Parameter-Free Selective Segmentation with Convex Variational Methods

Jack Spencer, Ke Chen, and Jinming Duan

***Abstract*—Selective segmentation methods involve incorporating user input to partition an image into a foreground and background. Often these methods are sensitive to some aspect of the user input in a counter intuitive manner, making their use in practice difficult. The most robust methods often involve laborious refinement on the part of the user, and sometimes editing/supervision. The proposed method reduces the burden on the user by simplifying the requirements on the input. Specifically, the fitting term does not depend on a distance function and so no selection parameter is introduced. Instead, we consider how the user input relates to some general intensity fitting term to ensure the approach is less sensitive to the decisions or intuition of the user. We give comparisons to existing approaches to show the advantages of the new selective segmentation model.**

I. Introduction

SEGMENTATION is the task of partitioning an image into regions of interest. In the case of two-phase segmentation, this consists of determining the foreground and background. Specifically, in the continuous setting, for an image $z(x) \in [0,1]$ in the domain $\Omega \subset \mathbf{R}^2$ the task is to compute disjoint subregions Ω_1 and Ω_2, such that $\Omega_1 \cup \Omega_2 = \Omega$, based on some partitioning criteria on the data $z(x)$. Notable examples include edge based methods, such as 'Snakes' [1] and Geodesic Active Contours [2], and region based methods, such as the two-phase piecewise-constant version of Mumford and Shah [3] from Chan and Vese [4]. However, in practice it is often useful to enhance such approaches by incorporating user input. An example of this, from the discrete setting, is Intelligent Scissors [5]. We refer to the task of two-phase segmentation with minimal user input as selective segmentation (although it is also described as interactive segmentation). Reliably incorporating simple constraints to determine the regions Ω_1 and Ω_2 is still often crucial, especially when manual segmentation approaches continue to be used or insufficient data is available to perform machine learning tasks. In cases where learning based methods are applicable, the work of Xu et al. [6] and Benard and Gygli [7] are state of the art approaches.

We consider two different types of input. One is to mark foreground/background regions. Important works that have utilised this input are graph cut approaches [8], geodesic methods [9], and Random Walks [10], based in the discrete setting, where each pixel in the image is treated as a node in a weighted graph. An energy function on this graph is minimised to extract a region of interest. We are interested in a more recent approach that uses the ideas discussed in [8], [9], and [10] in a continuous setting. That is the work of Nguyen et al. [11], an example of which can be seen in Fig. 1. They use Gaussian mixture models (GMMs) and distance terms to define a continuous data fitting term, with accurate results for many difficult cases.

The other type of approach is to consider n marker points, defined by the set $\mathcal{M} = \{x_i \in \Omega | i = 1, ..., n\}$, in order to apply geometrical constraints to the model. One of the first approaches to use this was from Gout et al. [12]. This was extended by Zhang et al. [13] and Rada and Chen [14] who incorporated area constraints on the polygon of the set $\mathcal{M}$, in an analogous way to Chan and Vese [4]. Also closely related is the recent work of Liu et al. [15]. These methods use the nonconvex framework of the level set method in a similar way to [4]. In [16], this work was extended to the convex relaxation framework of Chan, Esedoglu, and Nikolova [17], Bresson et al. [18], and many other more recent works. This combined a data fitting term from [4] with distance constraints based on $\mathcal{M}$. This had the advantage of improved robustness as for fixed intensity constants a global minimiser can be found. Approaches with foreground markers, $\mathcal{M}$, have the advantage of not needing user input in the background of the image, which is simpler and more intuitive. However, often results can be sensitive to the user input.

Here, we are interested in variational methods. Specifically, the convex relaxation approach of [17], [18] and many others, where a binary labelling of the foreground and background is determined based on minimising the following energy functional:

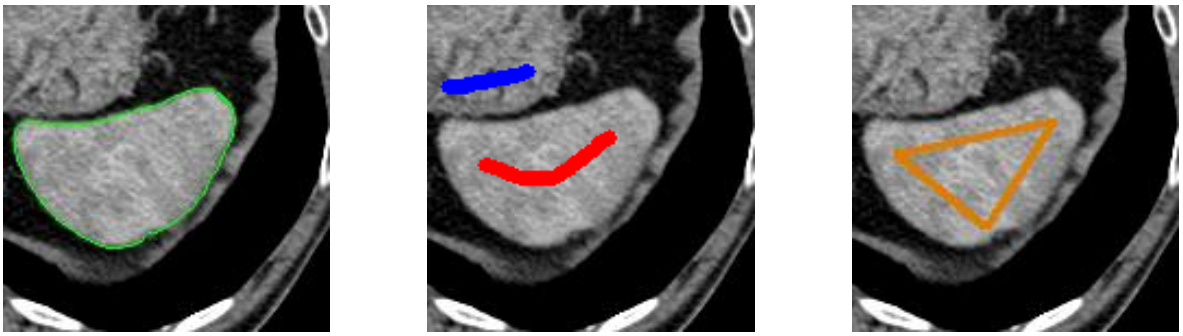
$$\min_{u \in [0,1]} \left\{ \int_\Omega g(x) |\nabla u(x)| \, dx + \lambda \int_\Omega f(x) u(x) \, dx \right\}, \quad (1)$$

involving a total variation (TV) regularisation term weighted by an edge function, $g(x)$, and some data fidelity term, $f(x)$. We will refer to this formulation throughout the paper, with different choices of $g(x)$ and $f(x)$ based on the model being considered. Minimisation of this energy is a well understood problem, but is not the focus of this work. Many methods are applicable in this case, such as Split-Bregman [19], the dual formulation ([18], [20], [21]), as well as many others. A thorough review of these methods in the continuous setting is given by Chambolle and Pock [22]. The focus of this paper is on the appropriate choice of $f(x)$ in order to effectively and robustly incorporate user input for selective segmentation

J. Spencer is with the Translational Research Exchange @ Exeter, Living Systems Institute, University of Exeter. e-mail: j.a.spencer@exeter.ac.uk.

K. Chen is with the Liverpool Centre for Mathematics in Healthcare, Department of Mathematical Sciences, University of Liverpool. e-mail: k.chen@liverpool.ac.uk.

J. Duan is with the Biomedical Image Analysis Group and MRC London Institute of Medical Sciences, Imperial College London. e-mail: j.duan@imperial.ac.uk.



(a) Image (b) Foreground/background regions (c) Markers $\mathcal{M}$

Fig. 1: Types of user input utilised in selective segmentation. a) Example image (ground truth in green). b) Foreground (red) and background (blue) regions used, e.g. in [11]. c) Polygon region $\mathcal{P}$ (orange) defined by markers $\mathcal{M}$, e.g. in [16].

problems, particularly with respect to medical examples from X-ray computed tomography data. Our main contribution is to propose an approach with minimal sensitivity to user input, based on not introducing additional parameters to the model. Here, it is applied in the context of piecewise-constant segmentation but can be extended to a wider range of examples.

An overview of recent developments in medical image segmentation is given by [23], providing a background for methods such as clustering and graph cuts. A survey of how deep learning has been applied to medical image analysis, including segmentation, is provided by Litjens et al. [24]. Details of how this is applied to specific anatomical areas is also discussed. Two recent examples of this type of application can be found for cardiac MR images [25] and cervical vertebrae in X-ray images [26].

