PAPER

Medical image segmentation started from a single point annotation: a novel variational model

To cite this article: Lu Tian *et al* 2025 *Phys. Scr.* **100** 115207

View the article online for updates and enhancements.

You may also like

Physica Scripta

PAPER

Medical image segmentation started from a single point annotation: a novel variational model

Lu Tian[1,2] , Liwen Zou[2] and Xiaoping Yang[2,*]

[1] Department of Mathematics, Nanjing University of Science and Technology Zijin College, Nanjing, 210023, People's Republic of China
[2] School of Mathematics, Nanjing University, Nanjing, 210093, People's Republic of China
[*] Author to whom any correspondence should be addressed.

E-mail: xpyang@nju.edu.cn

Abstract

Developing weak annotations is one of the effective methods to reduce the workload of labeling the training data and extract the valuable priori information. In this paper, we propose a novel weakly supervised variational segmentation model that requires only a single randomly selected point within the target region. First, We construct an anisotropic Riemannian metric that incorporates not only the texture patterns, the local directional information at the boundaries, but also the intensity heterogeneity through a conformal factor. This step is the cornerstone of our method, as it enables the effective integration of multi-scale image features. Second, we calculate two types of distances: geodesic distances based on the new Riemannian metric and Euclidean distances, from all the points in the image to the taken point. These pointwise distances are then integrated into the Chan-Vese model, enabling the simultaneous exclusion of regions with similar intensities both near and far from the target region. The distances serve as adaptive weighting parameters in the variational model, further enhancing segmentation accuracy. We employ the Douglas-Rachford algorithm for efficient numerical implementation. Experimental results show that our method achieves similar or better segmentation performance than other weakly annotated methods and the single-point annotation has better robustness.

1. Introduction

In the past decade, we have witnessed the revolutionary development of semantic segmentation in different domains, such as medical diagnosis, aerial imagery, autonomous driving and so on. Well-performing segmentation models generally require large amounts of high-quality training data or priori information. However, it is well known that acquiring annotations is very laborious and time-consuming, especially for some complex medical images. Therefore, weak annotation-based segmentation, which can quickly provide and extract relevant information about the target region by manually providing some inputs (point, scribble or bounding box annotation), seems to be one of the effective ways to reduce the workload of annotation and extract the valuable priori information [1–8].

Weak annotation based segmentation methods can be divided into two main categories: traditional optimization-based weakly supervised methods and modern deep learning-based weakly supervised methods. Traditional optimization based weakly supervised methods acquire the information about foreground and background through user interaction, build energy functionals using graph theory or variational methods, and obtain the segmentation results iteratively by minimizing the energy functionals, such as grab-cut [9–11], graph-cut [12–14], active contour-based selective segmentation [15–22]. Whereas modern deep learning based weakly supervised methods may utilize weak annotation in different contexts. Some of these approaches are to use weak annotations at test-time to refine or guide network predictions. These techniques are commonly referred to as interactive segmentation or semi-automated segmentation. Some other methods are only trained with incomplete but easy-to-obtain annotations and operate in fully-automatic fashion at inference stage
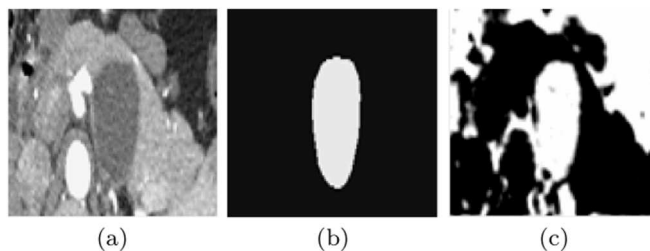
Figure 1. The three images show a case about the pancreatic cysts. (a): Original image. (b): Ground truth. (c): Result.

[18–22]. Actually, whether it is a traditional optimization-based weakly supervised approach or a modern deep learning-based weakly supervised approach, they all involve the crucial step of taking points (or scribbles or bounding boxes and so on) by the user to extract favorable information and guide the subsequent segmentation [23–27], which is also the focus of this paper. In this paper, we focus more on how to extract favorable information in the complex medical images.

Without utilizing additional prior knowledge, traditional models such as CV model may perform poorly when segmenting the complex images [28–31]. As shown in figure 1, CV model generally faces the following challenges: (1) Regions that are far away from the target but with similar intensities are mistaken as the target region. (2) Regions that are close to the target area and have similar intensities are also mistaken for the target region.

To alleviate the above difficulties, we propose a very weakly supervised variational segmentation model based on different distance information induced by one randomly choosing point within the target area. Firstly, we need to manually take only one random point in the target area, and then calculate two kinds of distances from all the points in the image to the taken point, geodesic distances based on the new Riemannian metric and Euclidean distances. The Euclidean distance is mainly used to exclude the areas that are similar in intensity to the target region but far away. While the geodesic distance is used to exclude the areas that are close to the target area and have a similar intensity. The most crucial step in the process of calculating the geodesic distances is to design the metric. We construct an anisotropic Riemannian metric by using the eigenvalues and eigenvectors of the structure tensor field, which takes into account the amount of intensity changes between pixels, the local edge direction and textural pattern, while multiplying the regional information as a conformal factor by it. Secondly, to analyze the effectiveness of the information extracted, we embed the two kinds of distances simultaneously into a classical variational model to guide the segmentation pointwisely. Finally, we experiment with the Douglas-Rachford algorithm.

Main contributions of our paper are summarized as follows.

(1) We construct a new anisotropic Riemannian metric that not only embeds the local edge direction and textural pattern extracted by the eigen-decomposition of structure tensor field, but also embeds the amount of the intensity heterogeneity through a conformal factor. Thus, we mitigate the challenges of intensity heterogeneity and weak boundaries, which can exclude the negative samples that are inconsistent and intrusive.

(2) We incorporate dual pointwise distance information into the variational model. First, this allows for the simultaneous exclusion of regions with similar intensities, both near and far from the target region, thereby addressing issues related to inhomogeneous intensity and weak boundaries. Second, this information serves as an adaptive weighting parameter within the variational model to further pointwisely facilitate segmentation.

(3) We just need to manually take only one point inside the target area as weakly supervised information, no need to take more points outside the target area or on the boundary, which provides both a variety of priori information such as localization and geometric structure, as well as a class of methods for reducing physician labeling.

The rest of this paper is organized as follows. In section 2, we briefly introduce CV model, the distance map, the structure tensor field and the Douglas-Rachford algorithm. In section 3, we present a detailed segmentation strategy, and the existence of the minimizer to the model is proved. In section 4, a numerical algorithm is derived. In section 5, we demonstrate the performance of our proposed method through different experiments. Conclusions and discussions are given in section 6.

2. Preliminary

2.1. CV model

The two-phase and piecewise constant Mumford−Shah segmentation functional proposed by [32], called the Chan-Vese (CV) model, has been widely studied and applied. Not only that, [33] further reformulate the

nonconvex CV model into a convex equivalent form, the energy functional can be expressed as

$$E_{CCV}(u, c_1, c_2) = \alpha \int_{\Omega} (I(x) - c_1)^2 u(x) + (I(x) - c_2)^2 (1 - u(x)) dx$$
$$+ \int_{\Omega} |\nabla u| dx, \tag{1}$$

where c_1 and c_2 are two constants that approximate the image intensity inside and outside the target area, respectively. $u \in [0, 1]$ can be interpreted as the probability that the pixel x belongs to the target region.

2.2. Distance map

A geodesic is a path, connecting two points, that is the globally minimizer of an energy functional, hence the geodesic is also called the minimal (or shortest) path. The energy functional is formulated as

$$E(\gamma) = \int_{\gamma} \mathcal{P}(\gamma(t), \gamma'(t)) dt, \tag{2}$$

where $\mathcal{P}$ is the potential function based on image features, $\gamma(t)$ is a smooth curve, $\gamma'(t)$ is the tangent of $\gamma(t)$. In this paper, the energy corresponded to the length of the curve $\gamma(t)$ is calculated in the sense of Riemannian metric, then the potential function is represented as $\mathcal{P}(\gamma(t), \gamma'(t)) = \sqrt{\gamma'(t)^T \mathcal{M}(\gamma(t)) \gamma'(t)}$ which essentially corresponds to the metric tensor $\mathcal{M}(\gamma(t))$(symmetric positive definite tensor field) [34, 35].

