Research Article Vol. 16, No. 5 / 1 May 2025 / Biomedical Optics Express 1999
Biomedical Optics EXPRESS

*gc*DLSeg: integrating graph-cut into deep learning for binary semantic segmentation

Hui Xie,[1] Weiyu Xu,[1] Ya Xing Wang,[2] John Buatti,[3] and Xiaodong Wu[1,*]

[1] *Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 52242, USA*
[2] *Beijing Institute of Ophthalmology, Beijing Tongren Hospital, Capital University of Medical Science, Beijing Ophthalmology and Visual Sciences Key Laboratory, Beijing, China*
[3] *Department of Radiation Oncology, The University of Iowa, Iowa City, IA 52242, USA*
[*] *xiaodong-wu@uiowa.edu*

Abstract: Binary semantic segmentation in computer vision is a fundamental problem. As a model-based segmentation method, the graph-cut approach was one of the most successful binary segmentation methods thanks to its global optimality guarantee of the solutions and its practical polynomial-time complexity. Recently, many deep learning (DL) based methods have been developed for this task and yielded remarkable performance, resulting in a paradigm shift in this field. To combine the strengths of both approaches, we propose in this study to integrate the graph-cut approach into a deep learning network for end-to-end learning. Unfortunately, backward propagation through the graph-cut module in the DL network is challenging due to the combinatorial nature of the graph-cut algorithm. To tackle this challenge, we propose a novel residual graph-cut loss and a quasi-residual connection, enabling the backward propagation of the gradients of the residual graph-cut loss for effective feature learning guided by the graph-cut segmentation model. In the inference phase, globally optimal segmentation is achieved with respect to the graph-cut energy defined on the optimized image features learned from DL networks. Experiments on the public AZH chronic wound data set and the pancreas cancer data set from the medical segmentation decathlon (MSD) demonstrated promising segmentation accuracy and improved robustness against adversarial attacks.

1. Introduction

Image semantic segmentation, which partitions images into multiple segments on the pixel level, plays a fundamental role in computer vision applications [1], such as scene understanding, remote sensing, autopilot, medical image analysis, robotic perception, video surveillance, augmented reality, and image compression.

Prior to the revolutionary rise of deep learning, the graph-cut method [2–4] had been one of the major image segmentation approaches [5–7]. It was proposed by Greig et al. [8] and Boykov et al. [4] to formulate binary semantic segmentation as a minimum *s*-*t* cut problem in an associated graph. The graph-cut method is ubiquitous in computer vision as a large variety of computer vision problems can be formulated as a min-cut/max-flow problem [9]. It has shown remarkable potential for solving challenging segmentation tasks [10], for exact or approximate energy minimization in low-level vision with a practical polynomial-time complexity [4,9,11]. However, the graph-cut method heavily relies on a "good" cost function map [12] and purely using low-level pixel intensity features [13] does not give a "good" representation of the cost function in complicated image contexts.

With superior data representation learning capacity, deep learning (DL) methods are emerging as a new generation of image segmentation alternatives with remarkably improved performance over traditional image segmentation algorithms [1,14–18], resulting in a paradigm shift in the field. However, DL segmentation algorithms often need extensive training data [1,19], which

#555206 https://doi.org/10.1364/BOE.555206
Journal © 2025 Received 15 Jan 2025; revised 3 Apr 2025; accepted 13 Apr 2025; published 23 Apr 2025

Biomedical Optics EXPRESS

poses significant challenges, especially, for medical image segmentation due to patient privacy and high cost. In addition, almost all widely used segmentation models, such as UNet [20], FCNs [21], and DeepLab [22], are classification-based in nature and the output probability maps are relatively unstructured, thus lacking the capability of capturing global structures of the target objects. To characterize the long-range data dependency, transformer [23–25] has been introduced for semantic image segmentation, such as TransUNet [26], SwinUNet [27], DS-TransUNet [28], and nnFormer [29], which, however, substantially increases the inference cost and memory complexity of the segmentation models. Recent research has demonstrated that, compared to the segmentation CNNs alone, the integration of a graphical model such as conditional random fields (CRFs) into CNNs enhances the robustness of the method to adversarial perturbations [30–32].

Very recently, large data models for image segmentation like SAM [33] and MedSAM [34] have been emerging quickly. However, their high demands on computation resources, e.g. 1600GB GPU memory for training, heavily restrict their applications in general scientific research settings. Building a small model, e.g. our proposed model, needs only 24GB GPU memory for training, with textbook quality data [35] for a specific application is still a major and practical research direction.

In this paper, we propose to seamlessly integrate the traditional graph-cut and deep learning methods for binary image segmentation, unifying the strengths of both methods while alleviating the drawbacks of each individual one. Our proposed framework makes use of deep learning networks to learn a high-level feature cost map, and then apply the graph-cut method to achieve a globally optimal segmentation while minimizing the graph-cut energy function defined on the learned cost map. As feature learning and graph-cut optimization are unified in a single deep learning network for end-to-end training, the learned features are tailored specifically for the graph-cut segmentation model with backward propagation.

The challenge in this integration is how to incorporate graph-cut algorithms into deep learning with effective backward propagation support for model training. The combinatorial nature of graph-cut algorithms hampers their applications in deep learning networks due to ineffective backward propagation, as min-cut module self is not differentiable. Some practical solutions to those combinatorial optimization problems [36–38] are to utilize a good approximation by leveraging the special structures in the problem [37,39]. With the approximation surrogate, a fundamental quandary persists, that is, the gradient of the optimal solution with respect to the (dynamic) input variables (e.g., the graph edge weights in the graph-cut problem) is frequently zero, which is not helpful for backward propagation to optimize the network parameters [40]. Many methods for solving this zero gradient quandary have been explored. Pogančić et al. proposed a differentiation of black box combinatorial solver [40], which, however, doubles the computation burden because of invoking the combinatorial algorithm two times to obtain the gradients. Elmachtoub and Grigas' SPO (Smart "Predict, then Optimize") [41] directly leverages the optimization problem structure – that is, its objective and constraints – for designing better prediction models. Mensch et al. developed a differentiable dynamic programming method [42–43] using a surrogate function for the operator of maximum, which is problem-specific. Khalil et al. proposed learning combinatorial optimization algorithms on graphs by using neurons to construct graphs dynamically, and heuristically exploring the optimal solution by reinforcement learning mechanism [44]. Gasse et al. used graph convolutional neural networks [45] to reformulate branch-and-bound as a Markov decision process for solving a mixed-integer linear programming problem. Both Khalil et al.'s and Gasse et al.'s methods directly used networks to "simulate" graphs for combinatorial optimization, they yet lost the practical polynomial-time complexity for solving the graph-cut problem [4,9,11].

We proposed a novel residual graph-cut loss and a quasi-residual connection to seamlessly unify a U-Net [20] for feature learning with a graph-cut module to achieve end-to-end training and

Biomedical Optics EXPRESS

optimization for binary semantic segmentation. The proposed method is termed as gcDLSeg. It effectively utilizes the backward propagation from the downstream graph-cut optimization module to guide feature learning, yielding statistically significant improved segmentation accuracy and improved robustness against adversarial attacks. Experiments on public AZH chronic wound [46] data set and pancreas cancer data set from MSD (medical segmentation decathlon [47]) demonstrated promising segmentation performance. To the best of our knowledge, this is the first work to integrate graph-cut within deep learning for end-to-end training and inference in medical image applications. This proposed method has the potential to be adapted for broader DL with graph-cut applications.

2. Methods

2.1. Problem formulation

Greig et al. [8] and Boykov et al. [4] developed the graph-cut approach to formulate the problem of binary semantic segmentation as a minimum s-t (source-sink) cut problem by minimizing an energy function defined on a graph.

Let $\mathcal{I}$ denote the input image and $\mathcal{N}$ denote its neighboring setting. A labeling $f : \mathcal{I} \mapsto \{0, 1\}$ is a segmentation of the image $\mathcal{I}$, where we should interpret 0 and 1 as standing for "background (bkg)" and "foreground (obj)", respectively. For each voxel $p \in \mathcal{I}$, we are given a *data consistency* function $\alpha_p : \{0, 1\} \mapsto \mathbb{R}$, where $\alpha_p(0)$ ($\alpha_p(1)$) represents some pre-computed penalty for assigning voxel p to the background (foreground). Thus, $\alpha_p(f_p)$ measures the fit of the label f at each voxel p to the foreground ($f_p = 1$) and background ($f_p = 0$). Similarly, we are given for each voxel pair $(p, q) \in \mathcal{N}$, a *pairwise regularization* function $\beta_{pq} : \{0, 1\} \times \{0, 1\} \mapsto \mathbb{R}$, which is the penalty for assigning labels f_p and f_q to two neighboring voxels p and q. The purpose is to penalize label differences for any two adjacent voxels. For binary segmentation, the Ising model [48] is generally used to design the function β_{pq}, with $\beta_{pq}(f_p, f_q) = \psi(p, q) \cdot \delta(f_p = f_q)$, where $\psi(p, q)$ is used to model appearance and smoothness between the pair of voxels p and q, and $\delta(\cdot)$ is an identify function giving 1 if labels f_p and f_q are the same and 0 otherwise. The segmentation problem seeks to find an optimal labeling f such that the Markov Random Field (MRF) energy function [49] $\mathcal{E}(f)$ is minimized, with

$$\mathcal{E}(f) = \sum_{p \in \mathcal{I}} \alpha_p(f_p) + \gamma \sum_{(p,q) \in \mathcal{N}} \psi(p, q) \cdot \delta(f_p = f_q), \tag{1}$$

where $\gamma > 0$ provides the relative weighting between the data consistency terms and the pairwise regularization terms.