The paper is organised as follows. In Section II we introduce some related work and discuss the limitations of related approaches. In Section III we detail the proposed method, to introduce selective segmentation without additional parameters. We discuss the new selection function, where the values of the intensity fitting term incorporate the user input. In Section IV we present some experimental results where we first compare the proposed method to the work of [11] and [16], examine its robustness to a variety of user input, and perform further tests on a large data set. The paper ends with some concluding remarks and discussion of future work.

II. Related Work

In the following we discuss related work involving different types of user input. We first highlight the approach of Nguyen et al. [11] and then recent work by Spencer and Chen [16]. We comment on their relation to similar approaches ([8], [9], [10], [12], [13], [14]) and mention some drawbacks of approaching the selective segmentation problem in this way.

A. Constrained Active Contours (CAC) [11]

The authors use a probability map, $P_1(x)$, from Bai and Sapiro [9] where the geodesic distances to the foreground/background regions are denoted by $D_F(x)$ and $D_B(x)$, respectively. An approximation of the probability that a point x belongs to the foreground is then given by $P_1(x)$, based on distance from the user input. An example of user input is shown in Fig. 1, leading to foreground/background GMM estimations. The normalised log likelihood that a point x belongs to the foreground and background is then given by $P_F(x)$ and $P_B(x)$, respectively. GMMs are widely used in selective segmentation ([5], [8], [9], [10]) and the authors in [11] define the following data fitting term:

$$h_1(x) = \alpha\left(P_B(x) - P_F(x)\right) + (1-\alpha)\left(1 - 2P_1(x)\right), \quad (2)$$

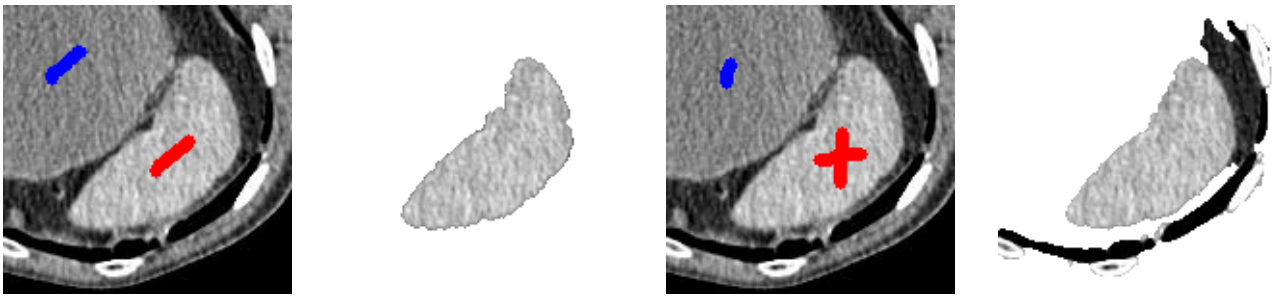
for a weighting parameter $\alpha \in [0,1]$, selected automatically. The TV regularisation is weighted by an edge function involving the following function, applied to the image $z(x)$:

$$g_0(x) = \frac{1}{1+\beta_0|\nabla z(x)|^2}. \quad (3)$$

In [11], they set $\beta_0 = 1$ and define g_1 as an enhanced edge function involving an automatically selected parameter, β. Thus, Nguyen et al. [11] define the Constrained Active Contours (CAC) Model as the formulation of (1) with $g(x) = g_1(x)$ and $f(x) = h_1(x)$. In Fig. 2 we present a result that illustrates a common drawback of this type of method. That is, the results can be quite sensitive to the user-defined foreground/background regions. For each case, we can see an object of interest marked by the red foreground region, with a nearby object marked with a blue background region. Fig. 2 demonstrates that the user input is similar, but that the result is very different in each case. The first result has accuracy ($a \in [0,1]$ defined by (13)) of $a = 0.952$ but the second has accuracy of $a = 0.562$. The aim of our paper is to move towards an arbitrary user input, such that results are less sensitive to intuition on the part of the user.

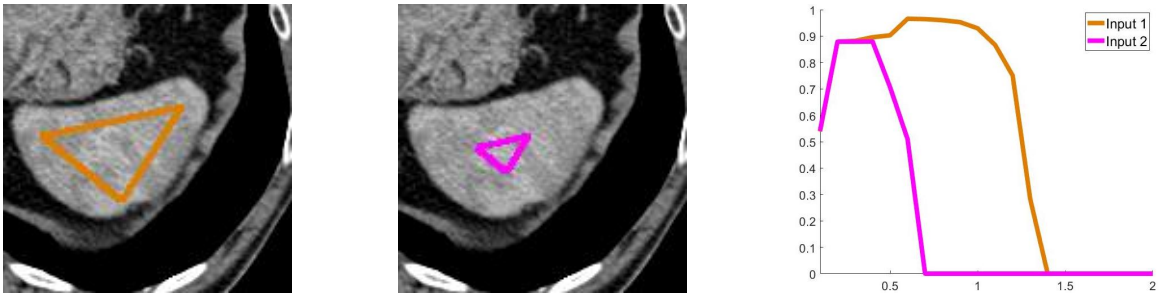
B. Convex Distance Selection (CDS) [16]

In a similar way to [12], [13] and [14] this approach requires the user to provide marker points $\mathcal{M} = \{x_i \in \Omega | i = 1, ..., n\}$. In Fig. 1, an example of this is shown for a case when $n = 3$. This allows the solution of (1) to be constrained by values associated with the polygon region, $\mathcal{P}$, associated with the set $\mathcal{M}$. In [16] the selection function is the normalised Euclidean



(a) Input 1 (b) Result 1 (c) Input 2 (d) Result 2

Fig. 2: An example of the dependence on user input for CAC [11]; minor changes produce very different results. Accuracy in each case: (b) $a = 0.952$, (d) $a = 0.562$. Similar behaviour can be observed in [8], [9], and [10].



(a) Input 1 (b) Input 2 (c) Parameter Dependence, $a(\theta)$

Fig. 3: An example of the dependence on user input in CDS [16]. Different marker sets, $\mathcal{M}$, often require a variation of the parameter, θ. The optimal parameter in terms of accuracy ($a \in [0, 1]$) in each case: Input 1 ($\theta^* = 0.4$), Input 2 ($\theta^* = 0.7$).

distance from $\mathcal{P}$, which we denote $P_2(x)$. The authors use the data fitting term

$$f_0(x) = (z - c_1)^2 - (z - c_2)^2, \tag{4}$$

as in Chan and Vese [4] and the corresponding convex formulation [17], where $c_1, c_2 \in \mathbf{R}$ are the average intenities in the foreground and background, respectively. This was extended to cases of intensity inhomogeneity in [27] and is consistent with other types of data fitting term, such as [28] and [29]. This selective fitting term is defined as follows, weighted by a selection parameter:

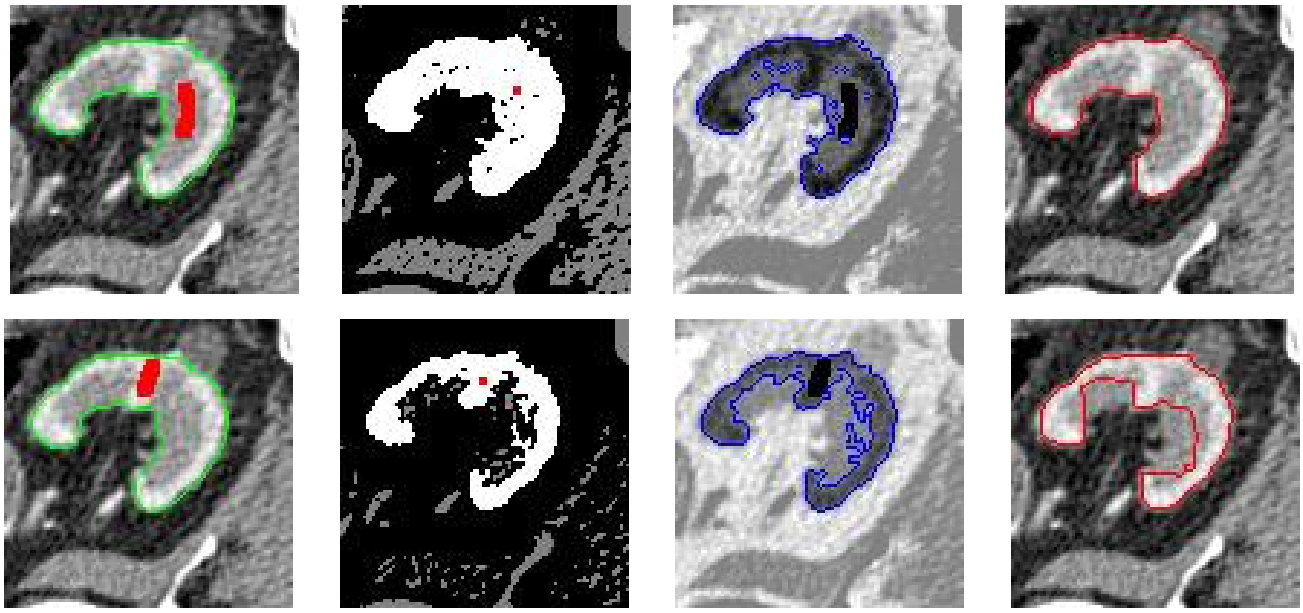
$$h_2(x) = f_0(x) + \theta P_2(x). \tag{5}$$

The Convex Distance Selection (CDS) Model [16] is then defined as (1), with $f(x) = h_2(x)$ and $g(x) = g_0(x)$ and a variable β_0. Whilst the results are often very good, the method can be overly dependent on the choice of $\mathcal{M}$ and θ. In Fig. 3 we include a particular example where the accuracy of the result ($a \in [0, 1]$) is given when varying θ for different sets of user input, $\mathcal{M}$. We can see that for slightly different choices of $\mathcal{M}$, the optimal selection parameter is $\theta^* = 0.4$ and $\theta^* = 0.7$, respectively. Whilst some guidance on the appropriate choice of θ can be provided, this feature makes methods that incorporate distance constraints to a marker set, $\mathcal{M}$, challenging. Again, the aim of this paper is to remove this sensitivity to the user input to increase the robustness of the approach.

C. Alternative Methods

In this section we briefly mention recent work relevant to the proposed model. First is the work of Deriche et al. [30], which involves Chan-Vese fitting after a geodesic or dynamic region merging based initialisation, using foreground/backgroud regions provided by the user. Another recent approach is that of Liu et al. [15] which works in a similar framework to Cai et al. [31], by introducing a new weighting function to their formulation. The focus of this approach is on applications to medical images, and requires user-markers on the boundary of the target object. Each method seems to be a competitive approach to interactive segmentation, and future work could establish how they perform in relation to the proposed method.

The work we intend to highlight in this section are recent approaches based on extensions of the random walks approach [10]. This includes Shen et al. [32], which is closely related to a more recent method known as Submarkov Random Walks (SRW) by Dong et al. [33]. SRW can be viewed as a traditional random walker with added auxiliary nodes, and is capable of achieving impressive results with user-defined foreground/background regions. We do not detail the method of SRW here as it uses a different framework to the proposed method. However, the user input is defined in the same way as CAC making it ideal for comparison and further details can be found in [33]. We compare SRW to our proposed approach on a large data set in Section IV.



a) Image b) Selection regions c) Fitting term d) Result

Fig. 4: Examples for the proposed model (different user inputs in each row). From left to right: an example image (green indicating ground truth, red indicating user input region w_f), the selection region ($w(x) = 1$ shown in white, μ in red), the fitting term $h(x)$ (blue indicating the zero threshold) based on (8), and the corresponding result (red contour).

III. Proposed Method

In this section we introduce details behind the proposed method. It is designed to prevent any additional parameter selection being required, and to simplify the user input such that it does not depend on intuition. In this section we introduce the method in the context of piecewise-constant segmentation, in a similar way to CDS, as many of the examples of interest contain target objects with approximately homogeneous intensities. However, it is consistent with related methods that do not include this assumption, such as [27], [28], [29], [34], and many others. First, we will introduce the details behind the fitting term we define to incorporate some foreground markers, $w_f(x) \in \{0, 1\}$. We will then detail the implementation of the proposed method, in relation to minor adjustments to the formulation and details behind obtaining a solution. We refer to the proposed method as Parameter-Free Selection (PFS), as we do not define any new parameters in the data fitting term.

A. Parameter-Free Fitting Term

Initially, the foreground region w_f is selected by the user (1 inside this region, 0 elsewhere). The fixed intensity approximations are then determined as follows:

$$\tilde{c_1} = \frac{\int_\Omega w_f z \; dx}{\int_\Omega w_f \; dx}, \quad \tilde{c_2} = \frac{\int_\Omega (1 - w_f) z \; dx}{\int_\Omega (1 - w_f) \; dx}. \tag{6}$$

From here we compute the intensity fitting term [4]:

$$\tilde{f}(x) = (z - \tilde{c_1})^2 - (z - \tilde{c_2})^2. \tag{7}$$

The new model depends primarily on the function $w(x) \in \{0, 1\}$, which we define as follows. We first determine a point $\mu \in \mathbf{R}^2$ chosen from the region(s), $r \subset \Omega$, defined by the intersection of $w_f = 1$ and $H(-\tilde{f}) = 1$ (where $H(\cdot)$ is the Heaviside function). In other words, μ is within the user-defined region, w_f, and the region indicated as being foreground by the fitting term, $\tilde{f}(x)$. Specifically, μ is automatically selected within r as the point furthest away from the boundary of r. This is illustrated in Fig. 4. The selection term $w(x)$ is defined as follows:

$$w(x) = \begin{cases} 1, & x \in \Omega, \; H(-\tilde{f(x)}) = 1, \text{ connected to } \mu \\ 0, & \text{else.} \end{cases}$$

This function indicates parts of the domain Ω that are most likely to be in the foreground based on the values of the intensity fitting term, $f(x)$, and the user input, w_f. The fitting term is then given by

$$h(x) = \tilde{f}(x) \left(1 - H(-\tilde{f}) + w\right). \tag{8}$$

In this way we essentially neutralise/suppress values of the fitting term that are unlikely to be in the foreground, based on the user input. The main consideration in terms of achieving a successful result is that the target object must be relatively distinct in terms of its intensity, as the function $w(x)$ relies on connected regions of negative values of $\tilde{f}(x)$. It should also be noted that this is similar, but not equivalent, to performing a segmentation using $\tilde{f}(x)$ directly and then defining regions of the binary solution connected to μ. This approach is less likely to successfully identify the correct region as minor differences in intensity will be exaggerated unless a large λ is chosen. Whilst it is likely to work in simple cases it is not robust enough for use in practice, and a parameter dependence is created though the choice of λ.