The distance map $\mathcal{D}: \Omega \to \mathbb{R}^+$ is the minimal energy integrated along a path between fixed point p and any point x in the domain Ω:

$$\begin{cases} \mathcal{D}(x) = \min_{\gamma \in \mathcal{A}_{p,x}} \{ \int_{\gamma} \mathcal{P}(\gamma(t), \gamma'(t)) dt \}, \\ \gamma(0) = p, \gamma(1) = x. \end{cases} \tag{3}$$

where $\mathcal{A}_{p,x}$ is the set of paths connecting x to p. The geodesic distance map $\mathcal{D}(x)$ admits the Eikonal equation

$$\begin{cases} \|\nabla \mathcal{D}(x)\|_{\mathcal{M}^{-1}(x)} = 1, & x \in \Omega, \\ \mathcal{D}(0) = p. \end{cases} \tag{4}$$

where $\|v\|_{\mathcal{M}^{-1}} = \sqrt{v^T \mathcal{M}^{-1} v}$ [36, 37].

In particular, if $\mathcal{M} = I$, where I is the identity matrix, $\mathcal{D}(x)$ is the Euclidean distance in this case. In addition, the eikonal equation(4) can be solved by Fast Marching Algorithm, see more details in [38–40].

2.3. Structure tensor field

The local autocorrelation function represents the change in the local intensity when the local image window is sliding locally along different directions, and is defined as follows:

$$E(\Delta x_1, \Delta x_2) = \sum_x \omega(x) [f(x_1 + \Delta x_1, x_2 + \Delta x_2) - f(x_1, x_2)]^2.$$

where $f(x)$ is the intensity value at the pixel $x = (x_1, x_2)$, $\omega(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$ is the Gaussian function.

Let

$$Q(x) = \sum_x \omega(x) \begin{pmatrix} f_{x_1} f_{x_1} & f_{x_1} f_{x_2} \\ f_{x_1} f_{x_2} & f_{x_2} f_{x_2} \end{pmatrix} = \omega * \begin{pmatrix} f_{x_1} f_{x_1} & f_{x_1} f_{x_2} \\ f_{x_1} f_{x_2} & f_{x_2} f_{x_2} \end{pmatrix}$$
$$= \begin{pmatrix} A & C \\ C & B \end{pmatrix}, \tag{5}$$

where $*$ denotes the convolution operation, f_{x_1}, f_{x_2} represent the horizontal and vertical gradients of the intensity of the image, respectively. $A = \omega * f_{x_1}^2$, $B = \omega * f_{x_2}^2$, $C = \omega * f_{x_1} f_{x_2}$. In addition, $Q(x)$ can be understood as the local covariance matrix around x, called the structure tensor field.

Perform a Taylor expansion of $f(x_1, x_2)$ and take the first-order approximation

$$E(\Delta x_1, \Delta x_2) \approx \begin{pmatrix} \Delta x_1 & \Delta x_2 \end{pmatrix} Q(x) \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix}. \tag{6}$$

Obviously, $Q(x)$ is a symmetric matrix, it can be diagonalized as follows: $Q(x) = \lambda_1(x) v_1(x) v_1^T(x) + \lambda_2(x) v_2(x) v_2^T(x)$, where $\lambda_1(x), \lambda_2(x)$ are the eigenvalues of $Q(x)$ and $v_1(x), v_2(x)$ are the eigenvectors of $Q(x)$ corresponding to $\lambda_1(x)$ and $\lambda_2(x)$, respectively [41–44].
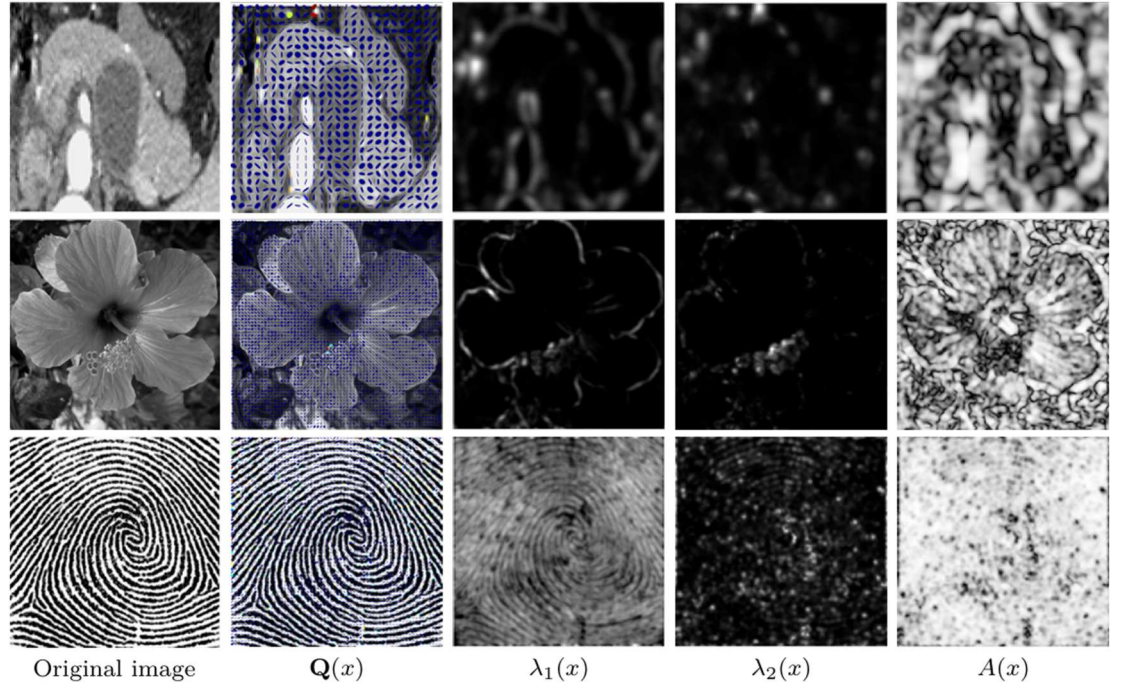
Figure 2. The first column is the original image. The second column is the structure tensor field $Q(x)$. To display a structure tensor field $Q(x)$, we draw an ellipse ε_x at each pixel x as the (scaled and translated) unit ball of the tensor $\varepsilon_x = \{\delta \in \mathbb{R}^2 : \delta^T Q(x)\delta \leqslant 1\}$. The third, fourth, and fifth columns are the eigenvalues $\lambda_1(x)$, $\lambda_2(x)$ and the anisotropy rate $A(x)$, respectively.

The autocorrelation function $E(\Delta x_1, \Delta x_2)$ can be approximated as a quadratic function.

$$E(\Delta x_1, \Delta x_2) \approx A\Delta x_1^2 + 2C\Delta x_1 \Delta x_2 + B\Delta x_2^2.$$

Let $E(\Delta x_1, \Delta x_2) = constant$, an ellipse can be used to depict this quadratic function. We visualize the anisotropy and orientation encoded in the tensor field $Q(x)$ in figure 2, the lengths of the short and long axes of the ellipse correspond to the eigenvalues $\lambda_1(x)$, $\lambda_2(x)$ of $Q(x)$, and the orientations of the short and long axes correspond to the eigenvectors of $v_1(x)$, $v_2(x)$ of $Q(x)$. $A(x) = \frac{\lambda_1(x) - \lambda_2(x)}{\lambda_1(x) + \lambda_2(x)}$ (assume that $\lambda_1(x) \leqslant \lambda_2(x)$) is the anisotropy, see more details in [45].

2.4. Douglas-rachford algorithm

To minimize the segmentation energy, we will make use of proximal splitting scheme. These scheme are adapted to solve structured non-smooth optimization problem. They basically replace the traditional gradient-descent step by proximal mappings, defined as [46]

$$Prox_F(z) = \underset{y}{\arg\max} \frac{1}{2}\|z - y\|^2 + F(y).$$

The Douglas-Rachford(DR) algorithm is an iterative scheme to minimize functionals of the form

$$\min_z F(z) + G(z),$$

where F and G are convex functions for which one is able to compute the proximal mapping $Prox_F$ and $Prox_G$. The important point is that F and G do not need to be smooth.

