Variational Wasserstein Model on Riemannian Manifolds for Image Segmentation

Jisui Huang
University of Liverpool

Yue Wu
University of Strathclyde

Ke Chen*
University of Strathclyde

Na Lei*
Dalian University of Technology

NEURAL INFORMATION PROCESSING SYSTEMS
2026

Related Work

Let $I \in L^1(\Omega, \mathbb{R}^C)$ denote an image with C channels on an open cubic domain $\Omega \subset \mathbb{R}^D$. We take I to be zero outside Ω.

Measure Theorem. Let X and Y be two Riemannian manifolds. For a Borel map $T : X \to Y$ and a measure μ on X, the push-forward measure $T_{\#}\mu$ is defined by

$$T_{\#}\mu(B) = \mu(T^{-1}(B))$$

for every measurable set $B \subset Y$. When no ambiguity arises, we also write $T_{\#}u$ by interpreting u as the density of a measure.

Optimal Transport on a Riemannian Manifold. Optimal transport (OT) distance between two probability distributions μ and ν on a Riemannian manifold M is defined as

$$\mathcal{T}(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{M\times M} c(x, y)\, \mathrm{d}\pi(x, y),$$

where $c(x, y)$ is the cost of transporting one unit of mass from x to y, and $\Pi(\mu, \nu)$ is the set of feasible transport plans with marginals μ and ν. When $c(x, y) = d(x, y)^2$, where d is the shortest distance on M, this OT distance gives the squared 2-Wasserstein distance.

The Kantorovich duality of the primal OT problem is given by

$$\mathcal{T}(\mu, \nu) = \sup_{\psi \in L^1(\nu)} \left\{ \int_M \psi^c(x)\, \mathrm{d}\mu(x) + \int_M \psi(y)\, \mathrm{d}\nu(y) \right\},$$

where the c-transform of ψ is defined by

$$\psi^c(x) = \inf_{y \in M} \{c(x, y) - \psi(y)\}. \tag{1}$$

Here ψ is referred to as the Kantorovich dual potential.

Papadakis–Rabin (PR) Model. The PR model [2] seeks a segmentation $u : \Omega \to [0, 1]$ such that $\{x \in \Omega : u(x) > 0.5\}$ denotes the foreground region:

$$\min_{u(x)\in[0,1]} \quad \gamma \int_\Omega |\nabla u|\, \mathrm{d}x + \mathcal{T}(I_{\#}u, \lambda^{\mathrm{f}}\nu^{\mathrm{f}}) + \mathcal{T}(I_{\#}(1-u), \lambda^{\mathrm{b}}\nu^{\mathrm{b}}),$$

where ν^{f} and ν^{b} are reference foreground and background probability measures on the color space, and

$$\lambda^{\mathrm{f}} = \int_\Omega u(x)\, \mathrm{d}x, \quad \lambda^{\mathrm{b}} = \int_\Omega (1 - u(x))\, \mathrm{d}x$$

can ensure that the two input measures of OT have the same mass.

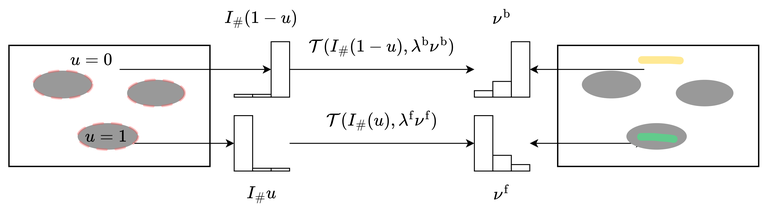

Figure 1. The PR model uses OT to match a region's color histogram with user-specified reference color histograms. The foreground and background OT produce competing forces to guide the segmentation contour evolution. However, since the PR model only conducts OT in RGB color space, the model usually fails for objects with similar colors.

Our Motivation

The PR model only considers color distributions and ignores spatial and edge information. Therefore, we propose a new model that

- incorporates color, edge, and spatial features; and
- compares features globally.

We address these problems by:

- applying geodesic distance on the feature manifold to the OT cost function.

