Local Intensity Order Transformation for Robust Curvilinear Object Segmentation

Tianyi Shi, Nicolas Boutry, Yongchao Xu, Thierry Géraud

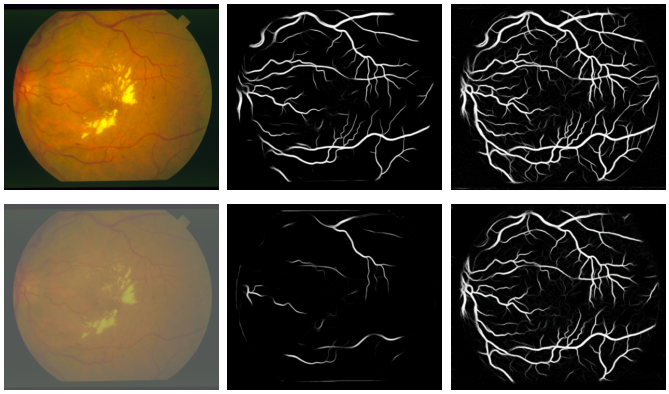
***Abstract*—Segmentation of curvilinear structures is important in many applications, such as retinal blood vessel segmentation for early detection of vessel diseases and pavement crack segmentation for road condition evaluation and maintenance. Currently, deep learning-based methods have achieved impressive performance on these tasks. Yet, most of them mainly focus on finding powerful deep architectures but ignore capturing the inherent curvilinear structure feature (*e.g.*, the curvilinear structure is darker than the context) for a more robust representation. In consequence, the performance usually drops a lot on cross-datasets, which poses great challenges in practice. In this paper, we aim to improve the generalizability by introducing a novel local intensity order transformation (LIOT). Specifically, we transfer a gray-scale image into a contrast-invariant four-channel image based on the intensity order between each pixel and its nearby pixels along with the four (horizontal and vertical) directions. This results in a representation that preserves the inherent characteristic of the curvilinear structure while being robust to contrast changes. Cross-dataset evaluation on three retinal blood vessel segmentation datasets demonstrates that LIOT improves the generalizability of some state-of-the-art methods. Additionally, the cross-dataset evaluation between retinal blood vessel segmentation and pavement crack segmentation shows that LIOT is able to preserve the inherent characteristic of curvilinear structure with large appearance gaps. An implementation of the proposed method is available at https://github.com/TY-Shi/LIOT.**

***Index Terms*—Curvilinear structure, segmentation, local intensity order, deep learning, generalizability**

I. Introduction

Curvilinear structures often appear in biomedical analysis. Their segmentation is very important in applications like retinal fundus disease screening [1], [2], early detection of vessel diseases, and in biometric authentication systems [3]–[5]. In parallel, curvilinear structures also appear in pavement crack segmentation, which is useful for road condition evaluation and road maintenance [6].

Compared to general object segmentation, curvilinear structure segmentation faces some particular challenges [7]: 1) thin, long, and tortuosity shapes; 2) inadequate contrast between curvilinear structures and the surrounding background; 3) uneven background illumination; 4) various image appearances. To cope with these challenges, classical curvilinear object segmentation methods [6], [8]–[22] mainly focus on designing specifically engineered features. They usually adopt filters or some morphological tools to capture one or more specific features from images. Such hand-crafted feature-based methods usually require careful parameter tuning, which is difficult to handle a wide variety of complex curvilinear structure segmentation.



(a) Input (b) Seg. on input (c) Seg. on LIOT

Fig. 1: An example of segmentation results on an image and the same image with changed contrast. Top row: segmentation on the original image by IterNet [23]; Bottom row: segmentation on contrast changed image. Different from direct segmentation on the original image, the proposed LIOT is invariant to contrast changes and thus yields robust segmentation results.

Most recent methods [24]–[40] leverage deep learning [41]–[44] for curvilinear object segmentation and have achieved significant improvement over previous methods. These methods mainly focus on designing various network architectures or loss functions to improve the segmentation performance. Despite the high in-dataset accuracy for the state-of-the-art methods, they usually do not generalize well to images with different image appearances. For instance, as illustrated in Fig. 1, one may achieve good results on the image similar to the training dataset. The segmentation accuracy greatly drops on the same image with slightly changed contrast. Yet, the inherent characteristic of curvilinear structure (*e.g.*, darker than the context) holds for both images. This implies that the generalizability of curvilinear object segmentation methods still remains challenging. Designing a robust method that captures the inherent property of curvilinear structure is of great interest.

In this paper, we aim to address the above issue and focus on improving the generalizability of current deep learning-based methods in segmenting curvilinear objects. Instead of directly

T. Shi is with the School of EIC, Huazhong University of Science and Technology (HUST), Wuhan, 430074, China (E-mail: shitianyihust@hust.edu.cn).

N. Boutry and T. Géraud are with Laboratoire de Recherche et Développement de l'EPITA (LRDE), Le Kremlin-Bicêtre, 94276, France (E-mail: {nicolas.boutry, thierry.geraud}@lrde.epita.fr).

Y. Xu is with the School of Computer Science, Wuhan University, Wuhan, 430070, China (E-mail: yongchao.xu@whu.edu.cn).

Corresponding author: Yongchao Xu.

operating on the original image, we introduce a novel image transformation called *Local Intensity Order Transformation (LIOT)*, which is a representation dedicated to curvilinear structure and invariant to the increasing change of image contrast. Specifically, we compare the intensity order of a given pixel to the values of its nearby neighbors (*e.g.* along with the four horizontal and vertical directions within a certain distance). In this way, we transform the original image into a new four-channel one, which does not depend on the absolute intensity value but the relative intensity order to capture the inherent characteristic of the curvilinear structure. An example of segmentation on LIOT is illustrated in Fig. 1. LIOT is robust to contrast changes, yielding very similar segmentation result for the original image and the one with contrast change.

The main property of the proposed LIOT is the robustness to curvilinear characteristic and contrast change. LIOT simply changes the input and can be plugged into any deep learning-based method. Since the main goal of this paper is not to develop a powerful pipeline that outperforms all other methods in segmenting curvilinear objects, we simply apply LIOT to the recent method IterNet [23] and conduct cross-dataset experiments on four datasets: DRIVE [45], STARE [46], CHASEDB1 [47], and CrackTree [6]. These cross-dataset experiments demonstrate that the proposed contrast-insensitive LIOT is simple yet effective to help deep learning-based methods to capture the characteristic of the curvilinear structure, and improves the generalizability of existing methods.

The main contribution of this paper is as follows: 1) We propose a novel image transform method that captures the inherent characteristic of curvilinear structures and is invariant to the increasing change of image contrast; 2) The spirit of using LIOT as input to the deep learning-based method enables the model to express a better generalization and to be more robust to non-targeted universal adversarial perturbations.

The rest of this paper is organized as follows. We shortly review some related works in Section II. The proposed method is then detailed in Section III, followed by extensive experimental results in Section IV. Finally, we conclude and give some perspectives in Section V.

II. Related work

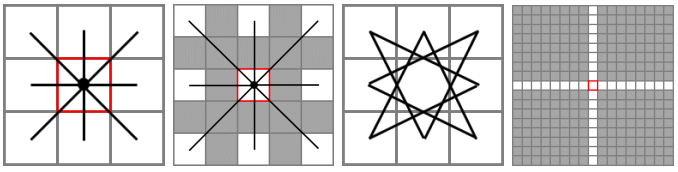
Curvilinear structure segmentation has been widely exploited recently. We first review some representative classical and deep learning-based methods for curvilinear object segmentation in Section II-A. A review of some related works on leveraging intensity order-based information, in particular variants of census transform and intensity order-based feature description, are discussed in Section II-B. The comparison of the proposed LIOT with some related works is depicted in Section II-C.

A. Curvilinear object segmentation

Before the era of deep learning, curvilinear object segmentation pipelines [6], [8]–[22] are usually based on different techniques ranging from hand-crafted filters to machine learning approaches. Most recent methods shift to deep neural networks to directly learn effective features from the training data. Some of these methods are categorized and detailed in the following.

Classical methods: Classical curvilinear object segmentation methods mainly rely on extracting engineered features. For instance, Koller *et al.* [8] introduce a method based on a non-linear combination of linear filters, with an edge-detection approach. Carlotto *et al.* [9] combine directional filter to enhance low-contrast curvilinear. In [10], the authors leverage pixel intensity and 2D Gabor wavelet transform for retinal blood vessel segmentation. In [11], the authors develop the line operator to construct a feature vector for supervised classification using the support vector machine. AI-Diri *et al.* [12] segment vessels in retinal images with a related active contour model. Marìn *et al.* [13] utilize a 7-D vector for pixel representation. Xiao *et al.* [15] propose to replace the low-level Gaussian kernel with a bi-Gaussian to detect curvilinear structures. The orientation analysis of gradient vector fields has been proposed for retinal blood vessel segmentation in [47]. Obara *et al.* [14] propose a contrast-independent approach to identify curvilinear structures based on oriented phase congruency. Krylov *et al.* [16] develop a stochastic approach for line segment extraction. Vicas *et al.* [18] use a structure tensor to extract meaningful low-level information for curvilinear structures. In [19], the authors rely on graph-based representations for reconstructing curvilinear networks. Merveille *et al.* [20] and [48] segment curvilinear structures by ranking the orientation responses of operators. Strisciuglio *et al.* [21] propose a novel operator for the delineation of curvilinear structures in images. Merveille *et al.* [22] propose a mixed gradient operator for segmentation task.

Deep learning-based methods: Deep learning-based methods learn features from the training data and significantly improve the performance of curvilinear object segmentation. For example, the neural network is used for retinal blood vessel segmentation in [24], which does not depend on the carefully engineered feature. Maninis *et al.* [25] propose a unified framework of retinal image analysis that jointly performs both retinal blood vessel and optic disc segmentation. In [27], Yan *et al.* propose a joint segment-level and pixel-wise loss for retinal blood vessel segmentation. Luan *et al.* [28] develop the Gabor convolution networks to reinforce the robustness of learned features against orientation and scale changes. Guo *et al.* [32] present a multi-scale deeply supervised network with short connections, achieving good cross-dataset results. In [33], Cherukuri *et al.* propose a deep model with regularization under geometric priors. Mosinska *et al.* [29] propose a multi-task learning of curvilinear structure segmentation and path classification. Wang *et al.* [30] propose a context-aware spatio-recurrent network for segmenting curvilinear structure. Lei *et al.* [31], [35] propose a channel and spatial attention network based on U-Net, effectively extracting curvilinear structures from three biomedical imaging modalities. Ma *et al.* [49] introduce a novel coarse-to-fine vessel segmentation network with the ability to separately detect thick and thin vessels. Dey [34] introduces a subpixel convolution and a feature fusion technique for retinal blood vessel segmentation. The distance field of tubular objects [50], [51] has been



(a) CT [57] (b) SP-CT [58] (c) SCT [59] (d) LIOT

Fig. 2: Comparison of LIOT with variants of census transform, which compare the intensity order between white pixels and the red one in the center for CT and Sparse CT (SP-CT).

