Dual structure-aware image filterings for semi-supervised medical image segmentation

Yuliang Gu, Zhichao Sun, Tian Chen, Xin Xiao, Yepeng Liu, Yongchao Xu, Laurent Najman

Medical Image Analysis (2024)

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Dual Structure-Aware Image Filterings for Semi-supervised Medical Image Segmentation

Yuliang Gu[a,b,c], Zhichao Sun[a,b,c], Tian Chen[a,b,c], Xin Xiao[a,b,c], Yepeng Liu[a,b,c], Yongchao Xu[a,b,c,*], Laurent Najman[d]

[a]*National Engineering Research Center for Multimedia Software, Wuhan University, Wuhan, China*
[b]*Institute of Artificial Intelligence, School of Computer Science, Wuhan University, Wuhan, China*
[c]*Hubei Key Laboratory of Multimedia and Network Communication Engineering, Wuhan University, Wuhan, China*
[d]*Univ Gustave Eiffel, CNRS, LIGM, Marne-la-Vallée, France*

ARTICLE INFO

Article history:

Keywords: Max-tree, Min-tree, Connected filtering, Semi-supervised medical image segmentation

ABSTRACT

Semi-supervised image segmentation has attracted great attention recently. The key is how to leverage unlabeled images in the training process. Most methods maintain consistent predictions of the unlabeled images under variations (*e.g.*, adding noise/perturbations, or creating alternative versions) in the image and/or model level. In most image-level variation, medical images often have prior structure information, which has not been well explored. In this paper, we propose novel dual structure-aware image filterings (DSAIF) as the image-level variations for semi-supervised medical image segmentation. Motivated by connected filtering that simplifies image via filtering in structure-aware tree-based image representation, we resort to the dual contrast invariant Max-tree and Min-tree representation. Specifically, we propose a novel connected filtering that removes topologically equivalent nodes (*i.e.* connected components) having no siblings in the Max/Min-tree. This results in two filtered images preserving topologically critical structure. Applying the proposed DSAIF to mutually supervised networks decreases the consensus of their erroneous predictions on unlabeled images. This helps to alleviate the confirmation bias issue of overfitting to noisy pseudo labels of unlabeled images, and thus effectively improves the segmentation performance. Extensive experimental results on three benchmark datasets demonstrate that the proposed method significantly/consistently outperforms some state-of-the-art methods. Source code is publicly available at https://github.com/GuGuLL123/DSAIF-SEMI.

1. Introduction

Accurate medical image segmentation plays an important role in computer-aided diagnosis (CAD) systems. Traditional supervised segmentation methods have achieved impressive results using a large amount of labeled data. Yet, the manual segmentation is laborious and time-consuming. Recently, semi-supervised segmentation methods have gained significant attention by utilizing easily accessible unlabeled images to improve the accuracy of segmentation models.

The mainstream semi-supervised segmentation methods are based on consistency regularization (Zhao et al., 2023; Wang et al., 2023b; Yang et al., 2023; Basak and Yin, 2023; Lei et al., 2022; Jin et al., 2022; Xiang et al., 2022; Basak et al., 2022; Lyu et al., 2022; Su et al., 2024; Adiga et al., 2024), which aims to produce consistent results under variations at image-level or/and model-level. In particular, many approaches aim to generate variations under image-level (Yu et al., 2019; Xu et al.,

*Corresponding author
e-mail: yuliang_gu@whu.edu.cn (Yuliang Gu), zhichaosun@whu.edu.cn (Zhichao Sun), tian.chen@whu.edu.cn (Tian Chen), xinxiao@whu.edu.cn (Xin Xiao), yepeng.liu@whu.edu.cn (Yepeng Liu), yongchao.xu@whu.edu.cn (Yongchao Xu), laurent.najman@esiee.fr (Laurent Najman)

2021; You et al., 2022a,b; Bai et al., 2023). A popular strategy for image variations utilizes the weak-to-strong paradigm (Fan et al., 2022; Liu et al., 2022b; Yang et al., 2023), where predictions generated from weakly-augmented versions are used to supervise the strongly-augmented versions. Augmented versions are usually generated by simple random augmentation (*e.g.*, Gaussian noise (Huang et al., 2022)), adversarial perturbation (Peiris et al., 2021; Wang et al., 2023a), and CutMix techniques (Chen et al., 2021; Yang et al., 2023). The model-level variations mainly adopt the Mean Teacher framework (Tarvainen and Valpola, 2017) or Co-training strategy (Qiao et al., 2018; Chen et al., 2021). In the Mean Teacher framework, the teacher network is usually obtained from the student network via Exponential Moving Average (EMA). The co-training strategy involves training two independent networks or decoders with different initializations and using each model's outputs to supervise the other's training in a mutual fashion.

Recently, the consistency regularization methods using pseudo labels for supervision have achieved impressive performance for semi-supervised segmentation (Chen et al., 2021; Yang et al., 2023; Basak and Yin, 2023; Lyu et al., 2022; Liu et al., 2022a). For instance, CPS (Chen et al., 2021) generates different pseudo labels by two networks with different initializations and applies mutual supervision between them. These methods have achieved impressive performance in natural images, thanks to effective strong image augmentation (*e.g.*, CutMix (Yun et al., 2019)) as image-level variations for avoiding the model overfit to incorrect pseudo-labels (Chen et al., 2021; Yang et al., 2023; Liu et al., 2022b). However, these existing image-level variations do not use the structural information, which is important for medical images. Moreover, the distribution variance in medical images is not as significant as in natural images, which makes the semi-supervised medical image segmentation more prone to overfit noisy pseudo-labels, due to confirmation bias (Arazo et al., 2020).

In this paper, we propose novel dual structure-aware image filterings (DSAIF), serving as the image-level variations to cope with the confirmation bias in semi-supervised medical image segmentation. For that, we aim to obtain two filtered images with diverse image appearances, while preserving the critical topological structure of the original image. Specifically, we resort to the dual contrast-invariant Max-tree and Min-tree (Salembier et al., 1998) representation, given by the inclusion relationship between connected components of upper and lower level sets, respectively. The topology of the tree structure encodes the topology of the image structure. Such structure-aware tree-based image representation is widely used to implement connected filterings (Salembier et al., 1998; Westenberg et al., 2007; Wilkinson et al., 2008; Ouzounis and Wilkinson, 2007; Xu et al., 2015) that do not create new edges. We propose a novel type of connected filtering that preserves the topological structure of the image. Precisely, we remove all nodes (*i.e.* connected components) having no siblings in the Max-tree and Min-tree, resulting in two simplified trees preserving topologically critical structure. The corresponding filters named upper/lower structure-aware image filtering (denoted as USAIF and LSAIF) give rise to two different images having the same topological structure as the original image.

To further cope with the confirmation bias issue on unlabeled medical images, we also propose to apply monotonically increasing contrast changes before performing the dual structure-aware image filterings. Since the Max-tree and Min-tree are invariant to such increasing changes, the resulting filtered images still preserve the topological image structure while having large diversity in image appearances. By incorporating the proposed DSAIF into mutually supervised networks, the consensus on incorrect predictions for unlabeled images is decreased. This helps to alleviate the confirmation bias issue, where models tend to overfit to noisy pseudo labels, thereby enhancing the performance of segmentation. Applying such dual structure-aware image filterings as the image-level variations decreases the consensus of erroneous predictions for unlabeled images. This helps to alleviate the confirmation bias issue of overfitting to noisy pseudo labels of unlabeled images, thereby enhancing the performance for semi-supervised medical image segmentation. We adopt the mutual supervision framework of CPS (Chen et al., 2021) and MC-Net (Wu et al., 2021) as the baseline models. The proposed DSAIF significantly boosts the performance of CPS and MC-Net baseline, and significantly/consistently outperforms some state-of-the-art methods on three benchmark datasets.

The main contribution of the paper is summarized as follows: 1) We propose novel dual structure-aware image filterings (DSAIF) as the image-level variations for semi-supervised medical image segmentation. DSAIF yields two images with quite different appearances while having the same topological structure as the original image. 2) We further leverage the contrast-invariance property of Max/Min-tree representation involved in DSAIF. We apply monotonically increasing contrast changes before performing DSAIF. This increases the appearance diversity while preserving topological image structure. 3) The proposed method significantly/consistently outperforms some state-of-the-art methods on three widely benchmark datasets. In particular, using only 20% of labeled images, the proposed method achieves similar (∼99.5%) segmentation performance with the use of full dataset.

The rest of this paper is organized as follows. We first review some related works in Section 2, followed by the detail of the proposed method in Section 3. We then present extensive experimental results in Section 4. Finally, we conclude in Section 5.

2. Related work

2.1. Semi-supervised Learning

Semi-supervised learning (SSL) aims to leverage limited annotated data and a large number of unlabeled data to improve the performance. Existing semi-supervised learning methods can be roughly grouped into two categories (Chen et al., 2022): self-training and consistency regularization. Self-training methods (Grandvalet and Bengio, 2005) learn from unlabeled data by assigning pseudo labels to unlabeled data and subsequently integrating them with manually labeled data for further retraining. Consistency regularization methods (Laine and Aila, 2016;

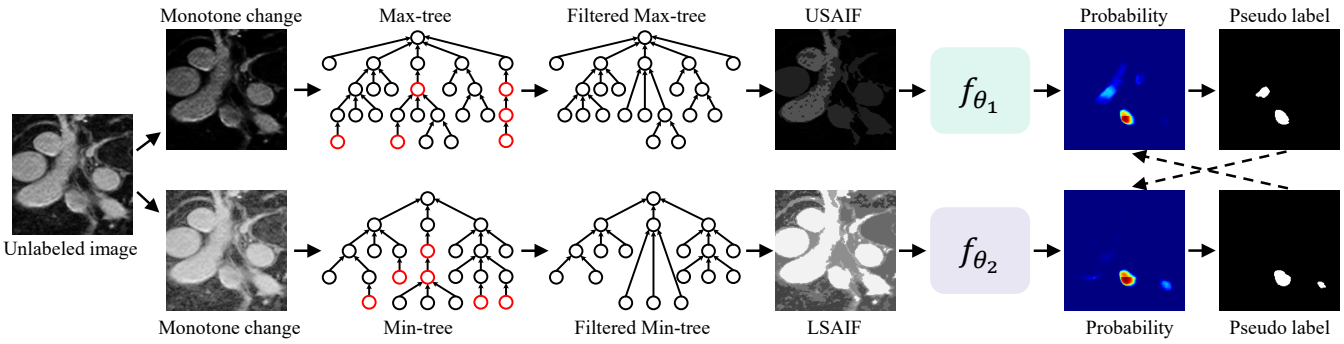


Fig. 1. The pipeline of the proposed DSAIF framework using mutual supervision of CPS Chen et al. (2021) as the model-level variations. We propose novel dual structure-aware image filterings (DSAIF) based on Max/Min-tree representation as the image-level variations. We remove every node (marked in red) without siblings in Max/Min-tree which is topologically equivalent to its ancestor node.

Tarvainen and Valpola, 2017) are mainly based on smoothness assumption (Chen et al., 2022), which aims to produce consistent results under small variations at image-level and/or model-level. The simple random augmentation (Sajjadi et al., 2016) and adversarial perturbation (Miyato et al., 2018) are representative works of image-level perturbations for consistency regularization. There are mainly three types of model-level perturbations: 1) Directly adding stochastic perturbation (*e.g.*, Gaussian noise (Rasmus et al., 2015) or dropout (Park et al., 2018)) to the model weights; 2) Mean Teacher (Tarvainen and Valpola, 2017) that ensembles model's parameters produced during training using exponential moving average (EMA) strategy; 3) Generating model variations via different decoders (Wu et al., 2022a) or networks (Qiao et al., 2018).

2.2. *Semi-supervised Medical Semantic Segmentation*

Semi-supervised learning is widely used in medical image segmentation tasks, thanks to its ability in alleviating the difficulty of manually annotating medical images.