Douglas-Rachford(DR) iteration: start at any z_0 and repeat for $k = 0, 1, \cdots$

$$\begin{cases} x_{k+1} = Prox_F(z_k), \\ \tilde{z}_{k+1} = z_k + Prox_G(2x_{k+1} - z_k) - x_{k+1}, \\ z_{k+1} = z_k + \rho(\tilde{z}_{k+1} - z_k)(0 < \rho < 2). \end{cases} \tag{7}$$

To speed up convergence, we do the relaxation in the last step of the iteration [47].

3. The proposed model

In this section, we will elaborate our proposed very weakly supervised segmentation framework which is illustrated in figure 3.
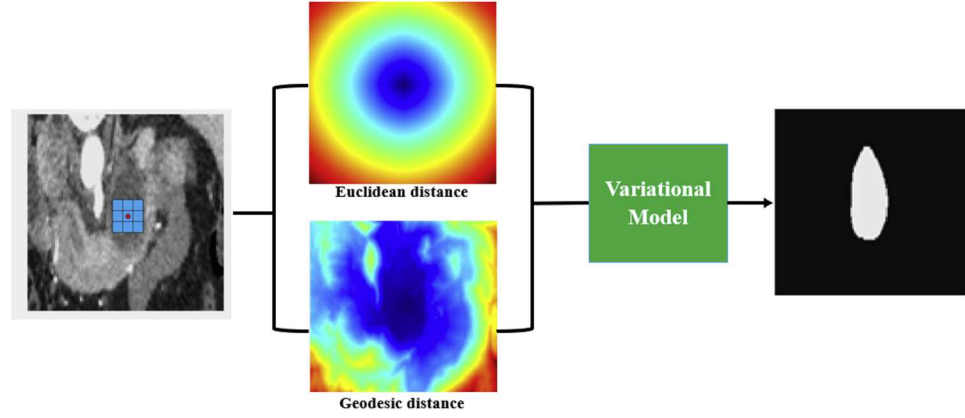
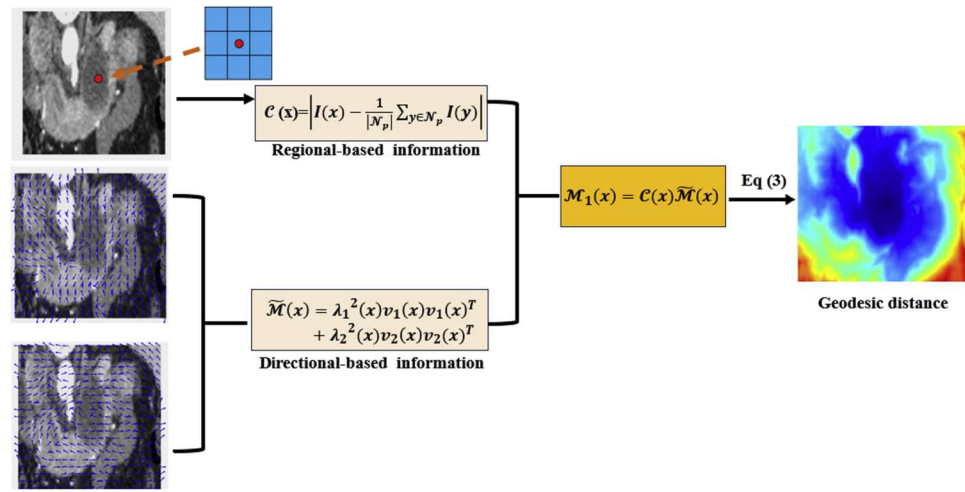**Figure 3.** Flowchart of our proposed weakly supervised segmentation framework.

Figure 4. Flowchart for constructing the Riemannian metric.

3.1. Geodesic distance

To exclude the areas that are close to the target area and have similar intensities, we design a novel Riemannian metric with both the regional information and the local geometrical information, see figure 4.

Firstly, we introduce an inhomogeneity quantity $\mathcal{C}(x)$. We manually take a random point p in the target region, and then compute the difference $\mathcal{C}(x)$ between the intensities of all the pixels x in the image and the average of the intensities within the patch $\mathcal{N}_p$ of the fixed point p, i.e., $\mathcal{C}(x) = \left| I(x) - \frac{1}{|\mathcal{N}_p|}\sum_{y \in \mathcal{N}_p} I(y) \right|$, where $\mathcal{N}_p$ denotes a 3^*3 square domain centered at pixel point p. Considering the intensity inhomogeneity in the target region, we compute the difference of all pixels in the image with respect to the average intensity within the patch $\mathcal{N}_p$ containing the point p, instead of with respect to a fixed point p.

Secondly, we construct a Riemannian metric $\widetilde{\mathcal{M}}(x)$. We observe that, the eigenvalues $\lambda_1(x)$, $\lambda_2(x)$ of the structure tensor field $Q(x)$ reflect the amount of intensity change along the direction of the respective corresponding eigenvectors, the eigenvectors $v_1(x)$, $v_2(x)$ reflect the local edge direction and textural pattern, see figure 2 and figure 5. Since $Q(x)$ is symmetric(not necessarily positive definite), it can be diagonalized as follows: $Q(x) = \lambda_1(x)v_1(x)v_1^T(x) + \lambda_2(x)v_2(x)v_2^T(x)$. However, the Riemannian metric must be definite positive, so we can modify the eigenvalues of $Q(x)$ to design the metric $\widetilde{\mathcal{M}}(x) = \lambda_1^2(x)v_1(x)v_1^T(x) + \lambda_2^2(x)v_2(x)v_2^T(x)$.

Finally, by after multiplying $\mathcal{C}(x)$ as a conformal factor by the Riemannian metric $\widetilde{\mathcal{M}}(x)$ constructed, we obtain a new local geodesic metric

$$\mathcal{M}_1(x) = \mathcal{C}(x)(\lambda_1^2(x)v_1(x)v_1^T(x) + \lambda_2^2(x)v_2(x)v_2^T(x)).$$

It has the amount of intensity changes, the local edge directional information and textural patterns. The new geodesic distance can be obtained by substituting the novel Riemannian metric $\mathcal{M}_1(x)$ into the distance
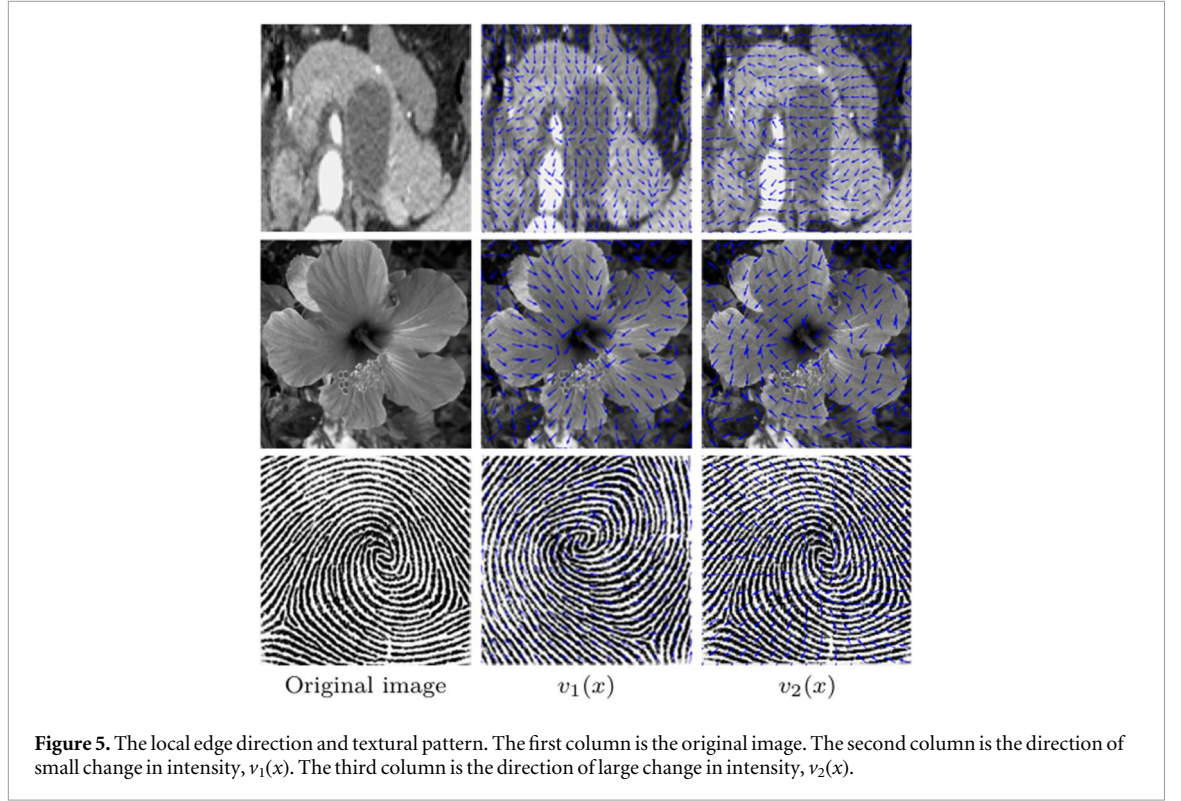
Figure 5. The local edge direction and textural pattern. The first column is the original image. The second column is the direction of small change in intensity, $v_1(x)$. The third column is the direction of large change in intensity, $v_2(x)$.

