Using Deep Image Prior to Assist Variational Selective Segmentation Deep Learning Algorithms

Liam Burrows[a], Ke Chen[a] and Francesco Torella[b]

[a]Department of Mathematical Sciences and Centre for Mathematical Imaging Techniques, University of Liverpool, United Kingdom;
[b]Liverpool Vascular & Endovascular Service, Liverpool University Hospitals, Liverpool L7 8XP, United Kingdom

ABSTRACT

Variational segmentation algorithms require a prior imposed in the form of a regularisation term to enforce smoothness of the solution. Recently, it was shown in the Deep Image Prior work that the explicit regularisation in a model can be removed and replaced by the implicit regularisation captured by the architecture of a neural network. The Deep Image Prior approach is competitive, but is only tailored to one specific image and does not allow us to predict future images. We propose to incorporate the ideas from Deep Image Prior into a more traditional learning algorithm to allow us to use the implicit regularisation offered by the Deep Image Prior, but still be able to predict future images.

Keywords: Image segmentation, Selective segmentation, Deep Image Prior.

1. INTRODUCTION

Image segmentation has been a well studied topic over the last few decades. The variational approach has been particularly successful in tackling a wide range of problems[1–3] without the need for large training sets or time consuming manual labels. A less studied variant is selective variational image segmentation,[4–6] which is the task of segmenting a particular object or objects, usually indicated by a set of marker points $\mathcal{M}$ input by the user. These marker points are typically placed inside the region of interest, and geometric constraints can be designed to penalise segmenting regions away from the intended object.

Deep learning methods for image segmentation[7–12] have become very popular over the last few years due to an increase in computing power and more data availability. A deep learning model tailored to a specific task is the gold standard for image segmentation currently. This is however dependent on a large dataset with ground truth labels, which typically can be rather expensive and/or time consuming to acquire.

The intersection of variational methods and deep learning methods can lead to a powerful tool to base a variational model in a deep learning framework, potentially reducing the need for many expensive ground truth labels in the form of semi-supervised or unsupervised algorithms, while still leveraging the impressive performance of learning methods. Such semi-supervised works often combine deep learning methods with variational methods.[13–16]

Another approach that can be used to tackle variational models using deep learning methods was proposed recently by Ulyanov et al.[17] in their work on the Deep Image Prior (DIP). Their method uses an untrained neural network to solve an inverse problem for a single image. No training data is required, and surprisingly the DIP approach is competitive with state of the art methods. They propose to remove the explicit regularisation term typically found in a typical inverse problem, and instead rely on the implicit regularisation offered by the architecture of a neural network.

A major limitation of the Deep Image Prior is that the network trained is tailored to the specific image it is trained on; therefore it is slow and too computationally expensive to be used in practical applications. Prediction

---

Further author information: (Send correspondence to Ke Chen.)
Ke Chen.: E-mail: k.chen@liv.ac.uk,

to other images of a similar class is not possible as in traditional deep learning methods, as the network is tailored to a specific image only. Furthermore early stopping is required to preserve the regularisation effect. Some work has been done to add more regularisation into the DIP method[18–20] to reduce the need for early stopping.

In[20] the authors used a learned regularisation term in the loss function, where they trained a standard DIP network on a small number of examples on networks individually, saved the weights of each of the networks, and used the prior information of the saved weights in the loss function as a regularisation term in order to train a new test image. Using this learned regularisation term reduces the reconstruction error when used to train new images, however it is still a disadvantage that for every new example a new network has to be trained.

In this work, we aim to use the idea of the Deep Image Prior method to implement a variational model without explicit regularisation in a deep learning model to achieve selective segmentation. We propose to utilise a network architecture that incorporates DIP, but allows for prediction of images in the same class without further training. Our network is provided with image and geometric information to aid the prediction.

2. RELATED WORKS

An inverse problem in imaging is typically formulated as the following:

$$u^* = \arg\min_u \left\{ \mathcal{E}(u; f) + \mathcal{R}(u) \right\},$$

where f is an observed image and u is the desired solution. In this case, $\mathcal{E}(u; f)$ is some data fidelity term relating to a specific task and $\mathcal{R}(u)$ is a regularisation term imposing some kind of prior information. It was shown by Ulyanov et al.[17] that the explicit regularisation term could be dropped by parametrising the problem using a neural network. The idea behind the Deep Image Prior is that we can write an equivalent problem as the following:

$$\Theta^* = \arg\min_\Theta \mathcal{E}(\varphi_\Theta(z); f), \qquad u^* = \varphi_{\Theta^*}(z),$$

where φ_Θ is a neural network parametrised by weights Θ and z is random noise as input. It was found that the implicit regularisation offered by the architecture of the network was enough to provide a smoothing effect. The implication is that convolutional models tend to fit to smooth signals before fitting to noise; early stopping is key in the DIP method, to prevent overfitting to noise.