Methods (Basak and Yin, 2023; Lei et al., 2022; Jin et al., 2022; Xiang et al., 2022; Basak et al., 2022; Wang et al., 2023b; Lyu et al., 2022; Adiga et al., 2024; Su et al., 2024) based on consistency regularization have achieved impressive performance for semi-supervised medical semantic segmentation. These methods usually use Mean Teacher framework (Tarvainen and Valpola, 2017) or Co-training strategy (Qiao et al., 2018) to generate variations in the model-level. Another way aims to generate diverse versions of the same image and enforce prediction consistency under image variations (Huang et al., 2022; Xu et al., 2021; Fan et al., 2022; Peiris et al., 2021; Wang et al., 2023a). A typical approach for image variations involves the weak-to-strong paradigm (Fan et al., 2022), where weakly-augmented and strongly-augmented images are employed to promote consistency. Methods (Peiris et al., 2021; Wang et al., 2023a) incorporate adversarial training strategy to generate adversarial perturbations on images and make the predictions robust to adversarial perturbations. Recently, an increasing number of methods enhance model performance by training unlabeled images with pseudo labels (Lyu et al., 2022; Qiao et al., 2022). Since there are inevitable noisy labels in the pseudo labels for unlabeled images, it is crucial to determine the confidence level of pseudo-labels (Qiao et al., 2022; Wang et al., 2021a). Moreover, some methods (Liu et al., 2022b) focus on pseudo rectifying during the training stage. Apart from these approaches, some methods (You et al., 2022b; Basak and Yin, 2023) exploit contrastive learning to achieve consistent feature representation.

Considering that objects of interest in medical images usually have specific shapes, some works (Li et al., 2020; Luo et al., 2021a; Meng et al., 2022; Wang et al., 2021b; Liu et al., 2022a) also incorporate shape information to alleviate the problem of insufficient labeled images in semi-supervised medical image segmentation. For instance, Li *et al.* (Li et al., 2020) leverage signed distance map (SDM) of object surfaces as a multi-task prediction jointly with semantic segmentation, and use an adversarial loss calculated by SDM as a geometric shape consistency constraint. A dual-task network is used in (Luo et al., 2021a) to jointly predict segmentation maps and level set representations that can capture global-level shape and geometric information of the target. Wang *et al.* (Wang et al., 2021b) extend the mean teacher architecture with foreground and background reconstruction task and signed distance field prediction task, to combine semantic information and shape information.

2.3. *Tree-based image representation*

Typically, an image is usually modeled as a discrete function defined on pixels or voxels over a 2D or 3D domain $V(\mathbb{R}^2 \text{or } \mathbb{R}^3)$. However, in the field of image processing and computer vision, many applications rely on interacting with some primitives of fundamental elements being more meaningful than the pixels. The tree-based image representation (Xu et al., 2014, 2015, 2016) is composed of a set of regions of the original image. These regions are either disjoint or have inclusion relationship between them, and thus can be encoded into a tree structure. Hierarchical segmentation and threshold decomposition are two main branches of tree-based image representations. A hierarchy of segmentation consists of a set of fine to coarse partitions. This hierarchy can be depicted as a tree structure, with the root node representing the entire image as a unified region, and the leaf nodes denoting the regions within the finest image partition. The intermediate nodes, situated between the root and the leaves, represent regions obtained through the fusion of all the regions represented by their child nodes. The α-tree (Soille,

2008) and the binary partition tree (BPT) (Salembier and Garrido, 2000) are two popular works of hierarchical segmentation.

Threshold decompositions developed in mathematical morphology are another widely used type of tree-based image representation. Image representations based on threshold decomposition rely solely on pixel-value ordering, rendering the generated tree structures invariant to monotonically increasing contrast changes. Embedding the set of upper level sets into a tree structure gives the Max-tree (Salembier et al., 1998). The root of the Max-tree represents the entire image domain, and the leaves correspond to the local regional maxima of the image. By duality, the lower level sets give rise to Min-tree representation (Salembier et al., 1998). The root of the Min-tree also represents the entire image domain, while the leaves correspond to the local regional minima of the image. The Max/Min-tree can be computed with quasi-linear complexity based on Union-Find process (Najman and Couprie, 2006; Carlinet and Géraud, 2014). Topographic map (Caselles et al., 1999), also known as tree of shapes (Monasse and Guichard, 2000), is another tree representation based on the threshold decomposition. It is derived by leveraging the inclusion relationship of the shapes, where a shape is defined as the connected component of upper or lower level sets with holes filled. The tree structures constructed through threshold decomposition are all contrast-invariant, offering a multi-scale representation, comprising a series of included or disjoint regions ranging from small to large scales (Xu et al., 2014, 2015). These trees are proved to be useful in many applications, such as lymphoma tumor segmentation from PET imaging Grossiord et al. (2020), local feature detection (Xu et al., 2014), or classification of high resolution satellite images (Luo and Zhang, 2013).

3. Method

3.1. Overview

Semi-supervised semantic segmentation task aims to enhance the performance of segmentation by leveraging a small set of labeled images $\mathcal{D}^l = \{(x^l, y^l)\}$ of N labeled images, along with a large collection of unlabeled images $\mathcal{D}^u = \{x^u\}$ of M unlabeled images, where $N \ll M$.

We follow the classical consistency regularization-based semi-supervised medical image segmentation framework, which is often composed of image-level variations and model-level variations on unlabeled images. For the image-level variations, we resort to the dual contrast-invariant Max-tree and Min-tree representations (see Sec. 3.2 for the construction) for connected filterings. We propose novel dual structure-aware image filterings (DSAIF) as the image-level variations. More specifically, we propose a novel type of connected filtering, that preserves only the topologically critical nodes of the Max/Min-tree. The corresponding filtering, named upper/lower structure-aware image filtering, and denoted as USAIF/LSAIF, yields two different images that have the same topological structure as the original one. We further leverage the invariance property of Max/Min-tree with respect to monotonically increasing contrast changes to further enforce the appearance diversity while preserving the topological image structure. For the model variations, we simply adopt cross pseudo supervision (CPS) method (Chen et al., 2021) as a baseline example to illustrate our method in Fig. 1. It is noteworthy that DSAIF can also be applied to other mutual supervision framework such as MC-Net (Wu et al., 2021), MC-Net+ (Wu et al., 2022a), Co-BioNet Peiris et al. (2023). The pipeline of the proposed framework using MC-Net, MC-Net+, Co-BioNet as baseline is depicted in the **Supplementary Material**.

3.2. Tree Construction

We utilize image threshold decompositions to build the Max/Min-tree representation. By thresholding a grayscale image x in descending order, starting from h_{max} to h_{min}, a sequence of nested upper level sets is obtained. Each upper level set at level h is denoted as

$$\mathcal{X}_h(x) = \{v \in V \mid x(v) \geq h\}. \tag{1}$$

Let $P_h^v(x)$ represents the binary connected operator of $\mathcal{X}_h(x)$ at point v, which gives the connected component of $\mathcal{X}_h(x)$ containing v if $v \in \mathcal{X}_h(x)$, and $\emptyset$ otherwise. Then, for any two connected components $P_{h_1}^{v_1}(x)$ and $P_{h_2}^{v_2}(x)$ at respectively level $h_1 \geq h_2$, we have either $P_{h_1}^{v_1}(x) \subseteq P_{h_2}^{v_2}(x)$, or $P_{h_1}^{v_1}(x) \cap P_{h_2}^{v_2}(x) = \emptyset$. Based on this inclusion relationship, a tree structure named Max-tree is formed, where nodes correspond to connected components. The parenthood between nodes corresponds to the inclusion relationship between the underlying connected components.

We use a water-covered surface analogy to better illustrate the process of Max-tree construction and the associated alterations in the level sets. For that, we suppose the surface is entirely submerged in water. With the level of water gradually decreasing, islands (regional maxima) emerge first to form the leaves of the tree. As the water level continues to drop, these islands expand, building the tree's branches. At certain levels, multiple islands fuse into a single connected piece, creating forks (i.e., the nodes of the tree with several children) in the tree structure. This process continues until all the water has evaporated, leaving behind a solitary landmass which forms the tree's root, representing the entirety of the image. By duality, a corresponding dual structure of the Max-tree, known as the Min-tree, is constructed based on the decomposition of lower level sets defined by

$$\mathcal{X}^h(x) = \{v \in V \mid x(v) \leq h\}. \tag{2}$$

A synthetic example of Max-tree and Min-tree is given in Fig. 2. The Max/Min-tree can be constructed efficiently using Union-Find-based algorithms (Najman and Couprie, 2006; Carlinet and Géraud, 2014), which have a quasi-linear complexity with respect to the number of pixels.

3.3. Dual Structure-Aware Image Filterings

The Max/Min-tree representation is equivalent to the original image in the sense that the image x can be reconstructed from the tree $\mathcal{T}$, composed of a set of nodes $\{\mathcal{N}\}$ with inclusion relationship encoded by *parent*. Specifically, we associate the graylevel h to the corresponding node on which the

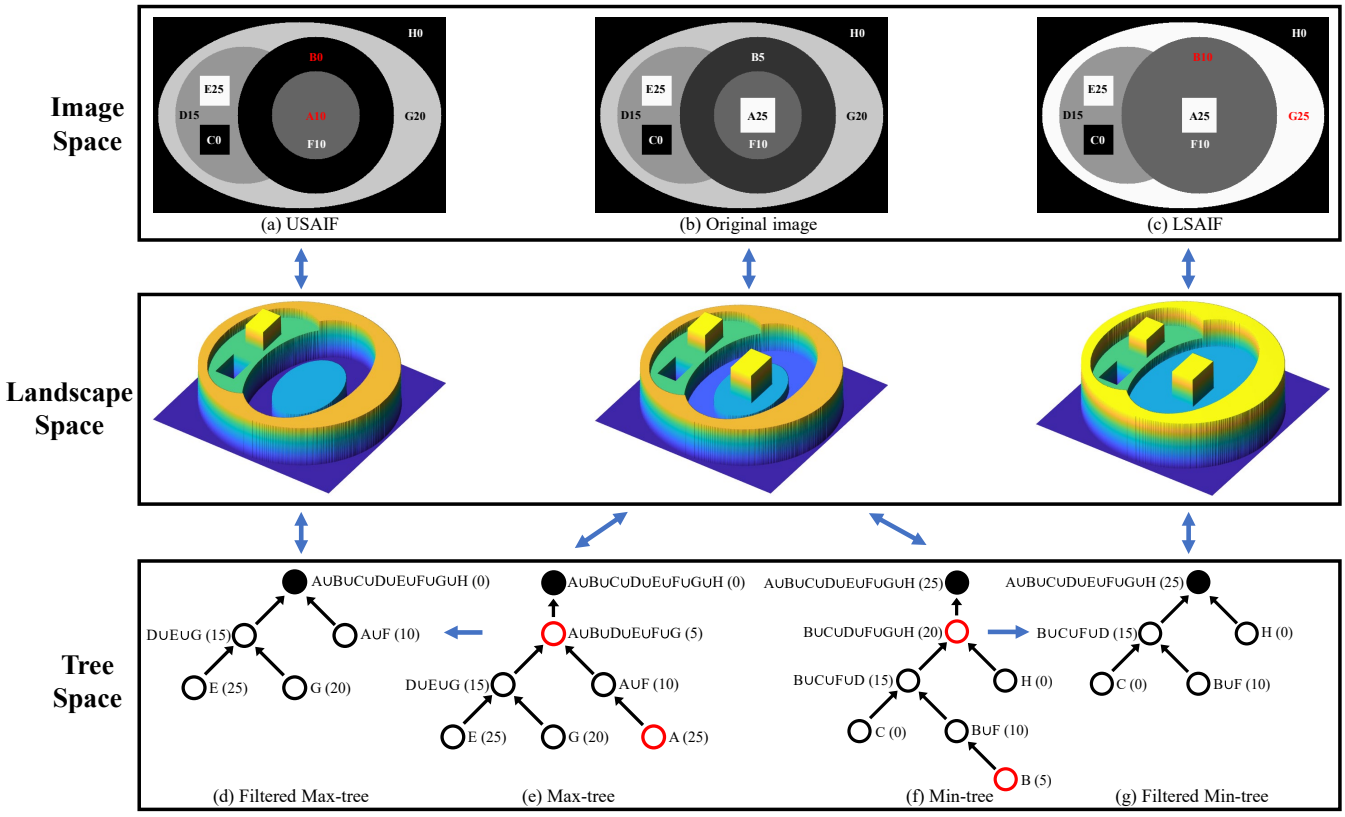


Fig. 2. An illustrative example of the proposed DSAIF. For the Max-tree (e) and Min-tree (f) built on the original image (b), we remove every node (marked in red) without siblings which is topologically equivalent to its ancestor node. The two images reconstructed from filtered Max/Min-tree denoted as USAIF (a) and LSAIF (c) have the same topological structure as the original image, but are of quite different appearances. The numbers after the letters in (a), (b), and (c) represent the gray level of the region. The numbers in parentheses in (d-e) (*resp.* (f-g)) means level h in Eq. (1) (*resp.* Eq. (2)).