In Fig. 4 we introduce two examples to illustrate how the proposed fitting term works in practice, and how the result depends on the user input in cases where there is some inhomogeneity in the target foreground. We consider an example with two different user input selections (rows 1 and 2, respectively). Here, the user-defined region w_f is shown in red in relation to the ground truth (green contour) in the left column. The second column illustrates the functions $H(-\tilde{f}(x))$ and $w(x)$ for these cases. Here, $H(-\tilde{f}(x)) = 0$ is given in black, $H(-\tilde{f}(x)) = 1$ is given in grey/white, and $w(x) = 1$ is given in white. The point μ is also shown in red. The third column is the fitting term as defined in (8). Darker regions (negative values) are more likely to be in the foreground, with lighter regions (positive values) more likely to be considered background. The zero threshold of this function is given in blue and is generally very close to the final result. The corresponding result is shown by the red contour in the right column. In this example the optimal intensity constants are $c_1^* = 0.71$ and $c_2^* = 0.35$. The proposed model is most likely to succeed if the region w_f is chosen such that $\tilde{c_1}$ and $\tilde{c_2}$ are close to c_1^* and c_2^*, respectively. However, because of the nature of the Chan-Vese fitting term some variation from this is acceptable. This allows the proposed method to be robust to user input except in cases of high levels of inhomogeneity, such as the image here. Typically, in the case that $c_1^* > c_2^*$, a user selection that underestimates the foreground intensity is superior to one that overestimates it. If $c_1^* < c_2^*$, this phenomenon is reversed. This is illustrated in Fig. 4. For the example in row 1 we have $\tilde{c_1} = 0.60$ and $\tilde{c_2} = 0.41$. For row 2 we have $\tilde{c_1} = 0.85$ and $\tilde{c_2} = 0.41$. It can be seen that the proposed fitting term successfully segments the target region in the first case.

The important aspect of this approach is that no new parameters are introduced and that the dependence on the user input is minimal, which we will return to later. CAC [11], despite often producing very good results, occasionally requires multiple foreground/background regions to define a viable fitting term. In the case of CDS [16] the parameter θ is introduced that can be difficult to optimise intuitively, and is therefore difficult to use in practice. The key aspect of our approach is that we do not require any new parameters, primarily because we do not introduce a distance function to the fitting term. The value introduced from the user selection, w_f (whether similar to foreground regions from [11] or markers, $\mathcal{M}$, from [16]), is the point μ and the values of $\tilde{c_1}$ and $\tilde{c_2}$, and is used to locate the region of negative fitting values of interest. It should also be noted that we do not require a background region to be selected, which simplifies the requirements on the user in practice. Results do not vary much depending on the input (assuming the foreground is approximately homogeneous), which is an important feature of any method depending on user input, and offers a clear advantage over the alternative methods discussed. It is also possible to incorporate the input to define alternative fitting terms to (7), which we use here because of the images under consideration, such that a wider range of examples can be successfully segmented.

B. New Model and its Implementation

We now define the new model with a parameter-free fitting term, and clarify some details concerning computing a solution. In this work we standardise the fitting terms by defining the maximum absolute value of $h(x)$, $m_h \in \mathbf{R}$. Then, the new scaled fitting term is defined as

$$\tilde{h}(x) = \frac{h(x)}{|m_h|}. \tag{9}$$

This allows the choice of λ to be more predictable, and in the tests below we fix it at $\lambda = 2$. It is worth noting here that some work exists based on optimising the fitting parameter λ automatically (Mylona et al. [35]), but in general this choice is well understood for fitting terms in a similar range. In our tests we set $g(x) = 1$ as in the convex variational setting we have found that the choice does not heavily influence the result. In nonconvex formulations ([1], [2], [14]) an edge function plays a crucial role as local minima are desirable. In this work, however, the quality of results is primarily dictated by the choice of $f(x)$ in (1). Additionally, poor choices of β_0 in (3) can actually reduce the quality of the results so it is omitted here. According to these choices, the problem then consists of a two-phase variational segmentation problem that we consider in a conventional manner, defined in Section I:

$$\min_{u\in[0,1]} \left\{ \int_\Omega |\nabla u(x)| \; dx + \lambda \int_\Omega \tilde{h}(x) u(x) \; dx \right\}. \tag{10}$$

We use the dual formulation of Bresson et al. [18], with further details provided in [20] and [21]. Other methods are also applicable, such as [19] and [36]. We do not detail them here, but the optimisation step is not the focus of this work. We have found that the fastest way to obtain a solution is to define the initialisation, $u_0(x)$, as

$$u_0(x) = \begin{cases} 1, & \text{for } x \in \tilde{h}(x) < 0 \\ 0, & \text{for } x \in \tilde{h}(x) \geq 0, \end{cases}$$

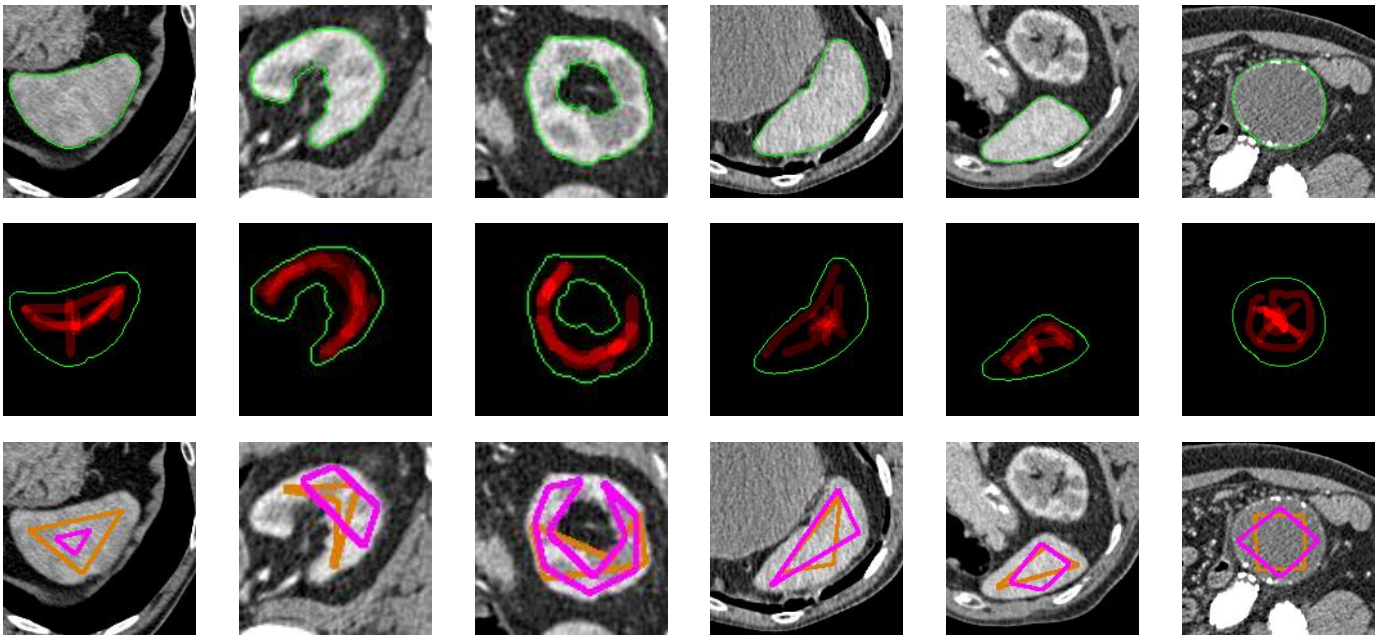
as this is quite close to the global minimum of (10). However, we note that for fixed $\tilde{h}(x)$ the solution is independent of initialisation. We define the solution as $u^*(x)$. Based on the work of [16], [17], [18], and others this will be approximately binary, such that any thresholding of the function will be a global minimiser of the original problem. We define the computed foreground region as follows:

$$\Omega_1^* = \{x \in \Omega |\; u^*(x) > \gamma\}. \tag{11}$$

We select $\gamma = 0.5$ (although other values $\gamma \in (0, 1)$ would yield a similar result). In the following we use the binary form of the solution, u^*, denoted Ω_1^*.

IV. Experimental Results

In this section we introduce some test examples to compare our approach to the work of CAC [11], CDS [16], and SRW [33]. In all the following tests, we use axial slices of CT data, with typical abdominal window width/levels. This is normalised and the resolution varies between 100x100 and 180x180, depending on the problem. We measure accuracy



a) Image 1 b) Image 2 c) Image 3 d) Image 4 e) Image 5 f) Image 6

Fig. 5: Test Examples displaying user input. Top row: Image (ground truth in green). Middle row: summation of five foreground regions (red) in relation to the ground truth, used with CAC. Bottom row: two polygon regions $\mathcal{P}$ (orange, pink), used with CDS. Left to right: Images 1-6, respectively. Each input type is used in PFS, defining $w_f(x)$.

with respect to some ground truth, given by a manual segmentation:

$$GT = \{x \in \Omega |\ x \in \text{foreground}\}.$$

Using the thresholding of the solution defined in Section III, we denote the computed foreground as Ω_1^*. We then define the accuracy $a \in [0, 1]$ of the solution (in terms of the Tanimoto Coefficient or Jaccard Index), u^*, as follows:

$$a = \frac{N(GT \cap \Omega_1^*)}{N(GT \cup \Omega_1^*)}, \tag{12}$$

where $N(\cdot)$ refers to the number of points in the enclosed region. We also provide a value that refers to the foreground consensus (FC); the percentage of the ground truth foreground that is successfully segmented by all variations of the user input. Specifically, if we denote the intersection of Ω_1^* for all sets of user input as Ω_F^*, the foreground consensus is given by

$$\text{FC} = \frac{N(\Omega_F^* \cap GT)}{N(GT)} \times 100. \tag{13}$$

This helps highlight the consistency of each method, but doesn't account for spurious background regions that are determined to be foreground. It is worth clarifying that a result could have an FC value of 100%, and an accuracy of $a < 1$. This is because FC only measures variation in the foreground of the ground truth. This can be useful to assess a method's dependence on user input even when the accuracy is good. These values are listed in Table I, and help establish how robust to user input the proposed method is.

First, we compare the proposed method, PFS, against CAC for five different user inputs in each image. We then compare PFS against the CDS approach using optimal parameters for different choices of $\mathcal{M}$. In Fig. 5 we show the different user inputs used in Tables I and II. It should be noted that in Fig. 5 we exclude the background regions as PFS only requires w_f. However, CAC requires background regions for each test and a representative example of the input used can be seen in Fig. 6 for Image 1. We then measure the robustness of the proposed approach to variations in user input, detailing the results in Table III. Finally, we compare the proposed method to CAC, CDS, and an additional method mentioned in Section II-C: SRW [33]. Here, we test each method on 100 CT images, using five variations of user input in each case, to assess their performance on a larger set of data.

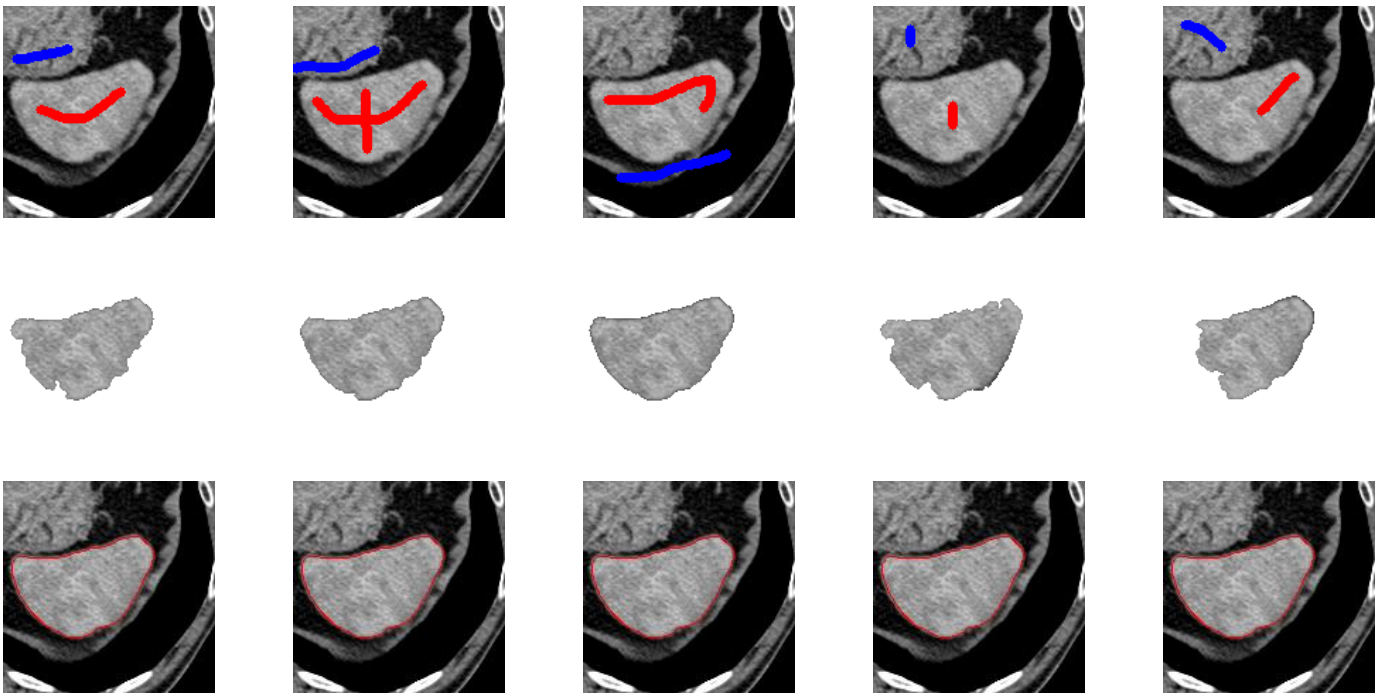
A. Comparisons with Related Work

In this section we compare the proposed method, PFS, to the methods of CAC and CDS by using identical user input for each set of tests. The results are displayed in Tables I and II, with additional detail provided in Figs. 6 and 7.

