FreeCOS: Self-Supervised Learning from Fractals and Unlabeled Images for Curvilinear Object Segmentation

Tianyi Shi, Xiaohuan Ding, Liang Zhang, Xin Yang†
School of EIC, Huazhong University of Science & Technology
{shitianyihust, dingxiaohuan, liangz, xinyang2014}@hust.edu.cn

Abstract

Curvilinear object segmentation is critical for many applications. However, manually annotating curvilinear objects is very time-consuming and error-prone, yielding insufficiently available annotated datasets for existing supervised methods and domain adaptation methods. This paper proposes a self-supervised curvilinear object segmentation method that learns robust and distinctive features from fractals and unlabeled images (FreeCOS). The key contributions include a novel Fractal-FDA synthesis (FFS) module and a geometric information alignment (GIA) approach. FFS generates curvilinear structures based on the parametric Fractal L-system and integrates the generated structures into unlabeled images to obtain synthetic training images via Fourier Domain Adaptation. GIA reduces the intensity differences between the synthetic and unlabeled images by comparing the intensity order of a given pixel to the values of its nearby neighbors. Such image alignment can explicitly remove the dependency on absolute intensity values and enhance the inherent geometric characteristics which are common in both synthetic and real images. In addition, GIA aligns features of synthetic and real images via the prediction space adaptation loss (PSAL) and the curvilinear mask contrastive loss (CMCL). Extensive experimental results on four public datasets, i.e., XCAD, DRIVE, STARE and CrackTree demonstrate that our method outperforms the state-of-the-art unsupervised methods, self-supervised methods and traditional methods by a large margin. The source code of this work is available at https://github.com/TY-Shi/FreeCOS.

1. Introduction

Automatically segmenting curvilinear structures (such as vascular trees in medical images and road systems in aerial photography) is critical for many applications, including retinal fundus disease screening [2, 14], diagnosing coronary artery disease [36], road condition evaluation and maintenance [49]. Despite a plethora of research works in the literature, accurately segmenting curvilinear objects remains challenging due to their complex structures with numerous tiny branches, tortuosity shapes, ambiguous boundaries due to imaging issues and noisy backgrounds.

Most recent methods [37, 10, 34, 40, 18, 28, 7, 27, 6, 33] leverage supervised deep learning for curvilinear object segmentation and have achieved encouraging results. However, those methods require a large number of pixel-wise manual annotations for training which are very expensive to obtain and error-prone due to poor image quality, annotator's fatigue and lack of experience. Although, there are several publically available annotated datasets for curvilinear object segmentation [35, 17, 49, 24], the large appearance variations between different curvilinear object images, e.g., X-ray coronary angiography images vs. retinal fundus images, yields significant performance degradation for supervised models across different types of images (even across the same type of images acquired using different equipments). As a result, expensive manual annotations are inevitably demanded to tune the segmentation model for a particular application. Potential solutions to alleviate the annotation burden include domain adaption [8, 31] and unsupervised segmentation [9, 19, 23, 11, 9, 1]. However, the effectiveness of domain adaptation is largely dependent on the quality of annotated data in the source domain and constrained by the gap between the source and target domain. Existing unsupervised segmentation methods [9, 19] can hardly achieve satisfactory performance for curvilinear objects due to their thin, long, and tortuosity shapes, complex branching structures, and confusing background artifacts.

Despite the high complexity and great variety of curvilinear structures in different applications, they share some common characteristics (i.e., the tube-like shape and the branching structure). Thus, existing studies [45, 44, 46] have demonstrated that several curvilinear structures (e.g., arterial trees of the circulation system) can be generated via the fractal systems with proper branching parameters to mimic the fractal and physiological characteristics, and some observed variability. These results motivate us to

use the generated curvilinear objects via the fractal systems to explicitly encode geometric properties and varieties (i.e., different diameters and lengths of branches, and different branching angles) into training samples and to assist feature learning of a curvilinear structure segmentation model. However, such formulas generated training samples can hardly mimic the appearance patterns within curvilinear objects, the transition regions between curvilinear objects and backgrounds, which are also key information for learning a segmentation model and contained in easily-obtained unlabeled target images. This paper asks the question, how to combine fractals and unlabeled target images to encode sufficient and comprehensive visual cues for learning robust and distinctive features of curvilinear structures?