underlying connected component is obtained. Each node represents a connected region in the image and has a gray level h, as shown in Fig. 2. Then, for each pixel $v \in V$, the grayscale value $x(v)$ is given by the associated graylevel of the smallest node containing v. For example, in the Max-tree of Fig. 2 (e), a pixel p_F within the region F is contained in the nodes $\mathcal{N}(A \cup B \cup C \cup D \cup E \cup F \cup G \cup H)$, $\mathcal{N}(A \cup B \cup D \cup E \cup F \cup G)$, $\mathcal{N}(A \cup F)$, where $\mathcal{N}(A \cup F) \subseteq \mathcal{N}(A \cup B \cup D \cup E \cup F \cup G) \subseteq \mathcal{N}(A \cup B \cup C \cup D \cup E \cup F \cup G \cup H)$. The node $\mathcal{N}(A \cup F)$ is the smallest node containing p_F. Hence, when converting the Max-tree back to an image, the grayscale value of the pixel p_F in the image is the grayscale level of the node $\mathcal{N}(A \cup F)$, which is 10. Removing nodes from the tree and updating the corresponding parenthood relationship results in a simplified tree, from which a filtered image is reconstructed. This is one of the most popular implementations of connected filters.

The topology of the tree encodes the topology of the image structure. The leaf nodes correspond to local regional maxima (*resp.* minima) in the Max-tree (*resp.* Min-tree). A node having more than one child signifies the fusion of two connected components, triggering a topological change of tree structure and thus image structure. A node having no siblings is topologically equivalent to its parent. Therefore, removing all nodes having no siblings does not change the topological structure of the image. This gives a simplified tree preserving topologically critical nodes. The filtered image reconstructed from the simplified tree has the same topological structure as the original image, but with different appearances. Such filter ψ is called upper/lower structure-aware image filter denoted as USAIF ψ_M and LSAIF ψ_m for the use of Max-tree and Min-tree, respectively. The filtered image by USAIF is no brighter than the original image, and the filtered image by LSAIF is no darker than the original image. Formally, for each pixel v in any image x, $\psi_M(x)(v) \leq x(v), \psi_m(x)(v) \geq x(v)$. Specifically, when deleting nodes with no siblings in the Max-tree as part of the USAIF process, it can lead to a change in the smallest node that contains certain pixels, thereby altering the grayscale values of these pixels during the restoration process from tree to image. For example, in area B of Fig. 2(b), a pixel q_B is included in the smallest node $\mathcal{N}(A \cup B \cup D \cup E \cup F \cup G)$ in the Max-tree (when converting the original Max-tree back to image, the grayscale value of the pixel q_B is 5). After USAIF filtering, the smallest node containing q_B in the filtered Max-tree becomes $\mathcal{N}(A \cup B \cup C \cup D \cup E \cup F \cup G \cup H)$, thus the grayscale value of q_B changes to 0. After removing certain nodes, the smallest node containing a pixel moves closer to the root node. Since the root node in a Max-tree has a smaller grayscale value, if a pixel's grayscale value changes due to USAIF filtering, it will only decrease. Conversely, since the root node in a Min-tree has a larger grayscale value, the image will change towards a higher grayscale value after LSAIF filtering. An illustrative example of the proposed dual structure-aware image filters (DSAIF) is

Algorithm 1: Structure-aware image filtering. Small regions with area less than τ may be caused by noise, and do not contribute to the topological changes. $\mathcal{T}$ is either a max-tree or min-tree.

Function STRUCT_AWARE_FILTER(x, τ)
  $\mathcal{T} \leftarrow Compute\_Tree(x)$; // max/min-tree computation. (in Sec. 3.2)
  foreach $\mathcal{N} \in \mathcal{T}$ **do**
    $numChildren(\mathcal{N}) \leftarrow 0$; // Initial number of child is 0
    $isRemoved(\mathcal{N}) \leftarrow$ **False** ; // Initialize as "not deleted"
  foreach $\mathcal{N} \in \mathcal{T}$ **do**
    if $area(\mathcal{N}) > \tau$ **then** // $area(\mathcal{N})$ = number of pixels contained in $\mathcal{N}$
      $++ numChildren(parent(\mathcal{N}))$;
      // Increase the number of children of the parent of $\mathcal{N}$
    else
      $isRemoved(\mathcal{N}) \leftarrow$ **True**; // Mark node $\mathcal{N}$ as deleted
  foreach $\mathcal{N} \in \mathcal{T}$ **do**
    // Mark node $\mathcal{N}$ with no-siblings as deleted.
    if $numChildren(parent(\mathcal{N})) = 1$ **then** $isRemoved(\mathcal{N}) \leftarrow$ **True**;
  foreach $v \in V$ **do**
    $\mathcal{N} \leftarrow Get\_Node(v)$; // Get smallest node that contains v
    while $isRemoved(\mathcal{N})$ **do**
      $\mathcal{N} \leftarrow parent(\mathcal{N})$;
    $x'(v) \leftarrow x(\mathcal{N})$; // Reconstruct the image from the tree.
  return x'

given in Fig. 2.

The dual structure-aware image filterings USAIF and LSAIF preserve the same topological structure as the original image, while generating diverse image appearances different from the original one. It is noteworthy that, different from classical monotonically increasing contrast changes (*e.g.*, Gamma correction) where pixels with the same graylevel have the same output graylevel, the proposed DSAIF may yield different output graylevels for the same input graylevel (see A and E in Fig. 2(a)). Since small regions may be caused by noise, and do not contribute much to the topological changes, we remove all nodes whose area is smaller than τ before performing DSAIF. The algorithm for the proposed structure-aware image filtering is given in Algorithm 1.

As illustrated in Fig. 2, an image can be viewed as a topological landscape with peaks and valleys. The topological structure of the landscape (*i.e.* image) is well reflected by the structure of Max-tree and Min-tree, where the leaf nodes represent peaks and valleys, respectively. Medical objects of interest often have some prior topological structure (*e.g.*, containing some peaks or valleys). The proposed DSAIF removes topologically equivalent nodes while preserving the critical ones whose merging triggers topological change. When the topological landscape has only one local minimum and one local maximum simultaneously, some regions with different gray levels may be merged into one in both USAIF and LSAIF. However, this is very rare in practice. Otherwise, either LSAIF or USAIF preserves the differentiated gray levels with the surrounding context. The use of both LSAIF and USAIF helps to effectively alleviate the confirmation bias problem during noisy pseudo label learning.

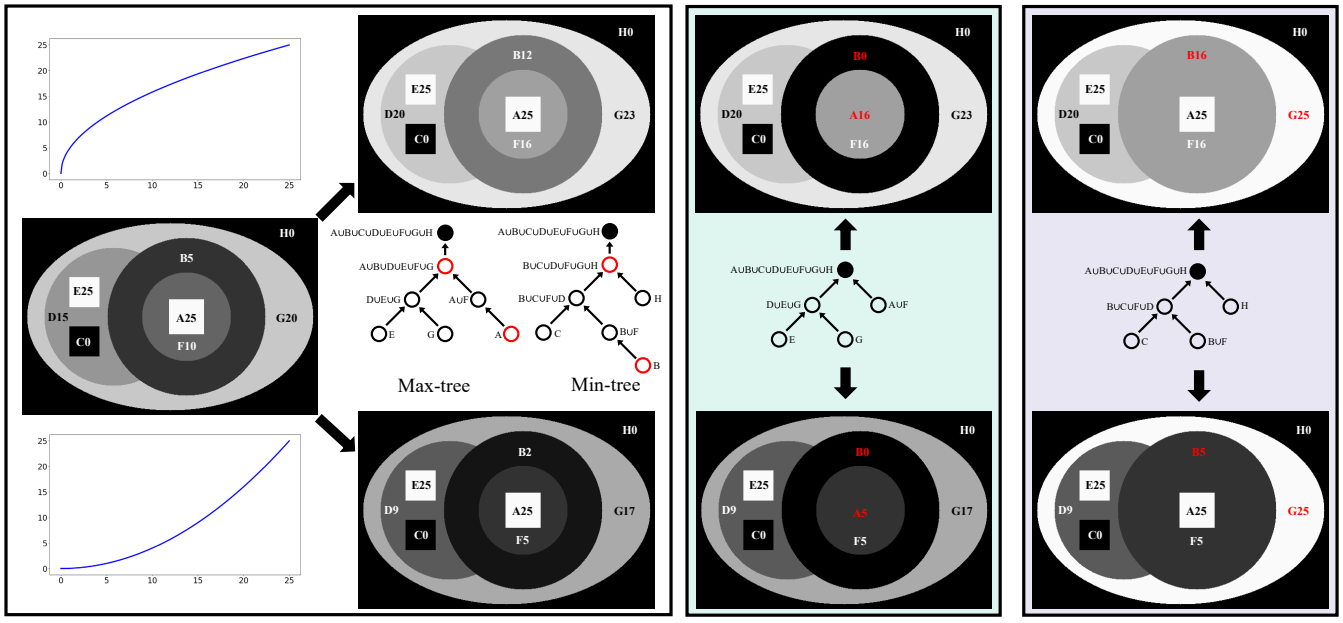
Since the Max-tree and Min-tree are invariant to monotonically increasing contrast changes, we further increase the appearance diversity while preserving the topological structure by applying some monotonically increasing contrast changes to the original image before performing DSAIF. Specifically, we use Gamma correction or monotonic Bézier Curve for each training image. For the Gamma correction augmentations, we independently random two Gamma values within [0.5, 1.5] to generate two different views of the image. A Bézier curve is a parametric curve defined by a set of control points. In this paper, we use two end points (P_0 and P_3) and two control points (P_1 and P_2) to generate cubic Bézier curves $B(t)$:

$$B(t) = (1-t)^3 P_0 + 3(1-t)^2 t P_1 + 3(1-t)t^2 P_2 + t^3 P_3,\ t \in [0,1], \quad (3)$$

where t is a fractional value along the length of the line. We set $P_0 = (-1,-1)$ and $P_3 = (1,1)$ as fixed points. Then, we set $P_1 = (-z, z)$ and $P_2 = (z, -z)$, where $z \in [0,1]$. In each iteration, we randomly choose z_1 and z_2 from $\{0, 0.5, 0.75\}$ to generate two transform functions B_1 and B_2 applied to the input image. As illustrated in Fig. 3, applying such monotonically increasing contrast changes to the image before performing DSAIF yields many different alternatives with diverse appearances while preserving the topological structure of the original image.

3.4. Mutual Supervision on Dual Structure-Aware Filtered Images

Network architecture: The model consists of two networks f_{θ_1} and f_{θ_2} with the same network architecture but different parameter initializations θ_1 and θ_2. For each image x, we apply the monotonically increasing contrast changes and the proposed DSAIF described in Sec. 3.3 to generate two different views x_1 and x_2 preserving the topological structure of the original image as the input for f_{θ_1} and f_{θ_2}, respectively. During the training process, we apply the DSAIF transformation to the images with a



(a) Invariance to monotonically increasing contrast changes (b) USAIF (c) LSAIF

Fig. 3. An illustrative example of leveraging the contrast-invariance property (a) of Max/Min-tree in DSAIF. Applying monotonically increasing contrast changes before DSAIF increases the appearance diversity while preserving the same topological structure as the original images.

probability of 0.5. In some iterations, both networks receive the original images as input, and then the pseudo-labels obtained by the two networks are used for mutual supervision. In other iterations, one network receives images filtered by USAIF, while the other receives images filtered by LSAIF.