function $\mathcal{D}_1(x)$, introduced in Section 2.2.

$$
\begin{cases}
\mathcal{D}_1(x) = \min_{\gamma \in \mathcal{A}_{p,x}} \{ \int_\gamma \mathcal{P}_1(\gamma(t), \gamma'(t)) dt \}, \\
\gamma(0) = p, \ \gamma(1) = x.
\end{cases}
\tag{8}
$$

where

$$
\mathcal{P}_1(\gamma(t), \gamma'(t)) = \sqrt{\gamma'(t)^T \mathcal{M}_1(\gamma(t)) \gamma'(t)}.
$$

According to our designed Riemannian metric with regional, geometric and textural information, when minimizing the geodesic distance, we can exclude the regions that are close to the target region but have similar intensities, see figure 8.

3.2. Weakly supervised variational segmentation model

To analyze the effectiveness of the information extracted with one point annotation, we embed the two kinds of distances simultaneously into a classical variational model to guide the segmentation pointwisely as follows.

$$
\begin{aligned}
E(u, c_1, c_2) = \ & \alpha \int_\Omega \mathcal{D}_1(x)(I(x) - c_1)^2 u(x) \\
& + (1 - \mathcal{D}_1(x))(I(x) - c_2)^2 (1 - u(x)) dx \\
& + \int_\Omega (\mathcal{D}_1(x) + \mathcal{D}_2(x) + 1) |\nabla u| dx,
\end{aligned}
\tag{9}
$$

where c_1 and c_2 are two constants that approximate the image intensity inside and outside the target area, respectively. $\mathcal{D}_1(x)$ and $\mathcal{D}_2(x)$ are the geodesic distance and Euclidean distance respectively, and are normalized, i.e., $\mathcal{D}_1(x) \in [0, 1]$ and $\mathcal{D}_2(x) \in [0, 1]$. To ensure the existence of the solution to model (9), we add a constant 1 to the sum of the two distances before the regularization term. $\mathcal{D}_1(x)$ is small when the point x belongs to the same region as the annotated point p. That is, even if the point x is relatively far from the fixed point p, $\mathcal{D}_1(x)$ remains small as long as they are in the same region. But the value of the Euclidean distance increases as the point x gets farther away from the fixed point p.

The model (9) with respect to u can be equivalently rewritten

$$E(u) = \alpha \int_{\Omega} [\mathcal{D}_1(x)(I(x) - c_1)^2 - (1 - \mathcal{D}_1(x))(I(x) - c_2)^2] u(x) dx$$
$$+ \int_{\Omega} (\mathcal{D}_1(x) + \mathcal{D}_2(x) + 1)|\nabla u| dx. \tag{10}$$

It is easy to prove that the minimizer of the model (10) exists. In addition, the first term of (10) is an affine function w.r.t u, and the second term is the BV semi-norm, they are both obviously convex w.r.t u. Therefore, the model (10) is convex with respect to u, thus the uniqueness of the minimizer is also naturally guaranteed.

In the CV model (1), the probability value $u(x)$ that the point x belongs to the target region is obtained by using only the differences between the intensity $I(x)$ at the pixel x and the mean intensity c_1 and c_2 inside and outside the target region as weights, respectively. The CV model usually deal with the images with piecewise constant values, for complex images it shows poor performance, see figure 1. Hence, we embed the priori knowledge obtained through the weakly supervised approach of one point annotation into the variational model to output more accurate probability values $u(x)$. It is emphasized that the annotation point p must be taken within the target area, see Section 5.5 for more analysis.

(1) If the difference in the intensity value, or the Euclidean distance or the texture information between point x and the fixed point p are relatively large, then the value of $\mathcal{D}_1(x)$ and $\mathcal{D}_2(x)$ will be large, and thus the probability value $u(x)$ will be predicted to be relatively small when minimizing the energy functional $E(u, c_1, c_2)$.

(2) If the difference in the intensity value, or the Euclidean distance or the texture information between point x and the fixed point p are relatively small, then the value of $\mathcal{D}_1(x)$ and $\mathcal{D}_2(x)$ will be small, and thus the probability value $u(x)$ will be predicted to be relatively large when minimizing the energy functional $E(u, c_1, c_2)$.

4. Numerical optimization

In this section, we provide an efficient algorithm based on Douglas-Rachford algorithm and discuss its convergence.

Let $w_1 = \mathcal{D}_1(x)(I(x) - c_1)^2 - (1 - \mathcal{D}_1(x))(I(x) - c_2)^2$ and $w_2 = \mathcal{D}_1(x) + \mathcal{D}_2(x) + 1$, then the model (9) is equivalent to the following optimization problem

$$\min_{z=(u,v)} F(z) + G(z),$$

where

$$F(u, v) = \alpha \int_{\Omega} w_1 u dx + \int_{\Omega} w_2 |v| dx + \delta_{[0,1]}(u),$$
$$G(u, v) = \delta_C(u, v),$$
$$\delta_{[0,1]}(u) = \begin{cases} 0, & u \in [0, 1], \\ +\infty, & \text{otherwise.} \end{cases}$$
$$C = \{z = (u, v)|v = \nabla u\}.$$

Proximal Operator of F

$$F(u, v) = F_0(u) + F_1(v),$$

where

$$\begin{cases} F_0(u) = \int_{\Omega} w_1 u dx + \delta_{[0,1]}(u), \\ F_1(v) = \int_{\Omega} w_2 |v| dx. \end{cases}$$

hence

$$Prox_F(u, v) = (Prox_{F_0}(u), Prox_{F_1}(v)).$$

By the definition of the proximal mapping,

$$Prox_{F_0}(u) = Proj_{[0,1]}(u - w_1),$$

where

$$Proj_{[0,1]}(w) = \max(0, \min(1, w)).$$
$$Prox_{F_1}(v) = S(v, \alpha),$$

where S is the shrinkage operator defined as

$$S(v, \alpha) = \frac{v}{\|v\|} * \max(\|v\| - \alpha, 0).$$

Proximal Operator of G

$$Prox_G(u, v) = Proj_C(u, v).$$

Finally, we define $Prox_F^*(x) = 2Prox_F(x) - x$, the DR iterations with relaxation in our numerical optimization can be written in a more concise equivalent form as follows,

$$\begin{cases} z_{k+1} = (1 - \frac{\rho}{2})z_k + \frac{\rho}{2}Prox_G^*(Prox_F^*(z_k)), \\ x_{k+1} = Prox_F(z_{k+1}). \end{cases} \tag{11}$$

Combining all these steps together, we obtain the proposed weakly supervised variational segmentation algorithm(WSVS) in algorithm 1 for solving the model (10). The stopping criteria in our algorithm for u is $\frac{\| u^{k+1} - u^k \|_2}{\| u^k \|_2} \leqslant \epsilon$, where ϵ is fixed to 10^{-4}.