In Table I we detail the comparison between PFS and CAC. For the six images given in Fig. 5 we use five different sets of user inputs. An example of the inputs we use is shown in Fig. 6 for Image 1. It is important to note that in the following tests we use the foreground regions (red) and background regions (blue) for CAC, but that for PFS we exclude the background input to define w_f based on the corresponding foreground input only. We are interested in the optimal result in terms of accuracy but also the robustness to varying the input, such that we provide the average accuracy for each image. From Table I we can see that PFS provides the optimal result (given

TABLE I: Accuracy results (in term of the Tanimoto Coefficient or Jaccard Index) comparing the proposed method, PFS, to CAC [11] for five different inputs and six images. The optimal result for each image is in bold. The average accuracy and foreground consensus (FC) percentage for different inputs is given for each method and image.

Input	Image 1		Image 2		Image 3		Image 4		Image 5		Image 6	
	CAC	PFS	CAC	PFS	CAC	PFS	CAC	PFS	CAC	PFS	CAC	PFS
1	0.920	0.967	0.857	0.905	0.874	0.943	0.952	0.932	0.938	0.934	0.529	0.973
2	0.962	0.967	0.889	0.909	0.895	**0.954**	0.562	0.938	**0.963**	0.935	0.871	**0.975**
3	**0.968**	**0.968**	0.906	**0.922**	0.871	0.941	0.959	0.937	0.961	0.937	0.886	**0.975**
4	0.923	0.964	0.887	0.901	0.889	0.946	**0.968**	0.946	0.939	0.935	0.909	0.973
5	0.792	0.967	0.907	0.894	0.868	0.950	0.953	0.939	0.678	0.934	0.908	0.973
Average	0.913	0.967	0.889	0.906	0.879	0.946	0.879	0.938	0.896	0.935	0.821	0.974
FC (%)	76.8	96.8	89.2	89.7	90.9	94.7	94.6	93.6	94.1	93.7	94.6	99.0



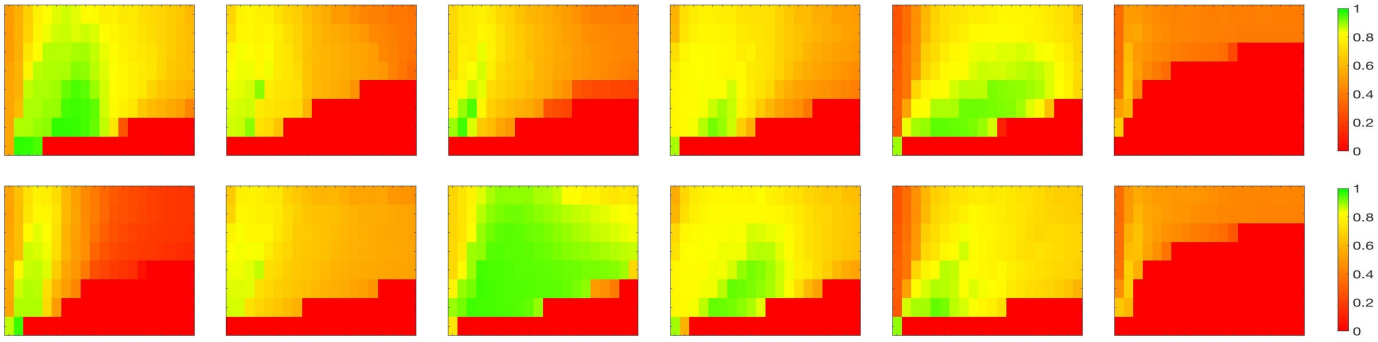
a) Input 1 b) Input 2 c) Input 3 d) Input 4 e) Input 5

Fig. 6: Results for Image 1. The top row gives the user input, the middle row gives the output of CAC [11], and the bottom row gives the result of PFS (based on the corresponding foreground regions defining $w_f(x)$). From left to right is Input 1-5. Accuracy results for each input are given in Table I.

TABLE II: Accuracy results comparing the proposed method, PFS, to CDS [16]. The optimal result for CDS is provided, including the corresponding parameters, (λ^*, θ^*).

Image	Input 1			Input 2		
	(λ^*, θ^*)	CDS(*)	PFS	(λ^*, θ^*)	CDS (*)	PFS
1	(0.1,0.2)	0.967	0.968	(0.1,0.2)	0.958	0.964
2	(2.0,0.4)	0.906	0.912	(2.0,0.4)	0.867	0.915
3	(0.5,0.2)	0.956	0.946	(0.5,0.3)	0.955	0.950
4	(0.5,0.5)	0.920	0.942	(1.0,0.7)	0.930	0.944
5	(0.5,0.5)	0.934	0.936	(0.5,0.5)	0.931	0.937
6	(0.5,0.1)	0.673	0.973	(2.0,0.2)	0.678	0.972

in bold) for Images 2, 3, and 6. For Images 4 and 5, CAC is superior in this sense but the result for PFS is still 0.946 and 0.937 respectively which is good for these cases. It is also worth noting that for every image PFS beats CAC in terms of the average accuracy, which is very important for this type of problem. For most examples the difference is noteworthy as well. The reasons for this vary, as can be demonstrated by highlighting two images. For Image 3 PFS comfortably outperforms CAC for Inputs 1-5 resulting in a higher average for the proposed method. For Image 4, CAC is noticeably better than PFS for Inputs 1-4. However, for Input 5 CAC fails completely ($a = 0.562$) whilst PFS performs well ($a = 0.938$) resulting in a superior average for PFS. Each case highlights an important aspect of the proposed method; reliability and consistent quality. CAC [11] and related methods ([8], [9], [10]) can perform very well but are often sensitive to minor changes to the input that isn't necessarily intuitive. The FC



a) Image 1 b) Image 2 c) Image 3 d) Image 4 e) Image 5 f) Image 6

Fig. 7: Results for CDS [16] for permutations of λ, θ. The colour represents the accuracy of each result, $a \in [0, 1]$, when varying λ (vertical axis) and θ (horizontal axis). The top row is for Input 1 and the bottom row is for Input 2, from Fig. 5.

(%) demonstrate that PFS is a more consistent method than CAC. Generally, for any foreground regions within the object PFS will produce the same result whereas CAC is more temperamental.