Training objective: For each labeled image x^l, we adopt the cross-entropy loss ℓ_{ce} and dice loss ℓ_{dc} as the supervised loss $\mathcal{L}_s$ given by:

$$\mathcal{L}_s = \ell_{ce}(p_1^l, y^l) + \ell_{dc}(p_1^l, y^l) + \ell_{ce}(p_2^l, y^l) + \ell_{dc}(p_2^l, y^l), \quad (4)$$

where p_1^l and p_2^l are the prediction output of the two networks, and y^l is the corresponding label. For each unlabeled image x^u, we use the pseudo label obtained from one network to supervise the output of another one. The loss $\mathcal{L}_u$ for the unlabeled image x^u is given by:

$$\mathcal{L}_u = \ell_{ce}(p_1^u, \hat{y}_2) + \ell_{ce}(p_2^u, \hat{y}_1), \quad (5)$$

where $\hat{y}_1$ and $\hat{y}_2$ are pseudo labels obtained from p_1^u and p_2^u, respectively. The overall training objective $\mathcal{L}$ is defined by:

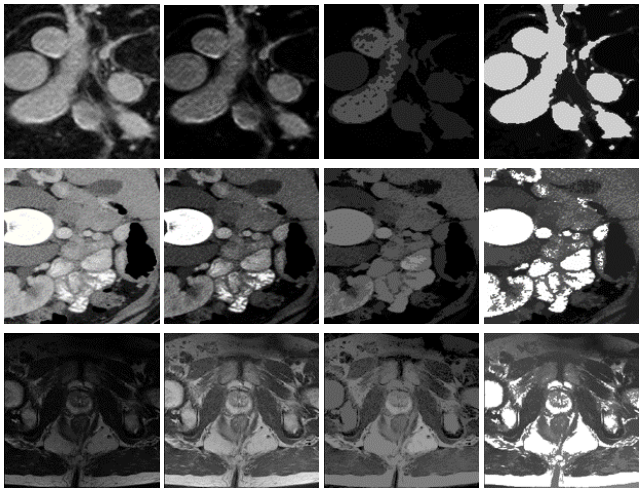
$$\mathcal{L} = \mathcal{L}_s + \lambda \times \mathcal{L}_u, \quad (6)$$

where λ balances the two loss terms.

4. Experiments

4.1. Dataset and Evaluation Protocal

Following some existing semi-supervised semantic segmentation methods, we mainly conduct experiments on the widely used 3D Left Atrium Segmentation MR Dataset (LA) (Xiong et al., 2021), Pancreas-NIH (Clark et al., 2013), and PROMISE12 dataset (Litjens et al., 2014).



(a) Image (b) Changed image (c) USAIF (d) LSAIF

Fig. 4. Some qualitative results of DSAIF on LA dataset (Xiong et al., 2021) (first row), Pancreas-CT (Clark et al., 2013) (middle row), and PROMISE12 (Litjens et al., 2014) (bottom row). The changed images in (b) are obtained by applying monotonically increasing contrast change to the original ones.

LA Dataset: 3D Left Atrial Segmentation Challenge dataset (Xiong et al., 2021) consists of 100 MRI scans. Following Wu *et al.* (Wu et al., 2022a), a fixed split is utilized, where 80 samples are designated for training and the remaining 20 samples are allocated for testing.

Pancreas-NIH Dataset: Pancreas-NIH dataset (Clark et al., 2013) consists of 82 3D abdominal contrast-enhanced CT scans. Following the commonly-used data split in Luo (Luo et al., 2021a), we take 62 samples for training and the rest 20 samples for testing.

PROMISE12 Dataset: PROMISE12 dataset (Litjens et al., 2014) consists of 50 transverse T2-weighted MRI scans. Following the data split in Liu *et al.* (Liu et al., 2022a), there are 35, 5, and 10 scans for training, validation, and testing. Due to the low cross-slice resolution, PROMISE12 dataset are segmented in 2D (slice by slice) (Liu et al., 2022a).

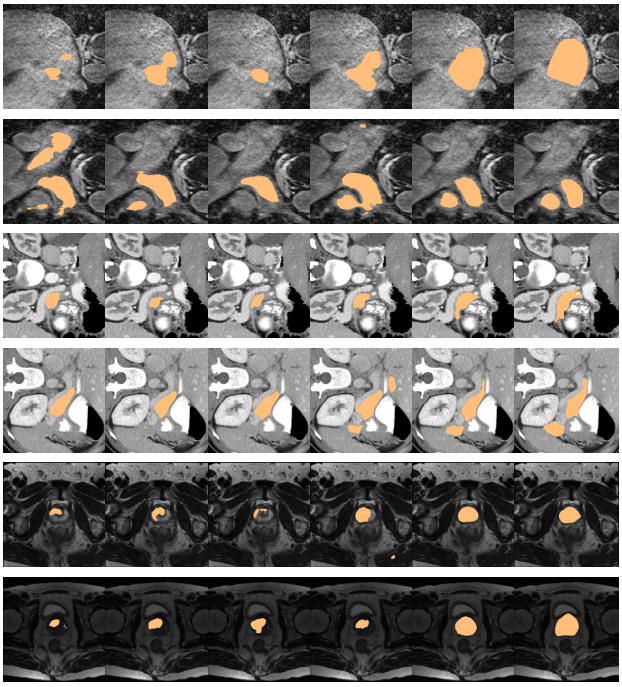
Evaluation protocol: The proposed method is evaluated with four widely used metrics in semi-supervised medical image segmentation: Dice coefficient (Dice), Jaccard Index (JAC), the 95% Hausdorff Distance (95HD), and the average surface distance (ASD).

4.2. Implementation Details

The SGD optimizer with a learning rate 10^{-2} and a weight decay factor 10^{-4} is used for all experiments. The loss weight λ in Eq. (6) is set as a time-dependent Gaussian warming-up function (Laine and Aila, 2016) using the same parameters as MC-Net+ (Wu et al., 2022a). We adopt the V-Net (*resp.* U-Net) model as the backbone for 3D (*resp.* 2D) segmentation tasks following the same settings in MC-Net+ (Wu et al., 2022a) for fair comparisons. The area threshold parameter τ in DSAIF is set to 50 in 2D segmentation experiments and 100 in 3D segmentation experiments. For the LA Dataset and Pancreas-NIH dataset, we use the same geometric transformation settings as MC-Net (Wu et al., 2021) . Specifically, the LA Dataset employs rotation, cropping, and flipping, while the Pancreas-NIH dataset only uses cropping. For the PROMISE12 Dataset, since MC-Net did not conduct experiment on this dataset, we use the same geometric transformation settings as SALCC (Liu et al., 2022a), including rotation, flipping, and resizing. All the experiments are conducted using the Pytorch framework with two NVIDIA GeForce RTX 3090 GPUs.

4.3. Qualitative Results of DSAIF

Some qualitative results of the proposed DSAIF are shown in Fig. 4. Both USAIF and LSAIF generate images with diverse appearances while preserving the same topological structure as the original image. Inheriting from the property of connected filters, the proposed DSAIF does not create any new contours. It is also noteworthy that monotonically increasing contrast change map pixels with the same graylevel to the same output graylevel. Differently, the output of DSAIF does not only depend on the input graylevel, but also the image structure. As shown in the first row of Fig. 4, for similar input graylevels on different pixels, USAIF may output very different graylevels on these pixels. Yet, the topological image structure is preserved.



(a) UA-MT (b) URPC (c) MC-Net+ (d) CPS (e) DSAIF (f) GT

Fig. 5. Some qualitative segmentation results of DSAIF on LA dataset (Xiong et al., 2021) (first two rows), Pancreas-CT dataset (Clark et al., 2013) (middle two rows), and PROMISE12 dataset (Litjens et al., 2014) (bottom two rows).

4.4. Comparative Results on Different Datasets

Some qualitative segmentation results on the three datasets are shown in Fig. 5, where we can observe that the proposed DSAIF achieves accurate segmentation results. We compare our proposed method with several state-of-the-art methods in the field of medical semi-supervised segmentation. **Among them, UA-MT (Yu et al., 2019), CVRL (You et al., 2022a), SS-Net (Wu et al., 2022b), SimCVD (You et al., 2022b), LLRU (Adiga Vasudeva et al., 2022), DUO-Net (Peiris et al., 2021), SCO-SSL (Xu et al., 2021), BCP (Bai et al., 2023), AC-MT (Xu et al., 2023), AAU (Adiga et al., 2024) also apply image-level variations.** The proposed DSAIF outperforms these image-level variations methods on three datasets, indicating the effectiveness of structure-aware image-level perturbations for semi-supervised medical image segmentation.

Results on LA Dataset: Tab. 1 depicts the quantitative evaluation of the LA dataset. Applying DSAIF achieves consistent improvement across four baselines. Specifically, when using 10% of labeled data, it enhances the Dice coefficient by 2.74% on MC-Net, 3.41% on CPS, 2% on MC-Net+, and 0.66% on Co-BioNet. Under the setting of using 10% labeled images, the proposed method outperforms the baseline CPS by 5.17% Jaccard index. Under the setting of using 20% labeled images, the application of DSAIF on MC-Net+ achieved the best performance, reaching 92.05% Dice coefficient and 85.32% Jaccard Index .

Results on Pancreas-NIH Dataset: The quantitative results on the Pancreas-CT dataset are shown in Tab. 2. Under the

Table 1. Quantitative evaluation on the LA dataset (Xiong et al., 2021). † represents the reproduced results based on the open-sourced implementation. We report the mean and standard deviation obtained over three runs.

L/U	Method	Dice (%)↑	JAC (%)↑	95HD↓	ASD↓
80/0	V-Net†	91.82	84.92	5.12	1.71
8/0	V-Net†	80.75	69.81	15.11	3.61
8/72 (10%)	CVRL (You et al., 2022a)	88.56	78.89	8.22	2.81
	SS-NET (Wu et al., 2022b)	88.55	79.62	7.49	1.90
	SimCVD (You et al., 2022b)	89.03	80.34	8.34	2.59
	LLRU (Adiga Vasudeva et al., 2022)	86.58	-	11.82	-
	AC-MT (Xu et al., 2023)	89.12	80.46	11.05	2.19
	BCP (Bai et al., 2023)	89.62	81.31	6.81	1.76
	AAU (Adiga et al., 2024)	86.58	-	11.82	-
	MC-Net (Wu et al., 2021)†	87.89±0.51	78.58±0.51	10.47±1.98	2.23±0.48
	MC-Net + DSAIF	**90.63±0.34**	**82.92±0.45**	**6.69±0.97**	**1.57±0.26**
	CPS (Chen et al., 2021)†	86.79±0.42	77.05±0.56	14.19±2.08	4.25±0.40
	CPS + DSAIF	90.20±0.14	82.22±0.23	6.72±0.19	1.77±0.12
	MC-Net+ (Wu et al., 2022a)†	88.37±0.45	80.01±0.50	8.16±1.87	1.99±0.43
	MC-Net+ + DSAIF†	90.37±0.31	82.55±0.41	7.47±0.95	1.74±0.24
	Co-BioNet (Peiris et al., 2023)†	89.05±0.63	79.19±0.57	7.09±1.97	2.12±0.53
	Co-BioNet + DSAIF†	89.71±0.46	81.18±0.44	7.36±1.33	2.19±0.48
16/0	V-Net†	88.41	79.43	10.05	2.40
16/64 (20%)	CVRL (You et al., 2022a)	90.45	83.02	6.56	1.81
	SimCVD (You et al., 2022b)	90.85	83.80	6.03	1.86
	LLRU (Adiga Vasudeva et al., 2022)	88.60	-	7.61	-
	MCF (Wang et al., 2023b)	88.71	80.41	6.32	1.90
	AC-MT (Xu et al., 2023)	90.31	82.43	6.21	1.76
	BCP (Bai et al., 2023)†	90.03	82.35	6.17	1.68
	AAU (Adiga et al., 2024)	88.60	-	7.61	-
	MC-Net (Wu et al., 2021)†	90.21±0.34	82.24±0.63	6.78±0.87	1.70±0.31
	MC-Net + DSAIF	91.63±0.26	84.61±0.42	5.33±0.58	**1.35±0.19**
	CPS (Chen et al., 2021)†	90.07±0.30	82.08±0.49	6.81±0.41	1.99±0.21
	CPS + DSAIF	91.20±0.21	83.90±0.35	5.49±0.37	1.57±0.08
	MC-Net+ (Wu et al., 2022a)†	91.03±0.29	83.47±0.52	5.99±0.69	1.86±0.25
	MC-Net+ + DSAIF†	**92.05±0.22**	**85.32±0.38**	**4.68±0.54**	1.44±0.16
	Co-BioNet (Peiris et al., 2023)†	91.04±0.48	82.73±0.79	5.68±0.99	1.52±0.49
	Co-BioNet + DSAIF†	91.25±0.43	83.97±0.67	5.14±0.79	1.40±0.37

Table 2. Quantitative evaluation on the Pancreas-NIH dataset (Clark et al., 2013). † represents the reproduced results based on the open-sourced implementation. ‡ represents Multi-scale MC-Net+. We report the mean and standard deviation obtained over three runs.