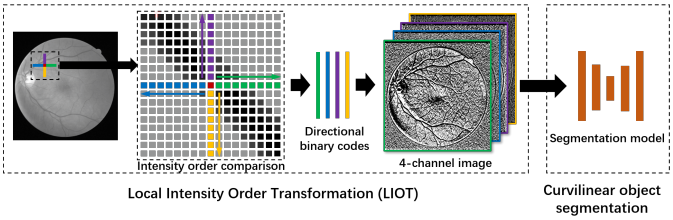


Fig. 3: Overview of the proposed LIOT for curvilinear object segmentation. We transform an input image to a 4-channel intensity-order-based image, which is then fed into a segmentation network.

proposed to improve the segmentation performance. Some recent methods [26], [52]–[55] aim to improve the topological mistakes by proposing topology-aware loss for curvilinear object segmentation. The method in [56] focuses on developing powerful network architecture for curvilinear object segmentation by leveraging boundary as a geometric constraint to refine feature.

B. Census transform and intensity order-based feature description

Census Transform (CT) [57] is a non-parametric local transform, which is originally defined as an ordered set of comparisons of pixel intensity in local 3×3 neighborhood. Froba *et al.* [60] propose the MeanCT that leverages the mean intensity of the neighborhood to increase the robustness. Ambrosch *et al.* [61] propose a SparseCT which uses a subset of pixels within the region of a larger neighborhood. Chang *et al.* [58] also propose a sparse CT Mini-census. In [62], Fife *et al.* propose generalized census transform (GCT) to define a family of masks in a 5×5 neighborhood with different levels of sparsity. Similarly, the center symmetric CT (CSCT) [63] compares a pair of pixels within the census window. The star census transform (SCT) [59] extends GCT by defining masks of symmetrical sequences of connected edges, of equal length, forming star-shaped scan-patterns around the center. Ahlberg *et al.* [64] propose a genetic algorithm to find a new and powerful Census Transform method. Yu *et al.* [65] adopt the census transform to calculate the illumination characteristics for tree reconstruction optimization. Lai *et al.* [66] also use census transform in medical imaging. Ramírez *et al.* [67] apply the census transform operator for audio-visual emotion recognition, owing to its robustness to monotonic changes.

Apart from the variants of census transform, some works [68]–[73] also share the similar idea with the proposed LIOT on leveraging the local intensity order information, but with different purposes (local feature description [68]–[70] and image classification [71]–[73]).

C. Comparison with related works

LIOT Versus Classical methods: Classical methods rely on engineered features to extract curvilinear structure. Therefore, these methods usually require careful parameter tuning and are hard to generate good performance in a wide variety of complex curvilinear object segmentation. The proposed LIOT can leverage a deep convolutional neural network (CNN) to learn more effective features. LIOT can be plugged into any CNN architecture for curvilinear object segmentation.

LIOT Versus Deep learning-based methods: Most deep learning-based methods for curvilinear object segmentation are mainly inspired by recent CNN-based object/image segmentation methods. They usually focus on improving the in-dataset segmentation accuracy by improving feature learning, and ignore the limited cross-dataset performance due to the appearance gap between different datasets. This hinders the transferability to other similar curvilinear object segmentation. The proposed LIOT aims to capture the inherent characteristic of curvilinear objects, and thus to improve the generalization performance in various curvilinear object segmentation for deep learning-based methods.

LIOT Versus Census transform and intensity order-based feature description: The proposed LIOT is in spirit similar to the census transform, which is a non-parametric transform that depends on the relative intensity ordering instead of absolute intensity values. This makes them invariant to increasing contrast changes. As shown in Fig. 2, compared with variants of census transform, the proposed LIOT is specifically designed for the curvilinear object that captures the curvilinear structure from four directions. Compared with the simple intensity order comparison between the center pixel and its surrounding neighborhood pixels for census transform, the horizontal and vertical long-range pixels capture better the inherent property of curvilinear objects having usually thin and long structures. The LIOT also differs with the intensity order-based feature description methods in how to define the pairs of pixels for intensity order comparison. Besides, to the best of our knowledge, we are the first to combine intensity order information and deep-learning for robust curvilinear structure segmentation, which enables the model to express a better generalization and to be more robust to non-targeted universal adversarial perturbations.

III. Proposed Method

A. Overview

Curvilinear objects have entirely different color distributions and backgrounds for different application scenarios. Such an appearance gap makes it difficult for the algorithm to accurately segment different curvilinear objects in different images. Thus, a general method to improve the generalizability for curvilinear segmentation algorithms is required. For that,

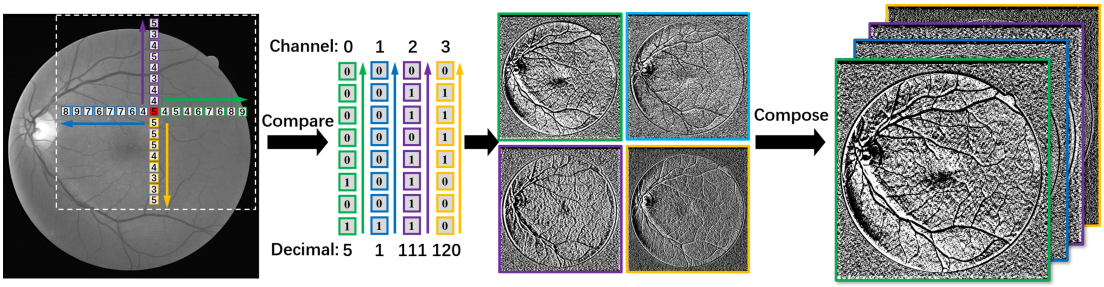


Fig. 4: Illustration of LIOT. For each pixel p, we compare its intensity with the value of every group of 8 neighboring pixels lying perpendicular with p, resulting in a 4-channel image with values in each channel ranging from 0 to 255 (8 bits).

we introduce a novel image transformation LIOT based on the relative intensity order between each pixel and its perpendicular neighbors. Such a transformation does not depend on the absolute value of each pixel, and is thus more robust to contrast changes. The overall pipeline of the proposed method is depicted in Fig. 3. We first convert the given image into a gray-scale one. Then we rely on the local intensity order to compute four directional binary codes, forming a 4-channel image that captures the curvilinear structure characteristic. We then feed this contrast invariant 4-channel image into a segmentation network.

B. Local intensity order transformation

Though different curvilinear object images may have various contrast and intensity distribution, the curvilinear structure is always darker (without loss of generality, we can invert the image when the curvilinear structure is brighter) than the context. Based on this property, we propose a novel and robust representation called LIOT that converts an input gray-scale image into a 4-channel intensity-order-based image. More precisely, as illustrated in Fig. 4, for each pixel p in the domain Ω of an image f, we compare the value of $f(p)$ with each group of 8 neighboring pixels $\{n_s^i \,|\, i = 1, \ldots, 8\}$ (with Euclidean distance to p ranging from 1 to 8) lying perpendicular to p, where $s \in \{l, r, t, b\}$ denotes left, right, top, and bottom side of p, respectively. This results in four 2D 8-bit images, where the i-th bit of each converted 2D image f'_s on p corresponds to the binary code given by the intensity order between $f(p)$ and $f(n_s^i)$. These four 2D images of directional binary codes are concatenated together, composing a 4-channel intensity-order-based image $F = [f'_l, f'_r, f'_t, f'_b]$. Formally, for each direction side $s \in \{l, r, t, b\}$, we compute the corresponding image f'_s of directional binary codes on each pixel p as follows:

$$f'_s(p) = \sum_{i=1}^{8} [f(p) > f(n_s^i)] \times 2^{i-1}, \tag{1}$$

where $[f(p) > f(n_s^i)]$ is 1 if the value $f(p)$ is larger than $f(n_s^i)$, otherwise 0.

As depicted in Eq. (1), the proposed transformation is irrespective of absolute intensity values, and only depends on the relative intensity order. LIOT captures the "darker than" property of curvilinear structure and is invariant to increasing change of contrast. This makes LIOT more robust to appearance differences for images from different applications. Therefore, LIOT can improve the generalizability of curvilinear object segmentation methods.

C. Network architecture

Since the major goal is to improve the generalizability of existing CNN-based methods, we do not aim to develop a powerful network that outperforms other methods. Instead, we simply apply LIOT with the recent IterNet [23] by changing its input channel to the 4-channel image given by LIOT. Iternet is a U-Net like encoder-decoder model that adopts U-Net as the base module and combines three mini U-Nets. Each mini U-Net uses feature maps from its precedent module with fewer parameters than U-Net. IterNet also adds skip-connections from the base U-Net to all mini U-Nets and the connections among the mini U-Nets. We simply change the first convolution layer channel from 3 to 4-channel for adapting LIOT inputs, and keep the following layers the same as the original IterNet [23].

D. Training objective

We leverage the network depicted in Section III-C to segment the curvilinear object. The loss is the same as that in IterNet [23]. The network parameters are optimized with the sigmoid cross-entropy loss, defined as:

$$L = \sum_i \big(- y_i log(p_i) - (1 - y_i) log(1 - p_i) \big), \tag{2}$$

where y_i represents the binary indicator (0 or 1) whether the pixel i belongs to the ground-truth curvilinear object, and p_i is the predicted probability that the pixel i is a foreground pixel.

During the testing phase, we convert the original gray-scale image to a 4-channel LIOT image, and feed LIOT into the segmentation network to compute the final probability map.

IV. Experiments

We evaluate the proposed method on four widely adopted datasets: DRIVE [45], STARE [46], CHASEDB1 [47], and CrackTree [6]. A short description of these datasets and adopted evaluation protocol are given in Section IV-A. Some implementation details are then depicted in Section IV-B. To demonstrate the generalizability of the proposed LIOT, we conduct four types of evaluations: 1) In-dataset evaluation

in Section IV-C to demonstrate that the proposed LIOT does not significantly decrease the performance on original images; 2) In-dataset evaluation on images with universal adversarial perturbations (see Section IV-D) to show that LIOT is more robust to universal adversarial attacks; 3) Cross retinal dataset evaluation in Section IV-E, showing that LIOT generalizes better to images with small domain gap; 4) Evaluation on cross-dataset between Retinal and CrackTree dataset (see Section IV-F), proving that LIOT is more robust to large domain changes. In Section IV-G, we also show some results of applying LIOT trained on the retinal dataset to images with very different types of curvilinear structures, further demonstrating the generalizability of LIOT.

A. Datasets and evaluation protocol

DRIVE [45]: The DRIVE dataset contains the curvilinear-shaped vessel. This dataset consists of 40 565×584 color retinal images, which are split into 20 training images and 20 test images.

