CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.SO-2024-0043

Vol. **7**, No. 1, pp. 178-206
March 2026

Double-Domain Driven Unet for Selective Segmentation of Medical Image

Zhi-Feng Pang[1], Lin Yang[1], Mingxiang Wu[1], Ziyu Niu[1],
Raymond Chan[2] and Xue-Cheng Tai[3,*]

[1] *College of Mathematics and Statistics, Henan University, Kaifeng 475004, China.*
[2] *School of Data Science, Lingnan University, Tuen Mun 999077, Hong Kong, SAR, China.*
[3] *Norwegian Research Centre, Nygardstangen, Bergen NO-5838, Norway.*

Received 9 September 2024; Accepted 13 June 2025

Abstract. Selective segmentation has been a notable increase in interest in the use of interactive deep learning-based methods. However, developing an efficient and accurate segmentation method for medical applications remains a formidable challenge, primarily due to the inherent complexity and diversity of medical image structures. To address these challenges, we introduce a novel deep learning approach, termed the double-domain driven Unet method (DDUM), designed for the selective segmentation of medical images. Our approach utilizes a threshold geodesic distance in conjunction with the original images as input to construct a parallel Unet architecture that captures information from both the image domain and the geodesic distance domain. To further enhance the accuracy and efficacy of the segmentation process, we employ soft threshold dynamics as a replacement for the sigmoid activation function in the final layer. The efficacy of our proposed DDUM is substantiated through extensive experiments conducted on multiple medical image datasets. In particular, the DDUM method exhibits exceptional performance in terms of both segmentation accuracy and robustness.

AMS subject classifications: 68U10, 65K05

Key words: Image segmentation, regions of interest, threshold geodesic distance, double-domain driven Unet method, soft threshold dynamics.

1 Introduction

Image segmentation plays a critical component in medical imaging, enabling precise diagnosis and treatment planning through the delineation of anatomical structures or

[*]Corresponding author. *Email addresses:* `zhifengpang@henu.edu.cn` (Z.-F. Pang), `yanglin_henu@163.com` (L. Yang), `mingxiangwu2022@163.com` (M. Wu), `niuziyu2892@163.com` (Z. Niu), `raymond.chan@ln.edu.hk` (R. Chan), `xtai@norceresearch.no` (X.-C. Tai)

pathological regions. Current segmentation methodologies can be broadly categorized into two principal approaches: semantic segmentation [10,25,35,37,56] and selective segmentation [13, 38, 47]. Semantic segmentation involves the comprehensive separation of all foreground objects from the background within an image, while selective segmentation focuses exclusively on isolating specific subsets of foreground objects or regions of interest (ROI). Despite substantial progress in selective segmentation techniques in recent decades, medical image analysis continues to face significant challenges. These include the presence of noise, imaging artifacts, low contrast variability, and substantial interpatient anatomical differences, all of which contribute to inconsistent segmentation accuracy, particularly in cases involving complex anatomical variations and intricate pathological presentations [44, 61]. Consequently, we focus here emphasizes the advancement of selective segmentation techniques for medical imaging, aiming to develop methods that deliver clinically applicable levels of accuracy and robustness.

In general, the selective segmentation method uses user annotations to guide the segmentation process, resulting in a faster and more accurate segmentation result for the ROI [4]. Traditional selective segmentation methods, such as graph cut-based methods [2, 5, 42, 60], grab cut-based methods [20, 24], random walk-based methods [17], geodesic-based methods [3, 9], and grow cut-based methods [51], typically rely on shallow image features and leverage annotation information to optimize an energy functional. However, such shallow features often fail to adequately capture the rich contextual information of the image, particularly in scenarios involving low contrast or significant noise. To address these limitations, recent advancements in active contour models [14] have emphasized the integration of prior annotation information, such as convexity priors [6, 29] or distance-based constraints [16, 38], to enhance segmentation accuracy and robustness. For example, Nguyen *et al.* [34] integrated the convex active contour [6] with the annotation information, computed the geodesic map to obtain the initial contour and achieved fast segmentation. Gout *et al.* [16] used the distance of some marked points as weights for the edge function and constructed an energy functional based on the level set method. Spencer and Chen [47] incorporated a regularization term based on Euclidean distance into the original active contour model for selective segmentation. Roberts *et al.* [38] proposed replacing the Euclidean distance with an edge-weighted geodesic distance, which adaptively increases near image edges, thereby offering enhanced suitability for selective segmentation tasks. Despite these advancements, achieving accurate segmentation remains heavily dependent on extensive user input and manually engineered features, limiting their scalability and practical applicability in complex scenarios.

In recent years, convolutional neural networks (CNNs) have demonstrated the capacity to achieve state-of-the-art performance in image segmentation, largely due to their ability to automatically learn advanced semantic features (for further details, please see references [1, 11, 28, 45]). The Unet architecture [39], with its distinctive encoder-decoder framework, has gained widespread adoption in medical image segmentation tasks. Nevertheless, a significant challenge persists in this domain: the extensive variability in the shapes and sizes of segmented objects inherently limits the network's ability to effectively model complex geometric transformations [21]. To assist networks in addressing complex boundaries or a specific objective, a number of deep learning-based interactive segmentation methods have been proposed in [22]. For example, image-free semantic segmentation (IFSeg) [43] employed user clicks and original images as input for an interactive convolutional neural network, whereas DeepCut [36] trained the neural network using boundary box annotations within the context of graph cut. However, these methods may encounter difficulties in extracting complex texture structures from images when only simple clicks and box annotations are used. To address this issue, Roth *et al.* [40, 41] proposed utilizing the random walker algorithm [17] to generate an initial segmentation based on extreme points, employing a noisy supervisory signal to train a full convolutional network (FCN). In [55], the authors introduced the deep extreme level set evolution (DELSE) method, which seamlessly integrates a robust convolutional neural network with the level set framework within a unified, end-to-end architecture. This approach is inherently interactive, as it incorporates user input in the form of clicks on extreme boundary points. However, the necessity of labeling extreme boundary points imposes a significant limitation on the operability and practicality of the DELSE method. To address this issue, an alternative strategy proposed in [54] leverages a dual deep CNN framework for interactive segmentation. Specifically, the first network generated an initial segmentation, while the second network refined the results by incorporating prior information and user interactions. Luo *et al.* [30] introduced a minimally interactive segmentation framework based on deep learning, termed MIDeepSeg. This approach utilized the exponentialized geodesic distance (EGD) transform as input to a Unet architecture to generate initial segmentation results, which were subsequently refined using additional annotation information. Such methodologies, often categorized as two-phase strategies [30,36,40,41,54,55], demonstrated promising segmentation performance. However, the integration of supplementary strategies may result in increased computational complexity and extended processing times, potentially complicating the user experience.

In order to address the aforementioned challenges, this paper proposes a novel framework, termed the double-domain driven Unet method, which facilitates selective segmentation of medical images by effectively leveraging complementary information from both the image domain and the geodesic distance domain. The key contributions of this work are summarized as follows:

(1) We propose an extension of the Eikonal equation to align with the specific characteristics of medical images. This results in the proposal of a threshold geodesic distance, which serves to guarantee consistency between the ROI and marked points, while simultaneously maintaining a clear distinction between the foreground and background.

(2) The DDUM leverages the strengths of two independently trained Unets, each dedicated to processing information from distinct domains: the image domain and the geodesic distance domain. By integrating these two Unets into a unified network architecture, the proposed framework achieves a synergistic enhancement, outperforming the individual Unet operating on separate domains.

(3) The soft threshold dynamics (STD) activation function, which incorporates both softmax and spatial information, can be employed to reinforce the resilience of the network architecture. Consequently, we use the STD for the processing of the output features derived from both the image domain and the geodesic domain. This approach enables the acquisition of more efficacious network features, thereby enhancing the accuracy of the segmentation.

The remainder of this paper is organized as follows. Section 2 defines the geodesic distance and its modified form, before introducing some descriptions of the well-known Unet. We then introduce the proposed segmentation framework, which incorporates information from both the image domain and the geodesic domain into the network architecture. This section also provides detailed explanations of the threshold geodesic distance, the network architecture, and the loss function. In Section 3, we compare our proposed method with several state-of-the-art methods to validate its effectiveness. In addition, we discuss ablation experiments in Section 4 and present our conclusions in Section 5.

2 Proposed method

This section presents fundamental concepts and methodologies essential for the subsequent development of our work. We begin with an overview of geodesic distance and its modified formulation, followed by the introduction of our proposed threshold geodesic distance (TGD) along with its theoretical motivation. In addition, we provide technical details on the Unet architecture and the modified STD activation function, which constitute critical components of our proposed framework.Then we provide the details of the network architecture and the associated loss function.

2.1 Geodesic distance

Let us consider an open and bounded domain, denoted by Ω, in the Euclidean space $\mathbb{R}^2$. We assume that Ω is equipped with a positive definite metric, which we denote by $\mathcal{H}:\Omega\to\mathbb{R}^{2\times 2}$. Then it can be shown that the Riemannian metric $\mathcal{H}(\cdot)$ allows us to define a global metric on the space Ω by using the shortest path.

Definition 2.1. *Given a Riemannian space $(\Omega,\mathcal{H})$, the geodesic distance* [33] *is defined by*

$$d(\mathrm{x},\mathrm{y})=\min_{\gamma\in\mathcal{P}(\mathrm{x},\mathrm{y})}\int_0^1\sqrt{\big(\gamma'(t)\big)^{\top}\mathcal{H}\big(\gamma(t)\big)\gamma'(t)}\mathrm{dt},$$

where $\mathcal{P}(\mathrm{x},\mathrm{y})$ denotes the set of all piecewise smooth curves which join x *and* y*, with each curve parameterized by $\gamma(t)$ satisfying $\gamma(0)=\mathrm{x}$ and $\gamma(1)=\mathrm{y}$.*

In general, there may be more than one geodesic curve between two points. In order to perform the numerical computation of geodesic distance, it is necessary to fix a marked point set $\mathcal{M}\subset\Omega$.

Definition 2.2. *The geodesic distance map of a point* $\mathrm{x}\in\Omega$ *to the marked point set* $\mathcal{M}$ *is defined by*

$$D(\mathrm{x})=\min_{\bar{\mathrm{x}}\in\mathcal{M}} d(\mathrm{x},\bar{\mathrm{x}}). \tag{2.1}$$

As illustrated in the distance map (2.1), it is evident that $D(\mathrm{x})=0$ if $\mathrm{x}\in\mathcal{M}$. It is established in [33] that the geodesic distance map $D(\mathrm{x})$ is the unique viscosity solution to the following generalised Eikonal equation:

$$\begin{cases} \|\nabla D(\mathrm{x})\|_{(\mathcal{H}(\mathrm{x}))^{-1}}=1, & \text{if} \quad \mathrm{x}\notin\mathcal{M}, \\ D(\mathrm{x})=0, & \text{if} \quad \mathrm{x}\in\mathcal{M}. \end{cases} \tag{2.2}$$

Remark 2.1. In general, the geodesic distance depends on the choice of the metric $\mathcal{H}(\mathrm{x})$. Here we only consider two cases as

- If we set the Riemannian metric $\mathcal{H}(\mathrm{x})$ to be the unit matrix $\mathcal{I}$, then the solution of the Eq. (2.2) corresponds to the Euclidean distance $D_e(\mathrm{x})$.
- If we set $d(\mathrm{x})$ to be a weight function and $\mathcal{H}(\mathrm{x})=d^2(\mathrm{x})\mathcal{I}$ to be the isotropic metric, then the solution of Eq. (2.2) corresponds to the general geodesic distance.