L/U	Method	Dice (%)↑	JAC (%)↑	95HD↓	ASD↓
62/0	V-Net†	82.68	71.05	5.19	1.41
6/0	V-Net†	68.69	55.03	13.47	3.63
6/56 (10%)	UA-MT (Yu et al., 2019)	66.44	52.02	17.04	3.03
	URPC (Luo et al., 2021b)	73.53	59.44	22.57	7.85
	DTC(Luo et al., 2021a)	66.58	51.79	15.46	4.16
	MC-Net+ (Wu et al., 2022a)‡	74.01	60.02	12.59	3.34
	BCP (Bai et al., 2023)†	75.57	61.35	27.29	8.16
	MLRP (Su et al., 2024)	75.93	62.12	9.07	**1.54**
	MC-Net (Wu et al., 2021)†	69.47±0.82	55.28±1.15	20.74±1.98	5.41±0.67
	MC-Net + DSAIF	70.49±0.84	56.63±1.27	13.03±2.12	2.48±0.84
	CPS (Chen et al., 2021)†	74.95±0.91	60.86±1.06	13.49±1.87	4.59±0.77
	CPS + DSAIF	**80.91±0.70**	**68.31±0.92**	**7.67±1.63**	2.18±0.25
	MC-Net+ (Wu et al., 2022a)†	70.13±0.76	55.72±1.08	15.95±1.74	3.79±0.59
	MC-Net+ + DSAIF†	73.34±0.65	59.60±1.01	13.02±1.67	1.66±0.53
	Co-BioNet (Peiris et al., 2023)†	77.80±0.78	64.54±1.16	9.36±1.78	2.46±0.61
	Co-BioNet + DSAIF†	79.63±0.67	66.77±0.91	8.26±1.62	2.19±0.34
12/0	V-Net†	76.91	63.86	8.16	2.07
12/50 (20%)	UA-MT (Yu et al., 2019)	76.10	62.62	10.84	2.43
	URPC (Luo et al., 2021b)	80.02	67.30	8.51	1.98
	DTC(Luo et al., 2021a)	76.27	62.82	8.70	2.20
	CVRL(You et al., 2022a)	76.68	61.16	8.24	3.19
	SimCVD (You et al., 2022b)	75.39	61.56	9.84	2.33
	MC-Net+ (Wu et al., 2022a)‡	80.59	68.08	6.47	1.74
	MCF (Wang et al., 2023b)	75.00	61.27	11.59	3.27
	BCP (Bai et al., 2023)	**82.91**	**70.79**	6.43	2.25
	MLRP (Su et al., 2024)	81.53	69.35	6.81	**1.33**
	MC-Net (Wu et al., 2021)†	78.26±0.35	65.12±0.42	11.90±2.04	3.25±0.83
	MC-Net + DSAIF	79.13±0.42	66.23±0.38	7.77±1.65	1.85±0.43
	CPS (Chen et al., 2021)†	79.63±0.14	66.77±0.19	8.77±1.46	2.33±0.62
	CPS + DSAIF	82.66±0.32	70.75±0.44	7.11±1.01	1.75±0.22
	MC-Net+ (Wu et al., 2022a)†	79.16±0.31	66.56±0.35	8.62±1.61	1.79±0.42
	MC-Net+ + DSAIF†	80.65±0.27	68.22±0.25	7.10±1.49	1.47±0.35
	Co-BioNet (Peiris et al., 2023)†	82.05±0.39	70.08±0.53	5.58±1.74	1.46±0.59
	Co-BioNet + DSAIF†	82.52±0.30	70.59±0.39	**5.47±1.53**	1.49±0.51

setting of using 10% labeled data, the proposed method significantly improves the baseline CPS (Chen et al., 2021) by 5.96% Dice coefficient and 7.45% Jaccard index, and significantly outperforms the other state-of-the-art methods. Using 20% labeled data under the CPS (Chen et al., 2021) baseline, the best result among our three experiments is 82.90 Dice, 71.10 JAC, and 1.60 ASD, which is comparable to the results of BCP (Bai et al., 2023). It is worth noting that MC-Net+ (Wu et al., 2022a) provides two kinds of results for the Pancreas-CT dataset: one without using the multi-scale strategy, consistent with other datasets, and another using the multi-scale strategy. Since the open-source code does not include the multi-scale version, we incorporate DSAIF based on the version without using the multi-scale strategy.

Results on PROMISE12 Dataset: On the PROMISE12 dataset, the proposed method achieves even more significant improvements across four baselines. In particular, as depicted in Tab. 3, the proposed method outperforms the CPS baseline by 18.27% Dice and 21.67% JAC (*resp.*, 11.93% Dice and 13.35% JAC) under the setting of using 10% (*resp.* 20%) labeled data. When using 20% of labeled data, it enhances the Dice coefficient (*resp.* Jaccard index) by 7.86% (*resp.* 8.16%) on MC-Net, 11.93% (*resp.* 13.35 %) on CPS, 12.23% (*resp.* 11.25%) on MC-Net+, and 5.56% (*resp.* 6.03%) on Co-BioNet. The more significant improvement on this dataset is probably because that the image variance within the dataset is more prominent, further demonstrating the effectiveness of the proposed DSAIF using structure information for semi-supervised medical image segmentation. Quantitative trend analyses of our DSAIF under CPS baseline for different labeled data portions are depicted in Fig. 6-Fig. 8.

4.5. Ablation Studies

We conduct ablation studies on LA dataset under the setting of using 10% labeled data. As depicted in Tab. 4, directly adopting monotonically increasing contrast changes and random rotation as data augmentations do not significantly improve the results (86.79% to 87.20% Dice). Applying the proposed DSAIF on the original images outperforms the baseline by 1.38% Dice and 1.97% Jaccard index. Besides, combining these data augmentations and DSAIF significantly boosts the segmentation results by 3.0% Dice and 4.54% Jaccard index. This demonstrates that the performance improvement is mainly brought by the proposed DSAIF.

We also conduct an ablation study on the area threshold τ involved in the proposed DSAIF. As shown in Fig. 5, different settings of τ slightly influence the results. Using too small values makes DSAIF sensitive to noise. Using too large values may filter out some important regions. Setting $\tau = 100$ gives the best result.

4.6. Domain Generalization Results

We conduct cross-domain experiments on prostate segmentation task in the semi-supervised setting to further verify the

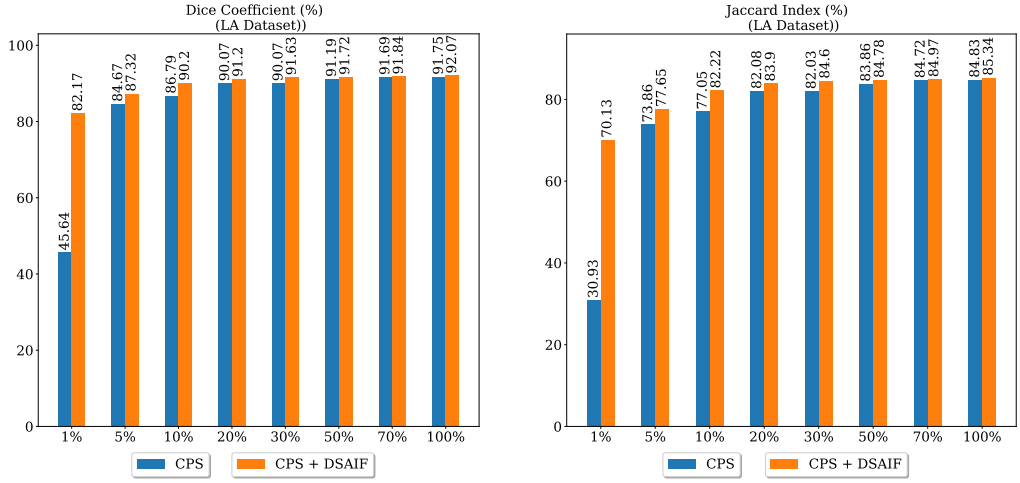


Fig. 6. Quantitative trend analysis on LA dataset under CPS baseline for different labeled data portions.

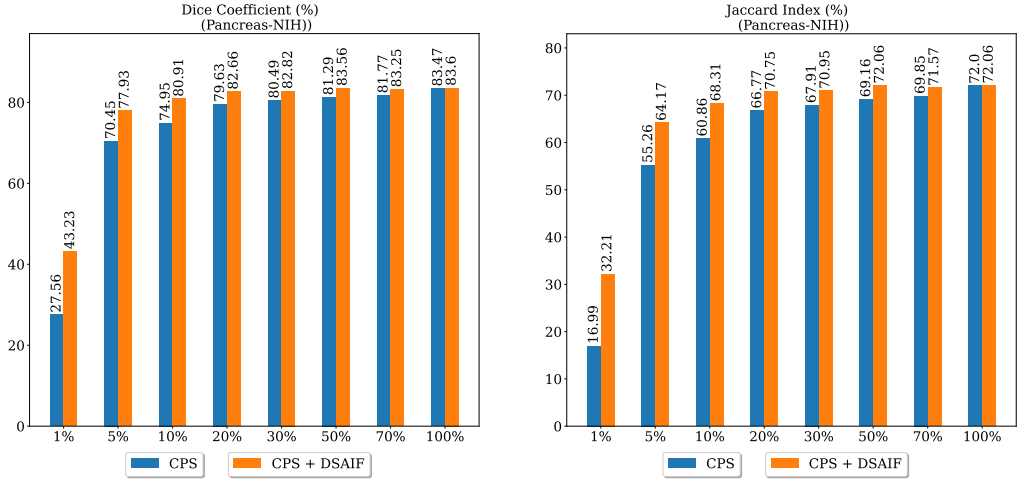


Fig. 7. Quantitative trend analysis on Pancreas-NIH dataset under CPS baseline for different labeled data portions.

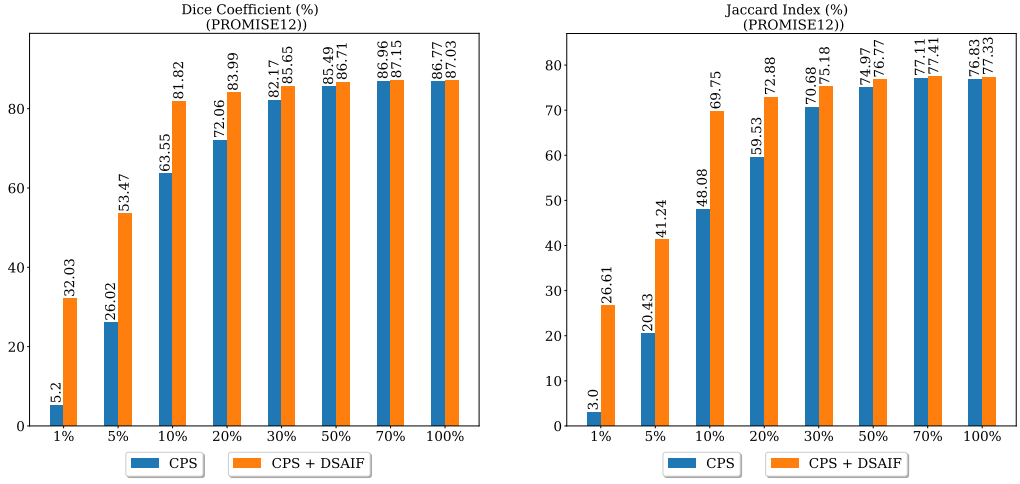


Fig. 8. Quantitative trend analysis on PROMISE12 dataset under CPS baseline for different labeled data portions.