The main contribution of this paper is a self-supervised segmentation method based on a novel Fractal-FDA synthesis (FFS) module and a geometric information alignment (GIA) approach. Specifically, curvilinear structures are synthesized by the parametric fractal L-Systems [45] and serve as segmentation labels of synthetic training samples. To simulate appearance patterns in the object-background transition regions and background regions, we apply Fourier Domain Adaptation [43] (FDA) to fuse synthetic curvilinear structures and unlabeled target images. The synthetic images via our FFS module can effectively guide learning distinctive features to distinguish curvilinear objects and backgrounds. To further improve the robustness to differences between intensity distributions of synthetic and real target images, we design a novel geometric information alignment (GIA) approach which aligns information of synthetic and target images at both image and feature levels. Specifically, GIA first converts each training image (synthetic and target images) into four geometry-enhanced images by comparing the intensity order of a given pixel to the values of its nearby neighbors (i.e., along with the up, down, left and right directions). In this way, the four converted images do not depend on the absolute intensity values but the relative intensity in order to capture the inherent geometric characteristic of the curvilinear structure, reducing the intensity differences between synthetic and target images. Then, we extract features from the 4-channel converted images and propose two loss functions, i.e., a prediction space adaptation loss (PSAL) and a curvilinear mask contrastive loss (CMCL), to align the geometric features of synthetic and target images. The PSAL minimizes the distance between the segmentation masks of the target images and synthetic curvilinear objects and the CMCL minimizes the distance between features of segmented masks and synthetic objects.

The FreeCOS based on FFS and GIA approaches applies to several public curvilinear object datasets, including XCAD [24], DRIVE [35], STARE [17] and CrackTree [49]. Extensive experimental results demonstrate that FreeCOS outperforms the state-of-the-art self-supervised [24, 21], unsupervised [9, 19], and traditional methods [13, 22]. To summarize, the main contributions of this work are as follows:

- We propose a novel self-supervised curvilinear feature learning method which intelligently combines tree-like fractals and unlabeled images to assist in learning robust and distinctive feature representations.
- We propose Fractal-FDA synthesis (FFS) and geometric information alignment (GIA), which are the two key enabling modules of our method. FFS integrates the synthetic curvilinear structures into unlabeled images to guide learning distinctive features to distinguish foregrounds and backgrounds. GIA enhances geometric features and meanwhile improves the feature robustness to intensity differences between synthetic and target unlabeled images.
- We develop a novel self-supervised segmentation network that can be trained using only target images and fractal synthetic curvilinear objects. Our network performs significantly better than state-of-the-art self-supervised /unsupervised methods on multiple public datasets with various curvilinear objects.

2. Related Work

2.1. Traditional Methods

Traditional curvilinear object segmentation methods [20, 25, 13, 38, 22] design heuristic rules and/or filters to capture features of the target curvilinear objects. For instance, Frangi et al. introduce the vesselness filter [13] based on the Hessian matrix to represent and enhance tube-like curvilinear objects. Khan et al. [20] further design B-COSFIRE filters to denoise retinal images and segment retinal vessels. Memari et al. [25] enhance image contrast via contrast-limited adaptive histogram equalization and then segment retinal vessels based on hand-crafted filters. In [38, 22], the authors propose optimally oriented flux (OOF) to enhance curvilinear tube-like objects. OOF exhibits better performance for segmenting adjacent curvilinear objects yet is sensitive to different sizes of curvilinear objects.

Traditional methods based on hand-crafted filters do not require any training yet they require careful parameter tuning for optimized performance. And the optimized parameter settings are usually data-dependent or even region-dependent, limiting their convenience in segmenting a wide variety of curvilinear objects.

2.2. Unsupervised Segmentation Methods

Unsupervised segmentation methods can be generally divided into two classes: clustering based [19, 23, 11] and ad-