The selective segmentation variational model which we will base our proposed approach is the model by Roberts and Spencer.[4] Suppose we have an image f defined on an image domain Ω, their model is the following:

$$u^* = \arg\min_u \mu \int_\Omega g|\nabla u| \, d\mathbf{x} + \lambda \int_\Omega \Phi(f, c_1) u \, d\mathbf{x} + \theta \int_\Omega \mathcal{D}_G u \, d\mathbf{x}, \tag{1}$$

where Φ is a reformulated Chan-Vese term introduced in,[4] $g = g(|\nabla f|) = \frac{1}{1 + \iota|\nabla f|^2}$ is an edge detector, c_1 is the average intensity of f inside the region of interest and $\mathcal{D}_G$ is the geodesic distance from $\mathcal{M}$ defined in.[5] The binary segmentation result Σ is found by thresholding the minimiser, i.e. $\Sigma = \{\mathbf{x} \in \Omega : u^*(\mathbf{x}) > \gamma\}$, where we usually set $\gamma = 0.5$. In,[14] the authors implemented a similar selective segmentation model to (1) in a variational framework.

In variational segmentation models, Total Variation (TV)[21] is a popular choice for regularisation. In,[22] Zhu et al. investigated using Euler elastica for regularisation in place of total variation for segmentation. While Euler elastica proved to be an improvement over total variation, minimising it is computationally difficult as the resulting equations are highly non-linear. Among other methods of regularisation, Total Generalised Variation (TGV)[23] has also been used as a regulariser in segmentation.[24] Our proposed approach investigates the effect of removing such explicit regularisation terms, using the Deep Image Prior method.

3. PROPOSED APPROACH

In this section we introduce our proposed approach, which combines the Deep Image Prior approach with a variational model, allowing us to remove the explicit regularisation while preserving the ability to predict other images of the same class. Our idea is to make use of the Deep Image Prior's ability to extract and learn from deep features in the image providing regularisation in the solution, to remove the need for explicit regularisation in the variational model. Combining it with a more traditional deep learning model allows us to use the combination to predict further images while utilising the regularisation from DIP.

Let us denote f as an image ($f \in \mathbb{R}^{N \times M}$), z as a fixed noise vector ($z \in \mathbb{R}^{N \times M \times 32} \sim U(0, \frac{1}{10})$) and $\mathcal{G}$ as geometric information associated to the image f. For this paper, we consider $\mathcal{G}$ to be either the binary mask formed by the marker set $\mathcal{M}$, or the geodesic distance $\mathcal{D}_G$. Our proposed approach is to have two networks: A variational method like network (VM Net) from[14] which we denote as ψ_{Θ_1}, and a deep image prior network which we denote as φ_{Θ_2} (architecture can be found in Figure 1), where Θ_1 and Θ_2 are the weights that parametrise each network respectively. Each network produces a segmentation output, u_{VM} and u_{DIP} i.e. $u_{VM} = \psi_{\Theta_1}(f, \mathcal{G})$ and $u_{DIP} = \varphi_{\Theta_2}(z)$. Note that the VM network takes in image information as input (f and $\mathcal{G}$) where the DIP network takes noise as input (z) as in the original.[17]
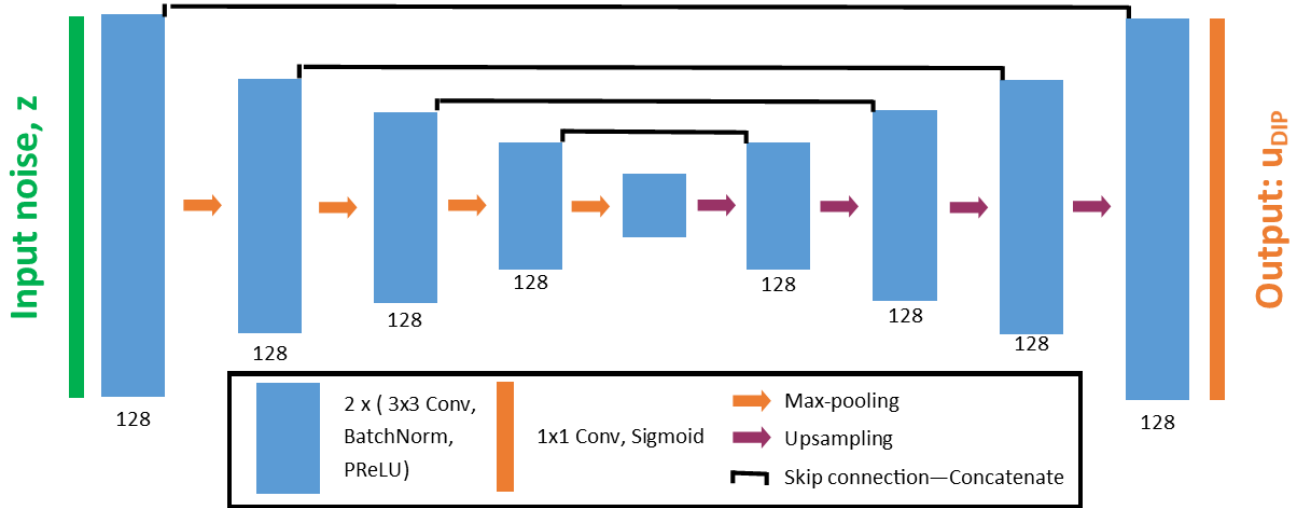

Figure 1: The architecture of the Deep Image Prior network (DIP Net) we use. A fixed noise vector z is used as input, and output is the segmentation result, u_{DIP}.