STARE [46]: The STARE dataset consists of 20 700×605 color retina images divided into 10 training and 10 test images.

CHASEDB1 [47]: The CHASEDB1 dataset is composed of 28 999×960 color retinal images, which are split into 20 training images and 8 test images.

CrackTree [6]: The CrackTree dataset contains 206 800×600 pavement images with different kinds of cracks having curvilinear structure. The whole dataset is split into 160 training and 46 test images. Following [26], we dilate the annotated centerlines by 4 pixels to form the ground-truth segmentation. As shown in Fig. 10, the multiple shadows and cluttered background make the segmentation a challenging task.

Images with curvilinear objects may have different image resolutions, and thus have varied thickness (in terms of pixels) of curvilinear structure. To cope with such a scale gap, for the retinal images, we first resize each image to a similar scale based on the size of the field of view (FOV) and image size. Specifically, we keep the image size of DRIVE unchanged, and resize the images in STARE, CHASEDB1 accordingly. Precisely, we resize STARE images from 700×605 to 554×479, and CHASEDB1 images from 999×960 to 584×561. The images in the CrackTree dataset are resized to 512×512. These size settings are used for both baseline methods and the proposed method. For retinal images, since LIOT requires a total order between pixel values, we thus convert the resized color image to a gray-scale one by selecting its green channel.

Evaluation protocol: Following the classical evaluation protocol for curvilinear object segmentation, we first compute the true positives (TPs), false positives (FPs), false negatives (FNs), and true negatives (TNs) for the segmentation result. Then, the classical accuracy (Acc), sensitivity (Se), specificity (Sp), area under the receiver operating characteristics curve (AUC), and F1-score are used to evaluate the performance. Besides, we also adopt the connectivity [74]: $1-\min\left(1, |\#_C(S_G)-\#_C(S)|/\#(S_G)\right)$, where $\#_C(S_G)$ and $\#_C(S)$ denotes the number of connected components in the ground-truth and segmented result, respectively. $\#(S_G)$ stands for the number of pixels in the ground-truth segmentation. This connectivity metric reflects the curvilinear structural continuity and is useful for branching analysis, further assessing the effectiveness of the proposed LIOT. We compute the metrics in the field of view area for retinal datasets: DRIVE, STARE , and CHASEDB1. For the CrackTree dataset, the metrics are computed on the whole image. For all methods on all involved datasets, we report the results by using the optimal segmentation threshold based on the F1-score.

B. Implementation details

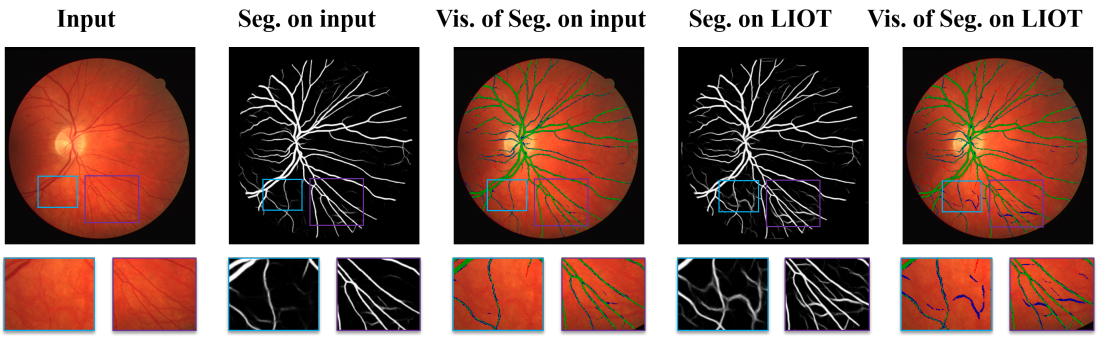
We adopt classical data augmentation strategy to increase the training data and avoid over-fitting. Specifically, we randomly rotate images from -180 to 180 degrees, and shear from -0.1 to 0.1. We also flip images with the horizontal and vertical direction. Then, images are randomly shifted from -0.1 to 0.1, and randomly zoomed from 0.8 to 1.2. Finally, we randomly crop the augmented images to 128×128.

Since the major goal is not to develop a powerful network that outperforms other methods, we simply apply LIOT with the recent IterNet [23]. Following [23], we employ a batch size of 32 to train the network using cross-entropy loss for 1000 epochs. To further demonstrate the usefulness of the proposed LIOT, we also conduct experiments by using the topological loss (Topo) [52]. We adopt Adam [75] with a learning rate of 0.001 to optimize the network. For Topo [52] method, we set different optimal weights for the topological loss term on different datasets: 0.005 for DRIVE, 0.001 for CHASEDB1 and STARE, and 0.01 for CrackTree. For census [57] method, we choose the window size 3×3 as the original census definition and do not set any other parameters because the census is a non-parametric local transformation. During inference, overlapping image patches are extracted with a stride equal to 8.

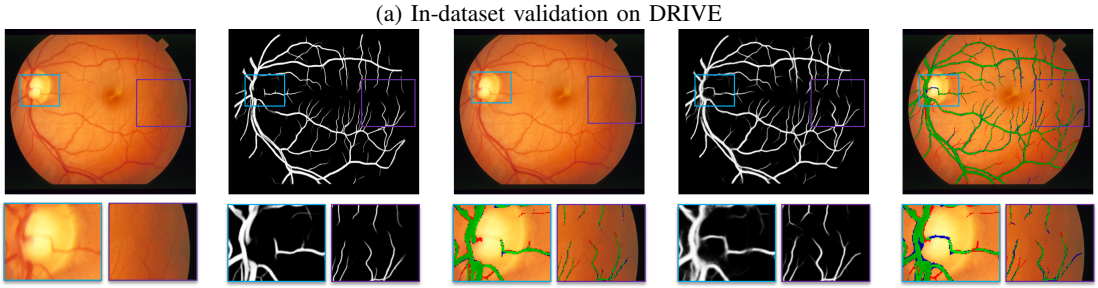
C. In-Dataset evaluation

The proposed LIOT converts the original image into a relative order-based representation, which may lose some information on the image content. From this point of view, though the proposed LIOT mainly focus on improving the generalization performance for deep learning-based methods, we first conduct experiments on in-datasets: DRIVE, STARE, CHASEDB1, and CrackTree to show that the proposed LIOT does not significantly decrease the performance for the ability of model itself.

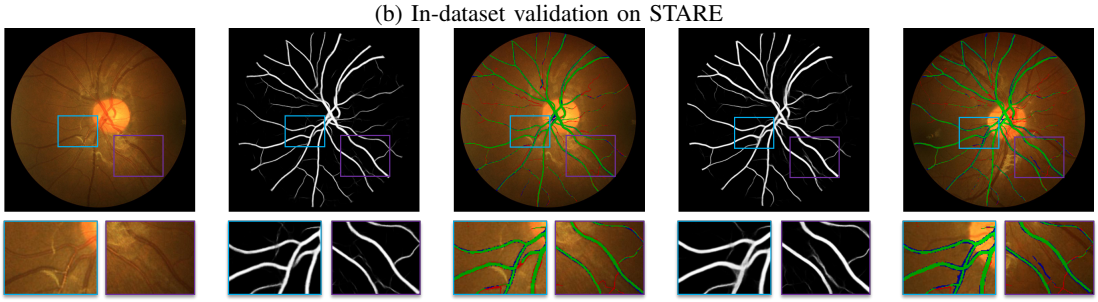
Some qualitative results are illustrated in Fig. 5. LIOT performs similarly with the baseline model, but usually has more false positives, which can find evidence in the original image, but are ignored in the ground-truth annotation. This is because that LIOT mainly focuses on the inherent characteristic (*e.g.*, darker than the context) of curvilinear structure, and thus captures more objects having curvilinear structure. This may also explain why LIOT achieves slightly worse results than the baseline model on retinal images in terms of quantitative evaluation depicted in Table I. On the retinal datasets, LIOT is slightly worse than the baseline model, but performs much



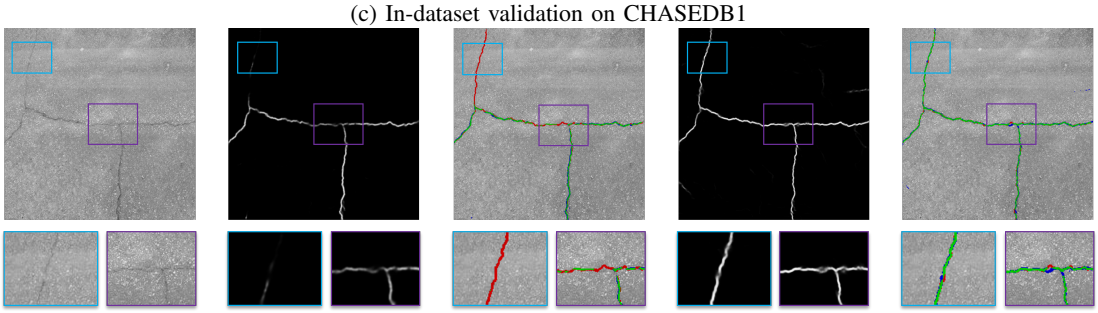
(a) In-dataset validation on DRIVE



(b) In-dataset validation on STARE



(c) In-dataset validation on CHASEDB1



(d) In-dataset validation on Cracktree

Fig. 5: Some segmentation results under in-dataset validation. Green pixels: TPs; Red pixels: FNs; Blue pixels: FPs. Some FPs achieved by LIOT can find evidence in the original image, which might be TPs ignored in the manual annotation. Best viewed by zooming in the electronic version.

better than the variant of classical census transform. On the CrackTree, LIOT is very competitive with the baseline model. On average, LIOT is still comparable with the baseline model.

D. In-dataset evaluation on images with perturbation

It is well known that deep learning-based models are highly vulnerable to adversarial perturbations, which may pose security problems for deep learning-based curvilinear object segmentation methods. In this section, we evaluate the robustness of the proposed LIOT against such adversarial perturbations. For that, we adopt the non-targeted universal perturbations by GAP [76] to generate adversarial perturbations. As shown in Fig. 6, these perturbed images resemble the original images. We apply LIOT on the perturbed images and feed it to the trained model. Some qualitative results are given in Fig. 6. Though the perturbed images mislead both the trained baseline model and the model trained on LIOT, LIOT is more robust.

The quantitative evaluation is depicted in Table II. The proposed LIOT outperforms the baseline model by a large margin (up to 47.8%) in terms of all metrics. The proposed LIOT can effectively retain the inherent characteristic of curvi-

TABLE I: Quantitative in-dataset evaluation of LIOT compared with the baseline model and Census transform.