In the context of image segmentation, pixels within the ROI generally demonstrate relatively minimal geodesic distances between them. Conversely, pixels situated on opposing sides of the segmentation boundary typically exhibit significantly larger geodesic distances. Nevertheless, it is crucial to acknowledge that distorted information or noise within the segmented image can accumulate along these geodesic paths, consequently compromising the accuracy and precision of the geodesic distance measurements. To this end, the Riemannian metric is commonly employed as the edge detection function, which is dependent on the gradient norm of the input image $I(\mathrm{x})$, as established in [38] by

$$d_r(\mathrm{x})=\varepsilon_1+\beta_1\|\nabla S(\mathrm{x})\|^2, \tag{2.3}$$

where ε_1 represents a small positive constant, while β_1 represents a large positive constant. The function $S(\mathrm{x})$ denotes a Gaussian filter as

$$S(\mathrm{x}):=G_\eta(\mathrm{x})*I(\mathrm{x}), \tag{2.4}$$

where the Gaussian kernel

$$G_\eta(\mathrm{x}):=\frac{1}{2\pi\eta^2}\exp\left(\frac{-\|\mathrm{x}\|^2}{2\eta^2}\right)$$

has a standard deviation of η and $*$ denotes the convolution operator. If we select $\mathcal{H}(\mathrm{x})=d_r(\mathrm{x})^2\mathcal{I}$ in (2.3), the partial differential equation (PDE) (2.2) can be rephrased as follows:

$$\begin{cases} \|\nabla D(\mathrm{x})\| = d_r(\mathrm{x}), & \text{if} \quad \mathrm{x} \notin \mathcal{M}, \\ D(\mathrm{x}) = 0, & \text{if} \quad \mathrm{x} \in \mathcal{M} \end{cases} \tag{2.5}$$

and denotes its solution as $D_r(\mathrm{x})$.

Remark 2.2. In the PDE (2.5), the geodesic distance $D_r(\mathrm{x})$ is predominantly governed by the parameter ε_1 in regions characterized by smooth intensity variations, where $\|\nabla S(\mathrm{x})\|^2 \approx 0$. Conversely, in proximity to image edges, where $\|\nabla S(\mathrm{x})\|^2 \gg 0$, the diffusion rate $d_r(\mathrm{x})$ is primarily influenced by the magnitude of $\beta_1 \|\nabla S(\mathrm{x})\|^2$. Notably, the choice of a larger value for β_1 induces a significant acceleration in the growth of the geodesic distance $D_r(\mathrm{x})$.

The fixed parameter β_1 in the formula (2.3) is challenging to describe the feature of the ROI, as a result, a suitable choice is typically selected adaptively. In order to achieve this, we introduce a threshold function $d_t(\mathrm{x})$ as

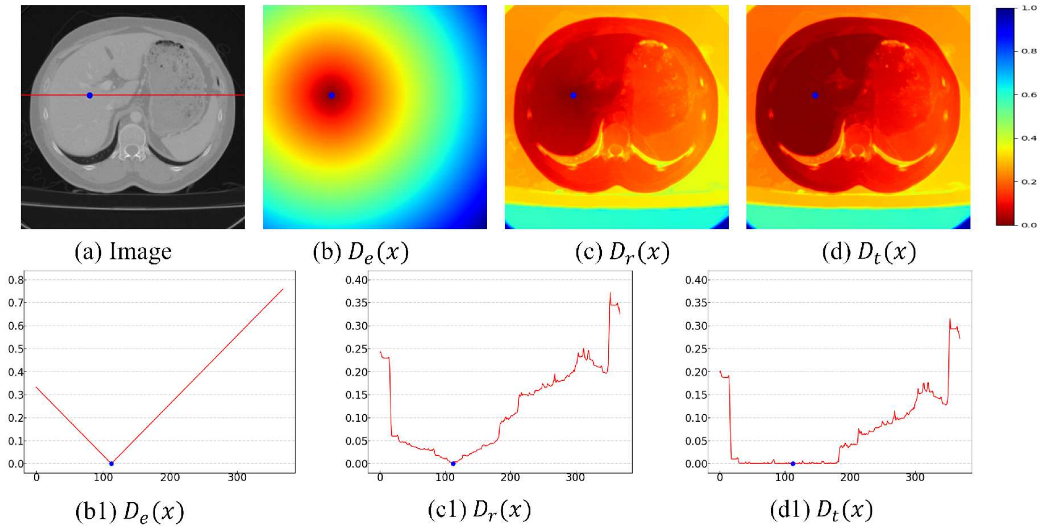
$$d_t(\mathrm{x}) = \begin{cases} \varepsilon_2 + \alpha(\mathrm{x}) \|\nabla S(\mathrm{x})\|^2, & \text{if} \quad \|\nabla S(\mathrm{x})\|^2 < t, \\ \beta_2 \|\nabla S(\mathrm{x})\|^2, & \text{otherwise}. \end{cases} \tag{2.6}$$

Here $S(\mathrm{x})$ is defined in the function (2.4), $\alpha(\mathrm{x}) = 1/(\delta + e^{-\sigma D_e(\mathrm{x})})$ integrates spatial contextual information into the weighting function through a distance-based penalty scheme, where pixel contributions are inversely weighted according to their Euclidean distance D_e from the marked points. Especially, σ controls the decay rate of the penalty function $\alpha(\mathrm{x})$, while δ serves as a regularization constant that ensures numerical stability. The threshold parameter t is used to distinguish between approximately flat regions and boundaries in the image. Based on above observations, we modify the PDE (2.5) as follows:

$$\begin{cases} \|\nabla D(\mathrm{x})\| = d_t(\mathrm{x}), & \text{if} \quad \mathrm{x} \notin \mathcal{M}, \\ D(\mathrm{x}) = 0, & \text{if} \quad \mathrm{x} \in \mathcal{M} \end{cases} \tag{2.7}$$

and denote its solution as the threshold geodesic distance $D_t(\mathrm{x})$.

In order to demonstrate our main motivation, we present a comparison of the geodesic distances as shown in Fig. 1 for $D_e(\mathrm{x})$ defined in Remark 2.1, $D_r(\mathrm{x})$ in PDE (2.5) and $D_t(\mathrm{x})$ defined in PDE (2.7). Specifically, we select the liver as the region of interest in the CT images and designate a blue point x_0 within this ROI. The pseudo-color maps presented in panels (b), (c), and (d) reveal that $D_t(\mathrm{x})$ provides superior discrimination of the ROI compared to the other two distance metrics. To further underscore the rationale for employing $D_t(\mathrm{x})$, we extract a slice from the geodesic map generated using each of the three metrics and display these in the second row of Fig. 1. It is evident that the geodesic distance $D_r(\mathrm{x})$ undergoes a gradual change as the pixel x moves away from the marked point x_0. The geodesic distance $D_r(\mathrm{x})$ exhibits a notable degree of variation even within the same target region. By incorporating the normalized Euclidean distance information $D_e(\mathrm{x})$ into the weight parameter $\alpha(\mathrm{x})$, the threshold geodesic distance $D_t(\mathrm{x})$ can achieve



(a) Image (b) $D_e(x)$ (c) $D_r(x)$ (d) $D_t(x)$

(b1) $D_e(x)$ (c1) $D_r(x)$ (d1) $D_t(x)$

Figure 1: The first row represents the initial image and the different geodesic distance from the image pixel point to the marked point x_0. (a) Original image (blue dot indicates the marked point x_0 in the ROI and red lines represent slices of row). (b) $D_e(\mathrm{x})$ is defined in Remark 2.1. (c) $D_r(\mathrm{x})$ is defined in PDE (2.5). (d) $D_t(\mathrm{x})$ is defined in PDE (2.7). The second row shows slices of three different geodesic distances $D_e(\mathrm{x}), D_r(\mathrm{x})$ and $D_t(\mathrm{x})$ from left to right. Here $\varepsilon_1 = \varepsilon_2 = 10^{-3}, \beta_1 = \beta_2 = 10^3, t = 0.05$ and $\sigma = 20$.

both the discrimination between the target and the background and the consistency of the distance values within the target region.

Remark 2.3. The PDE (2.7) is the boundary value problem associated with Eikonal equation. Here we use the fast sweeping method (FSM) to solve it as did in [31]. In general, the fast sweeping method is an iterative technique that utilizes the upwind difference for discretization and subsequently employs the Gauss-Seidel iteration with the alternating sweeping ordering to resolve the discretized Eikonal equation on a rectangular grid. The time complexity of the algorithm is $\mathcal{O}(m)$, where m represents the number of nodes in the computational grid, as demonstrated in references [50, 62]. Consequently, the algorithm exhibits enhanced efficiency and precision in practical applications.

2.2 Soft threshold dynamics activation function

The application of deep learning technology in the field of medical imaging has recently attracted considerable attention. The Unet architecture [39], has made a significant impact in the field of image segmentation, based on the architectural principles of FCNs and encoder-decoder architecture. The architecture comprises the repeated application of two 3×3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU), and a 2×2 max pooling operation with a stride of two for downsampling. In the upsampling path, a transposed convolution with a kernel of size 2×2 is used to reduce the number of feature channels. The Unet architecture also incorporates skip connections, which facilitate the transfer of features from the contracting path to the expanding path, thereby compensating for the loss of spatial information that occurs during downsampling. Consequently, the Unet architecture exhibits both expediency and precision in image segmentation [46,59].

In the context of image segmentation using Unet architecture, the sigmoid activation function is typically employed in the final layer to generate neural network predictions. However, this choice of activation function may not be optimal for feature extraction due to its inherent limitations, particularly the absence of a prior term which could potentially reduce the model's performance and interpretability. In light of this observation, Liu *et al.* [27] proposed to use the solution of the following optimization problem:

$$\begin{aligned}\min_{u(\mathrm{x})\in\{0,1\}}&\tau\int_\Omega \big(u(\mathrm{x})\ln u(\mathrm{x})+\big(1-u(\mathrm{x})\big)\ln\big(1-u(\mathrm{x})\big)\big)\mathrm{dx}\\&-\int_\Omega\big(o_1(\mathrm{x})u(\mathrm{x})+o_2(\mathrm{x})\big(1-u(\mathrm{x})\big)\big)\mathrm{dx}+\mu\int_\Omega\|\nabla u(\mathrm{x})\|_2\mathrm{dx}\end{aligned}\tag{2.8}$$

to replace the sigmoid activation function in the final layer in order to address the two-phase image segmentation problem. Here $o_1(\mathrm{x})$ and $o_2(\mathrm{x})$ are the feature outputs from the neural networks as shown in [27], τ and μ are the balance parameter and the indicator function $u(\mathrm{x})$ is defined by

$$u(\mathrm{x})=\begin{cases}0, & \text{if} \quad \mathrm{x}\in\mathcal{F},\\ 1, & \text{if} \quad \mathrm{x}\in\mathcal{B}.\end{cases}$$

Here $\mathcal{F}$ and $\mathcal{B}$ denote the region of foreground and background in the image, respectively. The principal numerical challenge in resolving problem (2.8) is the efficient treatment of the nonsmooth term

$$\int_\Omega\|\nabla u(\mathrm{x})\|_2\mathrm{dx}:=\int_\Omega\sqrt{\big(\nabla_{x_1}u(\mathrm{x})\big)^2+\big(\nabla_{x_2}u(\mathrm{x})\big)^2}\mathrm{dx}.$$

For dealing with the nonsmooth term, one classic method is the operator splitting methods [8,15], which include such as the alternating direction method of multipliers (ADMM) [57], the primal dual method [7], the Douglas-Rachford splitting method [19], and so forth. Recently, several works [12,49,52] proposed to approximate the nonsmooth term $\|\nabla u(\mathrm{x})\|_2$ in the image segmentation by a smoothed version via using threshold dynamics [23,26,52] and the convex relaxed method for $u(\mathrm{x})$. we adopt this established scheme, whereby the problem formulation presented in (2.8) can be expressed as follows:

$$\begin{aligned}\min_{u(\mathrm{x})\in[0,1]}&\tau\int_\Omega \big(u(\mathrm{x})\ln u(\mathrm{x})+\big(1-u(\mathrm{x})\big)\ln\big(1-u(\mathrm{x})\big)\big)\mathrm{dx}\\&-\int_\Omega\big(o_1(\mathrm{x})u(\mathrm{x})+o_2(\mathrm{x})\big(1-u(\mathrm{x})\big)\big)\mathrm{dx}\\&+\mu\int_\Omega u(\mathrm{x})G_{\bar\eta}(\mathrm{x})*\big(1-u(\mathrm{x})\big)\mathrm{dx},\end{aligned}\tag{2.9}$$

where the definition $G_{\bar\eta}(\mathrm{x})$ is similar to the definition of $G_\eta(\mathrm{x})$ in (2.4). In the problem (2.9), the last term is nonlinear. In order to propose an efficient numerical method, we use the linearization scheme as in [23, 26, 27, 52] to transform problem (2.9) into the following approximate optimization problem:

$$
\begin{aligned}
&-\int_\Omega \big(o_1(\mathrm{x})u(\mathrm{x})+o_2(\mathrm{x})(1-u(\mathrm{x}))\big)\mathrm{dx}+\mu\int_\Omega u(\mathrm{x})G_{\bar\eta}(\mathrm{x})*(1-2u^k(\mathrm{x}))\mathrm{dx}\\
=&\frac{\exp\Big(\big(o_1(\mathrm{x})-o_2(\mathrm{x})-\mu G_{\bar\eta}(\mathrm{x})*(1-2u^k(\mathrm{x}))\big)/\tau\Big)}{1+\exp\Big(\big(o_1(\mathrm{x})-o_2(\mathrm{x})-\mu G_{\bar\eta}(\mathrm{x})*(1-2u^k(\mathrm{x}))\big)/\tau\Big)}\\
=:&\mathcal{S}\big(u^k(\mathrm{x}),o_1(\mathrm{x}),o_2(\mathrm{x})\big).
\end{aligned}
\tag{2.10}
$$

Based on the above discussion, the soft threshold dynamics activation function can be summarized as the following Algorithm 1.