Overall, we propose to merge these two networks as shown in Figure 2. The idea is to combine each segmentation output, u_{VM} and u_{DIP} with a Hadamard product at the end to obtain u. This final u is used in the loss function.

Loss function: Our loss function is formed by taking the variational model (1) and removing the explicit regularisation term, inspired by the Deep Image Prior.[17] Moreover we include an additional term in order to enforce similarity between the outputs of the two networks. The loss function is given as follows:

$$\mathcal{L}(\Theta_1, \Theta_2) = \sum_{j=1}^{N} \lambda \int_{\Omega} \Phi(f^{(j)}, c_1^{(j)}) u^{(j)} d\mathbf{x} + \theta \int_{\Omega} \mathcal{D}_G^{(j)} u^{(j)} d\mathbf{x}$$
$$+ \frac{1}{2} \int_{\Omega} (u_{DIP}^{(j)} - u_{VM}^{(j)})^2 d\mathbf{x}, \tag{2}$$

where N is the total number of training images, $u^{(j)}$ denotes the output of the combined network of the $j'th$ training image, $u_{VM}^{(j)} = \psi_{\Theta_1}(f^{(j)}, \mathcal{G}^{(j)})$ and $u_{DIP}^{(j)} = \varphi_{\Theta_2}(z^{(j)})$. Therefore, $u_{VM} = u_{VM}(\Theta_1)$ and $u_{DIP} =$
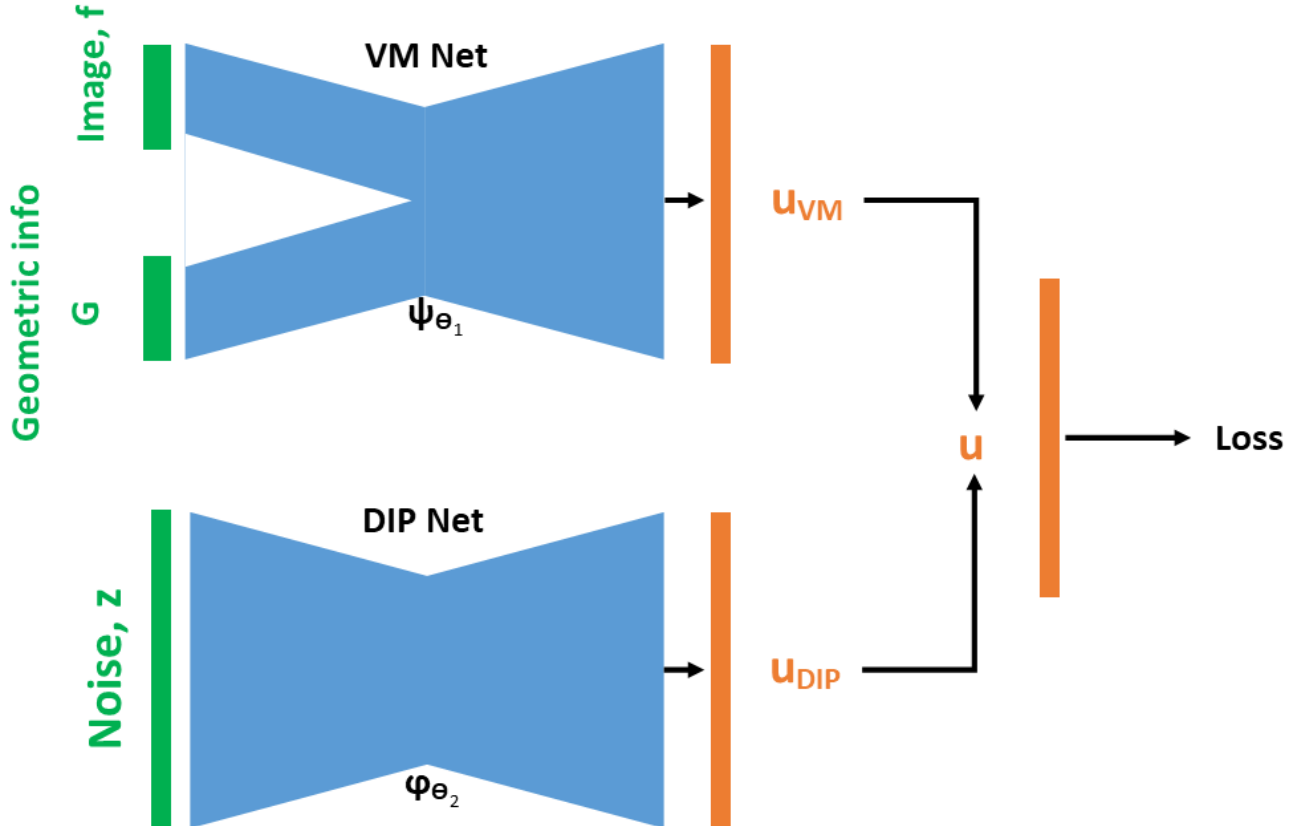
Figure 2: Our overall proposed workflow: Both networks produce a segmentation result, and the two results are combined using a Hadamard product.