Algorithm 1. Algorithm of WSVS

Initialization: $\alpha = 1, \rho = 1, c_1 = 1, c_2 = 0$.
   Compute $D_1(x)$ and $D_2(x)$ with Fast Marching Method.
   For $k = 1$ to max iterations
   Calculate ω_1, ω_2.
   Calculate $c_1 = \frac{\int_\Omega D_1(x)I(x)u(x)dx}{\int_\Omega D_1(x)u(x)dx}, c_2 = \frac{\int_\Omega (1 - D_1(x))I(x)u(x)dx}{\int_\Omega (1 - D_1(x))u(x)dx}$.
   Calculate $Prox_G^*(Prox_F^*(z_k))$.
   Update u^k to u^{k+1} using the DR iterations (11).
Output: u

---

5. Experiments

5.1. Dataset and evaluation measures

In this paper, we will validate the effectiveness of our method using a private dataset of pancreatic cysts and three public datasets about kidney tumors, lung tumors and liver tumors, respectively. The detailed introduction is as follows:

   Pancreatic Cyst Dataset: It contains 318 CT scans collected from Nanjing Drum Tower Hospital about pancreatic cysts with SCN (Serous Cystic Neoplasm, 85 cases), IPMN (Intraductal Papillary Mucinous Neoplasm, 95 cases), SPN (Solid Pseudopapillary Neoplasm, 88 cases), MCN (Mucinous Cystic Neoplasm, 70 cases). The CT scans have resolutions of 512×512 pixels with varying pixel sizes and slice thickness range from 1.5 mm to 5 mm. It is a private dataset.

   KiPA22 Dataset[3]: It contains 70 CT scans with labeled kidney and tumor masks from the Key Laboratory of Computer Network and Information Integration at Southeast University. The CT scans have resolutions of 512×512 pixels with varying pixel sizes and slice thickness range from 1 mm to 5 mm.

   MSD Lung Tumor Dataset[4]: It contains 96 CT scans with labeled lung tumor masks from the sixth task in the Medical Image Segmentation Decathlon Challenge. The CT scans have resolutions of 512×512 pixels with varying pixel sizes and slice thickness range from 1.5 mm to 5 mm.

   LiTS Liver Tumor Dataset[5]: It contains 131 CT scans with labeled liver tumor masks from seven different medical centers. The CT scans have resolutions of 512×512 pixels with varying pixel sizes and slice thickness range from 1 mm to 5 mm.

   In addition, we use four metrics to quantitatively evaluate the segmentation performance. Let G, R denote the ground truth and the segmentation result respectively. TP: true positive, FP: false positive, TN: true negative, FN: false negative, P: number of positive samples, N: number of negative samples.

   (1) Dice coefficient: It is used to evaluate the region overlapping ratio of predicted result and ground truth masks.

---

[3] KiPA22: https://kipa22.grand-challenge.org/

[4] MSD(Medical Segmentation Decathlon): http://medicaldecathlon.com/

[5] LiTS: https://competitions.codalab.org/competitions/17094#participate/

$$Dice = \frac{2|R \bigcap G|}{|R| + |G|}.$$

(2) Hausdorff distance: It is used to determine the maximum error margin.

$$HD(G, R) = \max\{h(G, R), h(R, G)\},$$

where

$$h(G, R) = \max_{a \in G}\{\min_{b \in R}\|a - b\|\},$$
$$h(R, G) = \max_{b \in R}\{\min_{a \in G}\|b - a\|\}.$$

(3) Accuracy: percentage of the number of correctly predicted samples to the total number of samples.

$$Accuracy = \frac{TP + TN}{P + N}.$$

(4) Precision: percentage of the number of correctly classified positive samples to the number of samples predicted to be positive.

$$Precision = \frac{TP}{TP + FP}.$$

Thus, for Dice coefficient, Accuracy and Precision, higher scores indicate better segmentation performance. For Hausdorff distance, lower scores indicate better segmentation performance.

5.2. Experimental design

In this section, we design four kinds of experiments to evaluate our proposed weakly supervised variational segmentation model.

Experiment 1: Ablation study. To illustrate the effect of geodesic distance $\mathcal{D}_1(x)$ and Euclidean distance $\mathcal{D}_2(x)$ in our model, we conduct three ablation experiments with different term combinations to demonstrate their performance.

Experiment 2: Robustness of Annotation Point. To verify the robustness of our method about the position of the annotation point, we choose five positions in the target region, three of which are taken inside the target area, one is taken on the boundary as well as one is taken outside the target area.

Experiment 3: Comparison with other weakly supervised methods. We compare different methods to validate the contributions of our proposed weakly supervised variational segmentation approach.

Experiment 4: Multi-objective segmentation. We show the application of our method to the simultaneous segmentation of multiple targets.

5.3. Experiment 1: ablation study

In order to analyze the individual impact of Euclidean distance and the geodesic distance in the model (9), we use only one of these distances as a weakly supervised variational model to segment pancreatic cysts, kidney tumors and lung tumors for comparison, respectively. Figure 6 and table 1 shows the segmentation results of our method on the three datasets. Furthermore, the computational time per case is recorded for all three datasets, as detailed in table 2.

Effectiveness of geodesic distance: without incorporating the geodesic distance $\mathcal{D}_1(x)$, pixels that are close to the target area and have similar intensity cannot be excluded, see the third column of figure 6.

Effectiveness of Euclidean distance: without considering the Euclidean distance $\mathcal{D}_2(x)$, pixels that are far away from the target area and have similar intensity cannot be excluded, see the fourth column of figure 6.

Effectiveness of Euclidean distance and geodesic distance: it can be observed that the baseline model(1) without the Euclidean and geodesic distance leads to the worst performance, indicating that the simultaneous actions of Euclidean and geodesic distance are important, see the second column and the fifth column of figure 6.

5.4. Experiment 2: robustness of annotation point

According to the previous analysis of the proposed model, the point annotation can theoretically be taken at any position within the target region. Therefore, we choose different positions in the target region to verify the robustness of our algorithm about the position of the annotated point. We experiment with the annotated points taken at five different locations, three of which are taken inside the target area, one is taken on the boundary as well as one is outside the target area, and show the segmentation results for the five different locations in figure 7. From table 3, it is observed that as long as the annotated points are taken inside the target region, there is not much difference in the comparison of segmentation results. However, if the annotated
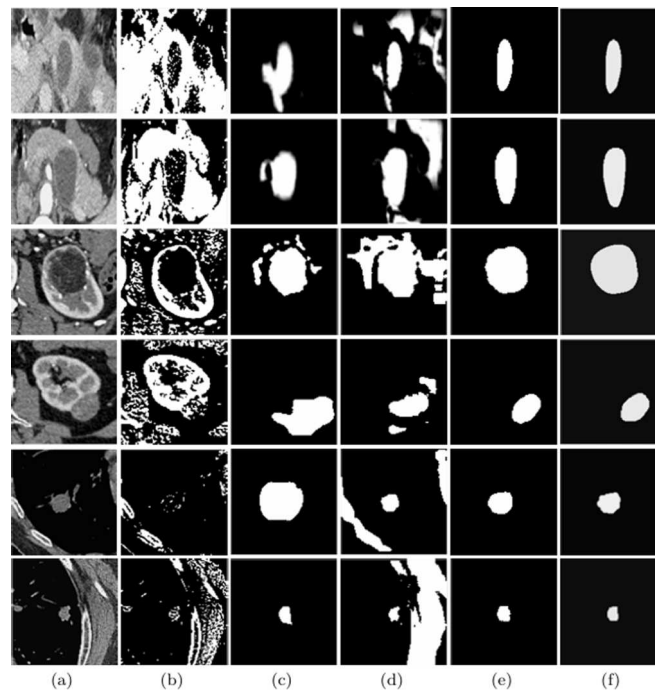
Figure 6. Ablation study. (a): Original images, (b): Segmentation results of baseline model (without geodesic distance and Euclidean distance, w/o GD & ED), (c): Segmentation results without geodesic distance (w/o GD), (d): Segmentation results without Euclidean distance (w/o ED), (e): Segmentation results with geodesic distance and Euclidean distance (ED&GD), (f): Ground truth.