Algorithm 1: Soft Threshold Dynamics Activation Function.

```
(STDAF) Input: The feature $o_1(\mathrm{x})$ and $o_2(\mathrm{x})$ and set parameters $\tau$ and $\eta$.
Initialization:
    $u^0(\mathrm{x})=\frac{\exp(o_1(\mathrm{x}))}{\exp(o_1(\mathrm{x}))+\exp(o_2(\mathrm{x}))}$.
for $k=0,1,2,\ldots$ do
    1: Compute the solution $u^{k+1}(\mathrm{x})$ by (2.10).
    2: Convergence check. If it is converged, end the algorithm.
end
Return: Output the result $\bar{u}(\mathrm{x}):=u^{k+1}(\mathrm{x})$.
```

Remark 2.4. In Algorithm 1, the feature representations $o_1(\mathrm{x})$ and $o_2(\mathrm{x})$ correspond to the intermediate outputs generated between the 1×1 convolutional layer and the final activation function, whose specific architectural configuration will be elaborated in the network framework illustrated in Fig. 2. In (2.8) or (2.9), if we set the parameters $\tau=1$ and $\mu=0$, then the activation function (2.10) can degenerate into the classic softmax activation function as

$$
\bar{u}(\mathrm{x})=\frac{\exp(o_1(\mathrm{x}))}{\exp(o_1(\mathrm{x}))+\exp(o_2(\mathrm{x}))}.
$$

In addition, to the optimization problem (2.8), we can add some prior information such as the L^2-norm term or the edge detection function to improve the activation function (2.10). In particular, in the later numerical comparisons we note that a number of iterations of 10 yields satisfactory numerical results, so we set $k=10$ for Algorithm 1 in the numerical implementations.

2.3 Double-domain driven Unet method (DDUM)

In the network stage, the objective is to integrate the geodesic distance with deep learning to achieve a more precise delineation of the target boundaries. In this approach, we adopt the concept of a dual-branch network [21], which employs two independent Unets: one to learn information about the image domain and another for learning information about the geodesic domain. In the Unet network architecture Fig. 2, once we get the feature $(\check{o}_1(\mathrm{x}),\check{o}_2(\mathrm{x}))$ and $(\bar{o}_1(\mathrm{x}),\bar{o}_2(\mathrm{x}))$ from the image domain and the threshold geodesic domain, by using Algorithm 1, we can get the predicted probability distribution as

$$F_{\theta_1}(\mathrm{x})=\mathcal{S}\big(\check{o}_1(\mathrm{x}),\check{o}_2(\mathrm{x})\big),\quad D_{\theta_2}(\mathrm{x})=\mathcal{S}\big(\bar{o}_1(\mathrm{x}),\bar{o}_2(\mathrm{x})\big). \tag{2.11}$$

Remark 2.5. With regard to the formulas (2.11), it is important to note that the geodesic domain is capable of learning distance information. This implies that the distance values are typically smaller within the ROI and larger in the background. The geodesic distance metric is used to quantify the proximity of each pixel to marked points, employing Riemannian metrics [18]. Therefore, $D_{\theta_2}(\mathrm{x})$ acquires information from the geodesic domain while simultaneously limiting the distance of each point to the marked point set $\mathcal{M}$. Currently, the image domain encompasses structural information in the foreground and background.

To help the network learn more discriminative features of segmentation objects, we consider the following loss function:

$$\begin{aligned}(\theta_1^*,\theta_2^*)=\underset{\theta_1,\theta_2}{\operatorname{argmin}}&\frac{1}{2}\|F_{\theta_1}(\mathrm{x})-Z(\mathrm{x})\|^2+\frac{1}{2}\big\|D_{\theta_2}(\mathrm{x})-\big(1-Z(\mathrm{x})\big)\big\|^2\\&+\lambda\|F_{\theta_1}(\mathrm{x})\|_1+\lambda\|D_{\theta_2}(\mathrm{x})\|_1+\omega\langle F_{\theta_1}(\mathrm{x}),D_{\theta_2}(\mathrm{x})\rangle,\end{aligned} \tag{2.12}$$

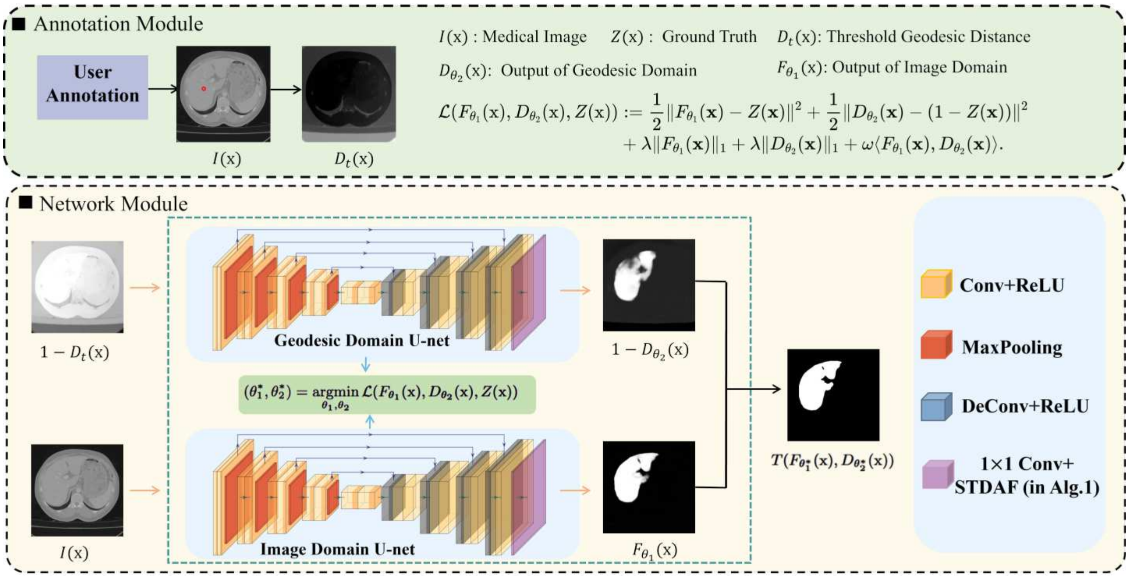

Annotation Module
User Annotation
$I(\mathrm{x})$
$D_t(\mathrm{x})$
$I(\mathrm{x})$: Medical Image
$Z(\mathrm{x})$: Ground Truth
$D_t(\mathrm{x})$: Threshold Geodesic Distance
$D_{\theta_2}(\mathrm{x})$: Output of Geodesic Domain
$F_{\theta_1}(\mathrm{x})$: Output of Image Domain
$\mathcal{L}(F_{\theta_1}(\mathrm{x}),D_{\theta_2}(\mathrm{x}),Z(\mathrm{x})):=\frac{1}{2}\|F_{\theta_1}(\mathbf{x})-Z(\mathbf{x})\|^2+\frac{1}{2}\|D_{\theta_2}(\mathbf{x})-(1-Z(\mathbf{x}))\|^2$
$+\lambda\|F_{\theta_1}(\mathbf{x})\|_1+\lambda\|D_{\theta_2}(\mathbf{x})\|_1+\omega\langle F_{\theta_1}(\mathbf{x}),D_{\theta_2}(\mathbf{x})\rangle.$
Network Module
$1-D_t(\mathrm{x})$
Geodesic Domain U-net
$1-D_{\theta_2}(\mathrm{x})$
$(\theta_1^*,\theta_2^*)=\underset{\theta_1,\theta_2}{\operatorname{argmin}}\,\mathcal{L}(F_{\theta_1}(\mathrm{x}),D_{\theta_2}(\mathrm{x}),Z(\mathrm{x}))$
$T(F_{\theta_1^*}(\mathrm{x}),D_{\theta_2^*}(\mathrm{x}))$
$I(\mathrm{x})$
Image Domain U-net
$F_{\theta_1}(\mathrm{x})$
Conv+ReLU
MaxPooling
DeConv+ReLU
1×1 Conv+ STDAF (in Alg.1)

Figure 2: The network architecture is constructed by concatenating the information from the image domain and the threshold geodesic distance domain. Under the shared loss function, the segmentation result can be learned separately from the image domain and the geodesic distance domain. For the purpose of convenient encoding of the network, images are normalised in the range [0,1].

where λ and ω are two empirically chosen parameters, and $Z(\mathrm{x})$ is the ground truth. The first two fidelity terms impose a penalty on the discrepancy between the predicted results and the labelled data. The third term provides an approximation of the sparsity of the network result $F_{\theta_1}(\mathrm{x})$ in the image domain and $D_{\theta_2}(\mathrm{x})$ in the geodesic domain. In particular, the geodesic domain result is primarily concerned with distance information, which explains why the prediction result of $D_{\theta_2}(\mathrm{x})$ is entirely antithetical to that of $F_{\theta_1}(\mathrm{x})$. Consequently, the fourth term employs an inner product to quantify the discrepancy between the predicted results, specifically $F_{\theta_1}(\mathrm{x})$ and $D_{\theta_2}(\mathrm{x})$. In this case, the simultaneous inclusion of information from both the image domain and the geodesic domain in the optimization problem (2.12) facilitates more effective edge detection, thereby leading to a satisfactory segmentation result.

Two networks are employed to update the network parameter θ_1 and θ_2 through a process of backpropagation. This is constrained by the spatial regularization and the same loss function. Each individual branch consists of approximately five convolutional layers serving as encoders and five deconvolutional layers serving as decoders. The batch normalization layer is replaced with the instance normalization layer, and the number of network feature layers is reduced by a quarter to achieve a balance between performance, memory consumption, and time costs [30]. During the testing phase, the image to be evaluated and the threshold geodesic distances are provided as inputs to the network. Once the network parameters have been trained, the network results can be obtained for both the image domain and the geodesic domain, namely $F_{\theta_1^*}(\mathrm{x})$ and $D_{\theta_2^*}(\mathrm{x})$. Upon completion of the training phase for the network parameters, the segmentation can be obtained using the following formula:

$$T\big(F_{\theta_1^*}(\mathrm{x}),D_{\theta_2^*}(\mathrm{x})\big)=\begin{cases}0, & \text{if} \quad \bar{\alpha}\big(1-F_{\theta_1^*}(\mathrm{x})\big)+(1-\bar{\alpha})D_{\theta_2^*}(\mathrm{x})\le 0.5,\\ 1, & \text{if} \quad \bar{\alpha}\big(1-F_{\theta_1^*}(\mathrm{x})\big)+(1-\bar{\alpha})D_{\theta_2^*}(\mathrm{x})>0.5.\end{cases} \tag{2.13}$$

Here $\bar{\alpha}\in[0,1]$ is the convex combination parameter and we set $\bar{\alpha}=0.5$ in this paper for the selective segmentation problem.