$u_{DIP}(\Theta_2)$, and so $u = u(\Theta_1, \Theta_2)$. Moreover, $\mathcal{D}_G^{(j)}$ is the geodesic distance calculated using the marker points $\mathcal{M}^{(j)}$ on the $j'th$ image, and $c_1^{(j)}$ is fixed as the average intensity inside the polygon formed by $\mathcal{M}^{(j)}$. Training a network with this loss function does not require any ground truth labels, and can be used to predict future unseen images unlike the original DIP work[17] which fundamentally cannot predict a new image.

We do not require an explicit regularisation term as we utilise the implicit regularisation offered by the DIP. During the training stage, we rely on the regularisation effect offered by the DIP net to influence the weights of the VM net, which is why the third term in (2) is present. We note that, as in the original work,[17] early stopping of our network is still required to preserve the regularisation effect. In our experiments, we find only a very small number of N is required (we use $N = 2$). As the principle of the Deep Image Prior is to underfit an overparametrised network to maintain a regularisation effect, we find that large values of N do not make sense. It is important that we therefore supplement the network with additional information, such as geometric information based on user input like the geodesic distance.

We stress that for each training image $f^{(j)}$ there is a unique fixed noise vector $z^{(j)}$ associated to that training image. At test stage, we only consider the output of the VM net for segmentation. That is to say, we fix the weights of the VM net and the DIP net is no longer needed as the regularisation effect from the DIP net has already been transferred to the VM net during training. Therefore to predict a test image we only require the VM net and the inputs f and $\mathcal{G}$ for the test image, $u = \psi_{\Theta_1}(f, \mathcal{G})$.

4. METHODS

The proposed method was implemented in Tensorflow 2.3.0 using Keras backend. For our networks, we used the Adam optimiser with a learning rate of 0.001. As inputs, for the VM Net we use the image f and geometric

4

information $\mathcal{G}$ as inputs. When $\mathcal{G}$ is $\mathcal{M}$, this is simply a binary image, where the region inside the polygon formed by the markers is 1, and 0 outside. When $\mathcal{G}$ is $\mathcal{D}_G$, we calculate the geodesic distance from $\mathcal{M}$ by solving the Eikonal equation detailed in.[5] For the DIP net, we use input $z \in \mathbb{R}^{N \times M \times 32} \sim U(0, \frac{1}{10})$, however as in[17] we vary the noise input at each epoch with noise drawn from a normal distribution. Therefore at each training epoch, the input to the DIP net is $z + \hat{z}$, where $\hat{z} \sim \mathcal{N}(0, \frac{1}{100})$, where $\hat{z}$ is different each epoch.

We will consider four deep learning methods as a comparison, where two use old ideas, and two use the proposed network and loss function. To introduce the four methods, we introduce a baseline loss function that includes explicit regularisation:

$$\mathcal{L}(\Theta) = \sum_{j=1}^{N} \quad \mu \int_{\Omega} g^{(j)} |\nabla u^{(j)}| \, d\mathbf{x} + \lambda \int_{\Omega} \Phi(f^{(j)}, c_1^{(j)}) u^{(j)} \, d\mathbf{x} + \theta \int_{\Omega} \mathcal{D}_G^{(j)} u^{(j)} \, d\mathbf{x}, \tag{3}$$

where $u = u(\Theta)$ is the output of the network, $g^{(j)}$ is an edge detector associated to the $j'th$ image, and $c_1^{(j)}$ and $\mathcal{D}_G^{(j)}$ are defined as in (2).

The four deep learning methods are:

- **M1**: The VM Net only, where $\mathcal{G} = \mathcal{M}$ with loss function (3). This method is most similar to the one proposed in.[14]

- **M2**: The VM Net only, where $\mathcal{G} = \mathcal{M}$ with loss function (3) where $\mu = 0$. I.e. without an explicit regularisation term.

- **M3**: The proposed network detailed in Figure 2, where $\mathcal{G} = \mathcal{M}$ with proposed loss function (2).

- **M4**: The proposed network detailed in Figure 2, where $\mathcal{G} = \mathcal{D}_G$ with proposed loss function (2).