In-Dataset	Methods	Se	Sp	Acc	AUC	F1	Connectivity
CrackTree ⇒ CrackTree	Baseline [23]	**0.813**	0.995	0.992	**0.994**	0.754	0.851
	Census [57]	0.493	0.996	0.989	0.898	0.566	0.563
	LIOT	0.780	**0.997**	**0.994**	0.990	**0.790**	**0.916**
	Topo [52]	**0.848**	0.995	0.993	**0.996**	0.769	0.870
	Topo+LIOT	0.802	**0.997**	**0.994**	0.988	**0.801**	**0.917**
DRIVE ⇒ DRIVE	Baseline [23]	**0.828**	0.973	**0.954**	**0.978**	**0.821**	**0.875**
	Census [57]	0.783	0.972	0.948	0.962	0.794	0.831
	LIOT	0.811	**0.974**	0.953	0.976	0.814	0.763
	Topo [52]	**0.828**	**0.973**	**0.954**	**0.980**	**0.822**	**0.878**
	Topo+LIOT	0.819	0.972	0.953	0.977	0.815	0.777
STARE ⇒ STARE	Baseline [23]	0.812	**0.982**	**0.964**	**0.984**	**0.826**	**0.886**
	Census [57]	0.763	0.975	0.953	0.964	0.771	0.641
	LIOT	**0.816**	0.977	0.960	0.983	0.810	0.803
	Topo [52]	**0.836**	**0.979**	**0.964**	**0.985**	**0.829**	**0.887**
	Topo+LIOT	0.817	0.976	0.959	0.982	0.807	0.803
CHASEDB1 ⇒ CHASEDB1	Baseline [23]	**0.833**	**0.978**	**0.965**	**0.986**	**0.811**	0.715
	Census [57]	0.738	0.979	0.957	0.963	0.756	0.682
	LIOT	0.816	0.976	0.961	0.983	0.792	**0.803**
	Topo [52]	**0.838**	**0.978**	**0.965**	**0.986**	**0.812**	0.735
	Topo+LIOT	0.819	0.977	0.963	0.983	0.799	**0.829**
Average	Baseline [23]	**0.822**	**0.982**	**0.969**	**0.986**	**0.803**	**0.832**
	Census [57]	0.694	0.980	0.962	0.947	0.722	0.679
	LIOT	0.807	0.981	0.967	0.983	0.801	0.821
	Topo [52]	**0.838**	**0.981**	**0.969**	**0.987**	**0.808**	**0.843**
	Topo+LIOT	0.814	**0.981**	0.967	0.983	0.806	0.832

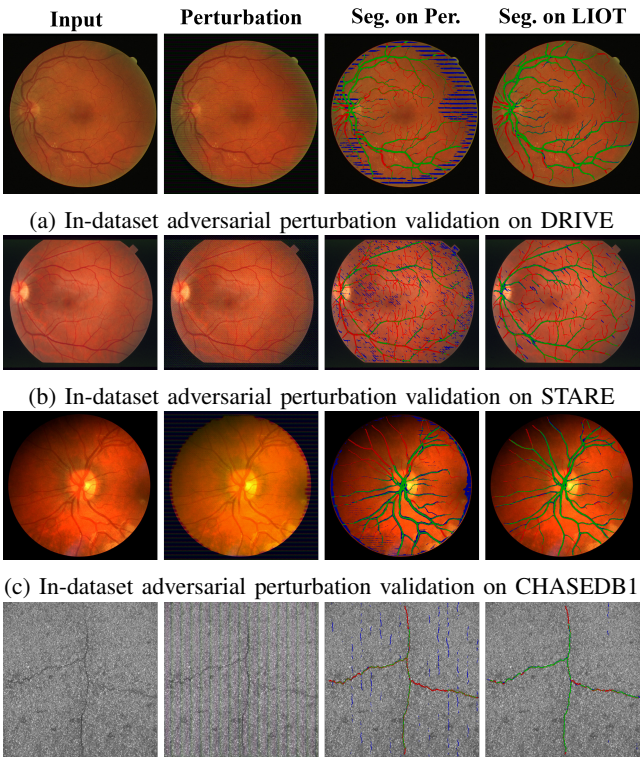
TABLE II: Quantitative in-dataset evaluation on images with non-targeted universal adversarial perturbation [76] of LIOT compared with the baseline model.

In-Dataset	Methods	Se	Sp	Acc	AUC	F1	Connectivity
CrackTree ⇒ CrackTree	Baseline [23]	0.453	0.993	0.985	0.963	0.474	0.342
	Census [57]	0.564	**0.996**	0.990	0.952	0.613	0.677
	LIOT	**0.668**	**0.996**	**0.992**	**0.976**	**0.699**	**0.788**
	Topo [52]	0.439	**0.995**	0.987	0.950	0.496	0.270
	Topo+LIOT	**0.625**	**0.995**	**0.990**	**0.971**	**0.641**	**0.748**
DRIVE ⇒ DRIVE	Baseline [23]	0.729	0.842	0.828	0.871	0.518	0.611
	Census [57]	0.591	0.778	0.755	0.762	0.380	0.621
	LIOT	**0.752**	**0.967**	**0.940**	**0.952**	**0.761**	**0.714**
	Topo [52]	0.715	0.863	0.844	0.876	0.538	0.574
	Topo+LIOT	**0.738**	**0.958**	**0.930**	**0.942**	**0.728**	**0.705**
STARE ⇒ STARE	Baseline [23]	0.429	0.880	0.833	0.754	0.389	0.373
	Census [57]	0.420	**0.976**	**0.918**	**0.842**	0.518	0.283
	LIOT	**0.461**	0.965	0.912	0.827	**0.523**	**0.396**
	Topo [52]	0.363	0.909	0.852	0.740	0.338	0.318
	Topo+LIOT	**0.447**	**0.966**	**0.912**	**0.813**	**0.514**	**0.345**
CHASEDB1 ⇒ CHASEDB1	Baseline [23]	0.633	0.899	0.875	0.867	0.477	0.322
	Census [57]	0.781	**0.977**	**0.960**	0.973	0.778	0.763
	LIOT	**0.788**	**0.977**	**0.960**	**0.978**	**0.779**	**0.776**
	Topo [52]	0.682	0.876	0.859	0.873	0.466	0.428
	Topo+LIOT	**0.778**	**0.978**	**0.960**	**0.978**	**0.779**	**0.776**
Average	Baseline [23]	0.561	0.904	0.880	0.864	0.455	0.412
	Census [57]	0.589	0.932	0.906	0.882	0.572	0.586
	LIOT	**0.667**	**0.976**	**0.951**	**0.933**	**0.691**	**0.668**
	Topo [52]	0.550	0.911	0.885	0.860	0.460	0.398
	Topo+LIOT	**0.647**	**0.974**	**0.948**	**0.926**	**0.665**	**0.644**

linear structure from the images with non-targeted universal adversarial perturbations, demonstrating the robustness of the proposed LIOT.

E. Cross retinal datasets evaluation

We now evaluate the proposed LIOT by conducting cross-retinal dataset validation between the DRIVE, STARE, and CHASEDB1 dataset.



(a) In-dataset adversarial perturbation validation on DRIVE

(b) In-dataset adversarial perturbation validation on STARE

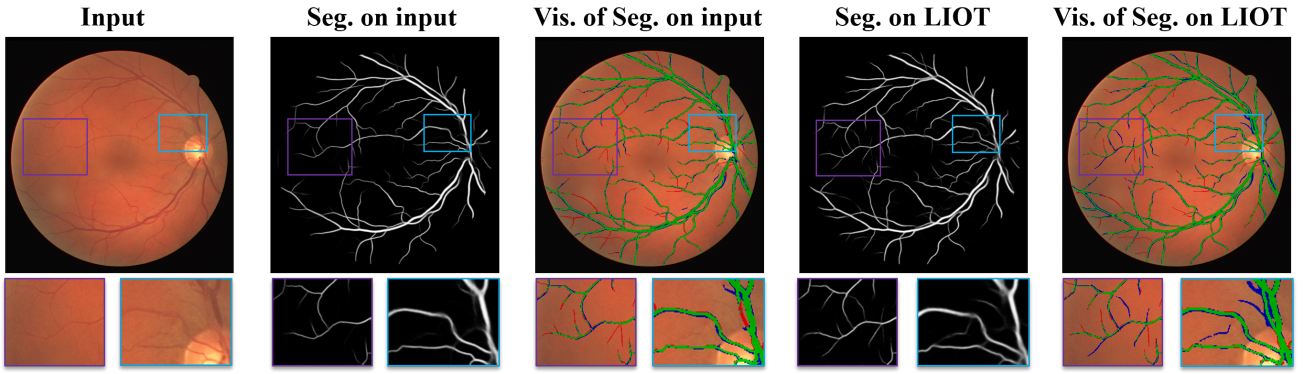
(c) In-dataset adversarial perturbation validation on CHASEDB1

(d) In-dataset adversarial perturbation validation on Cracktree

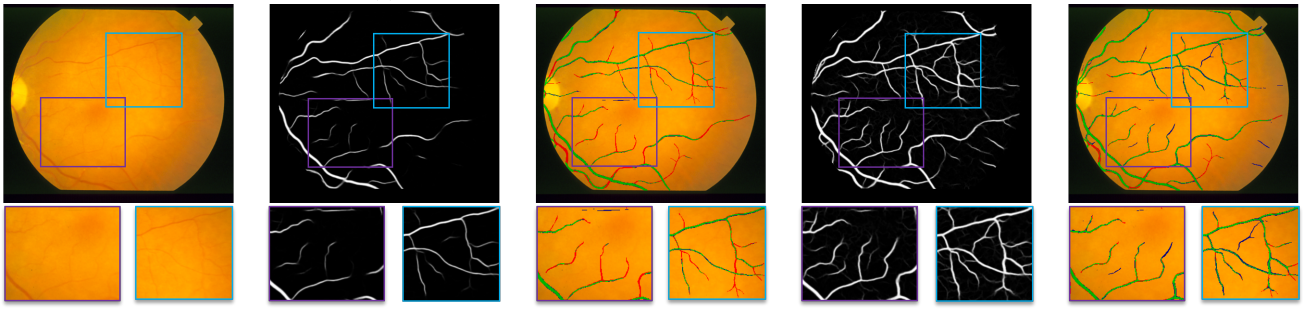
Fig. 6: Some segmentation results under in-dataset validation with non-targeted universal adversarial perturbation [76]. Green pixels: TPs; Red pixels: FNs; Blue pixels: FPs. Best viewed by zooming in the electronic version.