Table 1. Comparisons of the individual effects of different distances on three datasets ($p < 0.01$).

Methods	Pancreatic Cyst			
	Dice ↑	HD ↓	Accuracy ↑	Precision ↑
w/o GD	64.78±7.21	16.43±8.60	63.27±10.80	65.89±5.02
w/o ED	55.62±8.94	24.32±7.33	53.60±9.69	55.65±8.27
w/o GD&ED	41.76±5.76	37.49±3.59	43.21±8.83	44.49±7.54
GD&ED	**72.25±6.04**	**10.39±4.73**	**76.82±6.28**	**75.99±5.15**

Methods	Kidney Tumor			
	Dice ↑	HD ↓	Accuracy ↑	Precision ↑
w/o GD	68.44±5.20	9.37±6.96	72.05±11.47	73.85±7.65
w/o ED	56.74±6.67	13.65±5.73	62.65±10.35	64.52± 9.94
w/o GD&ED	41.41±10.82	20.16±10.95	48.52±11.06	51.24±13.62
GD&ED	**80.42±5.91**	**6.35±4.94**	**79.81±10.69**	**81.22±5.16**

Methods	Lung Tumor			
	Dice ↑	HD ↓	Accuracy ↑	Precision ↑
w/o GD	71.07±7.89	13.27±3.42	73.77±10.08	72.30±9.22
w/o ED	63.61±5.16	19.36±3.89	64.50±9.55	61.51± 8.87
w/o GD&ED	48.22±11.10	27.09±5.54	43.90±12.37	49.69±7.66
GD&ED	**76.59±6.43**	**8.73±2.10**	**80.72±10.13**	**77.50±6.85**

points are taken on the boundary of the target region, some of the results appear to be interfered by the surrounding tissues. In addition, if the annotated points are taken outside the target region, the segmentation results are very poor. Therefore, it is concluded that the point annotation in our proposed method can be taken anywhere within the target region, but not on the boundary or outside the target region, which is also consistent with our previous theoretical analysis.
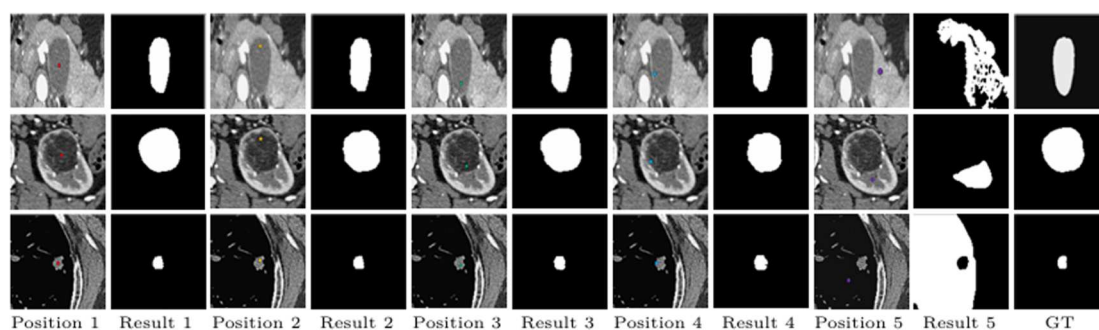
Figure 7. Robustness of Point Annotation. The first, third, fifth, seventh and ninth columns represent that the annotated points are located at different positions in the target region. The second, fourth, sixth, eighth and tenth columns represent the corresponding segmentation results, respectively. The last column is the corresponding ground truth.

Table 2. Average computation time for all cases on the three datasets respectively.

Dataset	Time(min/case)
Pancreatic Cyst	4.66
Kidney Tumor	5.35
Lung Tumor	5.89

Table 3. Segmentation results for annotated points taken at five different locations on three different datasets, in terms of Dice score statistic ($p < 0.01$).

Dataset	Result 1	Result 2	Result 3	Result 4	Result 5
Pancreatic Cyst	72.51±10.04	72.83±9.46	72.60±11.62	71.29±10.77	16.88±7.94
Kidney Tumor	81.71±8.86	80.67±9.09	80.84±9.72	79.50±10.39	28.16±15.89
Lung Tumor	77.29±7.85	76.34±5.85	76.23±8.83	74.76±7.07	10.32±9.57

5.5. Experiment 3: comparison with other weakly supervised methods

We compare the proposed weakly supervised variational segmentation framework with six other weakly supervised segmentation methods based on weak annotations: (1) Graph cut segmentation algorithm with scribbles drawn inside and outside the target region (Graph-cut) [14]. (2) Grab cut segmentation algorithm based on a rectangular bounding box containing the target region (Grab-cut) [9]. (3) Selective segmentation method incorporating variational model and the boundary point annotations (Selective-seg) [48]. (4) Weakly supervised learning segmentation using 7 point annotations (PA-seg) [21]. (5) Weakly supervised segmentation combining topological prior and deep tight prior with volume losses (DTP) [49]. (6) Segment anything model using pre-training and fine-tuning of large datasets (SAM) [50]. Furthermore, we compare the baseline model with the fully supervised model, where the baseline model is the CV model without embedding the information of the two distances, while the fully supervised model is nnUNet [51].

The first three segmentation methods are weakly supervised segmentation methods based on traditional optimization algorithms, while the last three are weakly supervised segmentation methods based on deep learning. In addition, the bounding box annotation for DTP uses the bounding box generated by the seven point annotations in the PA-seg approach. Since our proposed method takes the point only once and is not interactive, Graph-cut and Grab-cut also provide weak annotations only once for a fair comparison, but of course their segmentation results can be better if they take weak annotations further interactively. Except for SAM, whose weak annotation is also a single-point annotation, the weak annotations of the other comparison methods are harder to obtain than our method, especially Selective-seg and PA-seg need take multiple annotated points at the boundary, which are very time-consuming and laborious. However, SAM requires a large amount of datasets for pre-training in the early stage.

As can be seen from figure 8, the segmentation results of Graph-cut, Grab-cut and DTP all have significant errors at the boundaries because these methods only consider the voxel or topological information of the target
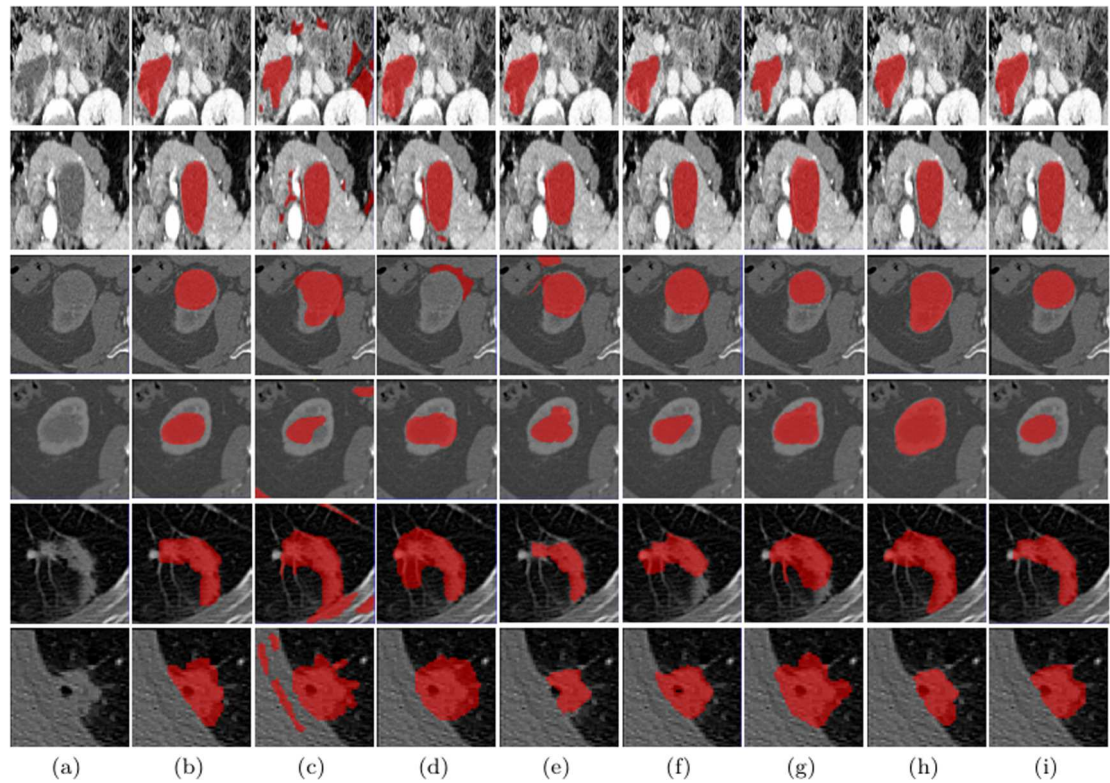
Figure 8. Comparison with the other weakly annotation-based methods about the segmentation of pancreatic cysts, kidney tumors and lung tumors. (a): Original images, (b): Ground truth, (c): Graph-cut, (d): Grab-cut, (e): Selective-seg, (f): PA-seg, (g): DTP, (h): SAM, (i): The proposed.