The two methods involving our proposed approach are **M3** and **M4**. These four methods are trained using training images and results on testing images are predicted after optimisation, with the weights of the network fixed. As a further comparison, we compare our work with a further three method whose methods require optimisation to be performed for each new image. We use a general form of the model (1):

$$u^* = \arg \min_{u} \mathcal{E}(u; f) + \mathcal{R}(u), \tag{4}$$

where $\mathcal{E}(u; f) = \lambda \int_{\Omega} \Phi(f, c_1) u \, d\mathbf{x} + \theta \int_{\Omega} \mathcal{D}_G u \, d\mathbf{x}$. The three further methods we compare with are:

- **TV-model:** The model (4) where $\mathcal{R}$ is the total variation regulariser,[21] i.e.

$$\mathcal{R}(u) = \int_{\Omega} |\nabla u| \, d\mathbf{x}.$$

We solve this model numerically using ADMM.

- **Elastica-model:** The model (4) where $\mathcal{R}$ is the Euler-Elastica regulariser,[22] i.e.

$$\mathcal{R}(u) = \int_{\Omega} \left(\alpha + \beta \left(\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right)^2 \right) \right) |\nabla u| \, d\mathbf{x}.$$

We solve this model numerically using ADMM.

- **DIP-like-model:** The model (4) optimised as in the original DIP work,[17] i.e. the eplxicit regularisation term $\mathcal{R}(u)$ is not needed, and we parametrise the problem using an untrained network φ_Θ, and the weights are optimised by minimising

$$\min_{\Theta} \mathcal{E}(\varphi_\Theta; f),$$

where early stopping is implemented to preserve a regularisation effect. We solve this model using Tensorflow using the conventional Deep Image Prior[17] setup.

(a) Image   (b) GT   (c) **M1**   (d) **M2**

(e) **M3**   (f) **M4**   (g) **TV-model**   (h) Elastica-model   (i) DIP-like-model

Figure 3: A sample result on the Lung data. We display the input image with the user input $\mathcal{M}$, the ground truth (GT) and results from the four methods. Moreover, we show comparisons with the model (4) solved in a variational framework with both Total Variation (TV) and Euler Elastica as explicit regularisation, as well as a comparison with the model solved in a Deep Image Prior framework.

(a) Image   (b) GT   (c) **M1**   (d) **M2**

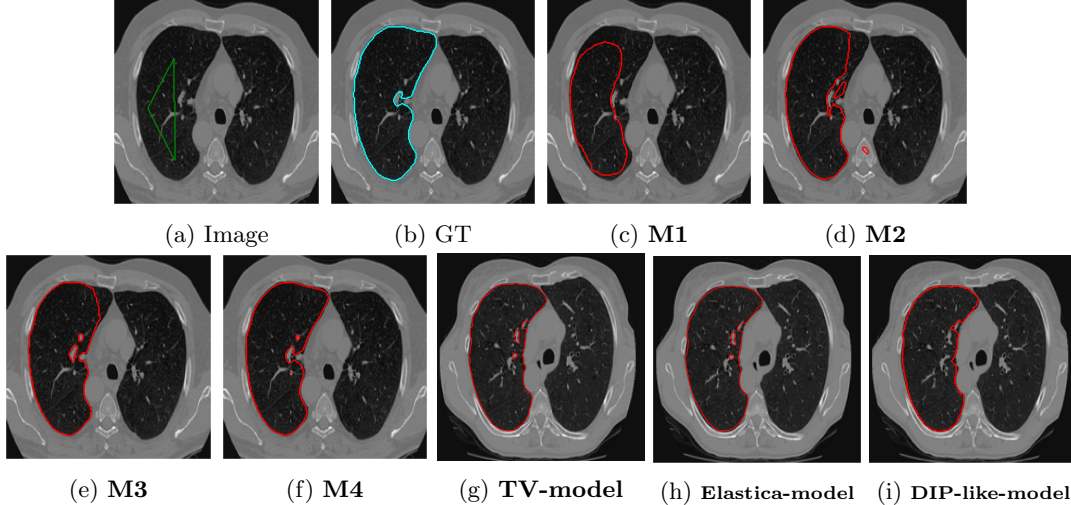(e) **M3**   (f) **M4**   (g) **TV-model**   (h) Elastica-model   (i) DIP-like-model