In comparing the proposed PFS to CDS we first look at the dependence of CDS on the parameters λ and θ in Fig. 7. Here we present the accuracy of CDS for varying θ (horizontal axis) and λ (vertical axis), with a total of 160 tests for each user input (shown in Fig. 5). We can then compare the optimal result for CDS against the proposed parameter-free method. Along the horizontal axis we vary the selection parameter θ linearly between 0.1 and 2. Along the vertical axis we vary the fitting parameter between the following choices: $\lambda \in \{0.1, 0.5, 1, 2, 5, 10, 50, 100\}$. The colour represents the accuracy of the corresponding solution, $a \in [0, 1]$. Visually, it can be seen from Fig. 7 that the results are very sensitive to the parameter choice and that this varies between different examples. This makes the method difficult to use in practice. The only case where this is not the case is Image 3, Input 2 where there is a large range for which a good result can be found (indicated by the green region). However, it is important to note that $\mathcal{M}$ uses $n = 11$ and $\mathcal{P}$ closely resembles the shape of the target object. Image 3, Input 1 uses a more reasonable $\mathcal{M}$ ($n = 5$) and consequently has a much narrower region where a good result can be achieved.

The comparison of CDS to PFS can be seen in Table II where we list the optimal result, CDS(*), and its corresponding selection parameters, (λ^*, θ^*), and the result of PFS using the input shown in Fig. 5, i.e. $w_f = \mathcal{P}$. From here, it can be seen that the optimal choice of the selection parameter, θ, varies between 0.1 and 0.7 and is often different for the same image with minor changes to the user input. We can see that the proposed parameter-free method is more accurate than the optimal result of CDS in all cases except Image 3. Even then, the difference is minor. For Image 6, the results are vastly superior for PFS. It is likely that increasing the number of marker points could enable a good result to be achieved for CDS but simplifying the requirements on the user is desirable in this context. The key conclusion from these tests is that for the same user input the proposed method, PFS, outperforms the best results of CDS despite requiring no parameter optimisation.

B. Robustness to User Input

One challenge in determining the robustness of selective segmentation methods is that each test requires user input, which is often difficult to generalise. For the proposed method, PFS, it is possible to mimic the user input as the foreground input produces three variables that determine the fitting term, $\tilde{h}$: $\tilde{c_1}, \tilde{c_2}$, and μ. In these tests, for each image we have the corresponding GT and therefore know the true intensity constants, c_1^* and c_2^*. By using reasonable approximations of these values for $\tilde{c_1}$ and $\tilde{c_2}$ and a random choice of μ within the foreground of GT, we can compute $\tilde{h}(x)$. In Table III we present the average PFS result for 200 simulations of user inputs for each image. We also include the optimal results for CAC and CDS detailed in the previous results section. This demonstrates that for Images 1, 2, and 6 the PFS average is better than the optimal result of CAC or CDS. For Images 3, 4, and 5 PFS is only beaten by the optimal result of one of the other methods; it is never the worst. Here, it can be seen that the proposed method holds up very well when the input is generalised.

C. Further Testing

In order to further establish the robustness of our method we now introduce the results of testing on a large data set. This consists of 100 CT images of structures such as the kidney, spleen, and abdominal aorta. Whilst the proposed method generally assumes homogeneity in the foreground, many of the images contain high levels of inhomogeneity. Some representative examples are shown in Fig. 8. In Fig. 9 we present boxplots of the average accuracy results of each image, i.e. each value is the average of five variations of user input for each method. We present two sets of results for the proposed method. The first is PFS (i), where we use the markers provided by the input for CDS (optimising θ for each result). The second is PFS (ii), where we use the user foreground/background regions used in CAC and SRW. It is worth noting that the inputs used are precisely the same between each method. The average for each method is 0.84

TABLE III: Average accuracy for 200 random choices of the user input with the proposed method, PFS. For each image, the optimal results for CAC [11] and CDS [16] are included.

	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6
CAC (*)	0.968	0.906	0.895	0.968	0.963	0.909
CDS (*)	0.967	0.906	0.956	0.930	0.934	0.678
PFS average	0.973	0.924	0.954	0.940	0.944	0.966

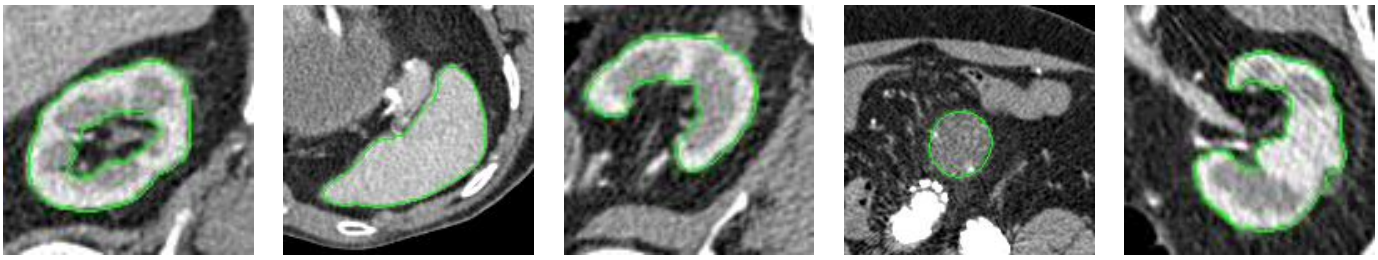


Fig. 8: Examples of images in the large data set tested, with results presented in Fig. 9.

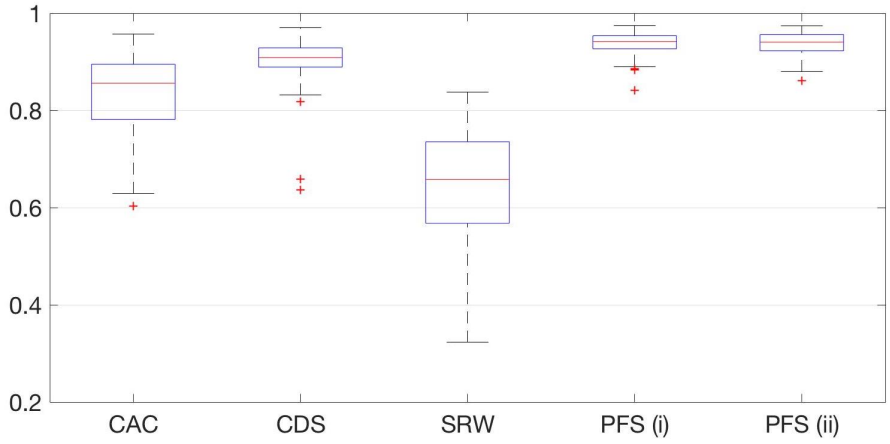


Fig. 9: Box plots of average accuracy results of varied user input on a large data set (100 CT images). The same user input was used for CAC, SRW, and PFS (ii), with a different set used for CDS and PFS (i) as described in Section II.

(CAC), 0.90 (CDS), 0.64 (SRW), 0.94 (PFS (i) & (ii)). It can be seen that the results for the proposed method are very consistent, with the worst result being 0.84 and 0.86 for PFS (i) and (ii), respectively. Notably, this is similar to the average result of CAC which seems to be quite inconsistent for this data set. The results for CDS are reasonably good. However, it assumes that the selection parameter θ is near optimal. This is not true in practice and if the choice was automated the quality and consistency of the results would suffer. The results for SRW indicate that the user input is insufficient to achieve good quality results for this data set. The results are very inconsistent, with few images near an acceptable standard. The model tends to be sensitive to variations of the user input in a similar way to the examples shown in Fig. 2.