Proposed Model

The feature space $M = F(\mathbb{R}^D) \subset \mathbb{R}^{D+C}$ is a submanifold in the ambient space $\mathbb{R}^{D+C}$ where the feature map $F\colon \mathbb{R}^D \to \mathbb{R}^{D+C}$ is defined by $F(x) = (\alpha x, \beta(K_\epsilon * I)(x))$, where $\alpha, \beta > 0$ balance spatial and intensity information; K_ϵ is a mollifier with kernel width ϵ.

The segmentation model $\min_{u\in\mathcal{A}} E(u)$ minimizes the energy

$$E(u) = \gamma \int_\Omega |\nabla u|\, \mathrm{d}x + \mathcal{T}(F_{\#}u, \lambda^{\mathrm{f}}F_{\#}m^{\mathrm{f}}) + \mathcal{T}(F_{\#}(1-u), \lambda^{\mathrm{b}}F_{\#}m^{\mathrm{b}}), \tag{2}$$

over the admissible space

$$\mathcal{A} = \{u \in L^1(\Omega) : \eta \le u(x) \le 1 - \eta \text{ for a.e. } x \in \Omega\}, \tag{3}$$

where the small constant $\eta > 0$ can prevent the vanishing of the derivative of the OT cost. Here $\mathcal{T}$ denotes the squared 2-Wasserstein distance on M; m^{f} and m^{b} are the discrete probability measures defined on user markers.

Solution Properties

Theorem 1 (Existence). *There exists a minimizer $u \in BV(\Omega)$ of the functional (2) over the admissible set $\mathcal{A}$ (3).*

Theorem 2 (Convexity). *The functional (2) is convex with respect to $u \in \mathcal{A}$ (3).*

The first two theorems show that no matter how u is initialized, all local minimizers lie in a convex set in the admissible space.

Theorem 3 (Stability). *Let u_n be a minimizer of (2) over $\mathcal{A}$ with hyperparameters $\alpha_n > 0$, $\beta_n > 0$, and $\epsilon_n > 0$. Suppose that*

$$\alpha_n \to \alpha, \quad \beta_n \to \beta, \quad \epsilon_n \to \epsilon.$$

Then any $L^1(\Omega)$ cluster point u of (u_n) in $\mathcal{A}$ (3) is a minimizer of (2) over $\mathcal{A}$ with parameters α, β and ϵ.

This implies that segmentation will not vary significantly with hyperparameters.

Optimization

In a primal-dual framework, the segmentation variable u in the OT is updated by

$$\begin{aligned} D\mathcal{T}(F_{\#}(u), \lambda^{\mathrm{f}}F_{\#}m^{\mathrm{f}}) &= ((\psi^{\mathrm{f}})^c \circ F)(x) + \int_\Omega \psi^{\mathrm{f}}(F(x))\, \mathrm{d}m^{\mathrm{f}}(x), \\ D\mathcal{T}(F_{\#}(1-u), \lambda^{\mathrm{b}}F_{\#}m^{\mathrm{b}}) &= -((\psi^{\mathrm{b}})^c \circ F)(x) - \int_\Omega \psi^{\mathrm{b}}(F(x))\, \mathrm{d}m^{\mathrm{b}}(x), \end{aligned} \tag{4}$$

where the intermediate Kantorovich potentials ψ^{f} and ψ^{b} are computed using semi-discrete OT [1].

In our setting, the OT admits a transport map (see Theorem 10.41 of [3]) on the Riemannian manifold given as

$$T(x) = \exp_x(-\frac{1}{2}\nabla\psi^c(x)),$$

which leads to the following differentiability property of the OT terms with respect to the segmentation variable u:

Theorem 4 (Differentiability). *The Fréchet derivatives of the foreground OT term $\mathcal{T}(F_{\#}(u), \lambda^{\mathrm{f}}F_{\#}m^{\mathrm{f}})$ and the background OT term $\mathcal{T}(F_{\#}(1-u), \lambda^{\mathrm{b}}F_{\#}m^{\mathrm{b}})$ with respect to the segmentation result u under the $L^1(\Omega)$ norm are $L^\infty(\Omega)$ functions and are given in (4).*

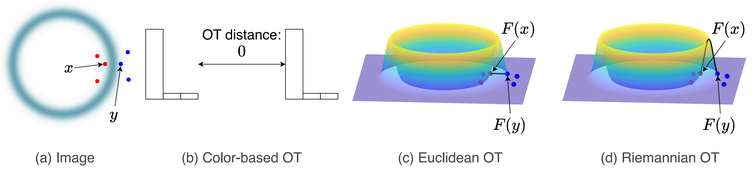

Figure 2. Different OT cost functions. (a): Input image. (b): Color-based OT in the PR model yields zero distance because both regions appear white. (c): Euclidean cost yields only a small distance because it ignores the edge slope. (d): Geodesic OT can simultaneously respect appearance similarity, spatial proximity, and boundary information, and separate the two regions well.