Figure 4: A sample result on the Lung data. We display the input image with the user input $\mathcal{M}$, the ground truth (GT) and results from the four methods. Moreover, we show comparisons with the model (4) solved in a variational framework with both Total Variation (TV) and Euler Elastica as explicit regularisation, as well as a comparison with the model solved in a Deep Image Prior framework.
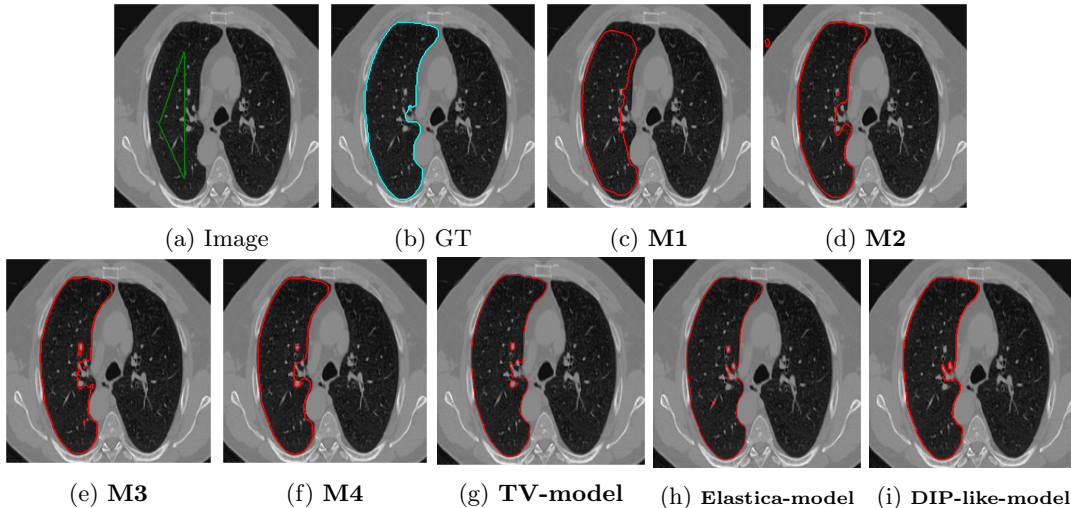
We conduct experiments on two datasets. The first selected data is images of the Lung from the Learn2Reg challenge dataset *.[25] We selected images of the lung to demonstrate the advantages of our method as lung segmentation with variational methods with explicit regularisation can be difficult to achieve a smooth result when nodules are present. We chose 18 images, one from each patient provided and used 2 to train and 16 to evaluate the performance. We aimed to segment the left lung by providing the some clicks on the region in order to provide the network with geometric information.

The second dataset we use is the Liver Tumour Segmentation Benchmark (LiTS) †,[26] focusing on only segmenting the lesions of the liver. We chose 29 images randomly, and used 2 images for training and 27 to

---

*Task 2: https://learn2reg.grand-challenge.org/Datasets/

†https://competitions.codalab.org/competitions/17094

(a) Image   (b) GT   (c) **M1**   (d) **M2**

(e) **M3**   (f) **M4**   (g) **TV-model**   (h) Elastica-model   (i) DIP-like-model

Figure 5: A sample result on the Liver data. We display the input image with the user input $\mathcal{M}$, the ground truth (GT) and results from the four methods. Moreover, we show comparisons with the model (4) solved in a variational framework with both Total Variation (TV) and Euler Elastica as explicit regularisation, as well as a comparison with the model solved in a Deep Image Prior framework.

(a) Image   (b) GT   (c) **M1**   (d) **M2**

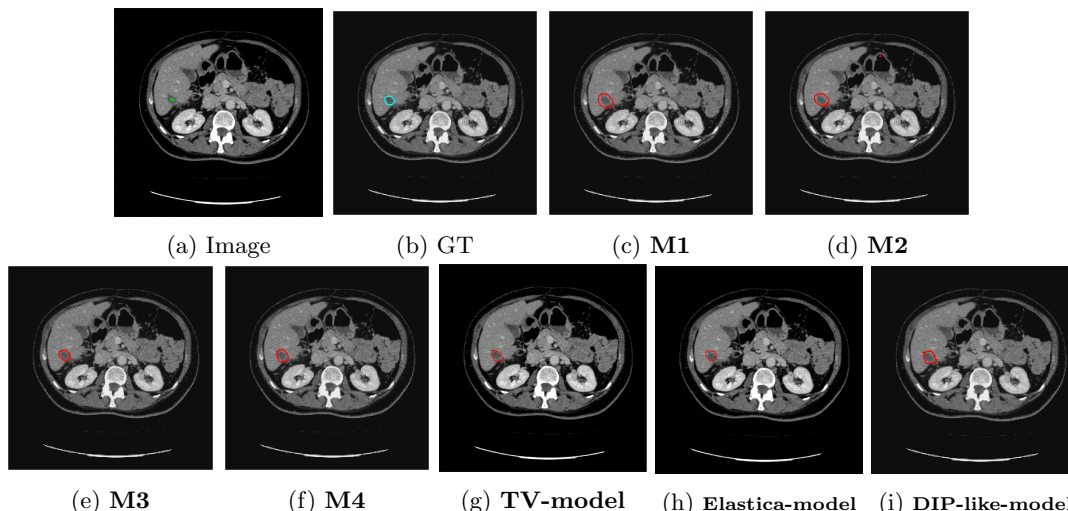(e) **M3**   (f) **M4**   (g) **TV-model**   (h) Elastica-model   (i) DIP-like-model