Table 4. Results of our method compared with other weakly supervised segmentation methods on the pancreatic cyst dataset for four evaluation metrics ($p < 0.01$).

Methods	Pancreatic Cyst			
	Dice ↑	HD ↓	Accuracy ↑	Precision ↑
Baseline	41.76±5.76	37.49±3.59	43.21±5.83	44.49±2.54
Full supervision	78.25±9.41	8.61±11.45	74.46±7.33	77.36±12.86
Graph-cut	57.34±12.15	21.63±8.82	66.76±10.44	67.52±11.12
Grab-cut	53.60±13.29	24.29±11.06	63.12±7.71	62.61±9.45
Selective-seg	64.18±9.61	19.81±9.39	69.83±9.36	69.83±8.64
PA-seg	67.65±11.73	15.98±7.62	72.55±7.05	72.72±12.73
DTP	65.76±10.80	17.31±9.03	70.72±12.25	66.64±7.03
SAM	70.39±11.48	12.53±8.27	75.87±9.91	70.06±11.57
The proposed	**72.25±6.04**	**10.39±4.73**	**76.82±6.28**	**75.99±5.15**

region, and do not take into account more image features such as intensity information and texture direction information. In addition, although both PA-seg and Selective-seg utilize the geodesic distance to generate the pseudo-labels or extract weakly supervised information, they only consider the magnitude of the gradient of the intensity at the boundary in the geodesic distance, and do not take into account the texture pattern and the orientation information of the intensity at the boundary, and thus the segmentation results of these two methods are still interfered by the surrounding tissues. SAM utilizes a large number of images to pre-train a large model, and although texture information can be well extracted from the target region, some errors still occur at the boundaries of the target region. In contrast, our proposed method can segment the target boundary more accurately. From tables 4–6, the proposed method gives similar or better results than other methods.

5.6. Experiment 4: multi-objective segmentation
Our method can also be applied to multi-objective segmentation , and can be applied to the detection and segmentation of cells and multiple tumors. As shown in figure 9, we randomly and manually provide one point
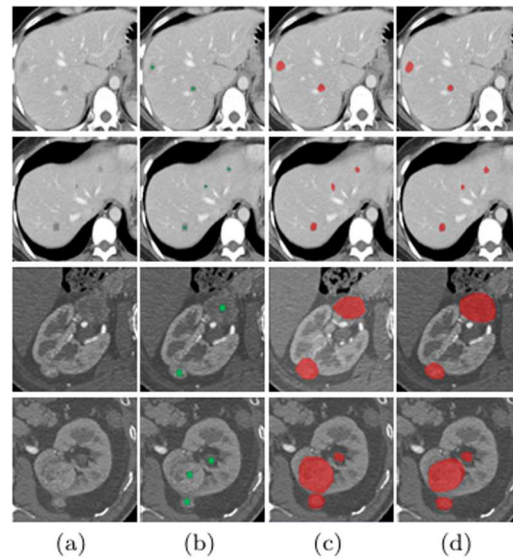
Figure 9. Multi-objective segmentation about liver tumors and kidney tumors. (a): Original images, (b): Point-annotation(Green), (c): Segmentation results, (d): Ground truth.

Table 5. Results of our method compared with other weakly supervised segmentation methods on the kidney tumor dataset for four evaluation metrics ($p < 0.01$).

Method	Kidney Tumor			
	Dice ↑	HD ↓	Accuracy ↑	Precision ↑
Baseline	41.41±12.82	20.16±10.95	48.52±11.06	51.24±13.62
Full supervision	87.40±6.08	4.25±2.81	82.86±5.37	86.09±6.20
Graph-cut	68.92±10.81	14.52±9.43	72.54±7.18	72.86±8.08
Grab-cut	64.43±9.25	15.31±10.55	66.36±8.90	68.22.±7.16
Selective-seg	71.86±7.94	11.65±5.27	77.25±9.42	77.43±9.49
PA-seg	76.47±6.63	9.05±8.83	79.06±10.19	80.99±4.97
DTP	73.05±7.82	10.75±5.08	75.65±7.73	75.63±10.37
SAM	75.60±9.44	8.78±12.11	76.63±10.90	77.85±11.24
The proposed	**80.42±5.91**	**6.35±4.94**	**79.81±10.69**	**81.22±5.16**

Table 6. Results of our method compared with other weakly supervised segmentation methods on the lung tumor dataset for four evaluation metrics ($p < 0.01$).

Method	Lung Tumor			
	Dice ↑	HD ↓	Accuracy ↑	Precision ↑
Baseline	48.22±3.10	27.09±5.54	43.90±12.37	49.69±17.66
Full supervision	80.13±10.42	3.69±4.35	85.15±4.06	82.30±5.17
Graph-cut	70.59±7.21	15.09±5.54	69.64±5.15	70.11±6.88
Grab-cut	69.72±6.58	18.29±3.65	71.60±4.63	66.51±4.25
Selective-seg	73.02±3.61	12.28±4.86	77.19±6.17	75.49±3.75
PA-seg	75.23±4.89	9.17±3.28	79.32±2.09	77.02±5.31
DTP	71.60±5.73	11.17±4.81	76.70±5.97	72.69±3.03
SAM	75.18±8.55	10.63±7.74	78.19±4.03	75.41±9.92
The proposed	**76.59±6.43**	**8.73±2.10**	**80.72±10.13**	**77.50±6.85**

within each target region. By using these points, we extract the textural features and intensity information specific to each tumor region. This allows us to exclude areas with similar intensity located near or far from the tumor, ultimately achieving satisfactory segmentation results, as shown in table 7.

Table 7. Results of multi-objective segmentation on the liver tumor and kidney tumor dataset for four evaluation metrics ($p < 0.01$).

Dataset	Dice ↑	HD ↓	Accuracy ↑	Precision ↑
Liver Tumor	73.55±14.73	11.20±15.61	72.73±12.28	74.82±16.97
Kidney Tumor	80.42±5.91	6.35±4.94	79.81±10.69	81.22±5.16

5.7. Limitations

We have successfully evaluated our proposed segmentation approach. But we admit that there are still some rooms for further improvements to our proposed method.

(1) In this study, two different kinds of distances and the variational model both act on the 2D images. Thus, some voxel information may be not extracted and utilized in this process. In the next study, the different distances and the variational model may be improved to 3D case.

(2) Due to space limitations, this paper focuses on how to extract the valuable information through a very weak supervision, and only incorporates a variational model to analyze the effectiveness of the information. As discussed earlier, it is also interesting to investigate how to better utilize the extracted information, such as incorporating AI algorithms and so on, which is a subsequent research topic.

(3) Our method is currently only applied to medical images, but in future work we will generalize our method to natural images. In addition, we will consider employing an automated point selection strategy to replace manual annotation for enhancing our approach.

6. Conclusion

We present a very weakly supervised variational segmentation model based on different distance information with a single point annotation within the target area, to joinly extract local appearances, global regional information, the local edge directional information and textural patterns. We firstly need to manually take only single random point in the target area, and then calculate two different kinds of distances from all points in the image to that point, Euclidean distance and geodesic distance. The Euclidean distance is mainly used to exclude the areas that are similar in intensity to the target region but are far away. While the geodesic distance is used to exclude the areas that are close to the target area and have a similar intensity. The most crucial step in the process of calculating the geodesic distance is the construction of the metric, which is an anisotropic Riemannian metric. We design a Riemannian metric by using the eigenvalues and eigenvectors of the structure tensor field, which takes into account both the local edge direction and textural pattern, while multiplying the regional information as a conformal factor by it. Secondly, to exploit the information extracted through our proposed one-point annotation, we embed the two kinds of distances simultaneously into a traditional variational model to guide the segmentation pointwisely. Experimental results show that our method achieves similar or better localization and segmentation performance than other weakly annotated methods.