The graph-cut segmentation problem can be then readily formulated as a minimum s-t cut problem in an edge-weighted graph [4]. First, a graph node p represents a voxel in the image $\mathcal{I}$ (to simplify the notation, we use p to denote both a voxel and its corresponding graph node), and edges are added between each pair of nodes (p, q) corresponding to neighborhood voxels with weight $\psi(p, q)$. These edges between neighbors are often referred to as "n-links." Next, two extra "terminal" nodes, s (the "obj" terminal) and t (the "bkg" terminal) are added to the graph and for each graph node p corresponding to a voxel in the image, two edges are added (often called "t-links" for "terminal"): (s, p) and (p, t). The weight of each edge between t and node p, denoted by $\phi_t(p)$, is given by the background cost value for the associated voxel p: $\alpha_p(f_p = \text{'}bkg\text{'})$, which indicates the coherence between nodes p and t. The weight of each edge between node p and "obj" terminal s, denoted by $\phi_s(p)$, is given by the object cost value for the associated voxel p, that is, $\alpha_p(f_p = \text{'}obj\text{'})$, which indicates the coherence between nodes p and s. When determining the minimum s-t cut, $(\{s\} \cup S, \{t\} \cup T)$, of this constructed graph, the set S of the nodes remaining connected to the "obj" terminal s correspond to the voxels belonging to the segmented object; while the nodes in T correspond to the voxels belonging to the background.

Biomedical Optics EXPRESS

Thus, the binary semantic segmentation problem, as illustrated in Fig. 1, is formulated to find an s-t cut $C = (\{s\} \cup S, \{t\} \cup T)$ in the constructed graph, whose capacity

$$||C|| = \sum_{p \in S} \phi_t(p) + \sum_{p \in T} \phi_s(p) + \gamma \sum_{\substack{p \in S \\ q \in T \\ (p,q) \in \mathcal{N}}} \psi(p, q) \tag{2}$$

is minimized.

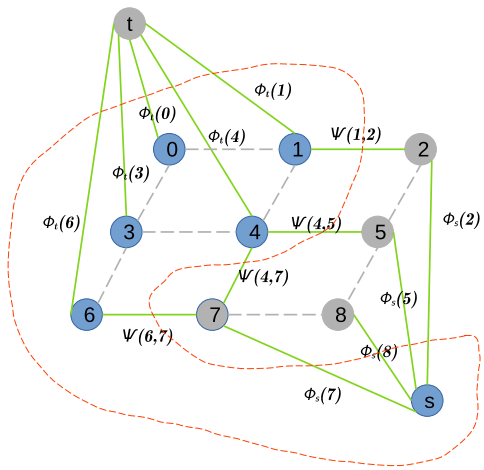

Fig. 1. The graph representation of a cut in a 4-neighbor graph system. Number 0,1,2,. . . ,8 indicate pixel nodes p_i in an image I where $i = 0, 1, 2, \ldots, 8$ for conciseness, nine pixel nodes are connected by gray dash lines, and s and t are virtual source and sink nodes. A red dash line cuts all nodes into two node sets: source node-set S blue, and sink node-net T gray. Green connections indicate all cut edges between S and T. ϕ_t and ϕ_s express t-links, and ψ expresses n-links. Other links that do not participate in current $S - T$ computations are not drawn on the figure for clarity.

To solve this binary semantic segmentation problem, the proposed *gc*DLSeg method first learns the edge weights $\phi_s(p)$ and $\phi_t(p)$ of t-links and $\psi(p, q)$ of n-links in the constructed graph via a U-Net model. Then, a max-flow algorithm [11] is used to find globally optimal binary labeling for each node p at the current edge weight setting of the graph learned from the network-in-training. The graph edge weight learning and the minimum s-t cut computation are implemented in a unified network for an end-to-end gradient backward propagation through a new residual graph-cut loss and a quasi-residual connection for better learning graph edge weights, further improving segmentation accuracy.

2.2. Network architecture of gcDLSeg

The proposed semantic segmentation network, *gc*DLSeg, is based on a U-Net architecture [20], as illustrated in Fig. 2, which consists of six convolution layers. This U-Net acts as a feature-extracting module for the downstream graph-cut segmentation head. The network starts with 48 feature maps in the first convolution layer. In each downsampling layer, a conv2d module followed by a 2x2 max-pooling doubles the feature maps, and then a cascade of three same conv2d modules with a standard residual connection [50]. The upsampling layers use one symmetric structure as the downsampling layers but with bilinear upsample modules. As we attempt to demonstrate the power of integrating the graph-cut segmentation method with deep learning, the U-Net used in this work was not specifically optimized. One may also choose any feature-extracting networks to replace the U-Net for specific applications.

Biomedical Optics EXPRESS

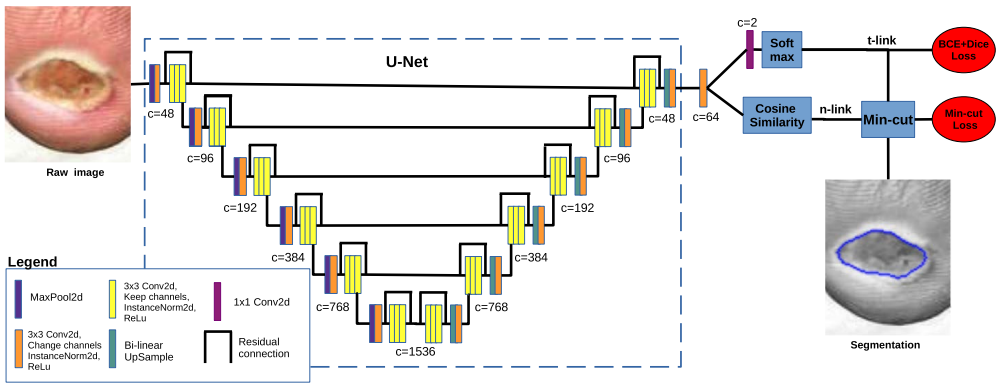

Fig. 2. The *gc*DLSeg combines U-Net with a graph-cut module for image semantic segmentation. The graph-cut module, with input t-link and n-link edge weights from the preceding feature learning network, solves the max-flow optimization problem for segmentation, outputs optimal segmentation results, and supports the gradient backward propagation of the graph-cut loss signal. c indicates the number of channels for each layer.

The graph-cut segmentation head in the proposed *gc*DLSeg has two branches (Fig. 2). The top t-link branch consists of a 1x1 conv2d and a soft-max module, which serves to learn the two t-link edge weights: $\phi_s(p)$ of the source-to-node edge and $\phi_t(p)$ of the node-to-sink edge for each node p. We set $\phi_s(p)$ to be the probability of each node p belonging to the source set S and $\phi_t(p)$ to be the probability of p belonging to the sink set T, with $\phi_s(p) + \phi_t(p) = 1.0$. According to the normalized cut method [13], which demonstrated that image segmentation based on low-level cues may not be able to produce a highly accurate segmentation, we compute the similarity of neighbor nodes with abstract high-level features, instead of using the standard Gaussian kernel over RGBXY low-level features [51,52]. Using the embedding similarity idea [53], the bottom n-link branch in Fig. 2 computes the cosine similarity between the feature vectors of two neighboring nodes (voxels), p and q, as follows:

$$\psi(p,q) = \frac{1}{2}(1.0 + \frac{<\vec{p},\vec{q}>}{||\vec{p}||\,||\vec{q}||}), \tag{3}$$

where $\vec{p}$ and $\vec{q}$ repersent the feature vectors of voxel nodes p and q, respectively.

Finally, the graph-cut module takes the input t-links and n-links to compute the minimum s-t cut $C_{min} = (\{s\} \cup S_{min}, \{t\} \cup T_{min})$, in which S_{min} defines the target object of segmentation.

2.3. Quasi-residual connection

We designed a quasi-residual connection, as shown in Fig. 3, with a residual graph-cut loss to solve the zero gradient quandary in our *gc*DLSeg model. The zero gradient quandary is, in some sense, similar to the gradient vanishing problem in RNN [54], but it is more difficult to resolve in practice. We consider a residual connection [50,55] to provide a pathway for gradients to back-propagate to early layers of the network to avoid vanishing gradient [56] or zero gradients. The graph-cut module exposes zero gradient to its weights input because of its essence of combinatorial algorithm. However, our residual connection restores the gradient signal from the downstream loss computation on the same weight input. We name it "quasi-residual connection" as it is not exactly the same as the $y = F(x) + x$ form of the vanilla identity residual connection [50,55]. The key observation is that the internal of the graph-cut module does not need backward gradient signals as there are no learning parameters within the module. However, the input (t-links and n-links) to the graph-cut module does need those signals from the downstream loss

Biomedical Optics EXPRESS

for feature learning, which is backward propagated via the quasi-residual connection to update the edge weights (Fig. 3) by bypassing the non-differentiable graph-cut module.

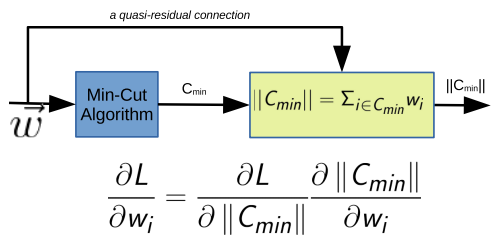

Fig. 3. Our proposed *gc*DLSeg makes use of the specialty of the graph structure and a quasi-residual connection to solve the zero gradient problem in combinatorial algorithms for deep learning. The minimum *s*-*t* cut algorithm outputs a minimum cut C_{min}, which is used in the following cut capacity computation. The $\vec{w}$ is a weight vector of all edges in the graph, $||C_{min}||$ is the min-cut capacity, and L is the loss from its downstream module.