Figure 6: A sample result on the Liver data. We display the input image with the user input $\mathcal{M}$, the ground truth (GT) and results from the four methods. Moreover, we show comparisons with the model (4) solved in a variational framework with both Total Variation (TV) and Euler Elastica as explicit regularisation, as well as a comparison with the model solved in a Deep Image Prior framework.
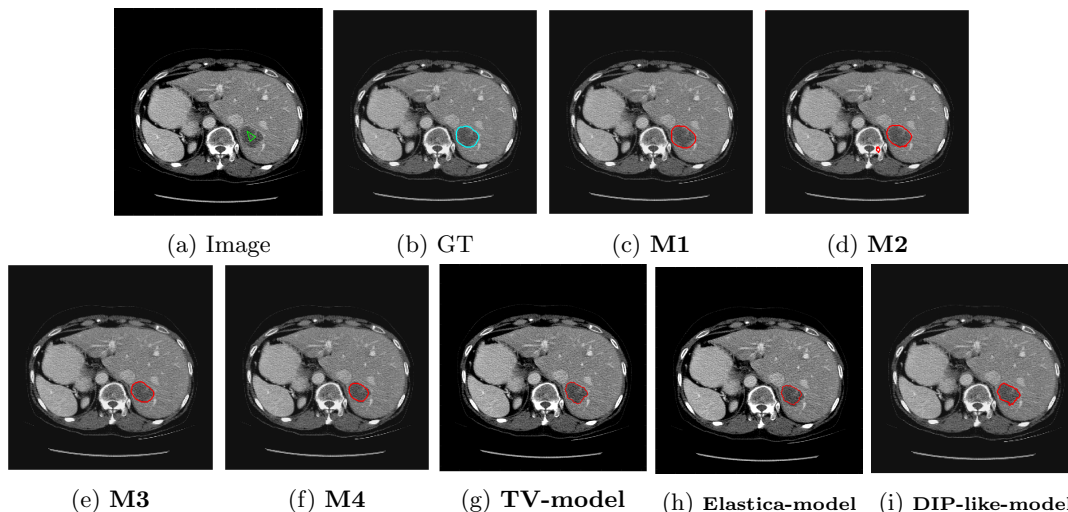
evaluate performance. Segmentation of the lesions was performed by providing the network with input inside the lesions for each image, so the network could be provided with geometric information.

	M1	M2	M3	M4	TV	Elastica	DIP
DICE mean	0.902	0.935	0.971	0.976	0.960	0.963	0.916
DICE std	0.047	0.048	0.021	0.024	0.031	0.015	0.143
JACCARD mean	0.825	0.882	0.944	0.953	0.925	0.929	0.830
JACCARD std	0.077	0.082	0.038	0.044	0.052	0.027	0.168

Table 1: Results for the 16 test images of the CT Lung data.

5. RESULTS

In this section we show results of the models discussed previously with the proposed approach. In order to show some quantitative results, we compute some common metrics to evaluate numerically the performance of the outputs of each method. We will display both the DICE coefficient and Jaccard coefficient. Given a segmentation result Σ for an image and an associated ground truth GT manually annotated by an expert, the DICE coefficient is given by:

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

A coefficient of 1 corresponds to a perfect result, whereas a value of 0 is the opposite. Similarly, the Jaccard coefficient is defined as:

$$JACCARD(\Sigma, GT) = \frac{|\Sigma \cap GT|}{|\Sigma| + |GT| - |\Sigma \cup GT|}.$$

In Figures 3 and 4 we show two select images from the test set for the Lung images to qualitatively show some results. To begin, we show some results of the model (1) solved with a Total Variation (TV)[21] regulariser, and with a Euler Elastica[22] regulariser. In addition, we show a result with the model (1) in a typical Deep Image Prior framework (i.e. training a network specifically for that image without the explicit regularisation in the loss function, employing early stopping). We see that the result from **M4** is an improvement over the TV and Elastica models, whereas results from the DIP example is comparable. However, the DIP example requires a new network trained specifically for the new image, whereas our result is acquired after training. In addition to the explicit regularisation comparison, we also display the results from the four methods (all of which are trained previously on 2 images and used for prediction). Moreover some quantitative results are shown in table 1, which shows the mean DICE score and standard deviation on the 16 images in the test set.

Similarly we show some results in Figures 5 and 6 from the LiTS dataset of all the methods. Quantitative results of the 27 images can be found in 2. Clearly for both datasets, **M3** and **M4** using the proposed ideas outperform **M1** and **M2**, as well as methods using explicit regularisation (TV and Elastica), and the original DIP method.

In addition, we can also compared with[14] who also used the LiTS dataset. The unsupervised approach trained on over 1000 images gave a mean DICE score of 0.671 with a standard deviation of 0.305 on 232 validation images. Using our method **M4** trained on just 2 images give a mean DICE score of 0.753 and standard deviation 0.213.