Some qualitative illustrations are shown in Fig. 7. As depicted in this figure, LIOT is able to correctly segment most retinal blood vessels, especially those of low contrast. The direct segmentation on the original image fails to retrieve those thin and low contrast retinal blood vessels. For instance, as depicted in Fig. 7(b), the model trained on DRIVE, using LIOT can accurately segment most retinal blood vessels, including the segments (blue pixels in the cropped region on the right side) of very low contrast and ignored even by the manual annotation. In fact, though these segments have very low contrast, we can still find cues in the original image and they are indeed retinal blood vessels. This demonstrates that LIOT is more robust to contrast changes. Indeed, as depicted in Fig. 1, the baseline method poorly segments the retinal blood vessels for the contrast changed image, on which LIOT still performs similarly with the original image. Thus, LIOT can improve the generalizability of some deep learning-based methods in segmenting retinal blood vessels.

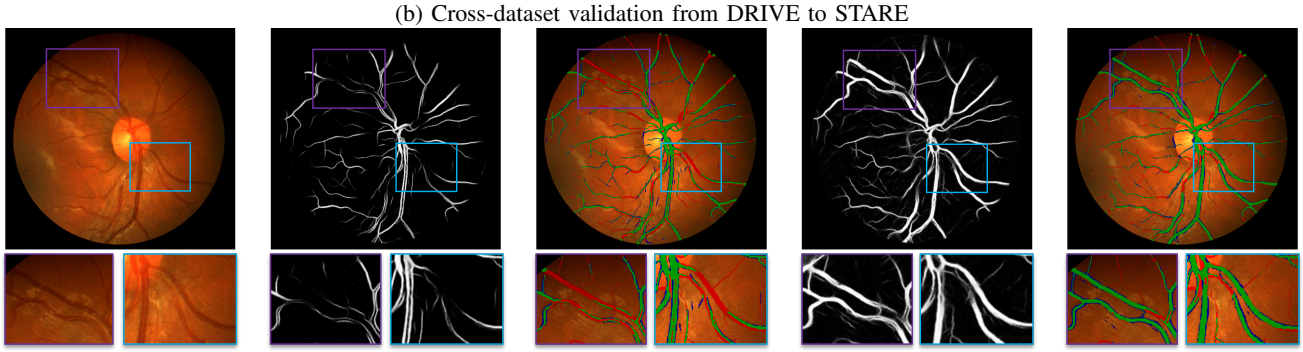
The quantitative cross retinal dataset evaluation of LIOT is depicted in Table III. We randomly divide three times DRIVE, STARE, and CHASEDB datasets into train/test images, and we train the corresponding model three times. The mean and standard errors are reported. Compared with the baseline method of IterNet [23] which directly operates on the original image, LIOT outperforms or is on par with the baseline model in terms of classical pixel-level metrics: Se, Acc, AUC, and



(a) Cross-dataset validation from CHASEDB1 to DRIVE



(b) Cross-dataset validation from DRIVE to STARE



(c) Cross-dataset validation from STARE to CHASEDB1

Fig. 7: Some segmentation results under cross-dataset validation. Green pixels: TPs; Red pixels: FNs; Blue pixels: FPs. Some FPs achieved by LIOT can find evidence in the original image, which might be TPs ignored in the manual annotation.

F1-score for all set of cross-dataset evaluation. LIOT has better mean values and smaller standard errors in most cases. LIOT is also competitive with other state-of-the-art methods [24], [27], [32], [33], [77] dedicated for cross-retinal dataset evaluation. On average, LIOT performs much better than the baseline model and census transform based on all metrics except Sp. In particular, LIOT features a high sensitivity regime, which is important in clinical usage. Some "false positives" that are retinal blood vessels of low contrast and ignored by the manual annotation (described above) may degrade a bit the other classical metrics, such as the Sp. Therefore, LIOT may even perform slightly better in practice.

It is noteworthy that LIOT not only improves the pixel-level metrics, but also boosts the segmentation performance on the vessel network level. Specifically, as shown in Table III, LIOT consistently outperforms the baseline models in terms of connectivity metric. This implies that LIOT improves the integrity of the segmentation effect because connectivity assesses the fragmentation degree between the segmentation result and the ground truth. Besides, the connectivity is also an important clue for clinicians to calculate the complexity and density of branching of the retinal vascular tree to measure patient's condition [78]. Therefore, LIOT is also potentially useful for clinical disease diagnosis.

Though the goal is not to demonstrate that simply changing the input image by the proposed LIOT on a baseline method can outperform all other methods, as depicted in Table III, the experimental results show that the proposed method is competitive with all other methods [24], [27], [32], [33], [77] under cross-dataset validation, demonstrating the potential of the proposed LIOT in segmenting retinal blood vessels for a wide daily clinical usage.

F. Cross CrackTree and retinal datasets evaluation

In addition to cross retinal dataset (with small domain gaps) validation depicted in Section IV-E, We also evaluate the

TABLE III: Quantitative comparison of LIOT and some other methods under cross-dataset evaluation on retinal images. The results obtained from re-implementation are marked with †, and ∗ indicates the reported results from the original papers.

Cross-dataset	Methods	Se	Sp	Acc	AUC	F1	Connectivity
STARE ⇒ DRIVE	[24]∗	0.727	0.981	0.949	0.968	–	–
	[77]∗	0.671	**0.992**	0.951	0.975	–	–
	[27]∗	0.729	0.982	0.949	0.960	–	–
	[33]∗	**0.772**	0.983	**0.956**	**0.977**	–	–
	[32]∗	0.745	0.978	0.950	0.971	–	–
	Baseline [23]†	0.781±**0.003**	**0.967±0.004**	**0.943±0.003**	0.959±0.003	**0.779±0.013**	0.779±**0.004**
	Census [57]†	0.724±0.013	0.955±0.006	0.926±0.007	0.932±0.009	0.716±0.022	0.721±0.020
	LIOT†	**0.791**±0.004	0.963±0.006	0.941±0.004	**0.962±0.002**	0.775±0.015	**0.788**±0.009
	Topo [52]†	0.770±**0.004**	**0.969±0.001**	**0.944±0.001**	**0.954±0.002**	**0.776**±0.006	0.772±**0.004**
	Topo+LIOT†	**0.790**±0.006	0.957±0.002	0.936±**0.001**	0.950±0.007	0.759±**0.001**	**0.792**±0.006
CHASEDB1 ⇒ DRIVE	[24]∗	**0.731**	**0.981**	**0.948**	**0.961**	–	–
	[32]∗	0.696	0.970	0.938	0.952	–	–
	Baseline [23]†	0.763±0.030	**0.971±0.001**	**0.945**±0.005	0.956±0.011	**0.779**±0.023	0.762±0.025
	Census [57]†	0.719±0.024	0.966±0.007	0.935±**0.003**	0.943±0.005	0.738±**0.007**	0.724±0.024
	LIOT†	**0.799±0.007**	0.963±0.005	0.942±0.005	**0.966±0.003**	**0.779**±0.019	**0.798±0.003**
	Topo [52]†	0.750±0.036	**0.968**±0.002	0.941±0.006	0.933±0.026	0.762±0.027	0.753±0.036
	Topo+LIOT†	**0.781±0.010**	0.965±**0.001**	**0.942±0.001**	**0.960±0.005**	**0.773±0.006**	**0.784±0.010**
DRIVE ⇒ STARE	[24]∗	0.703	0.983	0.955	0.967	–	–
	[77]∗	**0.845**	0.973	0.960	**0.985**	–	–
	[27]∗	0.721	0.984	0.957	0.971	–	–
	[33]∗	0.778	**0.986**	**0.971**	0.982	–	–
	[32]∗	0.719	0.982	0.955	0.969	–	–
	Baseline [23]†	0.760±0.038	**0.978**±0.002	0.956±0.003	0.962±0.010	0.782±0.018	0.731±0.052
	Census [57]†	0.766±0.011	0.970±0.005	0.949±0.005	0.966±0.003	0.759±0.020	0.720±0.058
	LIOT†	**0.812±0.007**	0.975±**0.001**	**0.958±0.001**	**0.981±0.002**	**0.800±0.006**	**0.806±0.010**
	Topo [52]†	0.713±0.031	**0.977**±0.002	0.949±0.004	0.915±0.034	0.745±0.023	0.673±0.046
	Topo+LIOT†	**0.799±0.005**	0.972±**0.001**	**0.955±0.001**	**0.977±0.003**	**0.785±0.006**	**0.796±0.006**
CHASEDB1 ⇒ STARE	[24]∗	**0.694**	**0.983**	**0.954**	**0.962**	–	–
	[32]∗	0.680	0.981	0.950	0.952	–	–
	Baseline [23]†	0.722±0.020	**0.976**±0.002	0.950±**0.001**	0.942±0.014	0.750±0.011	0.704±0.018
	Census [57]†	0.764±0.009	0.968±0.002	0.947±**0.001**	0.964±0.004	0.751±**0.003**	0.727±0.061
	LIOT†	**0.807±0.008**	0.971±**0.001**	**0.954±0.001**	**0.978±0.001**	**0.784**±0.007	**0.804±0.004**
	Topo [52]†	0.727±0.026	**0.974±0.002**	0.948±0.004	0.917±0.030	0.743±0.022	0.710±0.026
	Topo+LIOT†	**0.796±0.016**	**0.974±0.002**	**0.956±0.001**	**0.976±0.003**	**0.788±0.020**	**0.795±0.015**
DRIVE ⇒ CHASEDB1	[24]∗	**0.712**	**0.979**	0.943	**0.963**	–	–
	[32]∗	0.698	0.972	**0.944**	0.957	–	–
	Baseline [23]†	0.702±0.036	0.963±0.005	0.939±0.007	0.937±0.013	0.682±0.044	0.697±0.041
	Census [57]†	0.682±**0.005**	0.958±0.003	0.933±0.003	0.939±0.004	0.650±**0.012**	0.634±0.071
	LIOT†	**0.784**±0.011	**0.968±0.002**	**0.951±0.002**	**0.973±0.002**	**0.749**±0.020	**0.780±0.005**
	Topo [52]†	0.682±0.077	0.957±0.005	0.932±0.010	0.912±0.042	0.649±0.065	0.676±0.084
	Topo+LIOT†	**0.769±0.011**	**0.962±0.002**	**0.944±0.002**	**0.966±0.001**	**0.720±0.002**	**0.772±0.012**
STARE ⇒ CHASEDB1	[24]∗	**0.724**	**0.977**	**0.942**	**0.955**	–	–
	[32]∗	0.673	0.971	0.941	0.951	–	–
	Baseline [23]†	0.653±0.101	0.956±**0.003**	0.926±0.012	0.908±0.048	0.635±0.075	0.654±0.097
	Census [57]†	0.595±0.049	0.951±0.004	0.918±0.006	0.891±0.023	0.568±0.036	0.552±0.094
	LIOT†	**0.764±0.008**	**0.959±0.003**	**0.939±0.004**	**0.961±0.003**	**0.717±0.016**	**0.760±0.007**
	Topo [52]†	0.630±0.106	**0.961**±0.004	0.929±0.010	0.872±0.078	0.633±0.073	0.631±0.103
	Topo+LIOT†	**0.759±0.011**	0.960±**0.003**	**0.940±0.002**	**0.954±0.006**	**0.716±0.003**	**0.761±0.013**
Average	Baseline [23]†	0.730±0.038	**0.969±0.003**	0.943±0.005	0.944±0.017	0.735±0.031	0.721±0.040
	Census [57]†	0.708±0.019	0.962±0.005	0.935±0.004	0.939±0.008	0.697±0.017	0.680±0.055
	LIOT†	**0.793±0.008**	0.966±**0.003**	**0.947±0.003**	**0.970±0.002**	**0.767±0.014**	**0.789±0.006**
	Topo [52]†	0.712±0.047	**0.968**±0.003	0.940±0.006	0.917±0.035	0.718±0.036	0.702±0.050
	Topo+LIOT†	**0.782±0.010**	0.965±**0.002**	**0.945±0.001**	**0.964±0.004**	**0.757±0.004**	**0.783±0.010**

proposed LIOT by conducting cross-dataset validation between the retinal datasets and CrackTree dataset. To the best of our knowledge, no previous research has investigated capturing the inherent characteristic of curvilinear structure for such large appearance gaps.