2.4. Loss functions

Directly leveraging the objective and constraints of the optimization problem, the graph-cut algorithm module guarantees that, the capacity of the output minimum *s*-*t* cut from the graph-cut module, C_{min}, is always less than or equal to the capacity of the *ground truth cut*, that is, the *s*-*t* cut in the constructed graph corresponds to the ground truth segmentation, denote by C_{gt}. We proposed a novel residual graph-cut loss, which essentially converts the pixel-wise classification problem into a regression of the difference between the capacities of C_{min} and C_{gt}.

Without loss of generality, assuming the predicted *s*-*t* minimum cut C_{min} by the graph-cut module is unique, we have

$$\begin{aligned} &||C_{min}|| \leq ||C_{gt}||, \text{where} \\ &||C_{min}|| = \sum_{p \in S_{min}} \phi_t(p) + \sum_{p \in T_{min}} \phi_s(p) + \gamma \sum_{\substack{p \in S_{min} \\ q \in T_{min} \\ (p,q) \in \mathcal{N}}} \psi(p,q), \\ &||C_{gt}|| = \sum_{p \in S_{gt}} \phi_t(p) + \sum_{p \in T_{gt}} \phi_s(p) + \gamma \sum_{\substack{p \in S_{gt} \\ q \in T_{gt} \\ (p,q) \in \mathcal{N}}} \psi(p,q), \end{aligned} \tag{4}$$

where (S_{min}, T_{min}) is the node partition of the predicted minimum *s*-*t* cut C_{min}, while (S_{gt}, T_{gt}) is the node partition of C_{gt}. The *residual graph-cut loss* is defined, as follows:

$$L_{rGC} = \frac{1}{N_o}(||C_{gt}|| - ||C_{min}||), \tag{5}$$

where N_o is a normalization constant which equals the number of graph nodes plus the mean of the numbers of n-links in C_{min} and C_{gt}. When $L_{rGC} = 0$, the predicted min-cut C_{min} equals to the ground truth cut C_{gt}; otherwise, $L_{rGC}>0$.

The proposed L_{rGC} loss unifies two different optimization goals: the graph-cut optimization at the module level, and the ground truth-guided optimization over the whole network level. That is, L_{rGC} strives to guide the *gc*DLSeg network to output a minimum *s*-*t* cut C_{min} that is the same as the ground truth cut C_{gt}. In this way, we convert the pixel-wise segmentation (classification) problem into a regression of the difference between the capacities of C_{min} and C_{gt}. **This regression conversion enables to reformulate the non-differentiable binary node**

Biomedical Optics EXPRESS

labeling ("obj" or "bkg") in the graph-cut module as a comparison of cut-capacities in the continuous space, which makes the backward propagation feasible. The graph-cut module has no learning parameters, so the backward gradients can bypass the graph-cut module to be used for the updates of the input t-links and n-links for the module. With the help of the quasi-residual connection, as in Fig. 3, the backward gradients of L_{rGC} can be backward propagated to early layers of *gc*DLSeg and guide the upstream network to learn improved t-link and n-link weights to make a predicted minimum *s*-*t* cut approaching to the ground truth cut. The effectiveness of the proposed L_{rGC} loss for backward propagation will be proved in Section 2.5.

In addition to the proposed residual graph-cut loss L_{rGC}, we utilize a binary cross entropy loss L_{ce} and a generalized Dice loss L_{Dice} [57], as follows, to guide the learning of graph edge weights:

$$L_{ce} = \frac{-\sum_p \{g_p \ln \phi_s(p) + (1 - g_p) \ln \phi_t(p)\}}{N}, \tag{6}$$

where g_p is the ground truth probability that node p belongs to the foreground, and N is the number of total graph nodes.

$$L_{Dice} = 1.0 - 2.0 \frac{w_s \sum_p \{g_p \phi_s(p)\} + w_t \sum_p \{(1 - g_p)\phi_t(p)\}}{w_s \sum_p \{g_p + \phi_s(p)\} + w_t \sum_p \{(1 - g_p) + \phi_t(p)\}}, \tag{7}$$

where the coefficients w_s and w_t are used to provide invariance to foreground and background set properties, respectively, with $w_s = 1.0/(\sum_p g_p)^2$ and $w_t = 1.0/(\sum_p (1 - g_p))^2$.

To find a good set of weights for multiple loss terms in our loss function design, we use the coefficient of variations (relative standard deviation) [58] of component losses as the weight for each loss term. Considering the idea that a loss term is satisfied when its relative standard variance is decreased towards zero [58], we define the loss term weight $\alpha_\lambda^{(i)}$ for loss term i at time step λ, as follows. Here the time step λ indicates the iterative step for optimizing the network loss. For a sequence $\{b_1, b_2, \ldots, b_\kappa\}$, we use $\sigma(b_\kappa)$ and $\mu(b_\kappa)$ to denote the standard deviation and the mean of the sequence. Let $L_\lambda^{(i)}$ be the observed value of the i-th loss term ($i = 1, 2, 3$) at the time step λ. We define the loss ratio $r_\lambda^{(i)} = \frac{L_\lambda^{(i)}}{\mu(L_{\lambda-1}^{(i)})}$. Then,

$$\begin{aligned} \alpha_\lambda^{(i)} &= \frac{1}{z_\lambda} \cdot \frac{\sigma\left(r_\lambda^{(i)}\right)}{\mu\left(r_\lambda^{(i)}\right)}, \\ \text{where } z_\lambda &= \sum_i \frac{\sigma\left(r_\lambda^{(i)}\right)}{\mu\left(r_\lambda^{(i)}\right)}. \end{aligned} \tag{8}$$

Note that z_λ is a normalizing constant independent of the number of loss terms to ensure that $\sum_i \alpha_\lambda^{(i)} = 1.0$, which is important to decouple the loss term weights from the learning rate [58]. The total loss L_{total} at time step λ for this binary semantic segmentation network is defined, as follows:

$$L_{total} = \alpha_\lambda^{(1)} L_{ce} + \alpha_\lambda^{(2)} L_{Dice} + \alpha_\lambda^{(3)} L_{rGC}. \tag{9}$$

2.5. Effectiveness of the residual graph-cut loss

This section shows the effectiveness of the proposed residual graph-cut loss on the *gc*DLSeg model training by backward propagation. We prove that the minimum *s*-*t* cut capacity is differentiable almost everywhere over the graph edge weights. With the differentiability of the min-cut capacity, the backward gradient of the proposed residual graph-cut loss can be effectively used to update the edge weights in the graph to facilitate the predicted minimum *s*-*t* cut gradually converging

Biomedical Optics EXPRESS

to the ground truth cut during the model training. For the convenience of presentation, in this section, rather than using a node partition $(\{s\} \cup S, \{t\} \cup T)$ to represent an *s-t* cut C, we use the set of edges (u, v) with $u \in \{s\} \cup S$ and $v \in \{t\} \cup T$ to represent C.

Theorem 1 (**The derivatives of the min-cut capacity**) *Given an s-t graph $G = (\{s, t\} \cup V, E)$ with total $|V|+2$ nodes and $|E|$ edges, let $\vec{w} \in R^{|E|}$ be the edge weight vector, where each edge weight $w_i \geq 0$, and C_{min} be the set of edges in a minimum s-t cut.*

(1) The derivative of the min-cut capacity $||C_{min}||$ with respect to any edge weight w_i is, as follows:

$$\frac{\partial ||C_{min}||}{\partial w_i} = \begin{cases} 1, & \forall i \in C_{min}, \\ 0, & \forall i \notin C_{min}, \end{cases} \tag{10}$$

when the min-cut capacity $||C_{min}||$ is strictly smaller than the cut capacity of any other ones, which is guaranteed by the graph-cut algorithm.

(2) The min-cut capacity $||C_{min}||$ is differentiable over the vector $\vec{w}$ almost everywhere (except the minimum s-t cut C_{min} is not unique) in $R^{|E|}$.

Proof. We first consider the case where the min-cut capacity $||C_{min}||$ is strictly less than the cut capacity of any other cuts. Consider every possible weight $\vec{w} + \vec{\delta}$, where $\vec{\delta} \in R^{|E|}$ and $||\vec{\delta}||_2 \leq \epsilon$ with ϵ>0 being a small constant. When ϵ is small enough, under every such weight $\vec{w} + \vec{\delta}$, the set of original min-cut edges C_{min} under $\vec{w}$ will remain the set of min-cut edges under $\vec{w} + \vec{\delta}$. This is because under the new weight $\vec{w} + \vec{\delta}$, the cut capacity for C_{min} will at most increase by $\sqrt{|E|}\epsilon$ (by Cauchy-Schwarz inequality), while the cut capacity for any other cut will at most decrease by $\sqrt{|E|}\epsilon$. As long as $2\sqrt{|E|}\epsilon$ is smaller than the cut-capacity gap between the capacity of any other cut under $\vec{w}$ and the min-cut capacity under $\vec{w}$, C_{min} will remain the min-cut edge set under $\vec{w} + \vec{\delta}$.

Then, the new min-cut capacity under the edge weights $\vec{w} + \vec{\delta}$ is $||C_{min}|| = \sum_{i \in C_{min}} (w_i + \delta_i)$. So we can get the derivative of the min-cut capacity with regard to any edge weight w_i, as follows:

$$\frac{\partial ||C_{min}||}{\partial w_i} = \lim_{\delta_i \to 0} \frac{\Delta ||C_{min}||}{\delta_i} = \begin{cases} \lim_{\delta_i \to 0} \frac{\delta_i}{\delta_i} = 1, \forall i \in C_{min}, \\ \lim_{\delta_i \to 0} \frac{0}{\delta_i} = 0, \forall i \notin C_{min}, \end{cases} \tag{11}$$

where $\Delta ||C_{min}||$ expresses the change of min-cut capacity under perturbation $\vec{\delta}$.