3 Experiments and results

In this section, we introduce the implementation details, experimental datasets and evaluation metrics.

3.1 Implementation details

In this paper, we do not design a specialized network framework. To facilitate network decoding, we are normalized and scaled to the input image and the geodesic distance to the range [0,1], and invert the geodesic domain images during input. Specifically, the inputs and outputs of the geodesic domain are actually $1-D_t(\mathrm{x})$ and $1-D_{\theta_2}(\mathrm{x})$, respectively. This operation is only for the convenience of network learning, and the loss function still restricts the distance $D_{\theta_2}(\mathrm{x})$. For the parameters within the geodesic distance calculation, we follow the setting as in [47], i.e. $\varepsilon_1=\varepsilon_2=10^{-3}$ and $\beta_1=\beta_2=10^{-3}$. For other parameters, we set $\delta=10^{-3}$ and $\eta=0.5$. All experiments are conducted by using an AMD Ryzen 7 5800H CPU within the PyTorch framework. We perform 300 epochs of network training and utilize the Adam algorithm for optimization, with a mini-batch size of four and a weight decay of 10^{-4}. In addition, to show the robustness of the proposed DDUM, we compare it with several selective segmentation methods such as DEXTR [36], MIDeepSeg [30], IGMedSeg [48] and the deep learning method such as SDU-Net [58] in several standard medical image datasets.

3.2 Datasets

The numerical comparisons are based on three datasets: the CT liver dataset[1)], the MR left atrium dataset[2)], and the CT lung dataset[3)]. In the case of the liver dataset, all images are reconstructed using a matrix of dimensions 512×512, with an in-plane voxel resolution of between 0.7 and 0.8 mm and an interslice distance (ISD) of between 3 and 3.2mm. A total of 15 patients are selected, comprising 1562 slices in total, of which 540 are used for training and testing. The MR left atrium dataset is acquired on a 1.5 T Achieva scanner. The sequence acquires a non-angulated volume with the objective of covering the entire MR left atrium, with a voxel resolution of $1.25\times 1.25\times 2.7\,\mathrm{mm}^3$. The datasets are acquired during free breathing with respiratory gating and at end-diastole with ECG gating. A total of 15 patients are selected, comprising 710 slices, for training and testing on 200 slices from five patients. Regarding the lung data set, eight patients are selected for training, comprising 227 slices, while three patients are selected for testing, with 125 slices.

3.3 Annotation strategy

In the annotation stage, the objective is to compute the geodesic distance between the pixel points and the marked points. In general, the selection of these marked points should be as minimally interactive as possible while simultaneously meeting the demand for interactive segmentation of the ROI. Annotation strategies commonly employed in the field include line annotation [53], bounded box annotation [36], and point annotation [30,32,40,41,48]. Among these approaches, point annotation has emerged as a more efficient and practical solution, offering advantages in simplicity and speed compared to line and bounding box annotations. Roth *et al.* [41] introduced a methodology utilizing extremal points of target regions for annotation, though this technique presents limitations when applied to complex image analysis. The deep extreme cut method (DEXTR) [32] required that the marked points be on the target boundary, particularly

[1)] https://chaos.grand-challenge.org/

[2)] https://www.kaggle.com/datasets/adarshsng/heart-mri-image-dataset-left-atrial-segmentation

[3)] http://medicalsegmentation.com/covid19/

in the case of complex medical images, where a greater number of marked points are necessary. This strategy obviously required more precise marked techniques since it needed to determine how to mark the points on the boundary. MIDeepSeg, proposed by Luo *et al.* [30], required the marked points within the target region near the extreme points to allow a few pixels to be enlarged to obtain a bounding function that includes the target. Interactive segmentation framework based on a graph convolutional network (IGMedSeg) [48] required that the marked points were on the boundary and generally needed more marked points. However, these existing annotation methodologies often necessitate either an extensive number of annotation points or precise edge localization, which fundamentally contradicts the primary objective of point annotation to provide a simple and efficient annotation process.

In contrast to the aforementioned methods, our proposed DDUM demonstrates superior efficiency by requiring a significantly reduced number of annotation points while maintaining implementation simplicity. In the case of relatively simple segmentation targets, only a single marked point is required per image. The threshold function $d_t(x)$ defined in the formula (2.6) causes the geodesic distance within the ROI to approximate a constant value, thus exhibiting the property of piecewise constancy within the ROI. This mathematical property enables flexible placement of annotation points, as they can be positioned arbitrarily within the ROI without compromising segmentation accuracy. Nevertheless, based on empirical observations, we recommend positioning the annotation point near the centroid of ROI to optimize computational efficiency and segmentation precision. For more complex ROI segmentation scenarios that involve intricate structures or larger objects, our framework accommodates the incorporation of additional annotation points. This adaptive approach allows for enhanced segmentation accuracy while maintaining the method's fundamental simplicity. The number of required points scales proportionally with the target's complexity, ensuring optimal results across varying difficulty levels while preserving the method's core advantage of minimal annotation requirements. Fig. 3 displays examples of labelling in the ROI for different methods in one slice for three testing datasets. Furthermore, we present the average number of marked points required for the methods compared in different datasets, as shown in Table 1. Compared to alternative methodologies, DDUM necessitates the marking of a mere one to two points per slice. Even when dealing with multiple segmentation targets, a substantial number of markers is not required.

Table 1: The average number of marked points per slice for different methods in different datasets.

Dataset	The number of marked points (Slice)			
	DEXTR	MIDeepSeg	IGMedseg	DDUM
Liver	6.87	4.26	7.38	1.75
Left Atrium	4.38	4.08	5.62	1.91
Lung	9.86	5.12	10.27	2.68

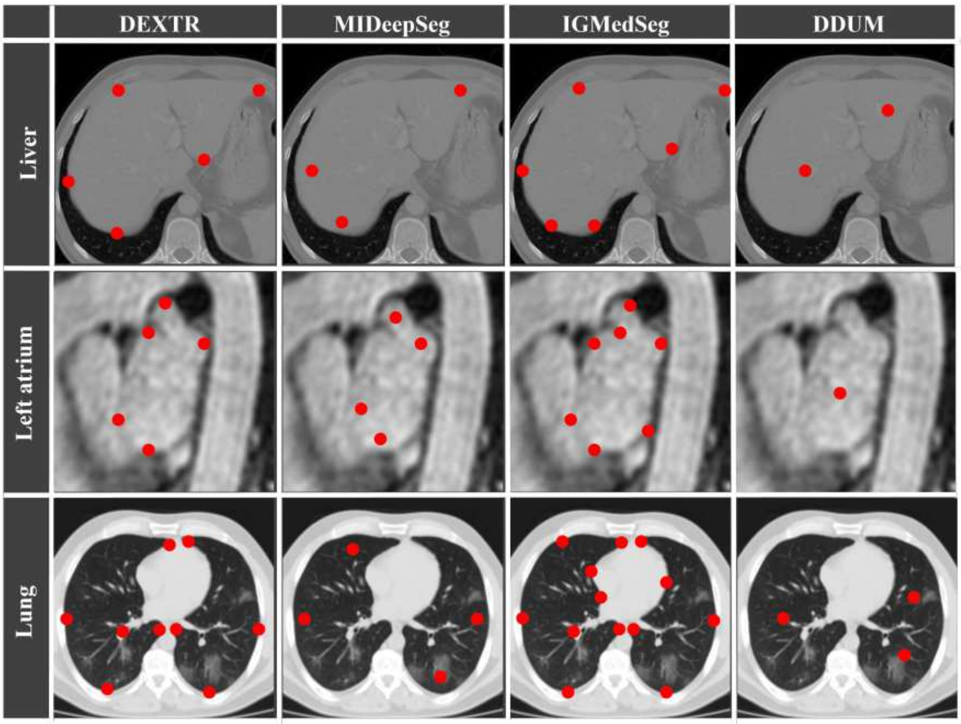


Figure 3: Some examples illustrate the number of marked points required for different methods across a range of datasets. It should be noted that ROI with complex structures require a greater number of marked points. The red points indicate the locations of the marked points.

3.4 Evaluation metrics

In order to evaluate the segmentation result, we select four evaluation metrics: Dice coefficient (Dice), Jaccard similarity (JS), Hausdorff distance (HD), and Average symmetric surface distance (ASSD) as follows:

$$\begin{aligned}&\text{Dice}=\frac{2\times|S_1\cap S_2|}{|S_1|+|S_2|},\quad \text{JS}=\frac{|S_1\cap S_2|}{|S_1\cup S_2|},\\&\text{HD}=\max\left\{\sup_{q\in\partial S_2}\inf_{p\in\partial S_1}d_e(q,p),\sup_{p\in\partial S_1}\inf_{q\in\partial S_2}d_e(p,q)\right\},\\&\text{ASSD}=\frac{1}{|S_1|+|S_2|}\left(\sum_{q\in S_2}d_e(q,S_1)+\sum_{p\in S_1}d_e(p,S_2)\right),\end{aligned}\tag{3.1}$$

where $d_e(\cdot,\cdot)$ is the Euclidean distance, $|\cdot|$ represents the number of pixels in the image domain, S_1 represents the target object region after the segmentation, S_2 represents the target object region of labelling, ∂S_1 represents the dividing boundary, and ∂S_2 represents the real boundary. It is obvious that larger (respectively smaller) values of the Dice and JS (respectively HD and ASSD) mean a more accurate segmentation result.

3.5 Single-organ segmentation

Here we perform a comprehensive series of experiments designed to evaluate the segmentation performance on CT liver and MR left atrium datasets. The visualization results of two different slices for these experiments are presented in Fig. 4, respectively. Our analysis reveals significant performance variations among different segmentation approaches. Specifically, SDU-Net [58] demonstrates limitations in accurately identifying boundary information when prior information is unavailable. Furthermore, our experimental results indicate that DEXTR [32] encounters substantial challenges in handling complex anatomical structures, particularly in precisely delineating the most distal points of intricate shapes. Compared to MIDeepseg [30] and IGMedseg [48], the results of the segmentation based on DDUM are found to be more closely aligned with the ground truth (GT). The quantitative evaluation presented in Table 2 provides evidence to support the superiority of DDUM. It is evident that the considerable deviation from the mean can be attributed to the presence of inter-individual variations between patients and the existence of weak boundaries in organ slices. This suggests that the accuracy of other methods is significantly influenced by the quality of the segmented image, resulting in considerable fluctuations in the segmentation error. In contrast, a smaller deviation indicates a greater robustness of the proposed DDUM in the segmenting of complex medical images.

In order to statistically assess the segmentation accuracy distribution across different methods applied to the CT liver dataset, a bootstrapping procedure is implemented. This method involves repeated random sampling with replacement, where 10 slices are extracted in each iteration. For each bootstrap sample, the mean values of both Dice and

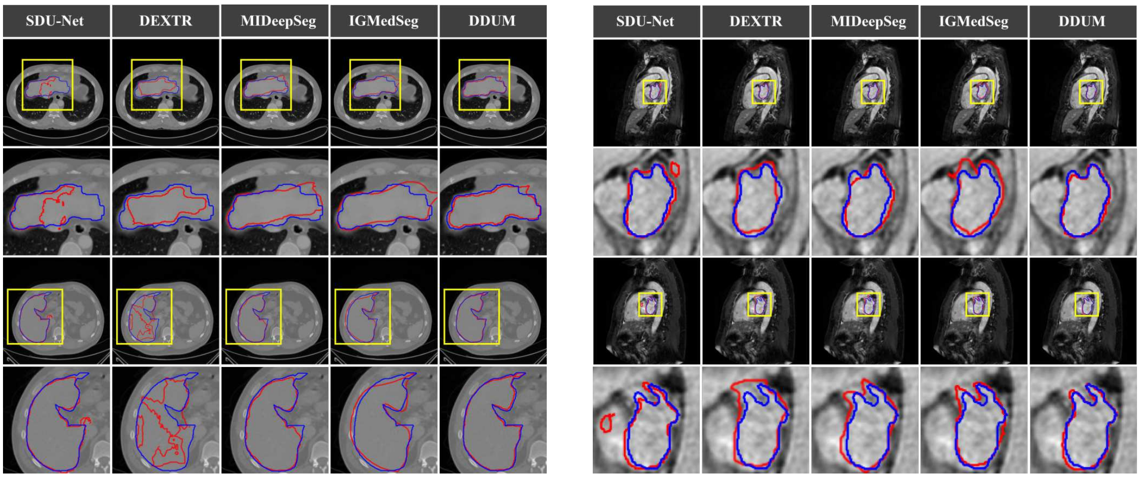


Figure 4: The results of the segmentation for two different slices of the CT liver (left) and the left atrium of MRI (right). The second and fourth rows are enlarged within the yellow-boxed area of the first and third rows. The blue curve represents the boundary of the ground truth, and the red curve represents the boundary of the segmentation result from different methods.