Cross-dataset validation from CrackTree to retinal datasets: As shown in Fig. 10, the CrackTree dataset has a great appearance gap with retinal images. We first apply the model trained on the CrackTree dataset to segment retinal blood vessels. Some qualitative retinal blood vessel segmentation results using the model trained on the CrackTree dataset are given in Fig. 8. The proposed LIOT roughly segment the curvilinear structure in retinal images. From this figure, we can observe that LIOT preserves the inherent characteristic of curvilinear structure in retinal images. LIOT is also able to segment the curvilinear structure from various contexts. On the contrary, the baseline method can not capture the curvilinear structure of such large appearance gaps. The direct segmentation on the original image fails to retrieve most retinal blood vessels. These are common results as most deep-learning

TABLE IV: Cross retinal and crack dataset evaluation of LIOT.

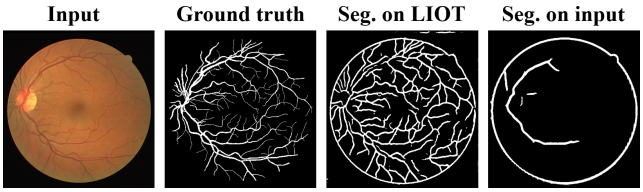
Cross-dataset	Methods	Se	Sp	Acc	AUC	F1	Connectivity
CrackTree ⇒ DRIVE	Baseline [23]	0.131	**0.965**	0.859	0.646	0.191	0.008
	Census [57]	0.532	0.896	0.850	0.812	0.474	0.487
	LIOT	**0.712**	0.895	**0.872**	**0.899**	**0.585**	**0.675**
	Topo [52]	0.013	**0.974**	0.852	0.504	0.023	0.014
	Topo+LIOT	**0.655**	0.902	**0.871**	**0.860**	**0.563**	**0.656**
CrackTree ⇒ STARE	Baseline [23]	0.130	**0.969**	0.882	0.626	0.185	0.006
	Census [57]	0.436	0.892	0.845	0.750	0.366	0.239
	LIOT	**0.640**	0.911	**0.883**	**0.898**	**0.530**	**0.505**
	Topo [52]	0.004	**0.991**	**0.889**	0.489	0.007	0.004
	Topo+LIOT	**0.539**	0.928	0.888	**0.814**	**0.497**	**0.541**
CrackTree ⇒ CHASEDB1	Baseline [23]	0.044	**0.981**	**0.897**	0.573	0.071	0.010
	Census [57]	0.369	0.922	0.872	0.753	0.342	0.334
	LIOT	**0.650**	0.890	0.877	**0.895**	**0.489**	**0.580**
	Topo [52]	0.003	**0.990**	**0.892**	0.486	0.005	0.003
	Topo+LIOT	**0.623**	0.908	0.879	**0.853**	**0.507**	**0.617**
DRIVE ⇒ CrackTree	Baseline [23]	0.037	**0.997**	0.984	0.643	0.061	0.015
	Census [57]	0.349	0.995	0.985	0.828	0.408	0.409
	LIOT	**0.485**	0.994	**0.987**	**0.911**	**0.515**	**0.527**
	Topo [52]	0.212	0.752	0.744	0.425	0.023	0.229
	Topo+LIOT	**0.389**	**0.995**	**0.986**	**0.905**	**0.444**	**0.383**
STARE ⇒ CrackTree	Baseline [23]	0.022	0.988	0.974	0.371	0.024	0.004
	Census [57]	0.040	**0.999**	**0.986**	0.696	0.075	0.103
	LIOT	**0.290**	0.995	0.985	**0.865**	**0.360**	**0.400**
	Topo [52]	0.042	0.970	0.957	0.505	0.027	0.040
	Topo+LIOT	**0.209**	**0.997**	**0.985**	**0.784**	**0.289**	**0.207**
CHASEDB1 ⇒ CrackTree	Baseline [23]	0.099	0.994	0.981	0.684	0.129	0.028
	Census [57]	0.178	**0.998**	0.986	0.726	0.272	0.254
	LIOT	**0.452**	0.996	**0.988**	**0.909**	**0.516**	**0.518**
	Topo [52]	0.150	0.977	0.965	0.733	0.110	0.147
	Topo+LIOT	**0.454**	**0.994**	**0.987**	**0.888**	**0.493**	**0.456**
Average	Baseline [23]	0.077	**0.982**	0.930	0.591	0.110	0.012
	Census [57]	0.317	0.950	0.921	0.761	0.323	0.304
	LIOT	**0.538**	0.947	**0.932**	**0.896**	**0.499**	**0.534**
	Topo [52]	0.160	0.942	0.883	0.524	0.033	0.073
	Topo+LIOT	**0.479**	**0.954**	**0.933**	**0.851**	**0.466**	**0.477**

methods struggle for large image appearance gaps.

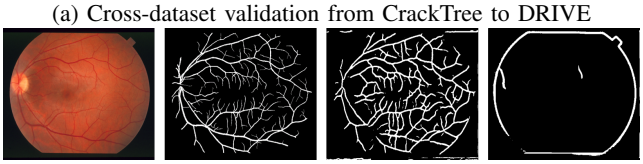
More precisely, as depicted in Fig. 9, for the model trained on CrackTree, using LIOT can accurately capture most retinal blood vessels, including those blue pixels which are of very low contrast and ignored by the manual annotation. Such an effect is difficult to achieve by general methods, especially in a simple way like LIOT. This demonstrates that LIOT is robust to contrast changes and is able to capture the inherent curvilinear structure.

The quantitative evaluation is shown in Table. IV. Compared with the baseline method, our proposed LIOT outperforms them by a large margin in terms of Se, AUC, and F1-score. Specifically, the proposed LIOT improves the baseline method by 34.5% to 54.0% in F1-score, significantly outperforming the baseline method. From the quantitative results given in Table V, the proposed LIOT makes the gap between CrackTree and the retinal dataset much smaller than the baseline method. Thus, LIOT significantly improves the generalization performance of the baseline method, especially for the dataset that has curvilinear structures with great appearance gaps.

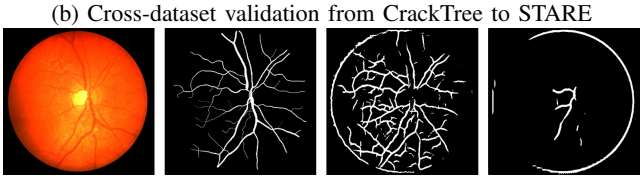
Cross-dataset validation from retinal to CrackTree dataset: We also evaluate the proposed LIOT on CrackTree using the model trained on the retinal datasets. The CrackTree dataset has a great appearance gap compared with the retinal datasets, and also has multiple shadows in the context. Though this makes such cross-dataset segmentation a challenging task, the proposed LIOT still successfully segments most curvilinear



(a) Cross-dataset validation from CrackTree to DRIVE

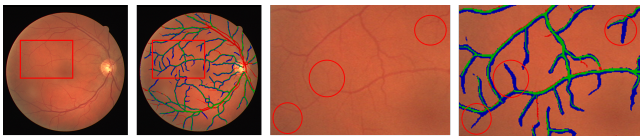


(b) Cross-dataset validation from CrackTree to STARE



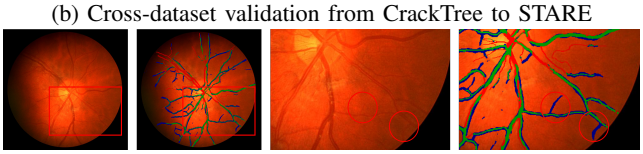
(c) Cross-dataset validation from CrackTree to CHASEDB1

Fig. 8: Visualization of the segmentation results under cross-dataset validation between CrackTree and retinal dataset.



(a) Cross-dataset validation from CrackTree to DRIVE

(b) Cross-dataset validation from CrackTree to STARE

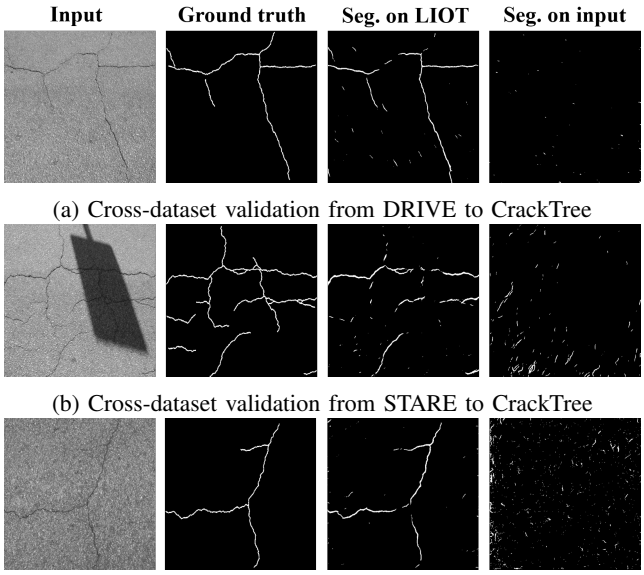


(c) Cross-dataset validation from CrackTree to CHASEDB1

Fig. 9: Visualization of some segmentation detail under cross-dataset validation between CrackTree and retinal datasets. Green pixels: TPs; Red pixels: FNs; Blue pixels: FPs. Some FPs achieved by LIOT can find evidence in the original image, which might be TPs ignored in the manual annotation.

structures in the CrackTree dataset. Some qualitative results are illustrated in Fig. 10. As shown in Fig. 10(b), LIOT is also able to extract curvilinear structure with the shadow influence. The quantitative compassion with the baseline method under this cross-dataset setting is given in Table. IV. LIOT achieves very competitive performance with the baseline method. Specifically, compared with the baseline model, LIOT achieves significant improvements, ranging from 26.2% to 45.4% in terms of F1-score. From the quantitative result given in Table V, the proposed LIOT makes the gap between retinal and crack images much smaller than the baseline method.