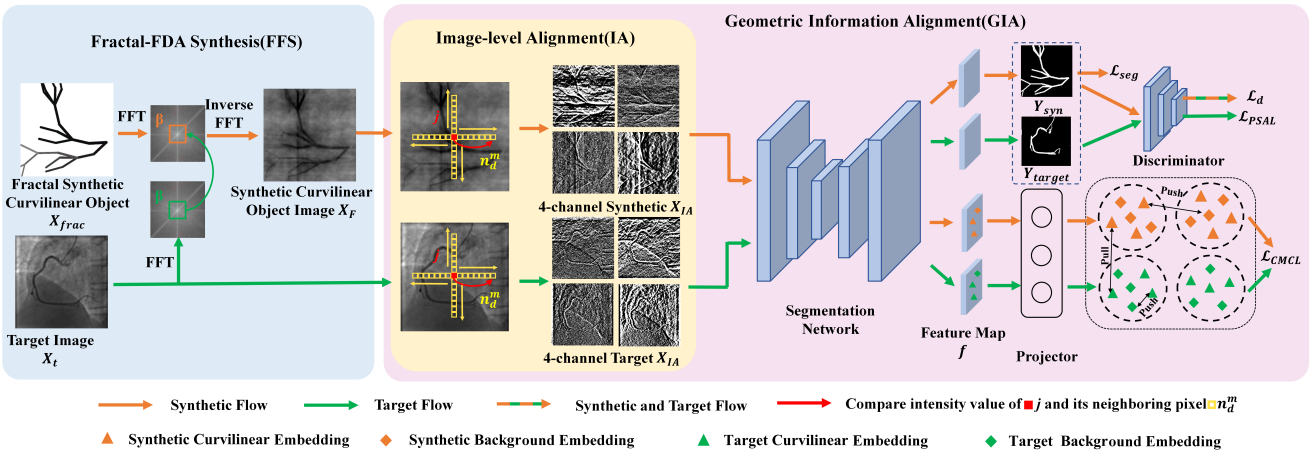

Figure 1. The pipeline of Self-Supervised Learning from Fractals and Unlabeled Images for Curvilinear Object Segmentation (FreeCOS).

versarial learning based [9, 1]. Xu et al. [19] propose Invariant Information Clustering (IIC) which automatically partitions input images into regions of different semantic classes by optimizing mutual information between related region pairs. Such a clustering-based method is more suitable for segmenting objects with aspect ratios close to one while becoming ineffective for curvilinear objects due to their thin, long, tortuous shapes. Redo [9] is based on an adversarial architecture where the generator is guided by an image and extracts the object mask, then redraws a new object at the same location with different textures/ colors. However, this adversarial learning-based unsupervised method only performs well for objects which are visually distinguishable from backgrounds. For the segmentation of curvilinear objects with complex and numerous tiny branching structures, embedded in confusing and cluttered backgrounds, the efficacy of such a method degrades significantly.

In contrast to unsupervised methods, our method explicitly encodes geometric and photometric characteristics, as well as some observed varieties of curvilinear objects in target application into synthetic images. Those synthetic images provide labels to effectively guide the model to learn robust and distinctive features and thus yield superior performance to state-of-the-art unsupervised methods [9, 19].

2.3. Self-supervised Learning Methods

Self-supervised learning methods construct pretexts from large-scale unsupervised data and utilize contrastive learning losses to measure the similarities of sample pairs in the representation space. To this end, various pretext tasks have been designed, including jigsaw [12], hole-fill [30] and transformation invariance [16]. Although existing self-supervised methods have achieved outstanding performance in classification [26, 4], detection [42, 5], and image translation [29, 41], few of them provide a suitable pretext task for curvilinear object segmentation. A potential solution is the pixel-level contrastive-based method [39, 3, 47], while it requires heavy manual annotations to prepare pixel-level positive and negative samples. Similar work to ours is [24, 21] which designs a self-supervised vessel segmentation method via adversarial learning and fractals. But this method requires clean background images as input (i.e., the first frame of the angiography sequence) for synthesis which greatly limits its applications. In addition, adversarial learning cannot explicitly and precisely enforce visual cues, which are important for learning segmentation-oriented features, being encoded in the synthetic images. As a result, although their synthetic images visually look similar to the target images, the segmentation accuracy is also quite low.

3. Method

Figure. 1 shows the framework of FreeCOS which consists of two main modules, i.e., Fractal-FDA Synthesis (FFS) and Geometric Information Alignment (GIA). In FFS, we generate synthetic curvilinear structures via the parametric Fractal L-Systems and use the generated structures as segmentation maps to guide self-supervised training of a segmentation network. The synthetic curvilinear structures are then integrated into unlabeled images via FDA to form synthetic images of curvilinear structures. The intensity distributions of synthetic images could deviate from those of real target images and hence yield poor feature robustness. To address this problem, our GIA module first reduces image-level differences between the synthetic and target images by converting intensity images into four-channel intensity order images. The converted images of both synthetic images and target images are then input into U-Net to extract feature representations. Our GIA further aligns features of synthetic images and target images via the prediction space adaptation loss (PSAL) and the curvilinear mask contrastive loss (CMCL). In the following, we present the details of FFS and GIA.