Table 3. Quantitative evaluation on the PROMISE12 dataset (Litjens et al., 2014). † represents the reproduced results based on the open-sourced implementation. We report the mean and standard deviation obtained over three runs.

L/U	Method	Dice (%)↑	JAC (%)↑	95HD↓	ASD↓
35/0	U-Net†	84.63	73.73	3.63	1.27
4/0	U-Net†	51.72	39.02	43.30	11.81
4/31 (10%)	UA-MT (Yu et al., 2019)†	49.81	37.52	74.03	15.95
	URPC (Luo et al., 2021b)†	51.70	38.30	75.19	22.32
	SCO-SSL (Xu et al., 2021)†	62.34	48.04	33.79	9.39
	DUO-Net (Peiris et al., 2021)†	45.02	32.64	69.36	29.62
	SALCC (Liu et al., 2022a)†	63.85	48.07	40.68	7.50
	SCP-Net (Zhang et al., 2023)	66.21	-	-	11.56
	BCP (Bai et al., 2023)†	77.93	64.35	27.39	6.65
	MC-Net (Wu et al., 2021)†	54.26±4.29	41.44±2.89	35.82±4.63	10.48±2.31
	MC-Net + DSAIF	78.26±1.02	65.22±0.86	11.26±1.12	4.58±0.68
	CPS (Chen et al., 2021)†	63.55±2.50	48.08±1.76	40.50±2.66	12.20±1.42
	CPS + DSAIF	**81.82±0.18**	**69.75±0.27**	**5.14±0.68**	**1.42±0.11**
	MC-Net+ (Wu et al., 2022a)†	55.40±3.65	42.01±2.47	31.79±3.57	9.51±1.99
	MC-Net+ + DSAIF†	72.75±1.21	58.87±0.96	11.94±1.42	2.97±0.84
	Co-BioNet (Peiris et al., 2023)†	66.25±2.76	50.92±1.83	23.41±2.89	8.42±1.61
	Co-BioNet + DSAIF†	80.82±0.45	68.18±0.72	8.84±1.03	2.15±0.67
7/0	U-Net†	65.63	52.44	12.36	2.98
7/28 (20%)	UA-MT (Yu et al., 2019)	61.55	-	-	13.94
	URPC (Luo et al., 2021b)	61.55	-	-	9.63
	SCO-SSL (Xu et al., 2021)†	74.60	61.72	18.40	5.02
	DUO-Net (Peiris et al., 2021)†	70.18	54.95	22.14	5.49
	SALCC (Liu et al., 2022a)	70.30	-	-	4.69
	SCP-Net (Zhang et al., 2023)	77.06	-	-	3.52
	BCP (Bai et al., 2023)†	80.26	67.46	9.67	4.87
	MC-Net (Wu et al., 2021)†	71.70±1.26	59.71±1.98	19.48±2.02	8.38±1.34
	MC-Net + DSAIF	79.56±1.67	67.87±2.12	15.32±2.48	6.17±1.54
	CPS (Chen et al., 2021)†	72.06±2.73	59.53±2.21	9.55±1.72	1.71±0.05
	CPS + DSAIF	**83.99±0.81**	**72.88±1.09**	**5.04±0.61**	**1.24±0.11**
	MC-Net+ (Wu et al., 2022a)†	66.91±1.72	55.26±2.31	12.74±1.93	2.41±1.51
	MC-Net+ + DSAIF†	79.14±1.13	66.51±1.75	10.60±1.83	2.47±1.20
	Co-BioNet (Peiris et al., 2023)†	76.32±1.64	63.74±1.91	7.23±1.74	1.72±1.17
	Co-BioNet + DSAIF†	81.88±1.26	69.77±1.87	5.50±1.68	1.49±1.02

Table 4. Abaltion study on LA dataset (Xiong et al., 2021) under 10% labeled data using CPS Chen et al. (2021) as baseline. We report the mean and standard deviation obtained over three runs.

CPS	Aug	DSAIF	Dice (%)	JAC (%)	95HD	ASD
✓			86.79±0.42	77.05±0.56	14.19±2.08	4.25±0.40
✓	✓		87.20±0.34	77.68±0.48	12.58±1.15	4.02±0.53
✓		✓	88.17±0.09	79.02±0.14	11.26±1.81	3.02±0.45
✓	✓	✓	**90.20±0.14**	**82.22±0.23**	**6.72±0.19**	**1.77±0.12**

generalization performance of the proposed DSAIF. Under the setting of using 10% (*resp.* 20%) labeled data, we use 4 (*resp.* 7) labeled images and 31 (*resp.* 28) unlabeled images in PROMISE12 dataset to train the model, and test the model on 2 different data sources with distribution shift: Site A and B are from NCI-ISBI13 dataset (Bloch et al., 2015). As depicted in Tab. 6, though the monotonically increasing contrast changes is helpful in domain generalization, the proposed DSAIF further significantly improves the baseline of using monotonically increasing contrast changes, demonstrating the effectiveness of DSAIF in domain generalization under semi-supervised setting. Specifically, under the setting of using 10% labeled data, DSAIF achieves 11.56% (*resp.* 9.16%) Dice (*resp.* JAC) improvement on Site A and 8.61% (*resp.* 7.53%) Dice (*resp.* JAC) improvement on Site B. Under the setting of using 20% labeled data, DSAIF achieves 4.77% (*resp.* 4.76%) Dice (*resp.* JAC) improvement on Site A and a 9.02% (*resp.* 8.39%) Dice (*resp.* JAC) improvement on Site B.

Table 5. Ablation study on the area threshold τ involved in the proposed DSAIF on LA dataset (Xiong et al., 2021) under 10% labeled data using CPS Chen et al. (2021) as baseline.

Threshold τ	0	50	100	150	200
Dice (%)	89.23	90.03	90.33	89.76	89.01

Table 6. Cross-dataset performance on prostate segmentation. We report the mean and standard deviation over three runs.

Setting		L/U: 4/31		L/U: 7/28	
		Dice (%)	JAC (%)	Dice (%)	JAC (%)
Site A	CPS (Baseline)	15.34±1.32	9.89±1.16	65.14±2.08	52.50±2.00
	CPS + Aug	28.80±5.39	18.73±4.02	70.43±2.33	57.31±1.29
	CPS + DSAIF	**40.36±2.62**	**27.89±2.17**	**75.20±0.33**	**62.07±0.61**
Site B	CPS (Baseline)	37.30±6.86	25.84±6.05	40.37± 6.91	33.45±4.21
	CPS + Aug	39.89±1.84	28.10±1.69	55.14±2.97	41.03±1.00
	CPS + DSAIF	**48.50±4.01**	**35.63±3.92**	**64.16±2.89**	**49.42±2.95**

4.7. Discussion

In the scenario with unlabeled data, two networks learn through mutual supervision. The pseudo-labels produced by f_{θ_1} serve as the supervised signal for f_{θ_2}, and vice versa. During the training process, three situations may arise: 1) Both networks produce correct pseudo-labels; 2) only one of the networks generates correct pseudo-label, while the other produces incorrect pseudo-label; 3) both networks produce incorrect outputs. When the first or second scenario occurs, the networks have the opportunity to gain useful knowledge from the unlabeled data. However, in the third scenario, where both networks make consistent incorrect predictions on some pixels, the mutual supervision between them may lead to a confirmation bias in the results. This makes the model overfit to noisy pseudo labels, yielding degenerated segmentation performance. Appropriate diversity between two networks' erroneous predictions helps to avoid such confirmation bias issue of overfitting to incorrect pseudo-labels. We define quantitative metrics D_e (*resp.* D_c) to characterize such diversity of erroneous predictions (*resp.* correct predictions) on unlabeled training images between the two mutually supervised networks. For that, let $\mathcal{E}^1$ and $\mathcal{E}^2$ denote the set of pixels with incorrect prediction of the first and second network, respectively. We compute D_e as the Dice score between $\mathcal{E}^1$ and $\mathcal{E}^2$ given by:

$$D_e = 2 \times |\mathcal{E}^1 \cap \mathcal{E}^2| / (|\mathcal{E}^1| + |\mathcal{E}^2|), \tag{7}$$

where $|\cdot|$ denotes the cardinality. Similarly, let C^1 and C^2 denote the set of pixels with correct prediction of the first and second network, respectively. We compute D_c as the Dice score between C^1 and C^2 given by:

$$D_c = 2 \times |C^1 \cap C^2| / (|C^1| + |C^2|). \tag{8}$$

The comparison of D_e and D_c for the baseline model and the proposed method during the training process is depicted in Fig. 9. Indeed, for a pair of different augmented versions of the same image, the more different they are, it is reasonable that the segmentation results differ more. This may yield more pronounced inconsistency on both incorrect and correct predictions. Thanks to the large appearance diversity between USAIF

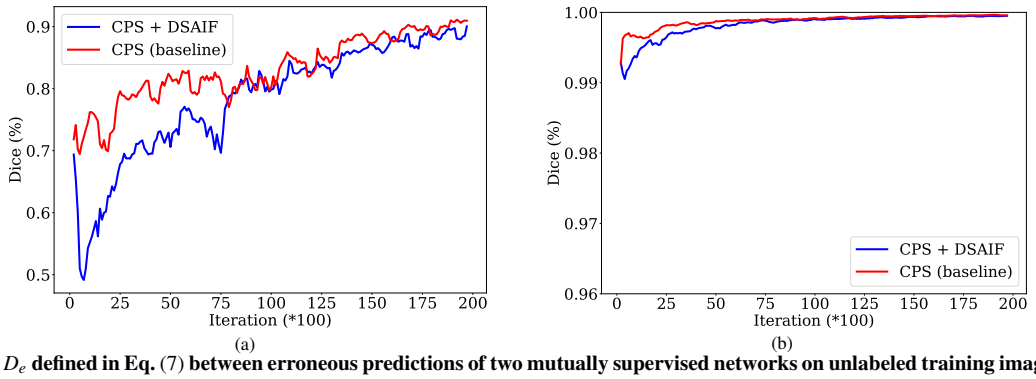


Fig. 9. (a) Dice score D_e defined in Eq. (7) between erroneous predictions of two mutually supervised networks on unlabeled training images of PROMISE12 Dataset (Litjens et al., 2014) during the training process. (b) Dice score D_r defined in Eq. (8) between correct predictions of two mutually supervised networks on unlabeled training images of PROMISE12 Dataset (Litjens et al., 2014) during the training process.

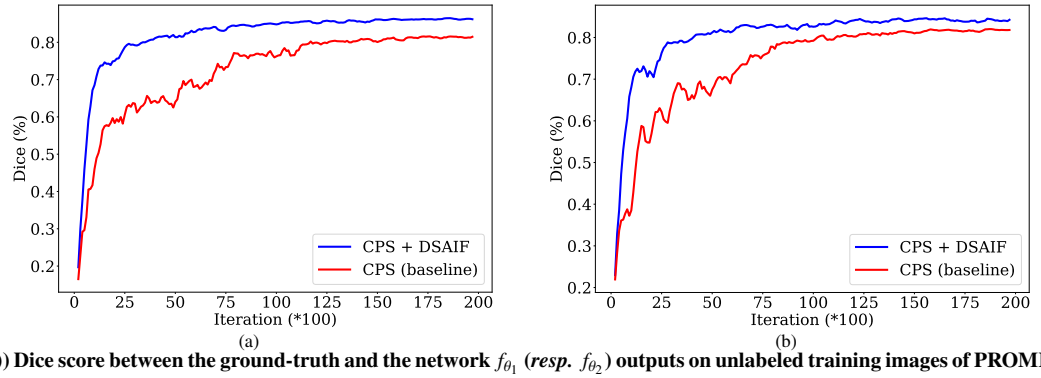


Fig. 10. (a) (*resp.* (b)) Dice score between the ground-truth and the network f_{θ_1} (*resp.* f_{θ_2}) outputs on unlabeled training images of PROMISE12 Dataset (Litjens et al., 2014) at different iterations in the training process.

and LSAIF while preserving the same topological structure as the original image, the proposed DSAIF has less consensus on the erroneous predictions of the two mutually supervised networks (see the left side of Fig. 9). The inconsistency between correct predictions (see the right side of Fig. 9) is in general less significant than that between erroneous predictions. Besides, in the early stages of the training process, the difference between the two Dice D_c curves on correct predictions for the baseline method and our method is greater than in the later stages. As training progresses, the two curves converge. By the time it reaches 10,000 iterations, the difference between them narrows to less than 0.001. This also indicates that, as long as one network predicts correctly, both networks will evolve towards the correct direction as training continues (see the almost 1.0 for the Dice D_c at the end of the training process). This helps to alleviate the confirmation bias issue of overfitting to noisy pseudo labels on unlabeled images, resulting in better pseudo labels of unlabeled images during the training process (see Fig. 10). Therefore, the proposed DSAIF is effective in improving the performance of semi-supervised medical image segmentation.