	M1	M2	M3	M4	TV	Elastica	DIP
DICE mean	0.841	0.767	0.879	0.868	0.862	0.864	0.840
DICE std	0.137	0.224	0.071	0.080	0.090	0.075	0.132
JACCARD mean	0.745	0.665	0.790	0.774	0.768	0.770	0.743
JACCARD std	0.174	0.241	0.112	0.116	0.129	0.122	0.165

Table 2: Results for the 27 test images of the LiTS liver tumour dataset.

6. CONCLUSION

In this paper we have introduced a new method that leverages the implicit regularisation offered by the Deep Image Prior method, while retaining the traditional deep learning power of predicting unseen data. Our method needs very limited training data (in both experiments only two training images were used), and is able to be used for prediction on new images of a similar class.

We evaluate our proposed method using two baseline methods: one which is a typical variational method-like network with explicit regularisation, and one without. We find that the two methods that use our proposed approach with a deep image prior network working with a variational method-like network outperforms the two baseline methods on both datasets used.

REFERENCES

[1] D. B. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on Pure and Applied Mathematics* , 1989.

[2] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on image processing* **10**(2), pp. 266–277, 2001.

[3] X. Cai, R. Chan, and T. Zeng, "A two-stage image segmentation method using a convex variant of the mumford–shah model and thresholding," *SIAM Journal on Imaging Sciences* **6**(1), pp. 368–390, 2013.

[4] M. Roberts and J. Spencer, "Chan–vese reformulation for selective image segmentation," *Journal of Mathematical Imaging and Vision* **61**(8), pp. 1173–1196, 2019.

[5] M. Roberts, K. Chen, and K. L. Irion, "A convex geodesic selective model for image segmentation," *Journal of Mathematical Imaging and Vision* **61**(4), pp. 482–503, 2019.

[6] J. Spencer and K. Chen, "A convex and selective variational model for image segmentation," *Communications in Mathematical Sciences* **13**(6), pp. 1453–1472, 2015.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[8] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence* **40**(4), pp. 834–848, 2017.

[9] B. A. Skourt, A. El Hassani, and A. Majda, "Lung ct image segmentation using deep neural networks," *Procedia Computer Science* **127**, pp. 109–113, 2018.

[10] L. Li, X. Weng, J. A. Schnabel, and X. Zhuang, "Joint left atrial segmentation and scar quantification based on a dnn with spatial encoding and shape attention," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 118–127, Springer, 2020.

[11] F. Cheng, C. Chen, Y. Wang, H. Shi, Y. Cao, D. Tu, C. Zhang, and Y. Xu, "Learning directional feature maps for cardiac mri segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 108–117, Springer, 2020.

[12] F. La Rosa, E. S. Beck, A. Abdulkadir, J.-P. Thiran, D. S. Reich, P. Sati, and M. B. Cuadra, "Automated detection of cortical lesions in multiple sclerosis patients with 7t mri," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 584–593, Springer, 2020.

[13] S. Chen, G. Bortsova, A. G.-U. Juárez, G. van Tulder, and M. de Bruijne, "Multi-task attention-based semi-supervised learning for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 457–465, Springer, 2019.

[14] L. Burrows, K. Chen, and F. Torella, "On new convolutional neural network based algorithms for selective segmentation of images," in *Annual Conference on Medical Image Understanding and Analysis*, pp. 93–104, Springer, 2020.

[15] S. Sedai, D. Mahapatra, S. Hewavitharanage, S. Maetschke, and R. Garnavi, "Semi-supervised segmentation of optic cup in retinal fundus images using variational autoencoder," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 75–82, Springer, 2017.

[16] T. Kalluri, G. Varma, M. Chandraker, and C. Jawahar, "Universal semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5259–5270, 2019.

[17] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9446–9454, 2018.

[18] J. Liu, Y. Sun, X. Xu, and U. S. Kamilov, "Image restoration using total variation regularized deep image prior," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7715–7719, IEEE, 2019.

[19] G. Mataev, P. Milanfar, and M. Elad, "Deepred: Deep image prior powered by red," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 0–0, 2019.

[20] D. Van Veen, A. Jalal, M. Soltanolkotabi, E. Price, S. Vishwanath, and A. G. Dimakis, "Compressed sensing with deep image prior and learned regularization," *arXiv preprint arXiv:1806.06438* , 2018.

[21] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena* **60**(1-4), pp. 259–268, 1992.

[22] W. Zhu, X.-C. Tai, and T. Chan, "Image segmentation using euler's elastica as the regularization," *Journal of scientific computing* **57**(2), pp. 414–438, 2013.

[23] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences* **3**(3), pp. 492–526, 2010.

[24] J. Duan, W. Lu, Z. Pan, and L. Bai, "New second order mumford–shah model based on γ-convergence approximation for image processing," *Infrared Physics & Technology* **76**, pp. 641–647, 2016.

[25] A. Hering, K. Murphy, and B. van Ginneken, "Learn2reg challenge: Ct lung registration - training data," 2020.

[26] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser, *et al.*, "The liver tumor segmentation benchmark (lits)," *arXiv preprint arXiv:1901.04056* , 2019.