Now let us consider the case where the min-cut capacity is not strictly less than the capacity of any other cuts. Suppose that there exist two different *s-t* cuts $C_{min}^{(1)}$ and $C_{min}^{(2)}$ with the same min-cut capacity, namely $||C_{min}^{(1)}|| = ||C_{min}^{(2)}||$. For some edge w_i where $i \in C_{min}^{(1)} \wedge i \notin C_{min}^{(2)}$, we consider a small perturbation $\vec{\delta}$ such that only δ_i is non-zero. Then, we have

$$\begin{cases} \lim_{\delta_i \to 0^+} \frac{\Delta ||C_{min}||}{\delta_i} = \frac{||C_{min}^{(2)}|| - ||C_{min}^{(2)}||}{\delta_i} = 0, \\ \lim_{\delta_i \to 0^-} \frac{\Delta ||C_{min}||}{\delta_i} = \frac{||C_{min}^{(1)}|| + \delta_i - ||C_{min}^{(1)}||}{\delta_i} = 1. \end{cases} \tag{12}$$

It shows that at some edge w_i where $i \in C_{min}^{(1)} \wedge i \notin C_{min}^{(2)}$, the derivative of the min-cut capacity with respect to w_i is undefined. We thus have the conclusion that the derivatives of the minimum *s-t* cut capacity are undefined for the weights of some edges when multiple cuts achieve the same min-cut capacity (when more than two *s-t* cuts achieve the same min-cut capacity, similar arguments apply by considering an edge belonging to some cuts achieving the min-cut capacity, but not belonging to the other cuts achieving the same min-cut capacity).

However, if $||C_{min}^{(1)}|| = ||C_{min}^{(2)}||$, we have $\sum_{i \in C_{min}^{(1)}} w_i = \sum_{i \in C_{min}^{(2)}} w_i$. Because the two sets $C_{min}^{(1)}$ and $C_{min}^{(2)}$ are distinct, this equation defines a hyperplane in the high-dimension space $R^{|E|}$. All hyperplanes have Lebesgue measure zero [59], and for an absolutely continuous distribution, the probability of hitting a set of zero Lebesgue measure is zero [60,61]. (The Lebesgue measure

Biomedical Optics EXPRESS

being zero means that $||C_{min}^{(1)}|| = ||C_{min}^{(2)}||$ is an improbable event, such that one can ignore this event for "most practical purposes" in the high-dimension space $R^{|E|}$. For example, the probability of choosing a specific value x, $0 \leq x \leq 1$, from a uniform distribution [0,1] is zero, that is, it is an unlikely event. There are at most $2^{|V|}$ cuts for this graph and at most $2^{2|V|}$ such equations (a finite number of hyperplanes) defining two cut capacities being equal. So the set of edge weights such that two cut-capacities are equal is of Lebesgue measure zero.

Therefore, we claim that the min-cut capacity is differentiable over the vector $\vec{w}$ almost everywhere (except when there exist two *s-t* cuts achieving the same min-cut capacity) in $R^{|E|}$. □

Theorem 2 (**The effectiveness of the residual graph-cut loss L_{rGC}**) *The backward gradients of the residual graph-cut loss L_{rGC} facilitate the weight decrease of cut edges unique in the ground truth cut and the weight increase of those edges unique in the predicted minimum s-t cut, simultaneously. It promotes the predicted min-cut in the next training iteration to gradually approach the ground truth cut.*

Proof. Using the quasi-residual connection, the gradient of the residual graph-cut loss L_{rGC} can be backward propagated to update the graph edge weights in the next training iteration. Recall that the residual graph-cut loss in formula (5), $L_{rGC} = \frac{1}{N_o}(||C_{gt}|| - ||C_{min}||)$, where C_{gt} is the ground truth cut and C_{min} is the predicted minimum *s-t* cut, and N_o is a normalization constant N_o>0. After canceling the common edges between C_{gt} and C_{min}, L_{rGC} can be further expressed, as follows:

$$L_{rGC} = \frac{1}{N_o}\left(\sum_{i \in C_{gt} \wedge i \notin C_{min}} w_i^{(gt)} - \sum_{i \notin C_{gt} \wedge i \in C_{min}} w_i^{(min)}\right), \tag{13}$$

where, for clearer presentation, $w_i^{(gt)}$ denotes the weight of a cut edge unique in C_{gt} and $w_i^{(min)}$ denotes the weight of a cut edge unique in C_{min}.

Using Theorem 1 (the derivatives of the min-cut capacity), we can get the backward gradient formulas, as follows:

$$\begin{aligned} \frac{\partial L_{rGC}}{w_i^{(gt)}} &= \frac{1}{N_o}, \forall i \in C_{gt} \wedge i \notin C_{min}; \\ \frac{\partial L_{rGC}}{w_i^{(min)}} &= -\frac{1}{N_o}, \forall i \notin C_{gt} \wedge i \in C_{min}. \end{aligned} \tag{14}$$

Using a standard gradient descent method, the edge weights are updated, as follows:

$$\begin{aligned} w_i^{(gt)} &\leftarrow w_i^{(gt)} - \alpha \frac{\partial L_{rGC}}{w_i^{(gt)}} \\ &= w_i^{(gt)} - \frac{\alpha}{N_o}, \forall i \in C_{gt} \wedge i \notin C_{min}, \\ w_i^{(min)} &\leftarrow w_i^{(min)} - \alpha \frac{\partial L_{rGC}}{w_i^{(min)}} \\ &= w_i^{(min)} + \frac{\alpha}{N_o}, \forall i \notin C_{gt} \wedge i \in C_{min} \end{aligned} \tag{15}$$

where α>0 is the learning rate during training, and N_o>0. The edge weight update formulas indicate that the backward gradient of the residual graph-cut loss L_{rGC} facilitates the decrease of the weight of cut edges unique in the ground truth cut, and promotes the increase of the weight of cut edges unique in the predicted minimum *s-t* cut, simultaneously. These updated edge weights decrease $||C_{gt}||$ while increasing $||C_{min}||$, which leads to $||C_{min}||$ converging to $||C_{gt}||$. In other words, the predicted minimum *s-t* cut in the next iteration during training gradually approaches the ground truth cut. When $||C_{gt}|| = ||C_{min}||$, $L_{rGC} = 0$, which implies the predicted minimum *s-t* cut overlaps with the ground truth cut. □

Biomedical Optics EXPRESS

3. Experiments

The proposed method was validated on two public data sets: the Advancing Zenith of Healthcare (AZH) chronic wound dataset [46] from the AZH Wound and Vascular Center, Milwaukee, Wisconsin, USA, and the pancreas cancer dataset from the medical segmentation decathlon (MSD) [47]. The irregular shapes, sizes, and inconsistent textures of chronic wounds make this segmentation extremely challenging. Segmenting pancreas cancer is the most challenging segmentation task among various MSD data sets.

The PyTorch version 1.12 [62] on Ubuntu Linux 20.04 was used for the experiments of the proposed method *gc*DLSeg. We chose GridCut [11] as the implementation of the minimum *s*-*t* cut algorithm [9]. To evaluate the segmentation performance, we followed the measurements of the compared methods, in which precision, recall, and Dice coefficient were adopted as the evaluation metrics [1]. In order to further demonstrate the performance of our proposed method, we also added the surface-based metrics: average surface distance (ASD), Hausdorff Distance (HD), and 95%HD. Similar to HD, the 95%HD is based on the calculation of the 95th percentile of the distances between boundary points in prediction and ground truth, whose purpose is to eliminate the impact of a very small subset of the outliers. For the purpose of the ablation study, we removed the graph-cut module from the proposed *gc*DLSeg network and trained the segmentation model with the loss of L_{ce} and L_{Dice}. The resulting model is termed as "NoGraph-Cut."

The experiments showed that our proposed *gc*DLSeg method outperformed the state-of-the-art methods in both metrics of Dice coefficient and recall, and all three surface-based metrics. The proposed *gc*DLSeg also demonstrated improved robustness against adversarial attacks.

3.1. AZH chronic wound segmentation

The AZH chronic wound data set [46] consists of 831 training images and 278 test images, each of which is of size 224×224 pixels with zero-padding. With various backgrounds, the raw images were taken by Canon SX 620 HS digital camera and iPad Pro under uncontrolled illumination conditions [46].

In our experiments, we randomly divided the 831 training images into training (706, 85%) and validation (125, 15%) sets and kept the original test set (278) untouched. We used 64 channels in the segmentation head, a batch size of 4, and an Adam optimizer with an initial learning rate of 0.0001 without weight decay. We used data augmentation on the fly, including random blurring/sharpening, color space transform, histogram equalization, color cast, white balance, flip, scale, rotation, and translation.

The proposed *gc*DLSeg method was compared to various deep learning segmentation models, as presented in Ref. [46]. In addition, we also compared to Zheng *et al.*'s method [10] termed as "GraphCutsLoss", in which the graph-cut energy function is used as part of the loss to boost model segmentation accuracy. However, their method does not explicitly integrate the graph-cut segmentation model into the network for model training and inference. Table 1 shows the performance of all compared methods. The proposed *gc*DLSeg method outperformed all compared methods with respect to the metrics of recall and Dice coefficient and ranked second with respect to the metric of precision (Table 1). Our proposed *gc*DLSeg method also exhibits explicit improvement on the surface position error compared with the NoGraph-Cut method, as in Table 2. The visual segmentation results from the AZH test set are shown in Fig. 4.