JS metrics are computed to quantify segmentation performance. This bootstrapping process is independently repeated 1,000 times to ensure robust statistical estimation of the accuracy distribution. As illustrated in Fig. 5, the results of the distribution indicate that the indicate the first and third quartile, and the whiskers above and below the box show the maximum and minimum values. In particular, the rounded points above and below the box represent outliers within the 5% and 95% ranges, respectively. As evidenced by the data presented in Fig. 6, the segmentation quality generated by DDUM is superior to that of other methods, as indicated by a higher median and a tighter distribution. The number of parameters and the mean inference time per slice for various methods are examined in Table 3. It is evident that, despite DEXTR having the fewest parameters and the lowest training time, its inference time is relatively high because of its complex annotation strategy. Compared to other methods, DDUM does not have the fewest parameters. Nevertheless, it attains superior segmentation metrics with reduced inference times.

Table 2: A comparative analysis of the proposed DDUM with other methods is presented for the CT liver and left atrium MR data sets, where the optimal values are marked in bold.

Dataset	Liver			
Method	Dice	JS	HD	ASSD
SDU-Net	0.7107±0.1934	0.6150±0.1849	81.7737±25.9399	14.9464±13.5453
DEXTR	0.8662±0.1138	0.7753±0.1744	43.3546±4.6453	7.3271±0.3618
MIDeepSeg	0.8778±0.0522	0.8183±0.1871	27.9066±5.9836	2.0203±0.1715
IGMedSeg	0.9055±0.1417	0.8248±0.2688	27.5898±6.8696	2.1321±0.2679
DDUM	**0.9302±0.0241**	**0.8697±0.0424**	**16.7778±5.5260**	**0.5526±0.1936**
Dataset	Left atrium			
Method	Dice	JS	HD	ASSD
SDU-Net	0.8193±0.1352	0.7516±0.1648	13.3498±12.4628	2.4605±4.1576
DEXTR	0.8709±0.2357	0.7719±0.2961	6.6349±17.1627	1.9867±5.4432
MIDeepSeg	0.9066±0.0715	0.8364±0.1246	5.3176±8.0458	0.2350±0.1143
IGMedSeg	0.8826±0.0205	0.7929±0.0287	5.4684±7.4328	1.7514±3.4767
DDUM	**0.9385±0.0082**	**0.8816±0.0137**	**2.8801±3.1078**	**0.8304±0.5997**

Table 3: The number of parameters, the average inference time per slice and the average training time per epoch for each method.

Dataset	Metrics		
	Parameter count	Inference time	Training time
SDU-Net	2468449	1.12 (s)	7778.87 (s)
DEXTR	155616	11.71 (s)	742.68 (s)
MIDeepSeg	34872109	3.87 (s)	4372.07 (s)
IGMedSeg	133277848	9.26 (s)	5464.14 (s)
DDUM	5492354	1.34 (s)	1384.12 (s)

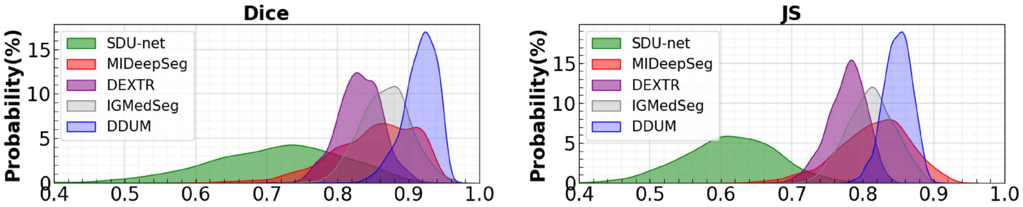


Figure 5: Bootstrap resampling is employed to generate the distribution plots of the various methods with respect to the values of Dice and JS on the CT liver dataset. The testing dataset is described in detail in Section 3.2.

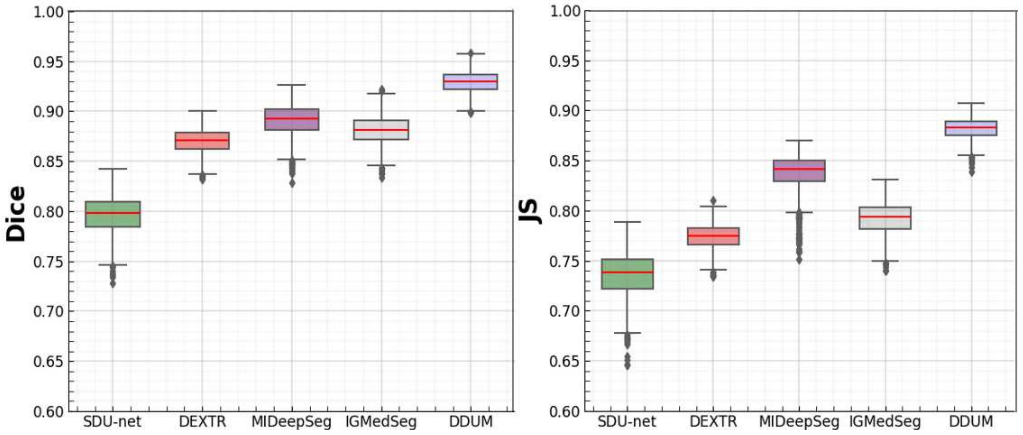


Figure 6: Box plots of Dice and JS for different methods on MR left atrium dataset on the testing dataset on 200 slices from five patients as described in Section 3.2.

3.6 Multi-organ segmentation

In order to demonstrate the feasibility of segmenting multi-organ structures with identical architectures, we conduct a series of comparisons utilizing a multiorgan CT lung dataset. As DEXTR [36] and IGMedSeg [48] are only capable of segmenting a single target, we employ a method whereby a single target is divided and the resulting segments are then aggregated to obtain the overall segmentation result. It is important to note that, in order to validate the generalization ability of the different models, only the transverse plane data is utilized during the training process. For testing purposes, the models are evaluated across the transverse, sagittal and coronal planes separately.

The results of the visual segmentation for the lung field are presented in Fig. 7 from three distinct viewpoint planes. It is evident that comparative methodologies demonstrate instances of undersegmentation in the transverse and coronal planes, as well as instances of oversegmentation in the sagittal plane. For comparison purposes, it is shown that DDUM is a more robust method for segmentation of the lung field, particularly in areas of complex structure and in regions of the image that are difficult to process, such as corners (see the zoom image for the sagittal plane). A quantitative assessment of the segmentation results for each method is provided in Fig. 8. It can be observed that the

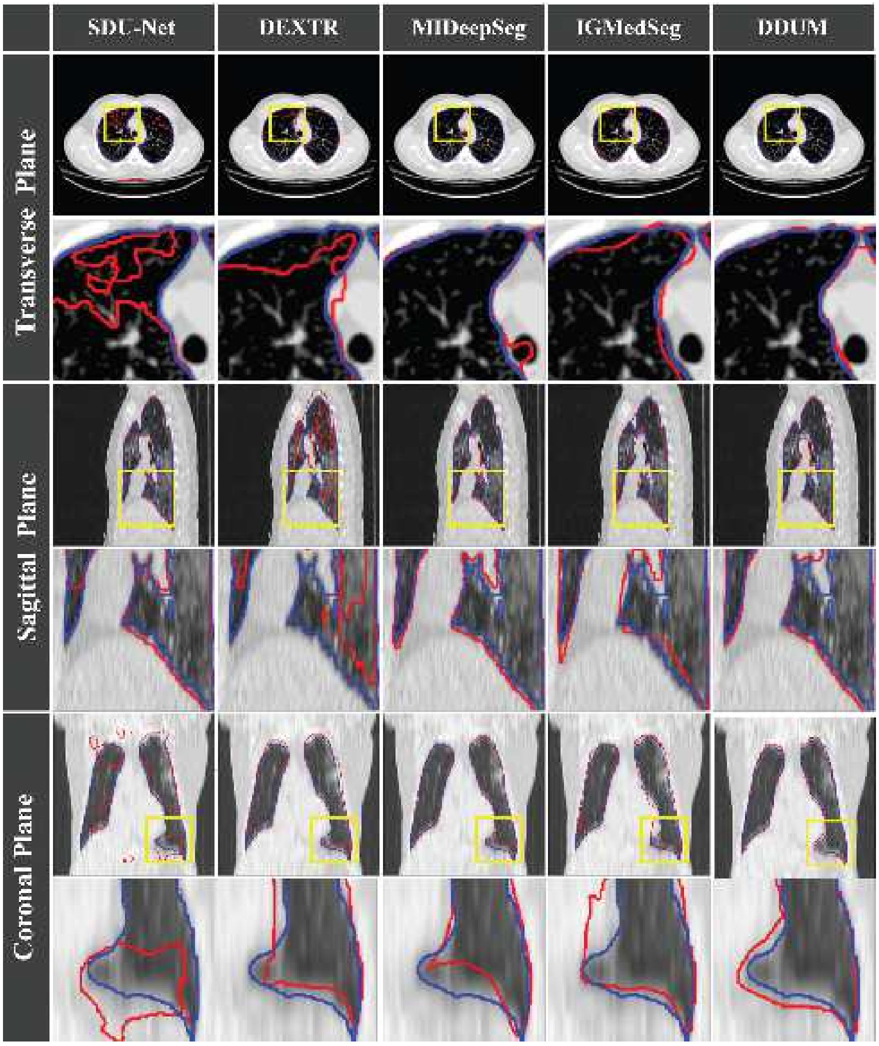


Figure 7: The segmentation results of different methods in lung lobes. The second, fourth and sixth rows are the enlarged images within the yellow boxed area in the first, third and fifth rows.

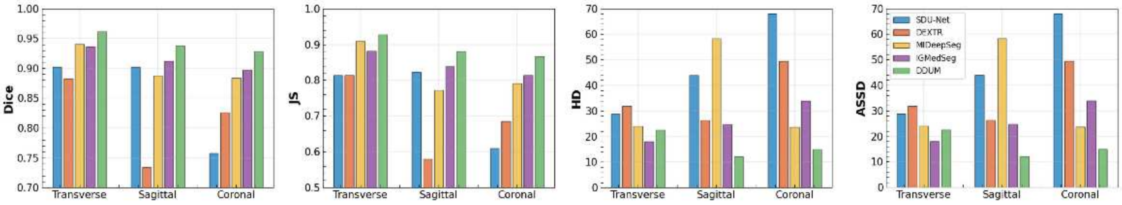


Figure 8: The performance metrics for lung lobe segmentation on different anatomical planes.

superior segmentation results in the transverse view, as compared to those in the sagittal and coronal views, are a consequence of the training phase utilizing transverse plane data. The SDU-Net displays reduced generalization capabilities due to the absence of prior information from the marked points. DEXTR exhibits inferior segmentation performance in images with complex edges. MIDeepSeg and IGMedSeg demonstrate some levels of generalization in the sagittal and coronal views, but they still fall below DDUM.

4 Ablation studies

This section presents a series of ablation experiments, conducted with the objective of investigating the impact of various factors on the experimental results and the threshold geodesic distance. These factors include the utilization of disparate encoding methodologies, the number and location of marked points, the dimensions of the training set, the parameters in the geodesic distance equation (2.6), the activation function (2.8), and the loss function (2.12). For the sake of convenience, all ablation experiments are conducted using the liver dataset.

