbb{R}^{N \times M \times 32} \sim U(0, \frac{1}{10})$ ,  $i = 1, 2$ .
- 3: **Calculate:**  $\mathcal{D}$  from  $\mathcal{M}$  on image  $T$  using (3).
- 4: **while** Not converged **do**
- 5: Set  $\hat{z}_1 \sim \mathcal{N}(0, \frac{1}{100})$  and  $\hat{z}_2 \sim \mathcal{N}(0, \frac{1}{30})$ .
- 6: Update  $\Theta_1^{k+1}$  by minimising (6) using back-propagation with network input  $z_1 + \hat{z}_1$ .
- 7: Update  $\Theta_2^{k+1}$  by minimising (7) using back-propagation with network input  $z_2 + \hat{z}_2$ .
- 8: **end while**
- 9: **Output:** Segmentation  $\theta(\mathbf{x})$  from  $\varphi_{\Theta_1}(z_1)$ , and registration map from  $\mathbf{u}(\mathbf{x})$  from  $\psi_{\Theta_2}(z_2)$



**Fig. 2.** An overview of the training of our deep image prior approach.

## 4 Numerical Experiments

In this section we conduct some experiments to show the two implementations of our model: the first as described in Sect. 2 with explicit regularisation implemented using gradient descent and time-marching, and the second as described in Sect. 3 implemented in a DIP framework. Additionally, as a further comparison we will use a model similar to the model by Ibrahim, Chen and Rada [11]. Their model makes use of a known segmentation of  $T$  given by a level set function  $\phi_0$ . This is slightly different from our approach, as we do not assume a known segmentation of  $T$ , and we use an indicator function  $\theta$  in the convex relaxed setting first introduced by Chan et al. [3]. In addition, the aim of [11] is to find the displacement map from the object in  $T$  to the object in  $R$  only, and



**Fig. 3.** Test images. Top row shows  $T$  with user input  $\mathcal{M}$ . Bottom row shows image  $R$ . From left to right, we refer to these pairs of images as: Lung 1 (L1), Lung 2 (L2), Lung 3 (L3), Brain 1 (B1), Brain 2 (B2), Brain 3 (B3).

not the entire image, whereas our approach finds the displacement map of the entire image. Therefore, for the sake of comparison, we introduce the following modified Ibrahim, Chen and Rada model (**M-ICR**):

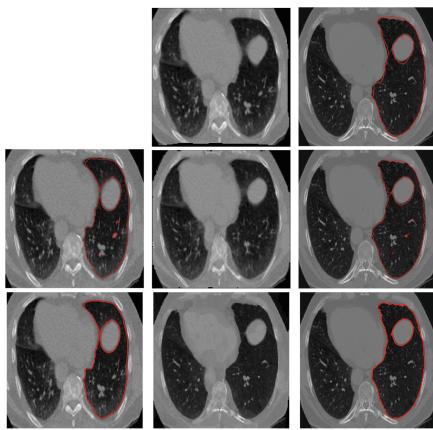
$$\begin{aligned} \min_{\mathbf{u}} \frac{1}{2} \int_{\Omega} (T(\mathbf{x} + \mathbf{u}(\mathbf{x})) - R(\mathbf{x})^2 d\Omega + \lambda \int_{\Omega} |R(\mathbf{x}) - c_1|^2 H(\phi_0(\mathbf{x} + \mathbf{u}(\mathbf{x}))) d\Omega \\ + \lambda \int_{\Omega} |R(\mathbf{x}) - c_2|^2 (1 - H(\phi_0(\mathbf{x} + \mathbf{u}(\mathbf{x})))) d\Omega + \alpha \int_{\Omega} (\Delta u_1)^2 + (\Delta u_2)^2 d\Omega, \end{aligned}$$

where  $\phi_0(\mathbf{x}) \in [0, 1]$  is the known ground truth segmentation of the object in  $T$ , and  $H$  is a heaviside function. Note as we assume a ground truth segmentation is given, in the quantitative results the **M-ICR** model will not have a DICE value for segmentation of  $T$ .