Our *gc*DLSeg model network also demonstrated improved robustness against adversarial attacks. We used untargeted white-box adversarial attacks to test this capability. Goodfellow *et al.*'s fast gradient sign method [64] was used to generate adversarial noise, as follows:

$$I_{adv} = I + \epsilon \cdot sign(\frac{dL}{dI}), \tag{16}$$

Biomedical Optics EXPRESS

where I is the original normalized input images, I_{adv} is the generated adversarial samples, and L is the network loss. The perturbation scale ϵ changes from 0.0 to 0.1 with step size 0.02. The precision, recall, and Dice coefficient evaluated on the AZH test set using different methods at

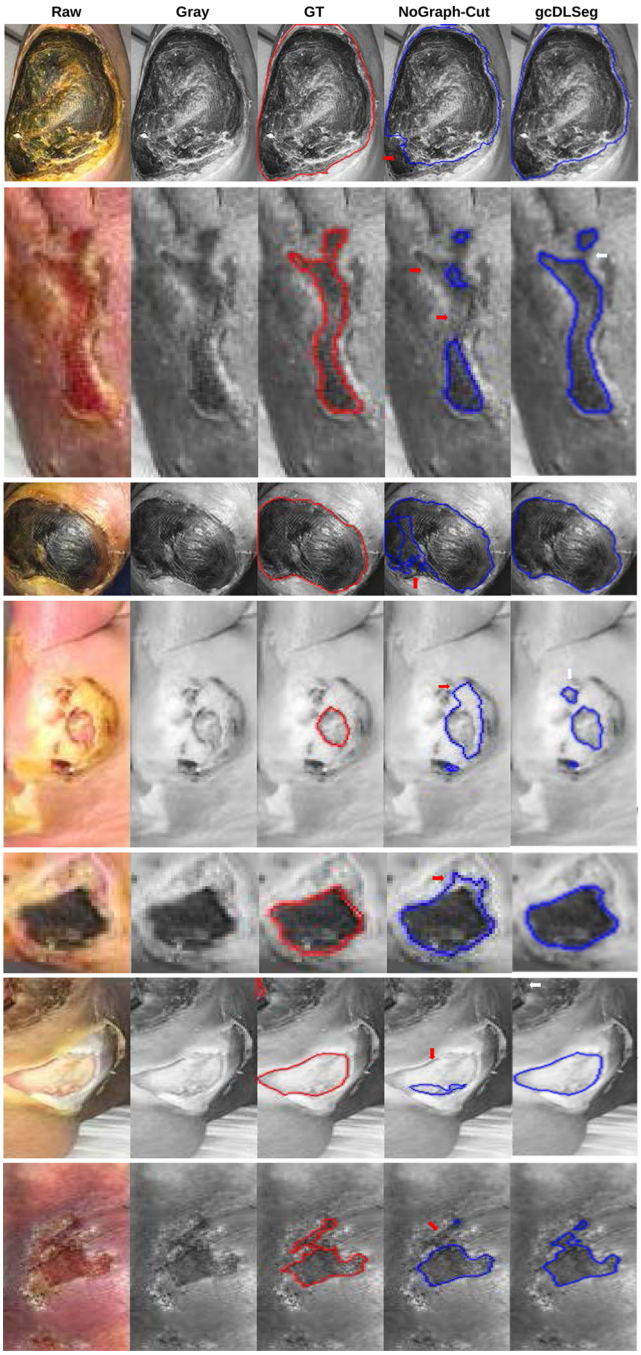

Fig. 4. Segmentation samples of seven cases from the AZH chronic wound test dataset. The red arrows indicate segmentation errors. The white arrows show that the results of our proposed *gc*DLSeg method are also not perfect.

Biomedical Optics EXPRESS

Table 1. Our proposed *gc*DLSeg method outperformed all other compared methods in the metrics of recall and Dice coefficient on the AZH test set. The best and second-best results are marked in bold red and black.[a]

Methods	Precision (%)	Recall (%)	Dice (%)
VGG16 [46]	83.91	78.35	81.03
SegNet [46]	83.66	86.49	85.05
U-Net [46]	89.04	**91.29**	90.15
Mask-RCNN [46]	**94.30**	86.40	90.20
MobileNetV2 [46]	90.86	89.76	90.30
MobileNetV2+CCL [46]	91.01	89.97	90.47
GraphCutsLoss [10]	91.45	90.24	**90.84**
*gc*DLSeg	**91.65**	**91.99**	**91.82**

[a]Notes: in medical disease measurement, recall is far more important than precision [63].

Table 2. Our proposed *gc*DLSeg method achieved reduced boundary errors, compared with the GraphCutsLoss method, in all three surface-based metrics on the AZH test set. The smaller the value, the better.

Methods	ASD (pixel)	HD (pixel)	95%HD (pixel)
GraphCutsLoss [10]	1.38	7.76	4.88
*gc*DLSeg	**1.30**	**6.56**	**4.00**
Improvement Rate	5.8%	15.5%	18.0%

different adversarial perturbations (ϵ) are shown in Table 3, and the dice decreasing curve is illustrated in Fig. 5. Their visual results are illustrated in Fig. 6.

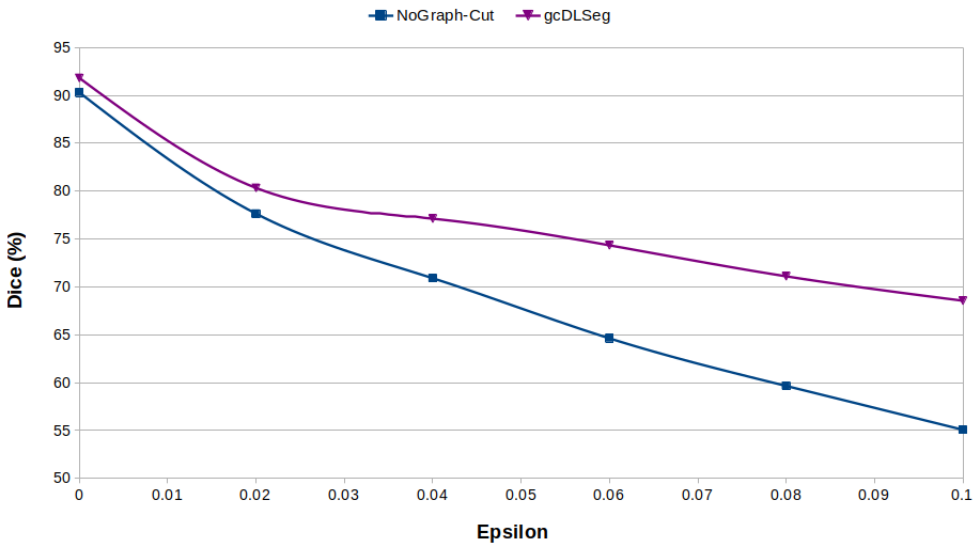

Fig. 5. The robustness performance of the proposed *gc*DLSeg method against adversarial attacks on the AZH test set. All Dice coefficients decrease with bigger-scale (ϵ) adversarial attacks. However, our proposed method shows a smaller decreasing slope.

Biomedical Optics EXPRESS

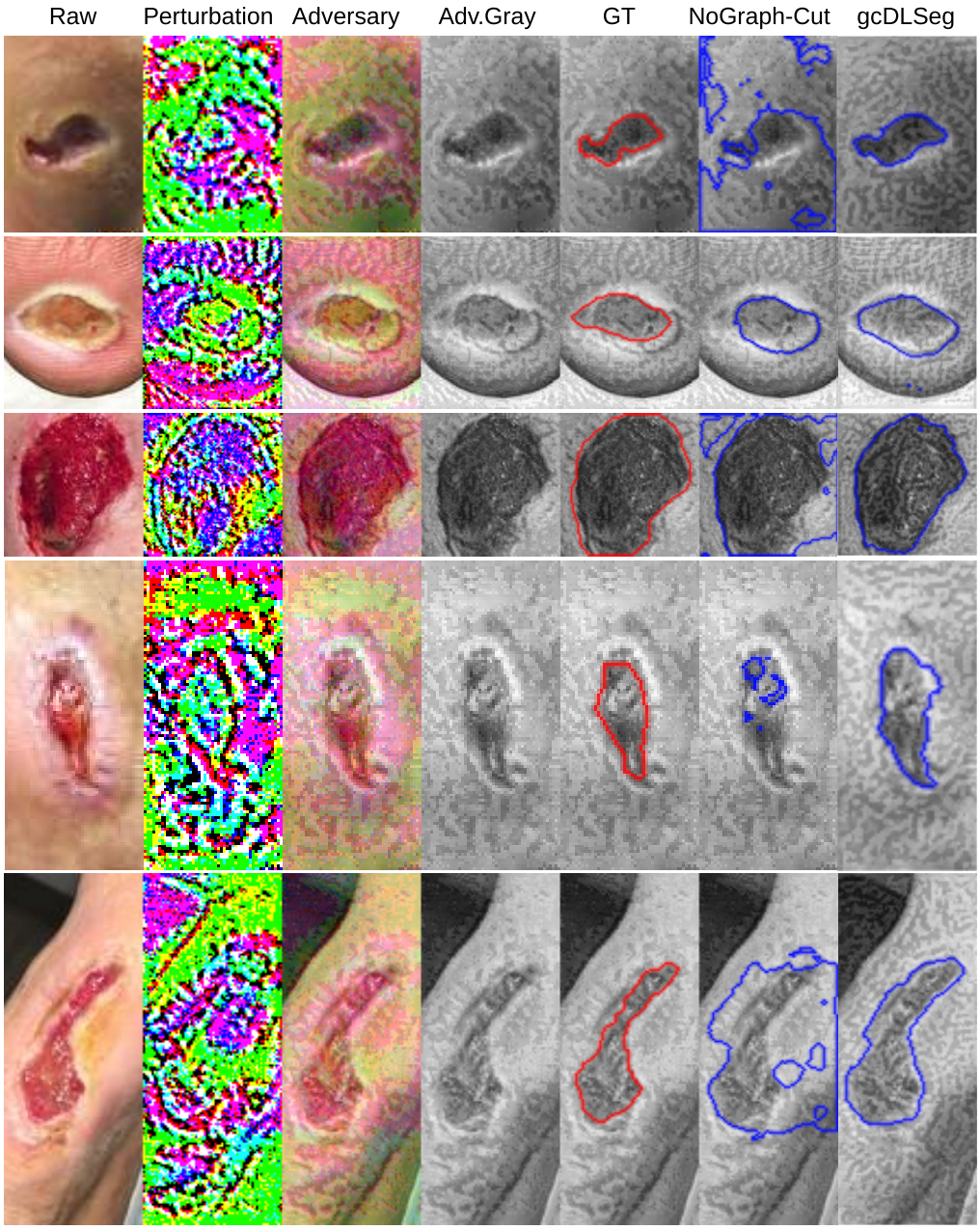

Fig. 6. Segmentation samples of five cases from the AZH test set using untargeted white-box adversarial attacks with $\epsilon = 0.1$. The adversarial samples (the 3rd column) were generated by adding $0.01 \cdot$ perturbation images (the 2nd column) to the original images (the first column) for the NoGraph-Cut method. For the proposed *gc*DLSeg method, we used its own training loss L_{total} (Eq. (9)) to generate the adversarial perturbations to maximize its attacks.