4.1 Threshold geodesic distances and marked points

To further validate the robustness of our proposed threshold geodesic distance $D_t(\mathrm{x})$, we replace it by other two geodesic distances such as the edge-weighted geodesic distance $D_r(\mathrm{x})$ defined in (2.5) and the exponential geodesic distance $D_w(\mathrm{x})$[4] introduced by Luo *et al.* [30] in the proposed network framework. In addition, to evaluate the correlation between the number and position of the marked points and the segmentation accuracy, we use five different marking strategies: marking a single point at the center, marking a single point at the center along with an additional point for supplemental information, randomly marking a single point, randomly marking two points, and randomly marking three points in the region of liver. These five strategies are abbreviated as C_1, C_1+, R_1, R_2, R_3 as shown in the first row of Fig. 9.

The second to fourth rows of Fig. 9 present the results of the segmentation of the liver region employing different geodesic distances $(D_w(\mathrm{x}),\ D_r(\mathrm{x}),\ D_t(\mathrm{x}))$ under different labelling strategies. It is evident that geodesic distances, specifically $D_w(\mathrm{x})$ and $D_r(\mathrm{x})$, are significantly affected by the different labelling strategies. In particular, when points are randomly masked, there are instances of under-segmentation or over-segmentation. In contrast, the proposed geodesic distance is less susceptible to the influence of the labelling strategy. In other words, the proposed DDUM is more likely to successfully segment complex regions, such as the lobular and hilar regions of the liver. The respective geodesic distances based on the marked points of Fig. 9 are represented by Fig. 10. It is readily apparent that the geodesic distances calculated by our proposed $D_t(\mathrm{x})$ exhibit

[4] Here exponentialized g geodesic distance $D_w(\mathrm{x})$ in [30] is defined by $D_w(\mathrm{x}) := \min_{\mathrm{y}\in\mathcal{M}} e^{-d_w(\mathrm{x},\mathrm{y})}$, where $d_w(\mathrm{x},\mathrm{y}) = \min_{\gamma\in\mathcal{P}(\mathrm{x},\mathrm{y})} \int_0^1 \sqrt{\nabla I(\gamma(t))\cdot\gamma'(t)/\|\gamma'(t)\|}\mathrm{d}t$

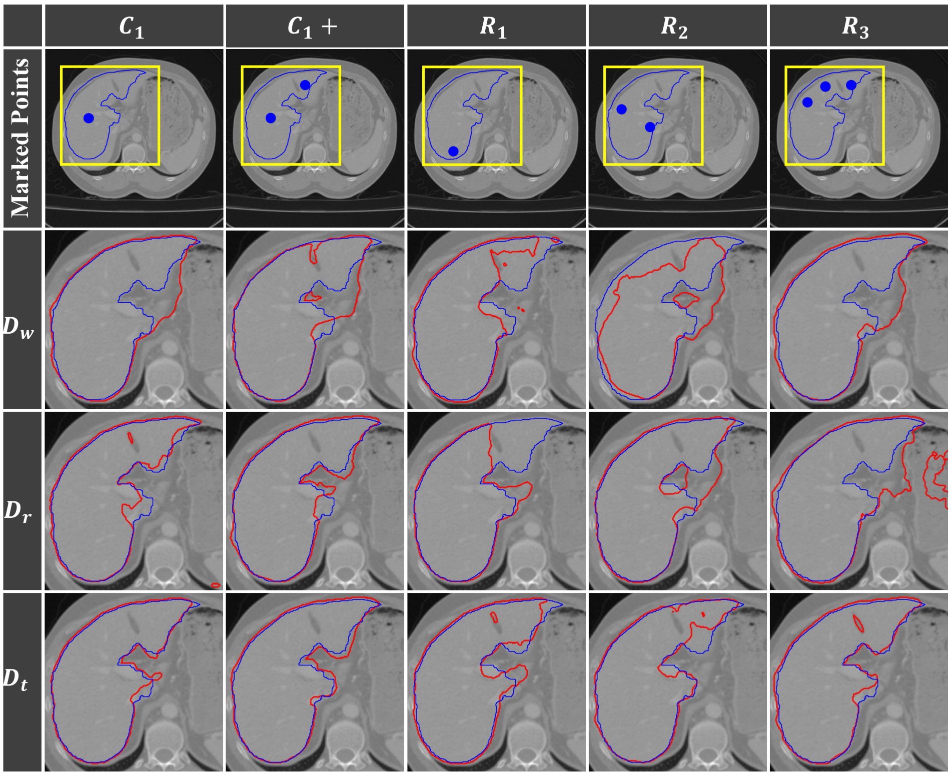


Figure 9: The first row shows different marked points inside the GT area. The second to the fourth rows show the segmentation results with different geodesic distances in the proposed network framework. Here the blue curve is the boundary of the ground truth, and the red curve is the boundary of the segmentation result. Note: The yellow box represents the ROI, and the blue points indicate the marked points.

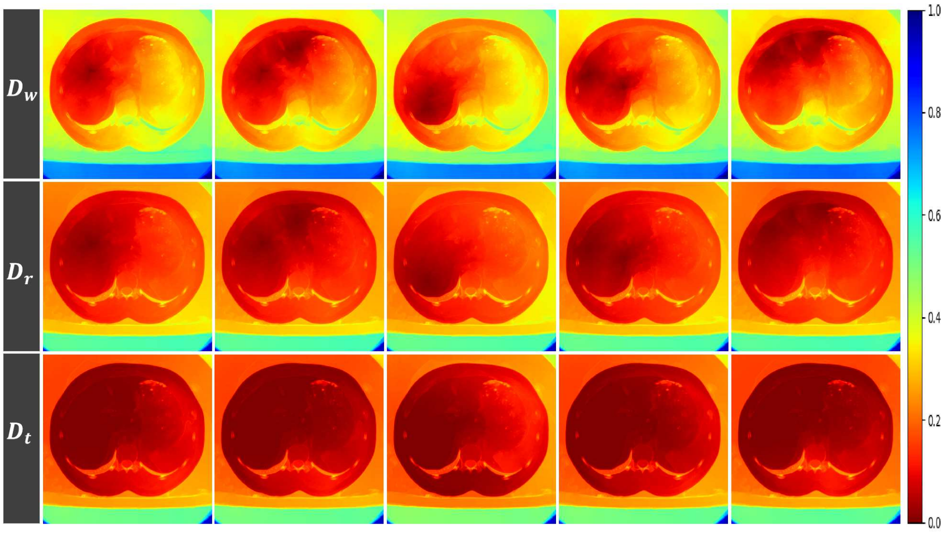


Figure 10: The geodesic distances corresponds to different marking strategies C_1, C_1+, R_1, R_2, R_3 as shown in the first row of Fig. 9.

minimal variation on the pseudo-color map as the marked points change. This observation supports the assertion that our proposed geodesic distance is robust. The line graphs in Fig. 11 demonstrate the trend of segmentation results with different marked strategies, thereby confirming the robustness of DDUM for variations in labelling strategies. For the sake of convenience, the labelling strategy is fixed during the experiments, whereby the target is marked at its center.

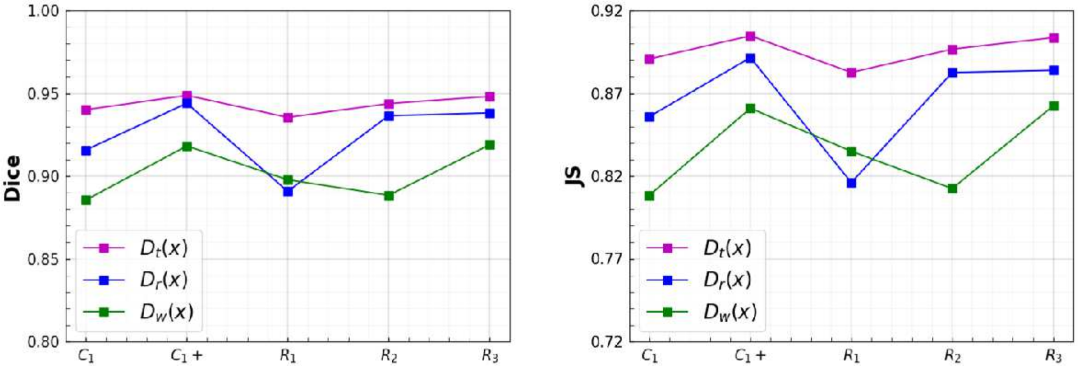


Figure 11: Each point in the line indicates the Dice (left) and JS (right) on the trend of segmentation results with different labelling strategies.

4.2 Influence of geodesic domain

The proposed network architecture employs a dual-branch Unet, integrating both image domain and threshold geodesic domain information, represented by $D_t(\mathrm{x})$, as the net inputs. In order to validate the rationale behind utilizing the geodesic domain as an input to the Unet, we conduct a series of experiments in which we replace $D_t(\mathrm{x})$ by both the image and the Euclidean norm of the image gradient. These methods are designated as DU-image and DU-gradient, respectively. Furthermore, in order to demonstrate the rationale behind the use of the double-branch Unet, we concatenate the image and the geodesic domain information as inputs to the Unet, resulting in the Unet-TGD method. Furthermore, our observations indicate that utilizing the STD as the activation function in the layer can enhance the accuracy of the segmentation. Consequently, a further comparison is conducted by replacing the STD with the sigmoid function, resulting in the DDUM-NoSTD. The results of the visualization are presented in Fig. 12. In comparison to the geodesic distance features, the DU-image and DU-gradient demonstrate a more pronounced undersegmentation due to the absence of feature guidance. The Unet-TGD model cannot effectively learn from both the image and geodesic domains, resulting in a lack of capability to capture boundaries compared to the DDUM. As quantitatively demonstrated in Table 4, both DU-image and DU-gradient configurations yield inferior metric performance. This comparative analysis reveals that geodesic features substantially enhance network segmentation accuracy, outperforming conventional image inten-

Table 4: The metric results of Unet-TGD, DU-Image, DU-Gradient, DDUM-NoSTD and DDUM on the liver dataset regarding Dice, JS, HD and ASSD.

Dataset	Liver			
Method	Dice	JS	HD	ASSD
Unet-TGD	0.9108±0.0643	0.8478±0.1062	26.4418±7.1198	0.6290±0.5125
DU-image	0.8906±0.1412	0.8240±0.1734	32.9328±26.3487	3.1904±2.2846
DU-gradient	0.9078±0.0418	0.8407±0.0863	35.7341±13.6670	0.6743±1.3936
DDUM-NoSTD	0.9266±0.0846	0.8621±0.1409	24.8475±5.6816	0.6032±0.1735
DDUM	**0.9302±0.0241**	**0.8697±0.0424**	**16.7778±5.5260**	**0.5526±0.1936**

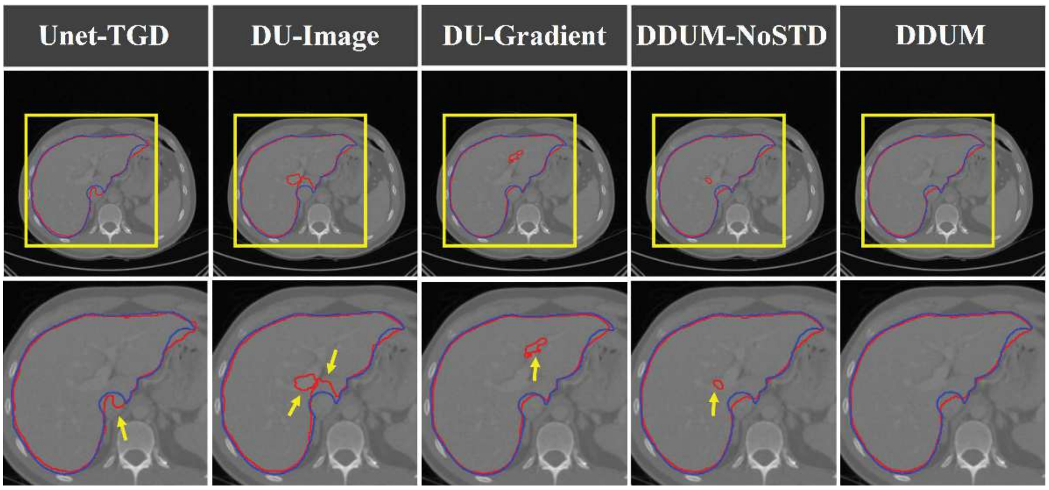
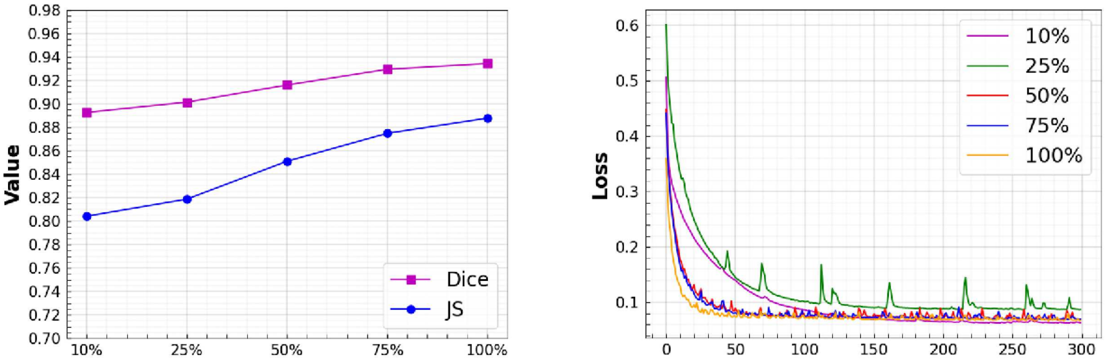