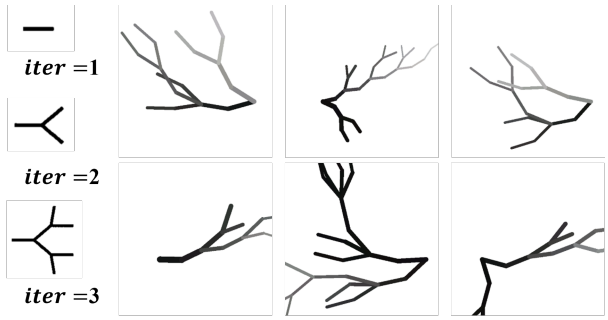

Figure 2. Exemplar results of the generated curvilinear structures from the basic and parametric Fractal L-system.

3.1. Fractal-FDA Synthesis

We first generate curvilinear structures via the parametric Fractal L-Systems and then integrate the synthetic objects into unlabeled target images to obtain synthetic training samples via FDA.

Fractal-based curvilinear structure generation. Fractals are simple graphic patterns rendered by mathematical formulas. In this work, we adopt the parametric Fractal L-systems proposed by Zamir et al. [45] to generate fractal tree structures and meanwhile select proper branching parameters according to physiological laws of curvilinear objects in target applications. Specifically, "grammar" for generating curvilinear structures with repeated bifurcations based on Fractal L-systems method is defined as follows:

$$\begin{aligned} \omega: &\quad F \\ rule: &\quad F \rightarrow F[-F][+F] \end{aligned} \tag{1}$$

The generated object by iteration $iter$:

$$\mathrm{Draw}(iter, F, rule) \tag{2}$$

where F represents a line of unit length in the horizontal direction, ω denotes an axiom, $iter$ denotes the iteration and $rule$ denotes the production rule. The square brackets represent the departure from ([) and return to (]) a branch point. The plus and minus signs represent turns through a given angle δ in the clockwise and anticlockwise directions, respectively. For example, the first three stages of a tree produced by the basic Fractal L-system, denoted by $iter$=1 $\sim$ 3, are given by the:

$$\begin{aligned} iter = 1 \quad & F \\ iter = 2 \quad & F[-F][+F] \\ iter = 3 \quad & F[-F][+F][-F[-F][+F]][+F[-F][+F]] \end{aligned} \tag{3}$$

The left part of Figure. 2 illustrates exemplar results of the generated curvilinear structures based on the basic Fractal system.

To synthesize curvilinear structures with various widths and lengths, we replace F using $F_{random}\,(w_i, l_i, c_i)$ which represents a line of width w_i, length l_i and intensity c_i. We set the initial width w_{init}, length l_{init} and decreased parameter γ by the repeated F_{random} in the $Draw$ grammar. For the index number i of F_{random}, w_i and l_i are given by:

$$\begin{aligned} w_i &= w_{init} \cdot \gamma^{i-1} \\ l_i &= l_{init} \cdot \gamma^{i-1} \end{aligned} \tag{4}$$

We randomly choose a value within the range of (0, 255) and set this value as the intensity c_i for the index i of F_{random}. For the branching angle δ, we replace δ using δ_{random} (defined by $\delta_{init} \pm \delta_{delta}$, δ_{init} within the range of $(20°, 120°)$), in which δ_{delta} is a random angle ranging from $10°$ to $40°$.

To mimic the geometric characteristics of a target application, we incorporate the above-mentioned parameters into the L-Systems. Specifically, we design a set of rules $ruleset: (rule_1, \cdots, rule_n)$ for different fractal structures, like $F_{random} \rightarrow F_{random} - F_{random}\,[+F_{random} - F_{random}]\,[-F_{random} + F_{random}]$. For each iteration $iter$, we randomly select a rule from the $ruleset$ as the production $rule$. In this way, we can increase the diversity of structures. The generated fractal curvilinear object $X_{frac} \in \mathbb{R}^{H \times W}$ by iteration is given by:

$$X_{frac} = \mathrm{Draw}(iter, F_{random}, ruleset, w_i, l_i, c_i) \tag{5}$$

FDA-based curvilinear object image synthesis. We further incorporate synthetic curvilinear objects into the target unlabeled images via FDA [43]. We define the target images as $X_t \in \mathbb{R}^{H \times W}$ and the synthetic images of fractal curvilinear object as X_{frac}. Let $\mathcal{F}^A :\in \mathbb{R}^{H \times W} \rightarrow\in \mathbb{R}^{H \times W}$ be the amplitude component of the Fourier transform $\mathcal{F}$ of a grayscale image. We define a $\beta \in (0, 1)$ to select the center region of the amplitude map.