3.2. Pancreas tumor segmentation

The pancreas cancer data set from the medical segmentation decathlon (MSD) [47] consists of 281 3D volumes of abdominal CT images each with 512×512 pixels per slice and its corresponding

Biomedical Optics EXPRESS

Table 3. Our proposed *gc*DLSeg model shows improved robustness against adversarial attacks. The precision, recall, and Dice coefficient were computed for the proposed *gc*DLSeg and the NoGraph-Cut methods applied on the AZH test set at different adversarial attack scales (ϵ).

Epsilon	NoGraph-Cut			*gc*DLSeg		
(ϵ)	Precision (%)	Recall(%)	Dice (%)	Precision (%)	Recall (%)	Dice (%)
0.00	90.11	90.51	90.31	91.65	91.99	91.82
0.02	75.20	80.28	77.65	77.20	83.75	80.34
0.04	64.69	78.45	70.91	72.39	82.52	77.13
0.06	54.54	79.30	64.63	67.11	83.34	74.35
0.08	47.77	79.44	59.66	62.16	83.06	71.11
0.10	42.46	78.39	55.08	58.40	83.00	68.56

Table 4. The proposed *gc*DLSeg method outperformed all other compared methods in all metrics on the pancreas cancer data set from MSD. The best results are marked in bold red. Blanks mean that the original literature didn't report the corresponding measurements.

Methods	Precision (%)	Recall (%)	Dice (%)
nnU-Net 3D [65][a]			52.74
K.A.V.athlon 3D[b]			43.00
nnU-Net 2D [65]			35.01
GraphCutsLoss [10]	40.74	39.52	40.12
*gc*DLSeg	**52.78**	**57.59**	**55.08**

[a]Top 1 method in MSD Grand Challenge 2018
[b]Top 2 method in MSD Grand Challenge 2018

ground truth labels. Segmenting pancreas cancer is the most challenging segmentation task among various MSD data sets. Figure 7 shows some challenging examples visually. We randomly divided the 281 volumes into three data sets (training, validation, and test) and deleted all slices without the pancreas region. Then we got the training set (1304 slices), the validation set (597 slices), and the test set (634 slices). We used 2D slices as input to our networks and predicted the 2D segmentation for each slice.

We used 64 channels in the segmentation head, a batch size of 4, and an Adam optimizer with an initial learning rate of 0.0001 without weight decay. Data augmentation on the fly was used, including random blurring/sharpening, slight rotation ($[-29°, 29°]$), salt and pepper noise, and speckle noise.

Among all comparison methods, our proposed *gc*DLSeg method improved more than 3% in Dice coefficient over the previous state-of-the-art nnU-Net method, as shown in Table 4. Our proposed *gc*DLSeg method also exhibits explicit improvement on the surface position error compared with the NoGraph-cut and GraphCutsLoss methods, as in Table 5. Their visual example segmentations from the pancreas test set are demonstrated in Fig. 7.

Our proposed min-cut model network exhibits better robustness against adversarial attacks. The pixel accuracy, IoU, precision, recall and Dice coefficient were evaluated on the pancreas cancer test set using different methods at different adversarial attack scales (ϵ), as illustrated in Table 6, and Fig. 8.

Biomedical Optics EXPRESS

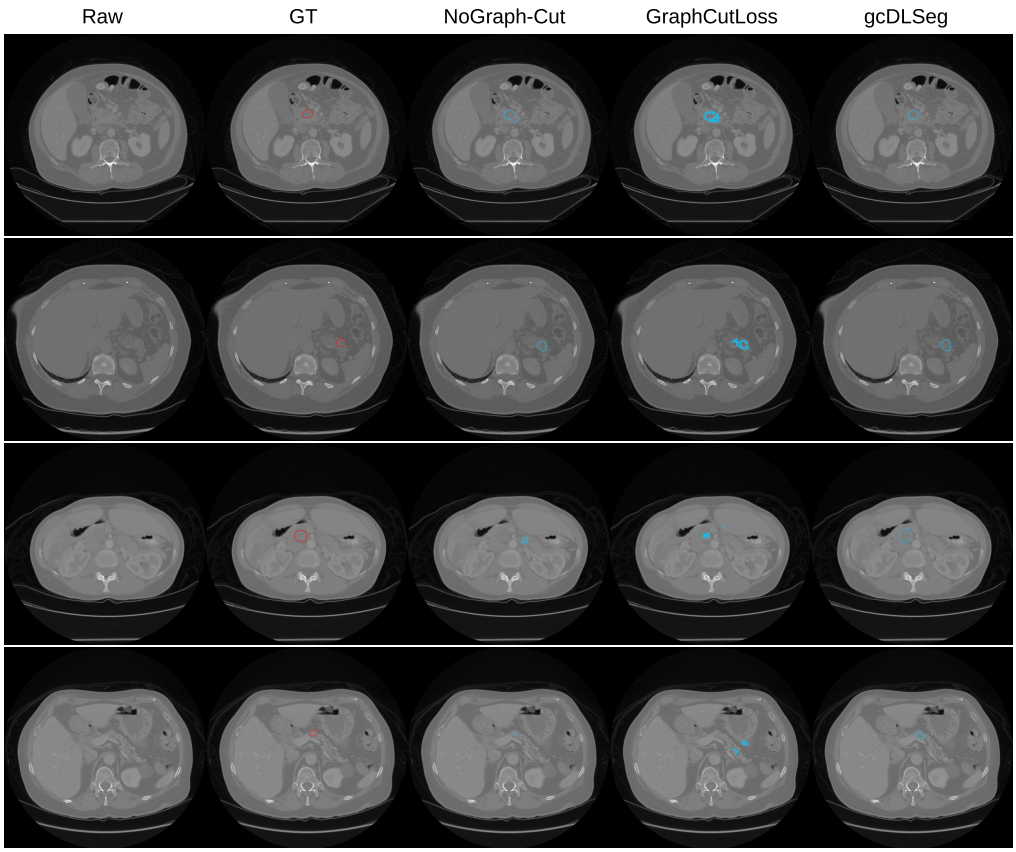

Fig. 7. Segmentation samples of four cases in the MSD pancreas cancer test dataset. The tumor boundary in both GT and predictions are drawn by a one-pixel boundary line. The GraphCutsLoss method has explicit un-smooth boundary phenomena, even it is drawn by the same one-pixel boundary. Please use the PDF magnifying glass tool to get a better view effect. We keep the full slice view, instead of the view of a region of interest (ROI), in order to show readers that segmenting tiny tumors in a messy abdominal background is very challenging.

Table 5. Our proposed *gc*DLSeg method achieved reduced boundary errors, comparing with GraphCutsLoss methods, in all three surface-based metrics on the pancreas cancer data set from MSD. The smaller the value, the better.

Methods	ASD (pixel)	HD (pixel)	95%HD (pixel)
GraphCutsLoss [10]	16.78	33.12	29.03
*gc*DLSeg	**12.30**	**23.21**	**21.96**
Improvement Rate	26.7%	29.9%	24.4%

Biomedical Optics EXPRESS

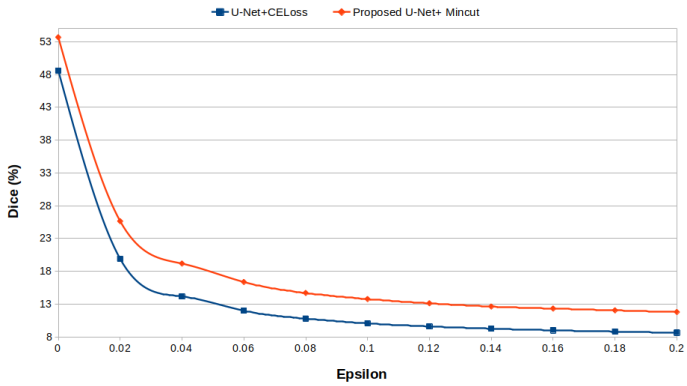

Fig. 8. Our proposed min-cut method exhibits better robustness against adversarial attacks, even though all Dice coefficients decrease along bigger-scale (ϵ) adversarial attacks in the pancreas cancer test set.

Table 6. Our proposed min-cut model network exhibits better robustness against adversarial attacks. The pixel accuracy, IoU, precision, recall and Dice coefficient were evaluated on the pancreas cancer test set using different methods at different adversarial attack scales (ϵ).