From the qualitative results shown in Fig. 8 and Fig. 10, and



(a) Cross-dataset validation from DRIVE to CrackTree

(b) Cross-dataset validation from STARE to CrackTree

(c) Cross-dataset validation from CHASEDB1 to CrackTree

Fig. 10: Visualization of the segmentation results under cross-dataset validation between retinal and CrackTree datasets.

TABLE V: Quantitative F1-score gap between in-dataset (Tab. I) and cross-dataset results (Tab. IV).

Cross-dataset	Baseline	LIOT	Topo	Topo+LIOT
CrackTree⇒ DRIVE	0.563	**0.205**	0.746	**0.238**
CrackTree⇒ STARE	0.569	**0.260**	0.762	**0.304**
CrackTree ⇒ CHASEDB1	0.683	**0.301**	0.764	**0.294**
DRIVE⇒ CrackTree	0.760	**0.299**	0.799	**0.371**
STARE⇒ CrackTree	0.802	**0.450**	0.802	**0.518**
CHASEDB1 ⇒ CrackTree	0.682	**0.276**	0.702	**0.306**

quantitative evaluations depicted in Table. IV and Table. V, the proposed LIOT is capable to capture the curvilinear structure in various curvilinear object datasets. This demonstrates the generalization ability of the proposed LIOT. Simply changing the original input images with LIOT can help deep learning-based methods express a better generalization in various curvilinear object segmentation.

G. Generalization to images with different curvilinear objects

To further demonstrate the generalizability of the proposed LIOT, we also apply the model trained on the DRIVE dataset to images with different types of curvilinear structures: material grains, ceiling grids, table lines, tree's growth rings, and hair. As illustrated in Fig. 11, though the adopted model is trained on the retinal images having very different appearances from these test images, LIOT can still successfully segment these curvilinear structures of different types. Whereas, the baseline model trained on the original images fails to extract many curvilinear structures and has many false positives. This illustrative experiment further implies that LIOT is able to capture the inherent characteristic of curvilinear objects, and can thus improve the generalizability of existing deep learning-based models for curvilinear object segmentation.

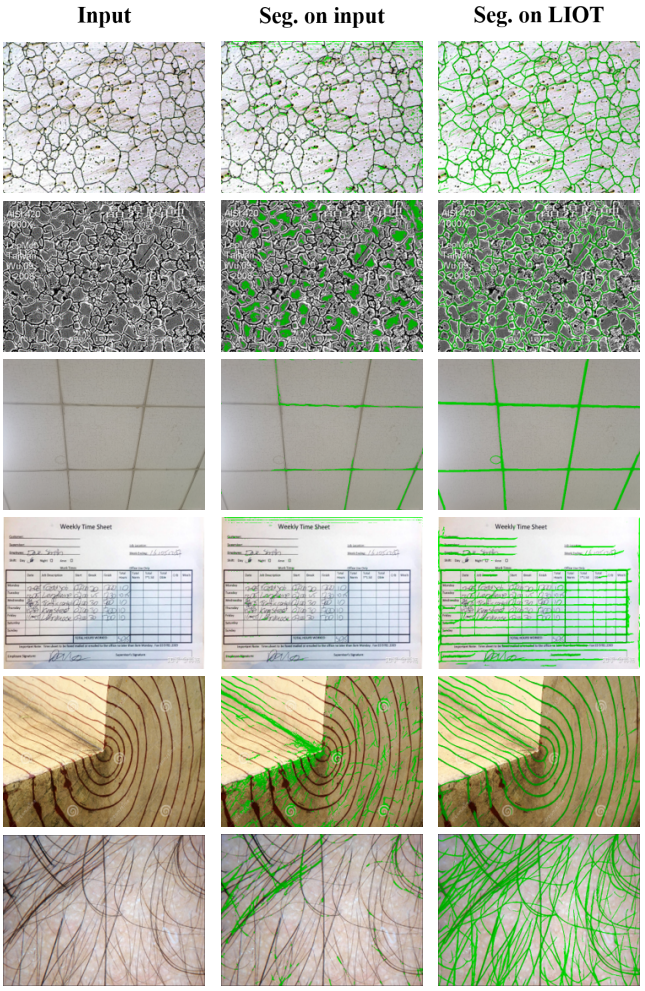


Fig. 11: Some segmentation results on different types of curvilinear images using the model trained on DRIVE dataset. From left to right: input image, segmentation on the original image using the baseline model, segmentation result by LIOT. Best viewed by zooming in the electronic version.

V. Conclusion

In this paper, we aim to improve the generalization ability of current deep learning-based curvilinear object segmentation methods. For that, we propose the LIOT that converts a gray-scale image to a novel representation that is invariant to increasing contrast changes. LIOT is built on the intensity order between pairs of pixels, and thus does not depend on the absolute intensity values. Such intensity-order based representation captures the inherent property of curvilinear objects (*e.g.*, curvilinear structure darker than the context). Despite its simplicity, LIOT is rather robust to non-targeted universal adversarial perturbations. Besides, extensive cross-dataset experiments on three widely adopted retinal blood vessel segmentation datasets and CrackTree dataset demonstrate that LIOT can improve the classical segmentation pipeline that directly operates on the original image. Therefore, LIOT forms a simple yet effective way to improve the generalization performance of different models. To the best of our knowledge, LIOT is a first simple attempt to improve the generalizability

of curvilinear object segmentation from the aspect of image transformation. More sophisticated image transformation is worthy of further investigation for robust curvilinear object segmentation.

REFERENCES

[1] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, 2010.

[2] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "Blood vessel segmentation methodologies in retinal images–a survey," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 1, pp. 407–433, 2012.

[3] E. Liu, H. Zhao, F. Guo, J. Liang, and J. Tian, "Fingerprint segmentation based on an adaboost classifier," *Frontiers of Computer Science*, vol. 5, no. 2, pp. 148–157, 2011.

[4] T. Laibacher, T. Weyde, and S. Jalali, "M2U-Net: Effective and efficient retinal vessel segmentation for real-world applications," in *Proc. of IEEE Intl. Conf. on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[5] K. Su, G. Yang, L. Yang, P. Su, and Y. Yin, "Non-negative locality-constrained vocabulary tree for finger vein image retrieval," *Frontiers of Computer Science*, vol. 13, no. 2, pp. 318–332, 2019.

[6] Q. Zou, Y. Cao, Q. Li, Q. Mao, and S. Wang, "Cracktree: Automatic crack detection from pavement images," *Pattern Recognition Letters*, vol. 33, no. 3, pp. 227–238, 2012.

[7] P. Bibiloni, M. González-Hidalgo, and S. Massanet, "A survey on curvilinear object segmentation in multiple applications," *Pattern Recognition*, vol. 60, pp. 949–970, 2016.

[8] T. M. Koller, G. Gerig, G. Szekely, and D. Dettwiler, "Multiscale detection of curvilinear structures in 2-D and 3-D image data," in *Proc. of IEEE Intl. Conf. on Computer Vision*, 1995, pp. 864–869.

[9] M. J. Carlotto, "Enhancement of low-contrast curvilinear features in imagery," *IEEE Trans. on Image Processing*, vol. 16, no. 1, pp. 221–228, 2006.

[10] J. V. Soares, J. J. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," *IEEE Trans. on Medical Imaging*, vol. 25, no. 9, pp. 1214–1222, 2006.

[11] E. Ricci and R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification," *IEEE Trans. on Medical Imaging*, vol. 26, no. 10, pp. 1357–1365, 2007.

[12] B. Al-Diri, A. Hunter, and D. Steel, "An active contour model for segmenting and measuring retinal vessels," *IEEE Trans. on Medical Imaging*, vol. 28, no. 9, pp. 1488–1497, 2009.

[13] D. Marín, A. Aquino, M. E. Gegúndez-Arias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE Trans. on Medical Imaging*, vol. 30, no. 1, pp. 146–158, 2010.

[14] B. Obara, M. Fricker, D. Gavaghan, and V. Grau, "Contrast-independent curvilinear structure detection in biomedical images," *IEEE Trans. on Image Processing*, vol. 21, no. 5, pp. 2572–2581, 2012.

[15] C. Xiao, M. Staring, Y. Wang, D. P. Shamonin, and B. C. Stoel, "Multi-scale bi-Gaussian filter for adjacent curvilinear structures detection with application to vasculature images," *IEEE Trans. on Image Processing*, vol. 22, no. 1, pp. 174–188, 2012.

[16] V. A. Krylov and J. D. Nelson, "Stochastic extraction of elongated curvilinear structures with applications," *IEEE Trans. on Image Processing*, vol. 23, no. 12, pp. 5360–5373, 2014.

[17] R. Annunziata, A. Kheirkhah, P. Hamrah, and E. Trucco, "Boosting hand-crafted features for curvilinear structure segmentation by learning context filters," in *Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention*, 2015, pp. 596–603.

[18] C. Vicas and S. Nedevschi, "Detecting curvilinear features using structure tensors," *IEEE Trans. on Image Processing*, vol. 24, no. 11, pp. 3874–3887, 2015.

[19] E. Türetken, F. Benmansour, B. Andres, P. Głowacki, H. Pfister, and P. Fua, "Reconstructing curvilinear networks using path classifiers and integer programming," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 12, pp. 2515–2530, 2016.

[20] O. Merveille, H. Talbot, L. Najman, and N. Passat, "Curvilinear structure analysis by ranking the orientation responses of path operators," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 2, pp. 304–317, 2017.

[21] N. Strisciuglio, G. Azzopardi, and N. Petkov, "Robust inhibition-augmented operator for delineation of curvilinear structures," *IEEE Trans. on Image Processing*, vol. 28, no. 12, pp. 5852–5866, 2019.

[22] O. Merveille, B. Naegel, H. Talbot, and N. Passat, "n D variational restoration of curvilinear structures with prior-based directional regularization," *IEEE Trans. on Image Processing*, vol. 28, no. 8, pp. 3848–3859, 2019.

[23] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks," in *Proc. of IEEE Winter Conf. on Applications of Computer Vision*, 2020, pp. 3656–3665.

[24] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE Trans. on Medical Imaging*, vol. 35, no. 1, pp. 109–118, 2015.

[25] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, "Deep retinal image understanding," in *Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention*, 2016, pp. 140–148.

[26] A. Mosinska, P. Marquez-Neila, M. Koziński, and P. Fua, "Beyond the pixel-wise loss for topology-aware delineation," in *Proc. of IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, 2018, pp. 3136–3145.