In summary, this research proposes an efficient very weak annotation approach for tumor detection, localization and segmentation, which can greatly reduce the workload of labeling the training data. Of course, it may be less accurate than per-pixel annotations but useful for creating large-scale datasets for screening tumors in various medical modalities.

Acknowledgments

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

References

[1] Chou Y C, Li B, Fan D P, Yuille A and Zhou Z 2024 *Mach. Intell. Res.* **21** 318–30
[2] Zhang S, Liew J H, Wei Y, Wei S and Zhao Y 2020 *Interactive Object Segmentation with Inside-outside Guidance Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 12234–44

[3] He R, Dong Q, Lin J and Lau R W 2023 Weakly-supervised camouflaged object detection with scribble annotations *Proceedings of the AAAI Conference on Artificial Intelligence vol 37 (AAAI)* 781–9

[4] Li W, Liu W, Zhu J, Cui M, Hua X S and Zhang L 2022 Box-supervised instance segmentation with level set evolution *European Conference on Computer Vision* (Springer) 1–18

[5] Dorent R, Joutard S, Shapey J, Bisdas S, Kitchen N, Bradford R, Saeed S, Modat M, Ourselin S and Vercauteren T 2020 *Scribble-based Domain Adaptation via Co-segmentation Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference* (Springer) 479–89

[6] Yang R, Song L, Ge Y and Li X 2023 *Boxsnake: Polygonal Instance Segmentation with Box Supervision Proceedings of the IEEE/CVF International Conference on Computer Vision* (IEEE) 766–76

[7] Li W, Yuan Y, Wang S, Zhu J, Li J, Liu J and Zhang L 2023 *Point2mask: Point-supervised Panoptic Segmentation via Optimal Transport Proceedings of the IEEE/CVF International Conference on Computer Vision* (IEEE) 572–81

[8] Yu Q, Du H, Liu C and Yu X 2024 *When 3d Bounding-box Meets Sam: Point Cloud Instance Segmentation with Weak-and-Noisy Supervision Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (IEEE) 3719–28

[9] Rother C, Kolmogorov V and Blake A 2004 *ACM Transactions on Graphics (TOG)* **23** 309–14

[10] Li Y, Zhang J, Gao P, Jiang L and Chen M 2018 Grab cut image segmentation based on image region *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)* (IEEE) 311–5

[11] He K, Wang D, Tong M and Zhu Z 2020 *Pattern Recognit* **103** 107292

[12] Boykov Y Y and Jolly M P 2001 Interactive graph cuts for optimal boundary & region segmentation of objects in nd images *Proceedings eighth IEEE International Conference on Computer Vision. ICCV 2001* 1 (IEEE) 105–12

[13] Yi F and Moon I 2012 Image segmentation: A survey of graph-cut methods *International Conference on Systems and Informatics (ICSAI2012)* (IEEE) 1936–41

[14] Kwatra V, Schödl A, Essa I, Turk G and Bobick A 2003 *Acm Transactions on Graphics (tog)* **22** 277–86

[15] Roberts M, Chen K and Irion K L 2019 *J. Math. Imaging Vision* **61** 482–503

[16] Rada L and Chen K 2013 *J. Algorithms Comput. Technol.* **7** 509–40

[17] Rahman A *et al* 2022 *IEEE Access* **10** 22344–58

[18] Zhao T and Yin Z 2020 *IEEE Trans. Med. Imaging* **40** 2736–47

[19] Zhang H, Burrows L, Meng Y, Sculthorpe D, Mukherjee A, Coupland S E, Chen K and Zheng Y 2023 Weakly supervised segmentation with point annotations for histopathology images via contrast-based variational model *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE) 15630–40

[20] Gao Z *et al* 2023 *Med. Image Anal.* **83** 102652

[21] Zhai S, Wang G, Luo X, Yue Q, Li K and Zhang S 2023 *IEEE Trans. Med. Imaging* **42** 2235–46

[22] Xie R, Yang Y and Chen Z 2023 *Pattern Recognit* **133** 108974

[23] Long Z, Gao Y, Meng H, Yao Y and Li T 2022 *Inf. Sci.* **600** 263–86

[24] Barath D and Matas J 2021 *IEEE Trans. Pattern Anal. Mach. Intell.* **44** 4961–74

[25] Feng J, Wang X and Liu W 2021 *Sci. China Inf. Sci.* **64** 1

[26] Bearman A, Russakovsky O, Ferrari V and Fei-Fei L 2016 What's the point: semantic segmentation with point supervision *European Conference on Computer Vision* (Springer) 549–65

[27] Chen T, Yao Y, Huang X, Li Z, Nie L and Tang J 2024 *IEEE Trans. Image Process.* **33** 1136–48

[28] Khadidos A, Sanchez V and Li C T 2017 *IEEE Trans. Image Process.* **26** 1979–91

[29] Cai Q, Qian Y, Zhou S, Li J, Yang Y H, Wu F and Zhang D 2021 *IEEE Trans. Image Process.* **31** 43–57

[30] Tian L, Zou L and Yang X 2023 *Phys. Med. Biol.* **68** 145005

[31] Zhang W, Wang X, Chen J and You W 2020 *IEEE Trans. Image Process.* **29** 7032–44

[32] Chan T F and Vese L A 2001 *IEEE Trans. Image Process.* **10** 266–77

[33] Chan T F, Esedoglu S and Nikolova M 2006 *SIAM J. Appl. Math.* **66** 1632–48

[34] Peyré G *et al* 2010 *Foundations and Trends® in Computer Graphics and Vision* **5** 197–397

[35] Chen D, Mirebeau J M, Shu M and Cohen L D 2023 *Proc. Natl. Acad. Sci.* **120** e2218869120

[36] Lions P L 1985 *SIAM Rev.* **27** 268–9

[37] Evans L C 2022 *Partial differential equations* 19 (American Mathematical Society)

[38] Sethian J A 1996 *Proc. Natl. Acad. Sci.* **93** 1591–5

[39] Benmansour F and Cohen L D 2011 *Int. J. Comput. Vision* **92** 192–210

[40] Mirebeau J M 2014 *SIAM J. Numer. Anal.* **52** 1573–99

[41] Harris C G and Stephens M J 1988 A combined corner and edge detector *Proceedings of the Alvey Vision Conference* (Alvey Vision Club) 1–6

[42] Kimmel R and Bruckstein A M 2003 *Int. J. Comput. Vision* **53** 225–43

[43] Law M W and Chung A 2010 An oriented flux symmetry based active contour model for three dimensional vessel segmentation *European Conference on Computer Vision* (Springer) 720–34

[44] Law M W K and Chung A C S Three dimensional curvilinear structure detection using optimally oriented flux *European Conference on Computer Vision* 368–82

[45] Weickert J 1998 *B.g.teubner Stuttgart* **16** 272

[46] Beck A 2017 First-order *methods in optimization (SIAM)* (https://doi.org/10.1137/1.9781611974997.ch11)

[47] Combettes P L and Pesquet J C 2007 *IEEE J. Sel. Top. Signal Process.* **1** 564–74

[48] Liu C, Ng M K P and Zeng T 2018 *Pattern Recognit* **76** 367–79

[49] Kervadec H, Dolz J, Wang S, Granger E and Ayed I B 2020 *Bounding Boxes for Weakly Supervised Segmentation: Global Constraints get close to full Supervision Medical Imaging with Deep Learning (PMLR)* **121** 365–81

[50] Kirillov A *et al* 2023 Segment anything *Proceedings of the IEEE/CVF International Conference on Computer Vision* (IEEE) 4015–26

[51] Isensee F, Petersen J, Kohl S A, Jäger P F and Maier-Hein K H 2019 arXiv:1904.08128