Figure 12: The visualization results of Unet-TGD, DU-Image, DU-Gradient, DDUM-NoSTD and DDUM on the liver dataset.

sity and gradient-based features by significant margins. Furthermore, the performance comparison between DDUM-NoSTD and the complete DDUM framework highlights the effectiveness of the STD module, DDUM exhibits high segmentation accuracy across all evaluation metrics. In summary, the DDUM framework is more rational and outperforms other frameworks in terms of metric scores.

4.3 Size of training set

In order to investigate the influence of the size of the training set on the segmentation result, the training set for CT liver data is sampled at different percentages such as 10%, 25%, 50%, 75% and 100%. Fig. 13(a) shows the values of Dice and JS for varying the size of the training set on a consistent testing set. As evidenced by the experimental results presented in Fig. 13(a), while the segmentation accuracy demonstrates a positive correlation with the size of the training dataset, our proposed DDUM maintains robust performance metrics even when trained with only 10% of the available training data. The loss conver-



(a) The size of training dataset (b) The iterations about epochs

Figure 13: The impact of the size of the training set on the segmentation accuracy and the convergence rate of the loss function (2.12).

gence characteristics, depicted in Fig. 13(b), reveal a significant relationship between the size of the training set and the efficiency of optimization. Specifically, the convergence rate of the loss function exhibits a substantial improvement with increazing training samples, suggesting that larger training sets facilitate more efficient parameter optimization during the network training process.

4.4 Influence of parameters

In our proposed DUMM, there are the parameters of Eq. (2.7), the parameters of the activation function (2.10) and the parameters of the loss function (2.12). The selection of these parameters is discussed below.

4.4.1 Parameters in Eq. (2.7)

In Eq. (2.7), there are four parameters: t,σ,ε_2, and β_2. In regard to the parameters ε_2 and β_2, we adopt the setting presented in [47] without further exploration or discussion. The influence of t and σ on the threshold geodesic distance is now considered using the cross-validation method. To this end, we initially fix the value of σ at 20 and then set the value of t to one of the following four values: $0,0.01,0.05$, and 0.1. Subsequently, t is fixed at 0.05, while the value of σ is set to $1,2,10,50$, and 100, respectively. From Fig. 14, it can be observed that the parameter t primarily influences the extent of the flat region. As the value of t increases, a greater number of pixels are included in the flat regions, resulting in a comparable diffusion speed. Additionally, the effect of parameter t on the geodesic distances of pixels in close proximity to the ROI is significantly constrained, while a substantial variation remains evident in the geodesic distances of pixels located at more distant positions from the ROI. Notably, when $t=0$, the relationship $d_t(\mathrm{x})\approx d_r(\mathrm{x})$ holds, leading to approximately equivalent values for the geodesic distances $D_t(\mathrm{x})$ and $D_r(\mathrm{x})$. Crucially, the geodesic distance metric employed in existing literature [38] can

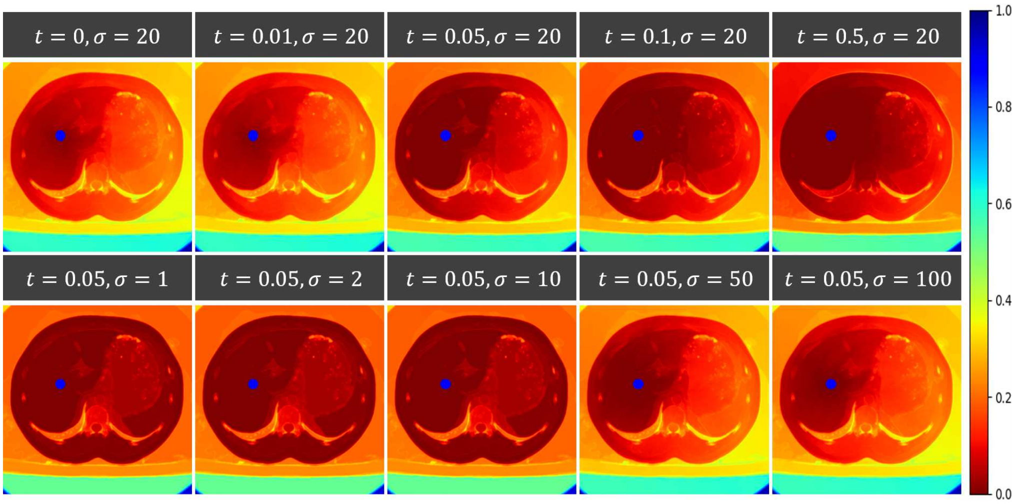


Figure 14: The influence of the parameters t and σ on the geodesic distance. (The blue dots represent the marked points).

be formally characterized as a special instance of the generalized theoretical framework introduced in this work.

The parameter σ mainly affects the weight $\alpha(\mathrm{x})$ of diffusion speed in the flat region. When σ is small, the weight $\alpha(\mathrm{x})$ is also small, insufficient to penalize the background region far from the marked point. When σ increases, the Euclidean distance provides spatial information, resulting in a greater penalty for the background region far from the marked point, and the distance grows faster. Meanwhile, the weight parameter $\alpha(\mathrm{x})$ of the target region tends towards 0, making the distance between the target region pixels and the marked point also close to 0.

4.4.2 Parameters in STD (2.10)

There are two parameters τ and μ in the activation function (2.10). We use cross-validation to choose them. More specifically, we first fix $\mu = 0.05$ and set $\tau = \{0,0.1,0.5,1.0,5.0\}$ and then fix $\tau = 0.05$ and set $\mu = \{0,0.1,0.5,1.0,5.0\}$. Table 5 presents the metric results under different parameter settings. By selecting the appropriate parameters, it is possible to effectively incorporate spatial regularization into the network. In this paper, the parameters are chosen as $\tau = 1$ and $\mu = 0.5$.

4.4.3 Parameters in the loss function 2.12

The proposed loss function in (2.12) incorporates two critical hyperparameters: λ, which serves as the regularization parameter balancing the geodesic and image domain constraints, and ω, which controls the cross-term influence. To optimize these parameters, we employ a systematic cross-validation strategy. Initially, we fix ω at 0.05 while varying λ across the set $\{0,0.01,0.05,0.1,0.5,1\}$. Subsequently, we fix λ at 0.05 and evaluated ω

Table 5: The influence of the parameters τ and μ in the activation function.

Parameters	Metrics		Parameters	Metrics	
	Dice	JS		Dice	JS
$\tau=0$, $\mu=0.5$	0.9261	0.8625	$\tau=1$, $\mu=0$	0.9286	0.8668
$\tau=0.1$, $\mu=0.5$	0.9275	0.8649	$\tau=1$, $\mu=0.1$	0.9237	0.9583
$\tau=0.5$, $\mu=0.5$	0.9251	0.8607	$\tau=1$, $\mu=0.5$	**0.9302**	**0.8697**
$\tau=1.0$, $\mu=0.5$	**0.9302**	**0.8697**	$\tau=1$, $\mu=1.0$	0.9286	0.8669
$\tau=5.0$, $\mu=0.5$	0.9292	0.8677	$\tau=1$, $\mu=5.0$	0.9284	0.8665

over the same range. The parameter sensitivity analysis, as depicted in Fig. 15, reveals several key insights:

(1) The regularization parameter λ exerts substantial influence on segmentation performance, where optimal regularization $\lambda=0.05$ yields superior metrics compared to both unregularized $\lambda=0$ and over-regularized scenarios.

(2) Excessive regularization $\lambda>0.1$ leads to significant performance degradation.

(3) The cross-term parameter ω demonstrates relatively stable impact across its range, with moderate values maintaining satisfactory performance. Based on this comprehensive analysis, we select $\lambda=0.05$ and $\omega=0.05$ as optimal values, achieving an effective balance between regularization strength and cross-term influence.

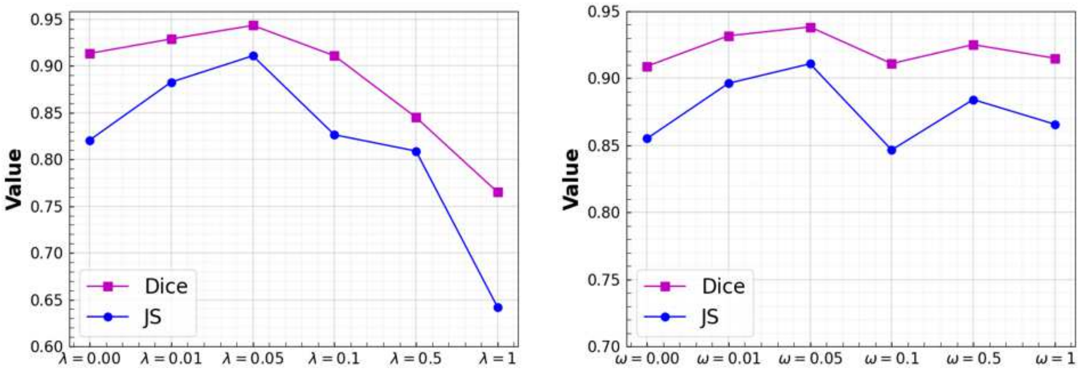


Figure 15: The influence of the parameters λ and ω in the loss function (2.12).

5 Conclusion and discussion

In this paper, we proposed a novel network architecture for the selective segmentation of medical images. In the proposed network architecture, the image domain and the geodesic domain information were concatenated into the hidden layer of the network.

Geodesic domain information was obtained by solving the classical Eikonal equation based on a suitable threshold parameter. The objective was to integrate information from both domains in order to achieve more precise and selective segmentation results for the medical image. Consequently, our network architecture consists of two independent Unets. Incorporating prior geometric information into the network architecture helps DDUM to have greater flexibility in the selection of marked points, thereby enhancing the adaptability of the segmentation method. Experiments conducted on three publicly available benchmark datasets for different modalities and organs demonstrated that the proposed DDUM exhibited superior performance compared to several state-of-the-art methods. However, the prior proposed in this paper is intended to distinguish boundary information from flat regions. When confronted with more complex medical images (such as tumor lesions), more generalized prior information is required. Additionally, this method is based on supervised learning and relies on a large amount of labeled data. In practical applications, obtaining carefully labeled and extensive datasets for training segmentation networks is a challenge. Therefore, future research will focus on the development of weakly supervised or unsupervised methods under more generalized geodesic distances to address the aforementioned limitations.