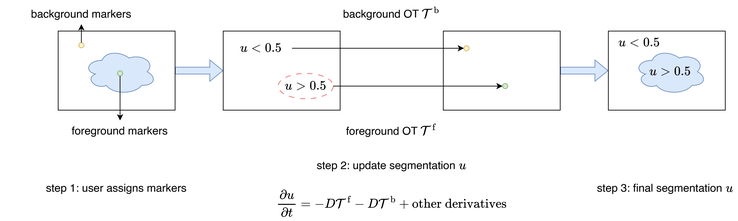

Figure 3. Pipeline of the proposed model. Given an image, the user can assign some foreground and background marker points. The model can directly update OT terms using Fréchet derivatives, while the TV term is handled via the primal-dual method. The final segmentation is the region $u > 0.5$.

Model Evaluation

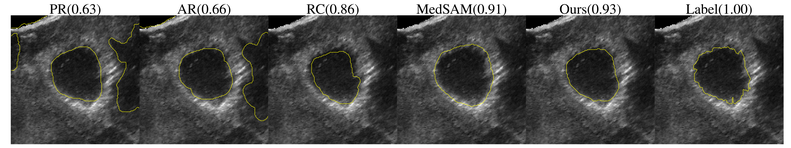

Figure 4. Segmentation comparison on RESECT. Yellow contours denote segmentation boundaries; numbers in parentheses indicate the Dice similarity coefficient (DSC).

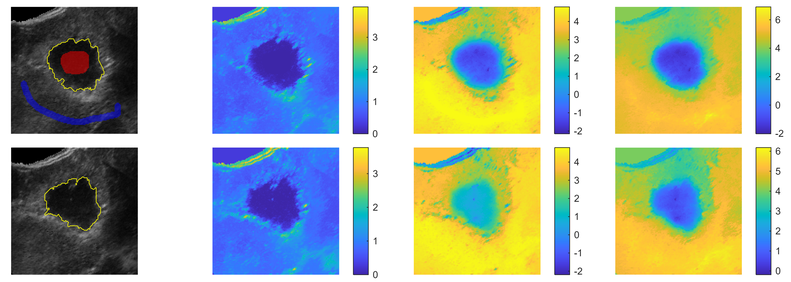

Figure 5. Intermediate variables in OT terms. From left to right: input with user markers, foreground OT derivative, background OT derivative, and their sum.

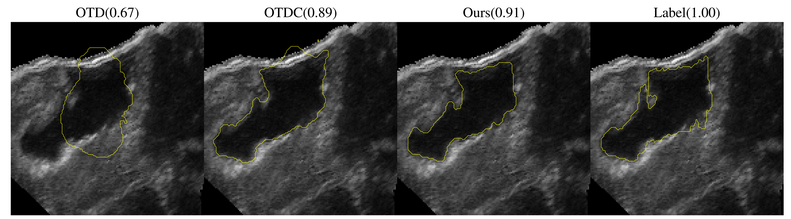

Figure 6. Ablation study. OTD: OT with spatial Euclidean distance only; OTDC: OT in joint color-spatial space using Euclidean distance, corresponding to Figure 2c; Ours: OT in joint spatial-color space using the geodesic distance, corresponding to Figure 2d.

References

[1] Dongsheng An, Yang Guo, Na Lei, Zhongxuan Luo, Shing-Tung Yau, and Xianfeng Gu. Ae-ot: A new generative model based on extended semi-discrete optimal transport. In *International Conference on Learning Representations*, 2020.

[2] Nicolas Papadakis and Julien Rabin. Convex histogram-based joint image segmentation with regularized optimal transport cost. *Journal of Mathematical Imaging and Vision*, 59:161–186, 2017.

[3] Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 2009.

*Corresponding authors

Ke Chen: ke.chen@strath.ac.uk

Na Lei: nalei@dlut.edu.cn