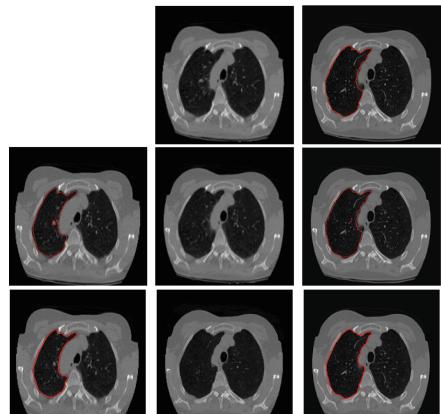
We show results on two different types of images: Lung CT scans from the Learn2Reg challenge dataset [10] and Brain MRI scans from the CUMC12 dataset [12]. For the lung images, we provide user input and aim to segment one of the lungs, and for the brain MRI scans we aim to segment the white matter, while in both cases registering the image to another. We show the 6 test images in Fig. 3, in the top row is  $T$  with the user input  $\mathcal{M}$  to indicate the region of interest, the bottom row displays the reference image,  $R$ .

In Figs. 4, 5 and 6 we display some results on the lung images. The image pairs are from the same patient at roughly the same slice, the difference is that  $T$  and  $R$  are from expiration and inspiration scans respectively. In all three examples, we find that the approach with explicit regularisation struggles to smooth out some nodules in the lung segmentation, which the DIP result does effectively. We note that, while increasing the parameter  $\mu$  in the model (1) solves this, doing so makes the contour at the edges of the object not as sharp as currently displayed.

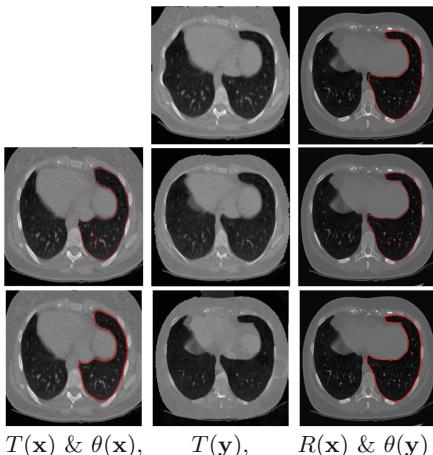
Moreover, in Figs. 7, 8 and 9 we display the results on the brain images. In these images, we aim to segment the white matter in the brain. Interestingly the model with explicit regularisation is competitive with the DIP approach for the segmentation task for this set of images, however for the registration task both visually and quantitatively (see Table 1), the DIP approach performs better.



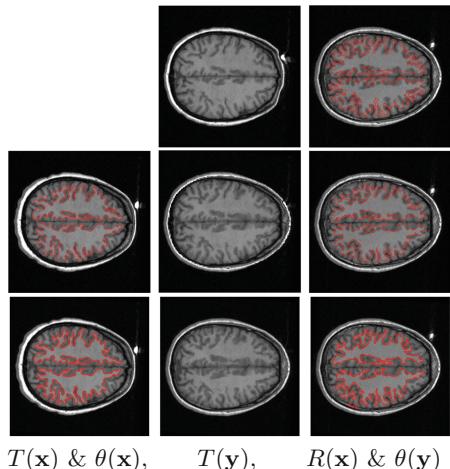
**Fig. 4.** Lung 1 Results: Top row - M-ICR model. Middle row - Explicit Regularisation. Bottom row - DIP framework.



**Fig. 5.** Lung 2 Results: Top row - M-ICR model. Middle row - Explicit Regularisation. Bottom row - DIP framework.



**Fig. 6.** Lung 3 Results: Top row - M-ICR model. Middle row - Explicit Regularisation. Bottom row - DIP framework.



**Fig. 7.** Brain 1 Results: Top row - M-ICR model. Middle row - Explicit Regularisation. Bottom row - DIP framework.

For all examples we provide quantitative results in Table 1. To determine the performance of the three models we use three evaluation metrics:

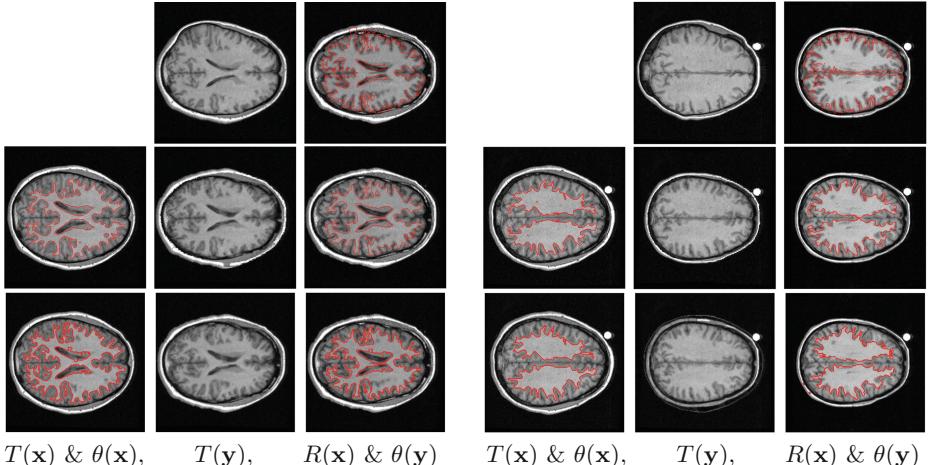
1. DICE similarity coefficient [7] to compare the segmentation result  $\theta(\mathbf{x})$  with the ground truth segmentation on  $T$ , and the segmentation result of  $\theta(\mathbf{x} + \mathbf{u}(\mathbf{x}))$  on  $R$ , which is defined by:

$$DICE(\Sigma, GT) = \frac{2|\Sigma \cap GT|}{|\Sigma| + |GT|}.$$

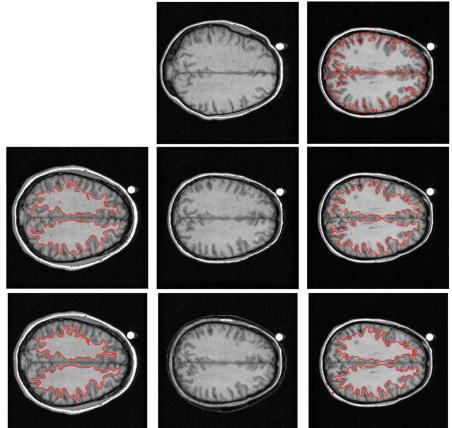
2. The structure similarity measure (SSIM) [23].
3. The relative sum of squared differences (Rel-SSD), defined by:

$$Rel\_SSD = \frac{\|T(\mathbf{x} + \mathbf{u}(\mathbf{x})) - R(\mathbf{x})\|^2}{\|T(\mathbf{x}) - R(\mathbf{x})\|^2}.$$

It is clear to see from Table 1, the DIP approach outperforms both the approach with explicit regularisation optimised using time-marching, and the **M-ICR** model. In general, the **M-ICR** displays competitive segmentation results, however this is largely due to the fact that it is supplied with the ground truth segmentation of  $T$ , whereas the other two approaches aren't.



**Fig. 8.** Brain 2 Results: Top row - **M-ICR** model. Middle row - Explicit Regularisation. Bottom row - DIP framework.



**Fig. 9.** Brain 3 Results: Top row - **M-ICR** model. Middle row - Explicit Regularisation. Bottom row - DIP framework.

**Table 1.** Comparison of M-ICR (M), Explicit Reg (ER) and the DIP approach for Lung (Lj) and Brain (Bj) images.

	Dice of T		Dice of R		SSIM			rel_ssd			
	ER	DIP	M	ER	DIP	M	ER	DIP	MR	ER	DIP
L1	0.95	0.95	0.97	0.97	<b>0.98</b>	0.46	0.57	<b>0.84</b>	0.86	0.32	<b>0.09</b>
L2	0.95	<b>0.96</b>	0.97	0.96	<b>0.99</b>	0.48	0.54	<b>0.73</b>	0.30	0.15	<b>0.08</b>
L3	0.96	0.96	0.97	0.97	<b>0.98</b>	0.45	0.55	<b>0.77</b>	0.63	0.18	<b>0.07</b>
B1	<b>0.87</b>	0.86	0.79	<b>0.84</b>	0.82	0.62	0.72	<b>0.81</b>	0.45	0.17	<b>0.07</b>
B2	0.84	<b>0.90</b>	0.77	<b>0.83</b>	0.82	0.59	0.71	<b>0.79</b>	0.40	0.12	<b>0.04</b>
B3	<b>0.84</b>	0.80	0.77	<b>0.88</b>	0.87	0.50	0.67	<b>0.71</b>	0.43	<b>0.06</b>	0.08

## 5 Conclusion

In this work, we proposed a new model for joint segmentation and registration. Our model assumes no prior knowledge, only user supplied input in the form of a marker set  $\mathcal{M}$  to indicate the region of interest. We then selectively segment the object of interest and simultaneously register the image  $T$  to  $R$ . We first introduce our model in the variational setting and then insert it into a deep image prior (DIP) framework. Numerical experiments show that the DIP approach generally outperforms the explicit regularisation approach, and an existing similar model (**M-ICR**).