V. Conclusions

,

In this paper we have introduced a parameter-free method for selective segmentation based on user input in the foreground of the image. We have compared it against the Constrained Active Contours (CAC) Method [11], the Convex Distance Selection (CDS) Method [16], and Submarkov Random Walk (SRW) [33] using a variety of input. We show that the proposed method is very consistent and generally outperforms the related approaches for the same input. Importantly, this is without any parameter choice or optimisation on behalf of the user, necessary in [16], [12], [13], [14]. We also demonstrate that the proposed method achieves good results for arbitrary user input. For 200 variations of the user input for each test image we show that the results are as good as, and often better than, the optimal results of CAC [11] and CDS [16]. This aspect of the method is particularly important for selective segmentation as it allows for this approach to be used in practice reliably, with minimal requirements on the user. In this work we incorporate the fitting term of [4], which assumes the foreground is approximately homogeneous. For the images under consideration here this is sufficient, but future work could focus on how to extend this to alternatives by generalising how the user input is defined in Section III.

ACKNOWLEDGEMENTS

This work was generously supported by the Wellcome Trust Institutional Strategic Support Award (204909/Z/16/Z). The authors would like to thank the Isaac Newton Institute for Mathematical Sciences for support and hospitality during the programme 'Variational methods and effective algorithms for imaging and vision' when work on this paper was undertaken. This work was supported by EPSRC Grant Number EP/K032208/1. The authors would also like to acknowledge the support of the EPSRC grant EP/N014499/1.

REFERENCES

[1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[2] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997.

[3] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *Communications on Pure and Applied Mathematics*, vol. 42, pp. 577–685, 1989.

[4] T. Chan and L. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[5] A. Falcao, J. Udupa, and F. Migazawa, "An ultrafast user-steered image segmentation paradigm: live wire on the fly," *IEEE Transactions on Medical Imaging*, vol. 19, no. 1, pp. 55–62, 2002.

[6] N. Xu, B. Price, S. Cohen, J. Yang, and T. Huang, "Deep interactive object selection," in *CVPR*, 2016.

[7] A. Benard and M. Gygli, "Interactive video object segmentation in the wild," 2017, arXiv:1801.00269.

[8] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM SIGGRAPH*, vol. 23, no. 3, pp. 1–6, 2004.

[9] X. Bai and G. Sapiro, "A geodesic framework for fast interactive image and video segmentation and matting," *IEEE International Conference on Computer Vision*, pp. 1–8, 2007.

[10] L. Grady, "Random walks for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006.

[11] T. Nguyen, J. Cai, J. Zhang, and J. Zheng, "Robust interactive image segmentation using convex active contours," *IEEE Transactions on Image Processing*, vol. 21, pp. 3734–3743, 2012.

[12] C. Gout, C. L. Guyader, and L. Vese, "Segmentation under geometrical conditions with geodesic active contours and interpolation using level set methods," *Numerical Algorithms*, vol. 39, pp. 155–173, 2005.

[13] J. Zhang, K. Chen, B. Yu, and D. Gould, "A local information based variational model for selective image segmentation," *Inverse Problems and Imaging*, vol. 8, no. 1, pp. 293–320, 2014.

[14] L. Rada and K. Chen, "Improved selective segmentation model using one level-set," *Journal of Algorithms & Computational Technology*, vol. 7, no. 4, pp. 509–540, 2013.

[15] C. Liu, M. K.-P. Ng, and T. Zeng, "Weighted variational model for selective image segmentation with application to medical images," *Pattern Recognition*, vol. 76, pp. 367–379, 2018.

[16] J. Spencer and K. Chen, "A convex and selective variational model for image segmentation," *Communications in Mathematical Sciences*, vol. 13, no. 6, pp. 1453–1472, 2015.

[17] T. Chan, S. Esedoglu, and M. Nikolova, "Algorithms for finding global minimizers of image segmentation and denoising models," *SIAM Journal on Applied Mathematics*, vol. 66, no. 5, pp. 1632–1648, 2006.

[18] X. Bresson, S. Esedoglu, P. Vandergheynst, J. P. Thiran, and S. Osher, "Fast global minimization of the active contour/snake model," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 2, pp. 151–167, 2007.

[19] T. Goldstein, X. Bresson, and S. Osher, "Geometric applications of the split bregman method," *Journal of Scientific Computing*, vol. 45, no. 1-3, pp. 272–293, 2010.

[20] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 89–97, 2004.

[21] J. F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture decomposition–modeling, algorithms, and parameter selection," *International Journal of Computer Vision*, vol. 67, no. 1, pp. 111–136, 2006.

[22] A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," *Acta Numerica*, vol. 25, pp. 161–319, 2016.

[23] E. Zanaty and S. Ghoniemy, "Medical image segmentation techniques: An overview," *International Journal of Informatics and Medical Data Processing*, vol. 1, no. 1, pp. 16–37, 2016.

[24] G. Litjens, T. Kooi, B. Bejnordi, A. Setio, F. Ciompi, M. Ghafoorian, J. van der Laak, B. van Ginneken, and C. Sanchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[25] J. Duan, J. Schlemper, W. Bai, T. Dawes, G. Bello, G. Doumou, A. De Marvao, D. O'Regan, and D. Rueckert, "Deep nested level sets: Fully automated segmentation of cardiac MR images in patients with pulmonary hypertension," in *MICCAI*, 2018, pp. 595–603.

[26] S. M. M. R. Al Arif, K. Knapp, and G. Slabaugh, "Fully automatic cervical vertebrae segmentation framework for X-ray images," *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 95–111, 2018.

[27] J. Spencer and K. Chen, "Stabilised bias field: Segmentation with intensity inhomogeneity," *Journal of Algorithms and Computational Technology*, vol. 10, no. 4, pp. 302–313, 2016.

[28] C. Li, C. Kao, J. Gore, and Z. Ding, "Minimization of region-scalable fitting energy for image segmentation," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1940–1949, 2008.

[29] D. Chen, M. Yang, and L. Cohen, "Global minimum for a variant Mumford-Shah model with application to medical image segmentation," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 1, no. 1, pp. 48–60, 2013.

[30] M. Deriche, A. Amin, and M. Qureshi, "Color image segmentation by combining the convex active contour and the chan vese model," *Pattern Analysis and Applications*, 2017, https://doi.org/10.1007/s10044-017-0632-9.

[31] 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*, vol. 6, no. 1, pp. 368–390, 2013.

[32] J. Shen, Y. Du, W. Wang, and X. Li, "Lazy random walks for superpixel segmentation," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1451–1462, 2014.

[33] X. Dong, J. Shen, and L. Shao, "Submarkov random walk for image segmentation," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 516–527, 2016.

[34] L. Tan, Z. Pan, W. Liu, J. Duan, and W. Wei, "Image segmentation with depth information via simplified variational level set formulation," *Journal of Mathematical Imaging and Vision*, vol. 60, no. 1, pp. 1–17, 2018.

[35] E. Mylona, M. Savelonas, and D. Maroulis, "Automated adjustment of region-based active contour parameters using local image geometry," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2757–2770, 2014.

[36] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, pp. 120–145, 2011.