It is noteworthy that the tree of shapes (Monasse and Guichard, 2000), known also as topographic maps (Caselles et al., 1999), provides another way to convert the image into a structure-aware tree space. This leads to a **single** filtered image while preserving the topological structure of the original image. Yet, the proposed method requires generating **dual** images with very different appearances to decrease the consensus of erroneous predictions on unlabeled images. Therefore, we choose the Max/Min-tree representation in our DSAIF. However, it would be interesting to explore the use of tree of shapes for more structure-aware filters in semi-supervised medical image segmentation. This is left for future work.

On the other hand, the prior knowledge about the topological structure of medical objects has also been explored in (Clough et al., 2020; Hu et al., 2019, 2021; Hu, 2022; Gupta et al., 2022; Singh et al., 2023) for medical image analysis. Most of them focus on designing topology-aware loss functions to incorporate the prior structure knowledge, helping to yield more plausible segmentation results. This is different from our DSAIF, which aims to generate dual images with different appearances while preserving the critical topological structure of the original image. This helps to cope with confirmation bias issue in semi-supervised medical image segmentation. It would also be interesting to combine DSAIF with these topological analysis tools in the future work.

A limitation of the current work is that the proposed DSAIF requires some extra time during the training process (but no extra runtime during inference). As shown in Alg. 1, the DSAIF is composed of 1) Min/Max-tree computation based on union-find process, 2) removing nodes with small areas and nodes without siblings and 3) reconstructing filtered image from Min/Max-tree. The union-find-based Min/Max-tree computation approach (Najman and Couprie, 2006; Carlinet and Géraud, 2014) would take $O(n\alpha(n))$ time, where n is the

total number of pixels and α is a very slow-growing "diagonal inverse" of the Ackermann's function. Removing nodes with small areas and nodes without siblings would take $O(N)$ time, where N is the total number of nodes of the tree which is smaller than n. Reconstructing a filtered image from Min/Max-tree would take $O(n)$ time. When using the online strategy for training, DSAIF processing is required for each iteration, which adds some extra time. For instance, on the PROMISE12 dataset, which uses 2D slices as inputs, the total training time for the baseline (*resp.*, our DSAIF) is about 4.5 hours (*resp.*, 7 hours). On the LA dataset, which uses 3D volumes as inputs, the total training time of the baseline (*resp.*, our DSAIF) is about 7.5 hours (*resp.*, 13 hours). It is noteworthy that the inference time remains the same as the baseline method, while requiring acceptable extra training time. Currently, we adopt CPU-based algorithm to build Max/Min-tree, which is not as efficient as GPU-based algorithm (Blin et al., 2022). Yet, this GPU-based algorithm (Blin et al., 2022) does not support 3D images. In the future, we plan to explore the implementation of DSAIF with GPU (Blin et al., 2022) to further accelerate the training process. An alternative solution is to compute the DSAIf offline, which does not increases the training time. This approach will be explored in the future.

5. Conclusion

We propose a novel image-level variation method named dual structure-aware image filterings (DSAIF) for semi-supervised medical image segmentation. Specifically, we leverage the dual Max-tree and Min-tree image representation, and remove all nodes having no siblings in the corresponding tree. This equals to remove all topologically equivalent regions while preserving topologically critical ones, resulting in two images with diverse appearances while having the same topological structure as the original image. By incorporating the proposed DSAIF into mutually supervised networks, the consensus on erroneous predictions for unlabeled images is decreased. This helps to alleviate the confirmation bias issue, where models tend to overfit to noisy pseudo labels, thereby enhancing the performance of segmentation. Extensive experimental results on three widely used benchmark datasets demonstrate that the proposed method significantly/consistently outperforms the state-of-the-art methods. In the future, we would like to explore DSAIF in more semi-supervised medical image segmentation frameworks, and using tree of shapes for more structure-aware filters. Combing DSAIF with other topological analysis tools is also an interesting direction to explore.

References

Adiga, S., Dolz, J., Lombaert, H., 2024. Anatomically-aware uncertainty for semi-supervised image segmentation. Medical Image Analysis 91, 103011.

Adiga Vasudeva, S., Dolz, J., Lombaert, H., 2022. Leveraging labeling representations in uncertainty-based semi-supervised segmentation, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 265–275.

Arazo, E., Ortego, D., Albert, P., O'Connor, N.E., McGuinness, K., 2020. Pseudo-labeling and confirmation bias in deep semi-supervised learning, in: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–8.

Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y., 2023. Bidirectional copy-paste for semi-supervised medical image segmentation, in: IEEE Conf. Comput. Vis. Pattern Recog., pp. 11514–11524.

Basak, H., Ghosal, S., Sarkar, R., 2022. Addressing class imbalance in semi-supervised image segmentation: A study on cardiac mri, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 224–233.

Basak, H., Yin, Z., 2023. Pseudo-label guided contrastive learning for semi-supervised medical image segmentation, in: IEEE Conf. Comput. Vis. Pattern Recog., pp. 19786–19797.

Blin, N., Carlinet, E., Lemaitre, F., Lacassagne, L., Géraud, T., 2022. Max-tree computation on GPUs. IEEE Transactions on Parallel and Distributed Systems 33, 3520–3531.

Bloch, N., Madabhushi, A., Huisman, H., Freymann, J., Kirby, J., Grauer, M., Enquobahrie, A., Jaffe, C., Clarke, L., Farahani, K., 2015. Nci-isbi 2013 challenge: automated segmentation of prostate structures. The Cancer Imaging Archive 370, 6.

Carlinet, E., Géraud, T., 2014. A comparative review of component tree computation algorithms. IEEE Trans. Image Process. 23, 3885–3895.

Caselles, V., Coll, B., Morel, J.M., 1999. Topographic maps and local contrast changes in natural images. Int. J. Comput. Vis. 33, 5–27.

Chen, X., Yuan, Y., Zeng, G., Wang, J., 2021. Semi-supervised semantic segmentation with cross pseudo supervision, in: IEEE Conf. Comput. Vis. Pattern Recog., pp. 2613–2622.

Chen, Y., Mancini, M., Zhu, X., Akata, Z., 2022. Semi-supervised and unsupervised deep visual learning: A survey. IEEE Trans. Pattern Anal. Mach. Intell. .

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al., 2013. The cancer imaging archive (TCIA): maintaining and operating a public information repository. Journal of Digital Imaging 26, 1045–1057.

Clough, J.R., Byrne, N., Oksuz, I., Zimmer, V.A., Schnabel, J.A., King, A.P., 2020. A topological loss function for deep-learning based image segmentation using persistent homology. IEEE Trans. Pattern Anal. Mach. Intell. 44, 8766–8778.

Fan, J., Gao, B., Jin, H., Jiang, L., 2022. UCC: Uncertainty guided cross-head co-training for semi-supervised semantic segmentation, in: IEEE Conf. Comput. Vis. Pattern Recog., pp. 9947–9956.

Grandvalet, Y., Bengio, Y., 2005. Semi-supervised learning by entropy minimization, in: Adv. Neural Inform. Process. Syst.

Grossiord, E., Passat, N., Talbot, H., Naegel, B., Kanoun, S., Tal, I., Tervé, P., Ken, S., Casasnovas, O., Meignan, M., et al., 2020. Shaping for pet image analysis. Pattern Recognition Letters 131, 307–313.

Gupta, S., Hu, X., Kaan, J., Jin, M., Mpoy, M., Chung, K., Singh, G., Saltz, M., Kurc, T., Saltz, J., et al., 2022. Learning topological interactions for multi-class medical image segmentation, in: Eur. Conf. Comput. Vis., pp. 701–718.

Hu, X., 2022. Structure-aware image segmentation with homotopy warping. Adv. Neural Inform. Process. Syst. 35, 24046–24059.

Hu, X., Li, F., Samaras, D., Chen, C., 2019. Topology-preserving deep image segmentation. Adv. Neural Inform. Process. Syst. 32.

Hu, X., Wang, Y., Fuxin, L., Samaras, D., Chen, C., 2021. Topology-aware segmentation using discrete morse theory. Int. Conf. Learn. Represent. .

Huang, W., Chen, C., Xiong, Z., Zhang, Y., Chen, X., Sun, X., Wu, F., 2022. Semi-supervised neuron segmentation via reinforced consistency learning. IEEE Trans. Medical Imaging. 41, 3016–3028.

Jin, Q., Cui, H., Sun, C., Zheng, J., Wei, L., Fang, Z., Meng, Z., Su, R., 2022. Semi-supervised histological image segmentation via hierarchical consistency enforcement, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 3–13.

Laine, S., Aila, T., 2016. Temporal ensembling for semi-supervised learning, in: Int. Conf. Learn. Represent.

Lei, T., Zhang, D., Du, X., Wang, X., Wan, Y., Nandi, A.K., 2022. Semi-supervised medical image segmentation using adversarial consistency learning and dynamic convolution network. IEEE Trans. Medical Imaging. .

Li, S., Zhang, C., He, X., 2020. Shape-aware semi-supervised 3d semantic segmentation for medical images, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 552–561.

Litjens, G., Toth, R., van de Ven, W., Hoeks, C., Kerkstra, S., van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al., 2014. Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. Medical Image Analysis 18, 359–373.

Liu, J., Desrosiers, C., Zhou, Y., 2022a. Semi-supervised medical image segmentation using cross-model pseudo-supervision with shape awareness and local context constraints, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 140–150.

Liu, Y., Tian, Y., Chen, Y., Liu, F., Belagiannis, V., Carneiro, G., 2022b. Perturbed and strict mean teachers for semi-supervised semantic segmentation, in: IEEE Conf. Comput. Vis. Pattern Recog., pp. 4258–4267.

Luo, B., Zhang, L., 2013. Robust autodual morphological profiles for the classification of high-resolution satellite images. IEEE Transactions on Geoscience and Remote Sensing 52, 1451–1462.

Luo, X., Chen, J., Song, T., Wang, G., 2021a. Semi-supervised medical image segmentation through dual-task consistency, in: AAAI, pp. 8801–8809.

Luo, X., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Chen, N., Wang, G., Zhang, S., 2021b. Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 318–329.

Lyu, F., Ye, M., Carlsen, J.F., Erleben, K., Darkner, S., Yuen, P.C., 2022. Pseudo-label guided image synthesis for semi-supervised covid-19 pneumonia infection segmentation. IEEE Trans. Medical Imaging. 42, 797–809.

Meng, Y., Zhang, H., Zhao, Y., Gao, D., Hamill, B., Patri, G., Peto, T., Madhusudhan, S., Zheng, Y., 2022. Dual consistency enabled weakly and semi-supervised optic disc and cup segmentation with dual adaptive graph convolutional networks. IEEE Trans. Medical Imaging. 42, 416–429.

Miyato, T., Maeda, S.i., Koyama, M., Ishii, S., 2018. Virtual adversarial training: A regularization method for supervised and semi-supervised learning. IEEE Trans. Pattern Anal. Mach. Intell. 41, 1979–1993.

Monasse, P., Guichard, F., 2000. Fast computation of a contrast-invariant image representation. IEEE Trans. Image Process. 9, 860–872.

Najman, L., Couprie, M., 2006. Building the component tree in quasi-linear time. IEEE Trans. Image Process. 15, 3531–3539.

Ouzounis, G.K., Wilkinson, M.H., 2007. Mask-based second-generation connectivity and attribute filters. IEEE Trans. Pattern Anal. Mach. Intell. 29, 990–1004.