## References

1. Badshah, N., Chen, K.: Image selective segmentation under geometrical constraints using an active contour approach. *Commun. Comput. Phys.* **7**, 759–778 (2010)
2. Balakrishnan, G., Zhao, A., Sabuncu, M., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. *IEEE TMI: Trans. Med. Imaging* **38**, 1788–1800 (2019)
3. Chan, T.F., Esedoglu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.* **66**(5), 1632–1648 (2006)
4. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Trans. Image Process.* **10**(2), 266–277 (2001)
5. Chen, K., Debroux, N., Guyader, C.L.: A survey of topology and geometry-constrained segmentation methods in weakly supervised settings. In: Chen, K., Schönlieb, C.B., Tai, X.C., Younes, L. (eds.) *Mathematical Models and Algorithms in Computer Vision and Imaging*. Springer (2022 to appear). <https://meteor.springer.com/project/dashboard.jsf?id=839&tab=About>
6. Cheng, Z., Gadelha, M., Maji, S., Sheldon, D.: A Bayesian perspective on the deep image prior. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5443–5451 (2019)
7. Dice, L.R.: Measures of the amount of ecologic association between species. *Ecology* **26**(3), 297–302 (1945)

8. Gandelsman, Y., Shocher, A., Irani, M.: “Double-dip”: unsupervised image decomposition via coupled deep-image-priors. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 6, p. 2 (2019)
9. Gout, C., Le Guyader, C., Vese, L.A.: Segmentation under geometrical conditions with geodesic active contour and interpolation using level set methods. *Numer. Algorithms* **39**(1), 155–173 (2005)
10. Hering, A., Murphy, K., van Ginneken, B.: Learn2reg challenge: Ct lung registration - training data (2020)
11. Ibrahim, M., Chen, K., Rada, L.: An improved model for joint segmentation and registration based on linear curvature smoother. *J. Algorithms Comput. Technol.* **10**(4), 314–324 (2016)
12. Klein, A., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* **46**(3), 786–802 (2009)
13. Laves, M.-H., Ihler, S., Ortmaier, T.: Deformable medical image registration using a randomly-initialized CNN as regularization prior. arXiv preprint [arXiv:1908.00788](https://arxiv.org/abs/1908.00788) (2019)
14. Le Guyader, C., Vese, L.A.: A combined segmentation and registration framework with a nonlinear elasticity smoother. *Comput. Vis. Image Underst.* **115**(12), 1689–1709 (2011)
15. Mataev, G., Milanfar, P., Elad, M.: Deepred: deep image prior powered by red. In: Proceedings of the IEEE International Conference on Computer Vision Workshops (2019)
16. Rada, L., Chen, K.: A new variational model with dual level set functions for selective segmentation. *Commun. Comput. Phys.* **12**(1), 261–283 (2012)
17. Roberts, M., Chen, K., Irion, K.L.: A convex geodesic selective model for image segmentation. *J. Math. Imaging Vis.* **61**(4), 482–503 (2019)
18. Roberts, M., Spencer, J.: Chan-vese reformulation for selective image segmentation. *J. Math. Imaging Vis.* **61**(8), 1173–1196 (2019)
19. 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](https://doi.org/10.1007/978-3-319-24574-4_28)
20. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9446–9454 (2018)
21. Unal, G., Slabaugh, G.: Coupled PDEs for non-rigid registration and segmentation. In: CVPR (2005)
22. Van Veen, D., Jalal, A., Soltanolkotabi, M., Price, E., Vishwanath, S., Dimakis, A.G.: Compressed sensing with deep image prior and learned regularization. arXiv preprint [arXiv:1806.06438](https://arxiv.org/abs/1806.06438) (2018)
23. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
24. Weickert, J.: Applications of nonlinear diffusion in image processing and computer vision. *Acta Mathematica Universitatis Comenianae* **LXX**, 33–50 (2001)
25. Weickert, J., Romeny, B.T.H., Viergever, M.A.: Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Trans. Image Process.* **7**(3), 398–410 (1998)
26. Zhu, W., Tai, X.-C., Chan, T.: Image segmentation using Euler’s elastica as the regularization. *J. Sci. Comput.* **57**, 414–438 (2013)