[27] Z. Yan, X. Yang, and K.-T. Cheng, "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1912–1923, 2018.

[28] S. Luan, C. Chen, B. Zhang, J. Han, and J. Liu, "Gabor convolutional networks," *IEEE Trans. on Image Processing*, vol. 27, no. 9, pp. 4357–4366, 2018.

[29] A. Mosinska, M. Koziński, and P. Fua, "Joint segmentation and path classification of curvilinear structures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 6, pp. 1515–1521, 2019.

[30] F. Wang, Y. Gu, W. Liu, Y. Yu, S. He, and J. Pan, "Context-aware spatio-recurrent curvilinear structure segmentation," in *Proc. of IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, 2019, pp. 12 648–12 657.

[31] L. Mou, Y. Zhao, L. Chen, J. Cheng, Z. Gu, H. Hao, H. Qi, Y. Zheng, A. Frangi, and J. Liu, "CS-Net: channel and spatial attention network for curvilinear structure segmentation," in *Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention*, 2019, pp. 721–730.

[32] S. Guo, K. Wang, H. Kang, Y. Zhang, Y. Gao, and T. Li, "BTS-DSN: Deeply supervised neural network with short connections for retinal vessel segmentation," *International Journal of Medical Informatics*, vol. 126, pp. 105–113, 2019.

[33] V. Cherukuri, V. K. BG, R. Bala, and V. Monga, "Deep retinal image segmentation with regularization under geometric priors," *IEEE Trans. on Image Processing*, vol. 29, pp. 2552–2567, 2019.

[34] S. Dey, "A subpixel residual U-Net and feature fusion preprocessing for retinal vessel segmentation," in *Proc. of European Conference on Computer Vision*, 2020, pp. 239–250.

[35] L. Mou, Y. Zhao, H. Fu, Y. Liu, J. Cheng, Y. Zheng, P. Su, J. Yang, L. Chen, A. F. Frangi *et al.*, "CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging," *Medical Image Analysis*, vol. 67, p. 101874, 2020.

[36] L. Ding, M. H. Bawany, A. E. Kuriyan, R. S. Ramchandran, C. C. Wykoff, and G. Sharma, "A novel deep learning pipeline for retinal vessel detection in fluorescein angiography," *IEEE Trans. on Image Processing*, vol. 29, pp. 6561–6573, 2020.

[37] J. Zhang, Y. Qiao, M. S. Sarabi, M. M. Khansari, J. K. Gahm, A. H. Kashani, and Y. Shi, "3D shape modeling and analysis of retinal microvasculature in OCT-angiography images," *IEEE Trans. on Medical Imaging*, vol. 39, no. 5, p. 1335, 2020.

[38] A. Nazir, M. N. Cheema, B. Sheng, H. Li, P. Li, P. Yang, Y. Jung, J. Qin, J. Kim, and D. D. Feng, "OFF-eNET: An optimally fused fully end-to-end network for automatic dense volumetric 3D intracranial blood vessels segmentation," *IEEE Trans. on Image Processing*, vol. 29, pp. 7192–7202, 2020.

[39] H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, and J. Qin, "SCS-Net: A scale and context sensitive network for retinal vessel segmentation," *Medical Image Analysis*, vol. 70, p. 102025, 2021.

[40] Z. Shen, H. Fu, J. Shen, and L. Shao, "Modeling and enhancing low-quality retinal fundus images," *IEEE Trans. on Medical Imaging*, vol. 40, no. 3, pp. 996–1006, 2021.

[41] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[42] N. He, L. Fang, and A. Plaza, "Hybrid first and second order attention Unet for building segmentation in remote sensing images," *Science China Information Sciences*, vol. 63, no. 4, pp. 1–12, 2020.

[43] J. Yue, L. Fang, H. Rahmani, and P. Ghamisi, "Self-supervised learning with adaptive distillation for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2021.

[44] J. Xie, L. Fang, B. Zhang, J. Chanussot, and S. Li, "Super resolution guided deep network for land cover classification from remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[45] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[46] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. on Medical Imaging*, vol. 19, no. 3, pp. 203–210, 2000.

[47] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–2548, 2012.

[48] J. Wang and A. C. Chung, "Higher-order flux with spherical harmonics transform for vascular analysis," in *Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention*. Springer, 2020, pp. 55–65.

[49] Y. Ma, H. Hao, J. Xie, H. Fu, J. Zhang, J. Yang, Z. Wang, J. Liu, Y. Zheng, and Y. Zhao, "ROSE: A retinal oct-angiography vessel segmentation dataset and new model," *IEEE Trans. on Medical Imaging*, vol. 40, no. 3, pp. 928–939, 2021.

[50] C. Wang, Y. Hayashi, M. Oda, H. Itoh, T. Kitasaka, A. F. Frangi, and K. Mori, "Tubular structure segmentation using spatial fully connected network with radial distance loss for 3D medical images," in *Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention*, 2019, pp. 348–356.

[51] Y. Wang, X. Wei, F. Liu, J. Chen, Y. Zhou, W. Shen, E. K. Fishman, and A. L. Yuille, "Deep distance transform for tubular structure segmentation in CT scans," in *Proc. of IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, 2020, pp. 3833–3842.

[52] X. Hu, L. Fuxin, D. Samaras, and C. Chen, "Topology-preserving deep image segmentation," in *Proc. of Advances in Neural Information Processing Systems*, 2019.

[53] X. Hu, Y. Wang, L. Fuxin, D. Samaras, and C. Chen, "Topology-aware segmentation using discrete morse theory," in *Proc. of International Conference on Learning Representations*, 2021.

[54] S. Shit, J. C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, A. Zhylka, J. P. Pluim, U. Bauer, and B. H. Menze, "clDice-a novel topology-preserving loss function for tubular structure segmentation," in *Proc. of IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, 2021, pp. 16 560–16 569.

[55] H. Wang, Y. Song, C. Zhang, J. Yu, S. Liu, H. Pengy, and W. Cai, "Single neuron segmentation using graph-based global reasoning with auxiliary skeleton loss from 3D optical microscope images," in *Proc. of Intl. Symp. on Biomedical Imaging*, 2021, pp. 934–938.

[56] M. Cheng, K. Zhao, X. Guo, Y. Xu, and J. Guo, "Joint topology-preserving and feature-refinement network for curvilinear structure segmentation," in *Proc. of IEEE Intl. Conf. on Computer Vision*, 2021, pp. 7147–7156.

[57] R. Zabih and J. Woodfill, "A non-parametric approach to visual correspondence," in *IEEE Trans. Pattern Anal. Mach. Intell.*, 1996.

[58] N. Y.-C. Chang, T.-H. Tsai, B.-H. Hsu, Y.-C. Chen, and T.-S. Chang, "Algorithm and architecture of disparity estimation with mini-census adaptive support weight," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 6, pp. 792–805, 2010.

[59] J. Lee, D. Jun, C. Eem, and H. Hong, "Improved census transform for noise robust stereo matching," *Optical Engineering*, vol. 55, no. 6, p. 063107, 2016.

[60] B. Froba and A. Ernst, "Face detection with the modified census transform," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 91–96.

[61] K. Ambrosch, C. Zinner, and H. Leopold, "A miniature embedded stereo vision system for automotive applications," in *IEEE Convention of Electrical and Electronics Engineers in Israel*, 2010, pp. 000 786–000 789.

[62] W. S. Fife and J. K. Archibald, "Improved census transforms for resource-optimized stereo vision," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 1, pp. 60–73, 2012.

[63] R. Spangenberg, T. Langner, and R. Rojas, "Weighted semi-global matching and center-symmetric census transform for robust driver assistance," in *International Conference on Computer Analysis of Images and Patterns*, 2013, pp. 34–41.

[64] C. Ahlberg, M. L. Ortiz, F. Ekstrand, and M. Ekstrom, "Unbounded sparse census transform using genetic algorithm," in *Proc. of IEEE Winter Conf. on Applications of Computer Vision*, 2019, pp. 1616–1625.

[65] S. Yu, Y. He, Z. Chen, C. Ru, and M. Pang, "Stereo matching method based on combination characteristic cost computing and unstable tree reconstruction optimization and its application in medical images," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 3, pp. 646–653, 2020.

[66] X. Lai, X. Xu, J. Zhang, Y. Fang, and Z. Huang, "An efficient implementation of a census-based stereo matching and its applications in medical imaging," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 6, pp. 1152–1159, 2019.

[67] J. Y. R. Cornejo and H. Pedrini, "Audio-visual emotion recognition using a hybrid deep convolutional neural network based on census transform," in *IEEE International Conference on Systems, Man and Cybernetics*, 2019, pp. 3396–3402.

[68] Z. Wang, B. Fan, and F. Wu, "Local intensity order pattern for feature description," in *Proc. of IEEE Intl. Conf. on Computer Vision*. IEEE, 2011, pp. 603–610.

[69] Z. Wang, B. Fan, G. Wang, and F. Wu, "Exploring local and overall ordinal information for robust feature description," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2198–2211, 2015.

[70] B. Fan, F. Wu, and Z. Hu, "Rotationally invariant descriptors using intensity order pooling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2031–2045, 2011.

[71] L. Shen, J. Lin, S. Wu, and S. Yu, "HEp-2 image classification using intensity order pooling based features and bag of words," *Pattern Recognition*, vol. 47, no. 7, pp. 2419–2427, 2014.

[72] Z. Lei, D. Yi, and S. Z. Li, "Local gradient order pattern for face representation and recognition," in *Proc. of Intl. Conf. on Pattern Recognition*. IEEE, 2014, pp. 387–392.

[73] G. Wang, B. Fan, Z. Zhou, and C. Pan, "Ordinal pyramid coding for rotation invariant feature extraction," *Neurocomputing*, vol. 242, pp. 150–160, 2017.

[74] M. E. Gegúndez-Arias, A. Aquino, J. M. Bravo, and D. Marín, "A function for quality evaluation of retinal vessel segmentations," *IEEE Trans. on Medical Imaging*, vol. 31, no. 2, pp. 231–239, 2011.

[75] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. of International Conference on Learning Representations*, 2015.

[76] O. Poursaeed, I. Katsman, B. Gao, and S. Belongie, "Generative adversarial perturbations," in *Proc. of IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, 2018, pp. 4422–4431.

[77] A. Oliveira, S. Pereira, and C. A. Silva, "Retinal vessel segmentation based on fully convolutional neural networks," *Expert Systems with Applications*, vol. 112, pp. 229–242, 2018.

[78] T. L. Torp, R. Kawasaki, T. Y. Wong, T. Peto, and J. Grauslund, "Temporal changes in retinal vascular parameters associated with successful panretinal photocoagulation in proliferative diabetic retinopathy: a prospective clinical interventional study," *Acta Ophthalmologica*, vol. 96, no. 4, pp. 405–410, 2018.