Acknowledgements

This work was supported in part by the Natural Science Foundation of China (Grant No. 12471398), by the Natural Science Foundation of Henan Province (Grant No. 232300420108), by the Key Scientific Research Projects in Universities in Henan Province (No. 24B110004), by the Henan Province International Science and Technology Cooperation (Project No. 242102520038), by the CityU1101120, CityU11309 922, NSFC/RGC (Grant No. NHKBU21419), by the CRF Grant C1013-21GF and by the NORCE Kompetanseoppbygging program, by the HKRGC (Grant Nos. CityU11309922, LU13300125), by the ITF (Grant Nos. MHP/054/22, LU BGR 105824), and by the InnoHK-Hong Kong Centre for Cerebro-cardiovascular Health Engineering.

References

[1] V. Badrinarayanan, A. Kendall, and R. Cipolla, *SegNet: A deep convolutional encoder-decoder architecture for image segmentation*, IEEE Trans. Pattern Anal. Mach. Intell., 39:2481–2495, 2017.

[2] E. Bae, X. Tai, and J. Yuan, *Maximizing flows with message-passing: Computing spatially continuous min-cuts*, in: Proceedings 10th International Conference Energy Minimization Methods in Computer Vision and Pattern Recognition, Springer, 15–28, 2015.

[3] X. Bai and G. Sapiro, *A geodesic framework for fast interactive image and video segmentation and matting*, in: 2007 IEEE 11th International Conference on Computer Vision, IEEE, 1–8, 2007.

[4] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, *Interactive Co-segmentation of Objects in Image Collections*, in: SpringerBriefs in Computer Science, Springer, 2011.

[5] Y. Boykov and M. Jolly, *Interactive graph cuts for optimal boundary & region segmentation of objects in ND images*, in: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, Vol. 1, IEEE, 105–112, 2001.

[6] X. Bresson, S. Esedoḡlu, P. Vandergheynst, J. Thiran, and S. Osher, *Fast global minimization of the active contour/snake model*, J. Math. Imaging Vision, 28:151–167, 2007.

[7] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40:120–145, 2011.

[8] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi, *Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists*, SIAM Rev., 65:375–435, 2023.

[9] A. Criminisi, T. Sharp, and A. Blake, *GeoS: Geodesic image segmentation*, in: European Conference on Computer Vision, Springer, 99–112, 2008.

[10] K. Ding, L. Xiao, and G. Weng, *Active contours driven by region-scalable fitting and optimized Laplacian of Gaussian energy for image segmentation*, Signal Process., 134:224–233, 2017.

[11] X. Dong and P. Chen, *Segmentation of liver tumors based on bottleneck residual attention mechanism U-net*, CT Theory and Applications, 30:661–670, 2021.

[12] S. Esedoḡlu and F. Otto, *Threshold dynamics for networks with arbitrary surface tensions*, Comm. Pure Appl. Math., 68:808–864, 2015.

[13] W. Feng, L. Ju, L. Wang, K. Song, and Z. Ge, *Unsupervised domain adaptation for medical image segmentation by selective entropy constraints and adaptive semantic alignment*, in: Proceedings of the AAAI Conference on Artificial Intelligence, AAAI Press, 37:623–631, 2023.

[14] R. Glowinski, S. Luo, and X. Tai, *Fast operator-splitting algorithms for variational imaging models: Some recent developments*, Handb. Numer. Anal., 20:191–232, 2019.

[15] R. Glowinski, S. Osher, and W. Yin, *Splitting Methods in Communication, Imaging, Science, and Engineering*, Springer, 2017.

[16] C. Gout, C. Le, and L. Vese, *Segmentation under geometrical conditions using geodesic active contours and interpolation using level set methods*, Numer. Algorithms, 39:155–173, 2005.

[17] L. Grady, *Random walks for image segmentation*, IEEE Trans. Pattern Anal. Mach. Intell., 28:1768–1783, 2006.

[18] T. Grandits, A. Effland, T. Pock, R. Krause, G. Plank, and S. Pezzuto, *GEASI: Geodesic-based earliest activation sites identification in cardiac models*, Int. J. Numer. Methods Biomed. Eng., 37:3505–3535, 2021.

[19] K. Guo, D. Han, and X. Yuan, *Convergence analysis of Douglas-Rachford splitting method for "strongly+weakly" convex programming*, SIAM J. Numer. Anal., 55:1549–1577, 2017.

[20] S. Han, W. Tao, D. Wang, X.-C. Tai, and X. Wu, *Image segmentation based on GrabCut framework integrating multiscale nonlinear structure tensor*, IEEE Trans. Image Process., 18:2289–2302, 2009.

[21] X. Hu and H. Yang, *DRU-net: A novel U-net for biomedical image segmentation*, IET Image Process., 14:192–200, 2020.

[22] Y. Hu, A. Soltoggio, R. Lock, and S. Carter, *A fully convolutional two-stream fusion network for interactive image segmentation*, Neural Netw., 109:31–42, 2019.

[23] F. Jia, J. Liu, and X. Tai, *A regularized convolutional neural network for semantic image segmentation*, Anal. Appl., 19(1):147–165, 2021.

[24] Y. Li, J. Zhang, P. Gao, L. Jiang, and M. Chen, *Grab cut image segmentation based on image region*, in: 2018 IEEE 3rd International Conference on Image, Vision and Computing, IEEE 311–315, 2018.

[25] G. Liu, J. Wang, D. Liu, and B. Chang, *A multiscale nonlocal feature extraction network for breast lesion segmentation in ultrasound images*, IEEE Trans. Instrum. Meas., 72:5011012, 2023.

[26] J. Liu, X.-C. Tai, H. Huang, and Z. Huan, *A fast segmentation method based on constraint optimization and its applications: Intensity inhomogeneity and texture segmentation*, Pattern Recognit., 44:2093–2108, 2011.

[27] J. Liu, X. Wang, and X.-C. Tai, *Deep convolutional neural networks with spatial regularization, volume and star-shape priors for image segmentation*, J. Math. Imaging Vision, 64:625–645, 2022.

[28] J. Long, E. Shelhamer, and T. Darrell, *Fully convolutional networks for semantic segmentation*, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 3431–3440, 2015.

[29] S. Luo, J. Chen, Y. Xiao, and X.-C. Tai, *A binary characterization method for shape convexity and applications*, Appl. Math. Model., 122:780–795, 2023.

[30] X. Luo et al., *MIDeepSeg: Minimally interactive segmentation of unseen objects from medical images using deep learning*, Med. Image Anal., 72:102102–102117, 2021.

[31] R. Malladi and J. Sethian, *Level set and fast marching methods in image processing and computer vision*, in: Proceedings of 3rd IEEE International Conference on Image Processing, Vol. 1, IEEE, 489–492, 1996.

[32] K. Maninis, S. Caelles, J. PontTuset, and L. Van, *Deep extreme cut: From extreme points to object segmentation*, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 616–625, 2018.

[33] J. Mirebeau, *Riemannian fast-marching on cartesian grids, using Voronoi's first reduction of quadratic forms*, SIAM J. Numer. Anal., 57:2608–2655, 2019.

[34] T. Nguyen, J. Cai, J. Zhang, and J. Zheng, *Robust interactive image segmentation using convex active contours*, IEEE Trans. Image Process., 21:3734–3743, 2012.

[35] P. Qiao, Z. Wei, Y. Wang, Z. Wang, G. Song, F. Xu, X. Ji, C. Liu, and J. Chen, *Fuzzy positive learning for semi-supervised semantic segmentation*, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 15465–15474, 2023.

[36] M. Rajchl et al., *DeepCut: Object segmentation from bounding box annotations using convolutional neural networks*, IEEE Trans. Med. Imag., 36:674–683, 2016.

[37] W. Ren, Y. Tang, Q. Sun, C. Zhao, and Q. Han, *Visual semantic segmentation based on few/zero-shot learning: An overview*, IEEE/CAA J. Autom. Sin., 11(5):1106–1126, 2024.

[38] M. Roberts, K. Chen, and K. Irion, *A convex geodesic selective model for image segmentation*, J. Math. Imaging Vision, 61:482–503, 2019.

[39] O. Ronneberger, P. Fischer, and T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, in: Medical Image Computing and Computer-Assisted Intervention, Vol. 9351, Springer, LNCS, 234–241, 2015,

[40] H. Roth, D. Yang, Z. Xu, X. Wang, and D. Xu, *Going to extremes: Weakly supervised medical image segmentation*, Mach. Learn. Knowl. Extr., 3:507–524, 2021.

[41] H. Roth, L. Zhang, D. Yang, F. Milletari, Z. Xu, X. Wang, and D. Xu, *Weakly supervised segmentation from extreme points*, in: Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention: International Workshops, Springer-Verlag, 42–50, 2019.

[42] C. Rother, V. Kolmogorov, and A. Blake, *"GrabCut" interactive foreground extraction using iterated graph cuts*, ACM Trans. Graph., 23:309–314, 2004.

[43] T. Sakinis et al., *Interactive segmentation of medical images through fully convolutional neural networks*, arXiv:1903.08205, 2019.

[44] N. Sharma and L. Aggarwal, *Automated medical image segmentation techniques*, J. Med. Phys., 35:3–14, 2010.

[45] D. Shen, G. Wu, and H. Suk, *Deep learning in medical image analysis*, Annu. Rev. Biomed. Eng.,

19:221–248, 2017.

[46] N. Siddique, S. Paheding, C. Elkin, and V. Devabhaktuni, *U-net and its variants for medical image segmentation: A review of theory and applications*, IEEE Access, 9:82031–82057, 2021.

[47] J. Spencer and K. Chen, *A convex and selective variational model for image segmentation*, Commun. Math. Sci., 13:1453–1472, 2015.

[48] L. Sun, Z. Tian, Z. Chen, W. Luo, and S. Du, *An efficient interactive segmentation framework for medical images without pre-training*, Med. Phys., 50:2239–2248, 2023.

[49] X.-C. Tai, H. Liu, and R. Chan, *Pottsmgnet: A mathematical explanation of encoder-decoder based neural networks*, SIAM J. Imaging Sci., 17:540–594, 2024.

[50] T. Tsai, L. Cheng, S. Osher, and H. Zhao, *Fast sweeping algorithms for a class of Hamilton-Jacobi equations*, SIAM J. Numer. Anal., 41:673–694, 2003.

[51] V. Vezhnevets and V. Konouchine, *"Growcut": Interactive multi-label N-D image segmentation by cellular automata*, in: GraphiCon 2005 Proceedings, Vol. 1, 150–156, 2005.

[52] D. Wang and X. Wang, *The iterative convolution-thresholding method (ICTM) for image segmentation*, Pattern Recognit., 130:108794, 2022.

[53] G. Wang et al., *Interactive medical image segmentation using deep learning with image-specific fine tuning*, IEEE Trans. Med. Imag., 37:1562–1573, 2018.

[54] G. Wang et al., *DeepIGeoS: A deep interactive geodesic framework for medical image segmentation*, IEEE Trans. Pattern Anal. Mach. Intell., 41:1559–1572, 2018.

[55] Z. Wang, D. Acuna, H. Ling, A. Kar, and S. Fidler, *Object instance annotation with deep extreme level set evolution*, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 7492–7500, 2019.

[56] D. Withey and Z. Koles, *Medical image segmentation: Methods and software*, in: 2007 Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging, IEEE, 140–143, 2007.

[57] C. Wu and X.-C. Tai, *Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models*, SIAM J. Imaging Sci., 3:300–339, 2010.

[58] S. Yin, H. Deng, Z. Xu, Q. Zhu, and J. Cheng, *SD-UNet: A novel segmentation framework for CT images of lung infections*, Electronics, 11:130–149, 2022.

[59] X.-X. Yin, L. Sun, Y. Fu, R. Lu, and Y. Zhang, *U-Net-based medical image segmentation*, J. Healthc. Eng., 2022:4189781, 2022.

[60] J. Yuan, E. Bae, X.-C. Tai, and Y. Boykov, *A spatially continuous max-flow and min-cut framework for binary labeling problems*, Numer. Math., 126:559–587, 2014.

[61] S. Zhai, G. Wang, X. Luo, Q. Yue, K. Li, and S. Zhang, *PA-Seg: Learning from point annotations for 3d medical image segmentation using contextual regularization and cross knowledge distillation*, IEEE Trans. Med. Imag., 72:2235–2246, 2023.

[62] H. Zhao, *A fast sweeping method for Eikonal equations*, Math. Comput., 74:603–627, 2005.