Epsilon	U-Net+CELoss					Proposed:U-Net+Mincut+CE-DiceLoss+mincutLoss				
(ϵ)	PixelAccuracy (%)	IoU (%)	Precision (%)	Recall (%)	Dice (%)	PixelAccuracy (%)	IoU (%)	Precision (%)	Recall (%)	Dice (%)
0.00	99.81	33.74	49.35	47.84	48.58	99.80	38.67	49.61	58.45	53.67
0.02	99.62	10.61	16.71	24.57	19.89	99.64	14.54	22.15	30.45	25.64
0.04	99.56	07.29	10.74	20.83	14.17	99.60	10.16	15.85	24.29	19.18
0.06	99.53	06.13	08.67	19.50	12.00	99.58	08.43	13.10	21.81	16.37
0.08	99.52	05.51	07.58	18.67	10.79	99.57	07.50	11.50	20.36	14.70
0.10	99.51	05.16	06.97	18.13	10.06	99.56	07.02	10.65	19.49	13.77
0.12	99.51	04.94	06.59	17.70	09.60	99.56	06.69	10.08	18.82	13.12
0.14	99.50	04.78	06.32	17.29	09.26	99.56	06.45	09.66	18.23	12.63
0.16	99.51	04.66	06.14	16.92	09.01	99.57	06.30	09.43	17.77	12.32
0.18	99.51	04.56	06.00	16.55	08.81	99.57	06.17	09.24	17.33	12.05
0.20	99.52	04.48	05.89	16.18	08.64	99.57	06.05	09.08	16.82	11.79

3.3. Ablation study

For the purpose of the ablation study, the NoGraph-Cut model by removing the graph-cut module from the proposed *gc*DLSeg network architecture was trained with the loss of L_{ce} and L_{Dice}. We also applied the graph-cut segmentation method as post-processing on the probability map output from the NoGraph-Cut model to investigate its performance. The ablation experiments were conducted on both the chronic wound [46] and the pancreas cancer [47]) data sets.

The performance of all three methods on both datasets is shown in Table 7 and Table 8, respectively. The segmentation results demonstrated that the proposed graph-cut module supported by backward propagation with the residual graph-cut loss within the deep learning network was able to significantly improve segmentation performance with respect to all metrics used, compared to the other two methods. For the method using graph-cut as postprocessing, the segmentation performance was comparable to its baseline method of NoGraph-Cut. In this scheme, feature learning by the NoGraph-Cut network is, in fact, disconnected from the graph-cut model; the learned features thus may not be truly appropriated for the graph-cut model. However, in the framework of our proposed *gc*DLSeg method, the graph-cut model is used to guide the feature learning directly with U-Net. We also performed experiments, which demonstrated that the graph-cut module with backward propagation in fine-tuning training can further improve probability maps. After pre-training of U-Net, the graph-cut module was added for network

Biomedical Optics EXPRESS

fine-tuning. The binary cross-entropy loss, which measures the distance between the predicted and ground truth probability distributions, was further reduced. It implies that the backward propagation of the residual graph-cut loss further improved the probability map. See (Table 9) and (Table 10).

Table 7. Ablation experiments showed the proposed graph-cut module integrated with U-Net for end-to-end training improved segmentation performance on the AZH test set. The best results are marked in bold red.

Ablation Experiments	Precision (%)	Recall (%)	Dice (%)
NoGraph-Cut	90.11	90.51	90.31
Graph-cut as post-processing	91.23	89.70	90.46
*gc*DLSeg	**91.65**	**91.99**	**91.82**

Table 8. Ablation experiments showed the proposed graph-cut module integrated with U-Net for end-to-end training improved segmentation performance on pancreas cancer data set from MSD. The best results are marked in bold red.

Ablation Experiments	Precision (%)	Recall (%)	Dice (%)
NoGraph-Cut	45.53	52.32	48.69
Graph-cut as post-processing	46.08	51.10	48.46
*gc*DLSeg	**52.78**	**57.59**	**55.08**

Table 9. Ablation experiments showed the proposed graph-cut module integrated with U-Net for end-to-end training reduced boundary error on the AZH test set. The smaller the value, the better.

Methods	ASD (pixel)	HD (pixel)	95%HD (pixel)
NoGraph-Cut	1.44	7.98	5.20
Graph-cut as post-processing	1.39	7.76	5.14
*gc*DLSeg	**1.30**	**6.56**	**4.00**

Table 10. Ablation experiments showed the proposed graph-cut module integrated with U-Net for end-to-end training reduced boundary error on the pancreas cancer data set from MSD. The smaller the value, the better.

Methods	ASD (pixel)	HD (pixel)	95%HD (pixel)
NoGraph-Cut	14.71	25.54	24.00
Graph-cut as post-processing	14.46	27.20	25.76
*gc*DLSeg	**12.30**	**23.21**	**21.96**

4. Discussion and conclusion

In this study, we developed a novel DL framework for binary semantic image segmentation, which leverages a new residual graph-cut loss and a quasi-residual connection to seamlessly integrate the graph-cut segmentation model with the U-Net segmentation network for end-to-end learning. The residual graph-cut loss essentially enables reformulating the pixel-wise classification problem as a regression problem for capturing the difference between the capacities of the min-cut and the ground truth cut. We theoretically proved the derivativity of the min-cut capacity and

the effectiveness of the proposed residual graph-cut loss for feature learning via backward propagation. The quasi-residual connection provides a pathway bypassing the combinatorial and non-differentiable graph-cut optimization module for the backward propagation of gradients to earlier layers of *gc*DLSeg. The global optimality ensured by the minimum *s*-*t* algorithm facilitates better feature learning efficiently during network training. In the inference phase, globally optimal segmentation is achieved with respect to the graph-cut segmentation model defined on the optimized image features from *gc*DLSeg.

Our proposed *gc*DLSeg method enhanced boundary refinement. While deep learning models excel at extracting high-level features for segmentation, they can sometimes struggle with precise boundary delineation, especially in medical images with fuzzy or low-contrast borders. Medical images often require domain-specific priors (e.g., shape, smoothness, and/or connection). Graph cuts can easily incorporate such constraints into the segmentation process, making them an ideal complement to deep learning. The proposed *gc*DLSeg approach seamlessly integrates graph-cut optimization with deep learning that leverages the global optimization capabilities of graph cuts with the powerful feature extraction of deep learning, enhancing accuracy in complex medical image segmentation tasks. We have validated *gc*DLSeg on two segmentation tasks: one is relatively easy and the other is highly challenging.

- For the easy segmentation task: AZH chronic wound segmentation, where object boundaries are well-defined, *gc*DLSeg shows only marginal improvement over other methods. Since deep learning models already perform well in such cases, further gains are naturally limited, as shown in Table 1 and 2.
- For the challenging segmentation task: pancreas tumor segmentation, where the target objects are small and the boundaries are weak and difficult to delineate, *gc*DLSeg demonstrates significant improvements. Specifically, it achieves much lower Hausdorff Distance (HD) and 95% HD, reflecting superior boundary precision and reduced extreme errors compared to other approaches, as show in Table 4 and 5.

Our proposed *gc*DLSeg method also enhanced robustness for Hausdorff Distance and adversarial attacks. A critical advantage of *gc*DLSeg is the robustness in both standard and adversarial settings:

- Lower both HD and 95% HD: Compared to baseline methods, *gc*DLSeg achieves significantly lower HD and 95% HD in pancreas cancer segmentation, indicating better structural integrity in segmented regions, as shown in Table 2 and 5.
- Resilience to adversarial attacks: Our experiments show that *gc*DLSeg is more robust to adversarial perturbations than other deep learning models. This is likely due to the explicit graph-cut optimization, which enforces a strong structural prior and makes the model less susceptible to minor input variations, as shown in Fig. 5 and 8.

Our proposed *gc*DLSeg method also demonstrated key advantages over GraphCutsLoss method. While GraphCutsLoss integrates graph-cut constraints into the loss function, *gc*DLSeg seamlessly incorporates graph-cut optimization into deep learning framework, which enforces the graph-cut model in both training and inference, leading to:

- More effective feature learning during training: *gc*DLSeg forces the deep learning model to learn features that align well with the graph-cut model, while GraphCutsLoss may require more iterations to achieve similar performance.
- Stronger boundary adherence during inference: Unlike GraphCutsLoss, which primarily influences training, *gc*DLSeg continues to enforce graph-cut principles during inference, leading to more stable and precise segmentations.

Biomedical Optics EXPRESS

These strengths confirm that *gc*DLSeg is not just an incremental improvement but a method with distinct advantages in robustness and boundary precision.

The major drawback of our proposed method is its training efficiency. During the network training, it is computationally intensive to run the minimum *s*-*t* cut algorithm to compute the optimal solutions with large training datasets and training epochs. The minimum *s*-*t* cut algorithm used in our current implementation is GridCut [11], which is an augmenting path algorithm and is challenging to parallelize in GPU. Further improvement includes using a parallel push-relabel min-cut algorithm implemented in GPU to improve training and inference efficiency. Another way is, during the early stages of training, to compute approximations to the minimum *s*-*t* cuts, and then refine the network with optimal *s*-*t* cuts in the final stages of training.

The proposed method was validated on the public AZH chronic wound data set [46] and the pancreas cancer data set from MSD [47]). Our experiments showed that the proposed *gc*DLSeg method outperformed the state-of-the-art methods in Dice, recall, and surface position error. Our proposed method also demonstrated improved robustness against adversarial attacks. We expect the developed *gc*DLSeg method would find broader applications in computer vision, which are involved in the minimum *s*-*t* cut algorithm. The techniques of integrating graph-cut into the deep learning networks can also be extendable for other combinatorial optimization methods.

Funding. National Science Foundation (CCF-1733742, ECCS-2000425); National Institutes of Health (1U54HL165442, 5R01AG067078).

Acknowledgments. Thanks to Patrick M Jensen at the Technical University of Denmark, who suggested using GridCut [11] as the implementation of the minimum *s*-*t* cut algorithm. This study was supported in part by NSF grants CCF-1733742 and ECCS-2000425 and in part by NIH grants 1U54HL165442 and 5R01AG067078.

Disclosures. The authors declare that there are no conflicts of interest related to this article.

Data availability. The data that support the findings of this study are available on request from the corresponding author.