Park, S., Park, J., Shin, S.J., Moon, I.C., 2018. Adversarial dropout for supervised and semi-supervised learning, in: AAAI.

Peiris, H., Chen, Z., Egan, G., Harandi, M., 2021. Duo-SegNet: adversarial dual-views for semi-supervised medical image segmentation, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 428–438.

Peiris, H., Hayat, M., Chen, Z., Egan, G., Harandi, M., 2023. Uncertainty-guided dual-views for semi-supervised volumetric medical image segmentation. Nature Machine Intelligence 5, 724–738.

Qiao, P., Li, H., Song, G., Han, H., Gao, Z., Tian, Y., Liang, Y., Li, X., Zhou, S.K., Chen, J., 2022. Semi-supervised ct lesion segmentation using uncertainty-based data pairing and swapmix. IEEE Trans. Medical Imaging. .

Qiao, S., Shen, W., Zhang, Z., Wang, B., Yuille, A., 2018. Deep co-training for semi-supervised image recognition, in: Eur. Conf. Comput. Vis., pp. 135–152.

Rasmus, A., Berglund, M., Honkala, M., Valpola, H., Raiko, T., 2015. Semi-supervised learning with ladder networks. Adv. Neural Inform. Process. Syst. 28.

Sajjadi, M., Javanmardi, M., Tasdizen, T., 2016. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. Adv. Neural Inform. Process. Syst. 29.

Salembier, P., Garrido, L., 2000. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. IEEE Trans. Image Process. 9, 561–576.

Salembier, P., Oliveras, A., Garrido, L., 1998. Antiextensive connected operators for image and sequence processing. IEEE Trans. Image Process. 7, 555–570.

Singh, Y., Farrelly, C.M., Hathaway, Q.A., Leiner, T., Jagtap, J., Carlsson, G.E., Erickson, B.J., 2023. Topological data analysis in medical imaging: current state of the art. Insights into Imaging 14, 58.

Soille, P., 2008. Constrained connectivity for hierarchical image partitioning and simplification. IEEE Trans. Pattern Anal. Mach. Intell. 30, 1132–1145.

Su, J., Luo, Z., Lian, S., Lin, D., Li, S., 2024. Mutual learning with reliable pseudo label for semi-supervised medical image segmentation. Medical Image Analysis , 103111.

Tarvainen, A., Valpola, H., 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Adv. Neural Inform. Process. Syst. 30.

Wang, G., Zhai, S., Lasio, G., Zhang, B., Yi, B., Chen, S., Macvittie, T.J., Metaxas, D., Zhou, J., Zhang, S., 2021a. Semi-supervised segmentation of radiation-induced pulmonary fibrosis from lung ct scans with multi-scale guided dense attention. IEEE Trans. Medical Imaging. 41, 531–542.

Wang, K., Zhan, B., Zu, C., Wu, X., Zhou, J., Zhou, L., Wang, Y., 2021b. Tripled-uncertainty guided mean teacher model for semi-supervised medical image segmentation, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 450–460.

Wang, P., Peng, J., Pedersoli, M., Zhou, Y., Zhang, C., Desrosiers, C., 2023a. CAT: Constrained adversarial training for anatomically-plausible semi-supervised segmentation. IEEE Trans. Medical Imaging. .

Wang, Y., Xiao, B., Bi, X., Li, W., Gao, X., 2023b. MCF: Mutual correction framework for semi-supervised medical image segmentation, in: IEEE Conf. Comput. Vis. Pattern Recog., pp. 15651–15660.

Westenberg, M.A., Roerdink, J.B., Wilkinson, M.H., 2007. Volumetric attribute filtering and interactive visualization using the max-tree representation. IEEE Trans. Image Process. 16, 2943–2952.

Wilkinson, M.H., Gao, H., Hesselink, W.H., Jonker, J.E., Meijster, A., 2008. Concurrent computation of attribute filters on shared memory parallel machines. IEEE Trans. Pattern Anal. Mach. Intell. 30, 1800–1813.

Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., Cai, J., 2022a. Mutual consistency learning for semi-supervised medical image segmentation. Medical Image Analysis 81, 102530.

Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J., 2022b. Exploring smoothness and class-separation for semi-supervised medical image segmentation, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 34–43.

Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L., 2021. Semi-supervised left atrium segmentation with mutual consistency training, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 297–306.

Xiang, J., Qiu, P., Yang, Y., 2022. FUSSNet: Fusing two sources of uncertainty for semi-supervised medical image segmentation, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 481–491.

Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., et al., 2021. A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Medical Image Analysis 67, 101832.

Xu, X., Sanford, T., Turkbey, B., Xu, S., Wood, B.J., Yan, P., 2021. Shadow-consistent semi-supervised learning for prostate ultrasound segmentation. IEEE Trans. Medical Imaging. 41, 1331–1345.

Xu, Y., Carlinet, E., Géraud, T., Najman, L., 2016. Hierarchical segmentation using tree-based shape spaces. IEEE Trans. Pattern Anal. Mach. Intell. 39, 457–469.

Xu, Y., Géraud, T., Najman, L., 2015. Connected filtering on tree-based shape-spaces. IEEE Trans. Pattern Anal. Mach. Intell. 38, 1126–1140.

Xu, Y., Monasse, P., Géraud, T., Najman, L., 2014. Tree-based morse regions: A topological approach to local feature detection. IEEE Trans. Image Process. 23, 5612–5625.

Xu, Z., Wang, Y., Lu, D., Luo, X., Yan, J., Zheng, Y., Tong, R.K.y., 2023. Ambiguity-selective consistency regularization for mean-teacher semi-supervised medical image segmentation. Medical Image Analysis 88, 102880.

Yang, L., Qi, L., Feng, L., Zhang, W., Shi, Y., 2023. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation, in: IEEE Conf. Comput. Vis. Pattern Recog., pp. 7236–7246.

You, C., Zhao, R., Staib, L.H., Duncan, J.S., 2022a. Momentum contrastive voxel-wise representation learning for semi-supervised volumetric medical image segmentation, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 639–652.

You, C., Zhou, Y., Zhao, R., Staib, L., Duncan, J.S., 2022b. SimCVD: Simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation. IEEE Trans. Medical Imaging. 41, 2228–2237.

Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A., 2019. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 605–613.

Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y., 2019. CutMix: Regularization strategy to train strong classifiers with localizable features, in: IEEE Conf. Comput. Vis. Pattern Recog., pp. 6023–6032.

Zhang, Z., Ran, R., Tian, C., Zhou, H., Li, X., Yang, F., Jiao, Z., 2023. Self-aware and cross-sample prototypical learning for semi-supervised medical image segmentation, in: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention.

Zhao, Z., Yang, L., Long, S., Pi, J., Zhou, L., Wang, J., 2023. Augmentation Matters: A simple-yet-effective approach to semi-supervised semantic segmentation, in: IEEE Conf. Comput. Vis. Pattern Recog., pp. 11350–11359.

Supplementary Material on the pipeline of DSAIF, based on MC-Net Wu et al. (2021), MC-Net+ Wu et al. (2022a), and Co-BioNet Peiris et al. (2023)

MC-Net. The pipeline of the proposed framework based on MC-Net (Wu et al., 2021) is depicted in Fig. 11. MC-Net (Wu et al., 2021) comprises one shared encoder and two different decoders with distinct up-sampling strategies. Different from CPS, MC-Net (Wu et al., 2021) introduces a mutual consistency constraint between the probability output of one decoder and the soft pseudo labels of the other decoder. DSAIF enables the shared encoder to receive both Max-tree and Min-tree filtered images. The two decoders probabilistically receive features of images generated by either filtered Max-tree or filtered Min-tree in every iteration. When one decoder takes in features of the image generated with filtered Max-tree, the other decoder processes features of the same image generated with filtered Min-tree, and vice versa.

MC-Net+. The pipeline of the proposed framework based on MC-Net+ (Wu et al., 2022a) is depicted in Fig. 12. MC-Net+ (Wu et al., 2022a) comprises one shared encoder, and three different decoders with distinct up-sampling strategies. DSAIF enables the shared encoder to receive Max-tree, Min-tree filtered images, and original images. The first and third decoders probabilistically receive features of images generated by either filtered Max-tree or filtered Min-tree in every iteration, and the second decoder receive features of original images.

Co-BioNet. The pipeline of the proposed framework based on Co-BioNet (Peiris et al., 2023) is depicted in Fig. 13. Co-BioNet (Peiris et al., 2023) comprises two networks. DSAIF enables the two networks to receive both Max-tree and Min-tree filtered images, and other settings are consistent with Co-BioNet.

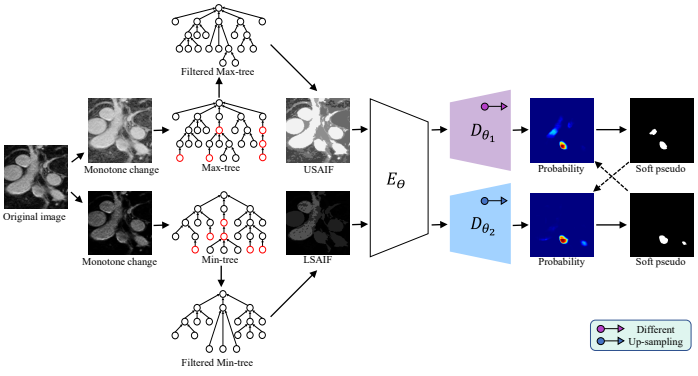


Fig. 11. The pipeline of the proposed DSAIF framework using mutual supervision of MC-Net (Wu et al., 2021) as the model-level variations. The pipeline composed of image-level variations and model-level variations on images. We propose novel dual structure-aware image filterings (DSAIF) based on Max/Min-tree representation as the image-level variations.

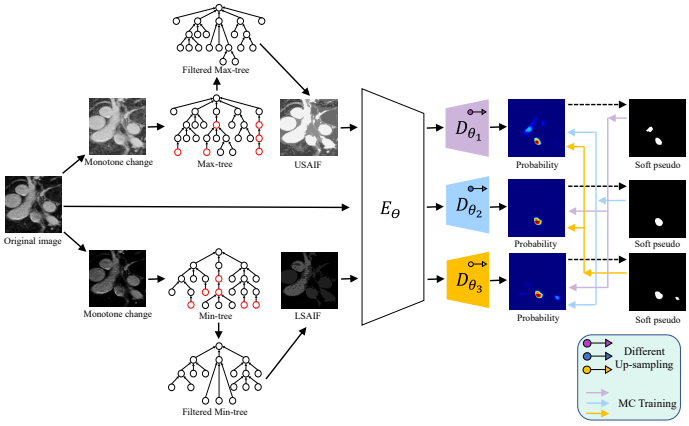


Fig. 12. The pipeline of the proposed DSAIF framework using mutual supervision of MC-Net+ (Wu et al., 2022a) as the model-level variations. The pipeline composed of image-level variations and model-level variations on images. We propose novel dual structure-aware image filterings (DSAIF) based on Max/Min-tree representation as the image-level variations.

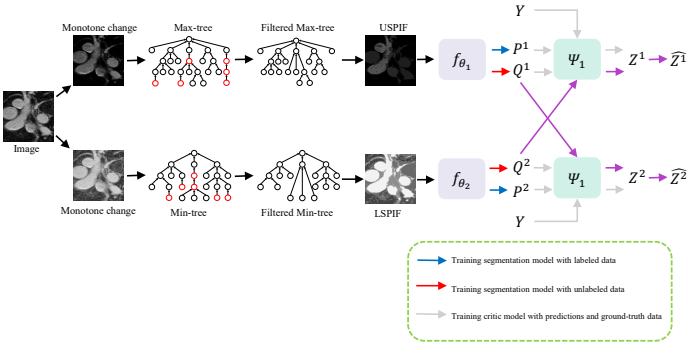


Fig. 13. The pipeline of the proposed DSAIF framework using mutual supervision of Co-BioNet (Peiris et al., 2023) as the model-level variations. The pipeline composed of image-level variations and model-level variations on images. We propose novel dual structure-aware image filterings (DSAIF) based on Max/Min-tree representation as the image-level variations.