References

1. S. Minaee, Y. Y. Boykov, F. Porikli, *et al.*, "Image segmentation using deep learning: A survey," IEEE Trans. Pattern Anal. Mach. Intell. **44**, 3523–3542 (2021).
2. Y. Y. Boykov and M.-P. 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, 2001), pp. 105–112.
3. Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," IEEE Trans. Pattern Anal. Mach. Intell. **23**(11), 1222–1239 (2001).
4. Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," IEEE Trans. Pattern Anal. Mach. Intell. **26**(9), 1124–1137 (2004).
5. L. A. Vese and T. F. Chan, "A multiphase level set framework for image segmentation using the mumford and shah model," International journal of computer vision **50**(3), 271–293 (2002).
6. C. Li, C. Xu, C. Gui, *et al.*, "Distance regularized level set evolution and its application to image segmentation," IEEE Trans. on Image Process. **19**(12), 3243–3254 (2010).
7. C. Li, R. Huang, Z. Ding, *et al.*, "A level set method for image segmentation in the presence of intensity inhomogeneities with application to mri," IEEE Trans. on Image Process. **20**(7), 2007–2016 (2011).
8. D. M. Greig, B. T. Porteous, and A. H. Seheult, "Exact maximum a posteriori estimation for binary images," Journal of the Royal Statistical Society: Series B (Methodological) **51**(2), 271–279 (1989).
9. P. M. Jensen, N. Jeppesen, A. B. Dahl, *et al.*, "Review of serial and parallel min-cut/max-flow algorithms for computer vision," arXiv (2022).
10. Z. Zheng, M. Oda, and K. Mori, "Graph cuts loss to boost model accuracy and generalizability for medical image segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2021), pp. 3304–3313.
11. O. Jamriška, D. Sỳkora, and A. Hornung, "Cache-efficient graph cuts on structured grids," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, (IEEE, 2012), pp. 3673–3680.
12. M. Bleyer and M. Gelautz, "Graph-based surface reconstruction from stereo pairs using image segmentation," in *Videometrics VIII*, vol. 5665 (SPIE, 2005), pp. 288–299.
13. J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 888–905 (2000).
14. G. Litjens, T. Kooi, B. E. Bejnordi, *et al.*, "A survey on deep learning in medical image analysis," Med. Image Anal. **42**, 60–88 (2017).

Biomedical Optics EXPRESS

15. D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," Annu. Rev. Biomed. Eng. **19**(1), 221–248 (2017).
16. Y. Guo, Y. Liu, T. Georgiou, *et al.*, "A review of semantic segmentation using deep neural networks," Int. J. Multimed. Info. Retr. **7**(2), 87–93 (2018).
17. X. Liu, L. Song, S. Liu, *et al.*, "A review of deep-learning-based medical image segmentation methods," Sustainability **13**(3), 1224 (2021).
18. Y. Mo, Y. Wu, X. Yang, *et al.*, "Review the state-of-the-art technologies of semantic segmentation based on deep learning," Neurocomputing **493**, 626–646 (2022).
19. N. Tajbakhsh, L. Jeyaseelan, Q. Li, *et al.*, "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," Med. Image Anal. **63**, 101693 (2020).
20. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, (Springer, 2015), pp. 234–241.
21. 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*, (2015), pp. 3431–3440.
22. L.-C. Chen, G. Papandreou, I. Kokkinos, *et al.*, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2017).
23. A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need," in *Advances in neural information processing systems*, (2017), pp. 5998–6008.
24. R. Strudel, R. Garcia, I. Laptev, *et al.*, "Segmenter: Transformer for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, (2021), pp. 7262–7272.
25. H. Thisanke, C. Deshan, K. Chamith, *et al.*, "Semantic segmentation using vision transformers: A survey," Engineering Applications of Artificial Intelligence **126**, 106669 (2023).
26. J. Chen, Y. Lu, Q. Yu, *et al.*, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv (2021).
27. H. Cao, Y. Wang, J. Chen, *et al.*, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *European conference on computer vision*, (Springer, 2022), pp. 205–218.
28. A. Lin, B. Chen, J. Xu, *et al.*, "Ds-transunet: Dual swin transformer u-net for medical image segmentation," IEEE Trans. Instrum. Meas. **71**, 1–15 (2022).
29. H.-Y. Zhou, J. Guo, Y. Zhang, *et al.*, "nnformer: Interleaved transformer for volumetric segmentation," arXiv (2021).
30. C. Xie, J. Wang, Z. Zhang, *et al.*, "Adversarial examples for semantic segmentation and object detection," in *Proceedings of the IEEE international conference on computer vision*, (2017), pp. 1369–1378.
31. A. Arnab, O. Miksik, and P. H. Torr, "On the robustness of semantic segmentation models to adversarial attacks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2018), pp. 888–897.
32. L. Chen, W. Ruan, X. Liu, *et al.*, "Seqvat: Virtual adversarial training for semi-supervised sequence labeling," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (2020), pp. 8801–8811.
33. A. Kirillov, E. Mintun, N. Ravi, *et al.*, "Segment anything," arXiv (2023).
34. J. Ma and B. Wang, "Segment anything in medical images," arXiv (2023).
35. S. Gunasekar, Y. Zhang, J. Aneja, *et al.*, "Textbooks are all you need," arXiv (2023).
36. K. A. Smith, "Neural networks for combinatorial optimization: a review of more than a decade of research," INFORMS Journal on Computing **11**(1), 15–34 (1999).
37. Y. Bengio, A. Lodi, and A. Prouvost, "Machine learning for combinatorial optimization: a methodological tour d'horizon," European Journal of Operational Research **290**(2), 405–421 (2021).
38. J. Kotary, F. Fioretto, P. Van Hentenryck, *et al.*, "End-to-end constrained optimization learning: A survey," arXiv (2021).
39. R. C. Larson and A. R. Odoni, *Urban operations research*, Monograph (Englewood Cliffs, N.J., 1981).
40. M. V. Pogančić, A. Paulus, V. Musil, *et al.*, "Differentiation of blackbox combinatorial solvers," in *International Conference on Learning Representations*, (2019).
41. A. N. Elmachtoub and P. Grigas, "Smart "predict, then optimize"," Management Science **68**, 1–808 (2021).
42. A. Mensch and M. Blondel, "Differentiable dynamic programming for structured prediction and attention," in *International Conference on Machine Learning*, (PMLR, 2018), pp. 3462–3471.
43. H. Xie, W. Xu, and X. Wu, "A deep learning network with differentiable dynamic programming for retina oct surface segmentation," Biomed. Opt. Express **14**(7), 3190–3202 (2023).
44. E. Khalil, H. Dai, Y. Zhang, *et al.*, "Learning combinatorial optimization algorithms over graphs," Advances in neural information processing systems **30**, 6531 (2017).
45. M. Gasse, D. Chételat, N. Ferroni, *et al.*, "Exact combinatorial optimization with graph convolutional neural networks," Advances in Neural Information Processing Systems **32**, 1 (2019).
46. C. Wang, D. Anisuzzaman, V. Williamson, *et al.*, "Fully automatic wound segmentation with deep convolutional neural networks," Sci. Rep. **10**(1), 21897 (2020).
47. M. Antonelli, A. Reinke, S. Bakas, *et al.*, "The medical segmentation decathlon," Nat. Commun. **13**(1), 4128 (2022).
48. S. G. Brush, "History of the lenz-ising model," Rev. Mod. Phys. **39**(4), 883–893 (1967).
49. N. Komodakis, N. Paragios, and G. Tziritas, "MRF energy minimization and beyond via dual decomposition," IEEE Trans. Pattern Anal. Mach. Intell. **33**(3), 531–552 (2010).

Biomedical Optics EXPRESS

50. K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), pp. 770–778.
51. M. Tang, A. Djelouah, F. Perazzi, *et al.*, "Normalized cut loss for weakly-supervised cnn segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2018), pp. 1818–1827.
52. M. Tang, F. Perazzi, A. Djelouah, *et al.*, "On regularized losses for weakly-supervised cnn segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018), pp. 507–522.
53. J. Gibert, E. Valveny, and H. Bunke, "Graph embedding in vector spaces by node attribute statistics," Pattern Recognition **45**(9), 3072–3083 (2012).
54. S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," Int. J. Unc. Fuzz. Knowl. Based Syst. **06**(02), 107–116 (1998).
55. K. He, X. Zhang, S. Ren, *et al.*, "Identity mappings in deep residual networks," in *European conference on computer vision*, (Springer, 2016), pp. 630–645.
56. M. S. Ebrahimi and H. K. Abadi, "Study of residual networks for image recognition," in *Intelligent Computing*, (Springer, 2021), pp. 754–763.
57. C. H. Sudre, W. Li, T. Vercauteren, *et al.*, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, (Springer, 2017), pp. 240–248.
58. R. Groenendijk, S. Karaoglu, T. Gevers, *et al.*, "Multi-loss weighting with coefficient of variations," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, (2021), pp. 1469–1478.
59. C. Goffman and G. Pedrick, "A proof of the homeomorphism of lebesgue-stieltjes measure with lebesgue measure," Proc. Am. Math. Soc. **52**(1), 196–198 (1975).
60. A. A. Borovkov, "Random variables and distribution functions," in *Probability Theory*, (Springer, 2013), pp. 31–63.
61. C. R. Rao and M. M. Lovric, "Testing point null hypothesis of a normal mean and the truth: 21st century perspective," J. Mod. App. Stat. Meth. **15**(2), 2–21 (2016).
62. A. Paszke, S. Gross, F. Massa, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," Advances in neural information processing systems **32**, 8026–8037 (2019).
63. D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," arXiv (2020).
64. I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," arXiv (2014).
65. F. Isensee, P. F. Jaeger, S. A. Kohl, *et al.*, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," Nat. Methods **18**(2), 203–211 (2021).