Given two randomly sampled synthetic image X_{frac} and target image X_t, we follow FDA [43] to replace the low-frequency part in the amplitude map of X_{frac} by the FFT [15], denotes as $\mathcal{F}^A\,(X_{frac})$, with that of the target image X_t, denoted as $\mathcal{F}^A\,(X_t)$. Then, the modified spectral representation of X_{frac}, with its phase component unaltered, is mapped back to image $X_{frac \rightarrow t}$ by the inverse FFT [15], whose content is the same as X_{frac}, but will resemble the appearance of a sample X_t. After that, we apply a Gaussian blur to the output synthetic image of FDA to obtain synthetic images from FFS, denoted as X_F.

3.2. Geometric Information Alignment

Synthetic images generated by FFS could still have non-trivial intensity differences from that of real images. To address this problem, we further align synthetic images and real target images at both image and feature levels.

Image-level Alignment. We aim to explicitly remove the dependency on raw intensity values for both synthetic and real images and meanwhile enhance the geometric characteristics of curvilinear structures. To this end, for each image pixel j we compare its intensity value $X(j)$ with its 8 neighboring pixels $\{n_d^m \mid m = 1, \ldots, 8\}$ (yellow box in Figure. 1) lying perpendicular to j (red box in Figure. 1), where directional d denotes left, right, top, and bottom side of j. If $X(j) > X(n_d^m)$, we set the corresponding value for n_d^m to 0. Otherwise, we set the value for n_d^m to 1. Such operation produces four 8-bit images for each input image, where the m-th bit on the d-th image denotes the relative intensity order between j and its neighbor m along the d-th direction. For each input image X (denote as X_F and X_t), we concatenate four 8-bit images $\left(X_{IA}^1, \cdots, X_{IA}^d\right)$, where $X_{IA}^d(j)$ is represented as

$$X_{IA}^d(j) = \sum_{m=1}^{8} \left[X(j) > X\left(n_d^m\right)\right] \times 2^{m-1} \tag{6}$$

Such local intensity order transformation [33] can capture the intrinsic of the curvilinear object and meanwhile reduces the intensity gap between synthetic and real images.

Feature-level Alignment. Given the 4-channel transformed images, we utilize U-Net [32] to extract features. We align curvilinear objects' features of synthetic images and unlabeled target images based on two loss functions, i.e., the prediction space adaptation loss (PSAL) and the curvilinear mask contrastive loss (CMCL).

1) **Prediction space adaptation loss** We utilize adversarial learning to explicitly align the prediction space distribution of target images and synthetic images. We denote the prediction segmentation masks of the target images and the synthetic images as $Y_{target} \in \mathbb{R}^{H \times W}$ and $Y_{syn} \in \mathbb{R}^{H \times W}$, respectively. We input Y_{target} and Y_{syn} into a fully-convolutional discriminator D as [48] trained via a binary cross-entropy loss $\mathcal{L}_d$:

$$\mathcal{L}_d = \mathbb{E}\left[\log\left(D\left(Y_{\text{syn}}\right)\right)\right] + \mathbb{E}\left[\log\left(1 - D\left(Y_{target}\right)\right)\right] \tag{7}$$

Accordingly, the PSAL is computed as:

$$\mathcal{L}_{PSAL} = \mathbb{E}\left[\log\left(D\left(Y_{target}\right)\right)\right] \tag{8}$$

2) **Curvilinear mask contrastive loss**. The CMCL aims to reduce the distance between features of synthetic images and true target images, and meanwhile to improve the feature distinctiveness between curvilinear objects and backgrounds. To this end, we take the feature maps (denoted as $f \in \mathbb{R}^{H \times W \times C}$) from the final decoder layer of U-Net, the mask of synthetic image (denoted as $G_{syn} \in \{0,1\}^{H \times W}$) and prediction mask of target image (denoted $Y_{target} \in [0,1]$) as input. We process f using a lightweight contrastive encoder and a contrastive projector as [39] to map f to the feature space where the pixel-level contrastive loss is applied for $Z \in \mathbb{R}^{H \times W \times C}$.

We denote I as $H \times W$ spatial location of the projected feature maps Z, then for a location $i \in I$, we can obtain a feature vector z_i at location i from feature map Z, label values g_i and y_i at i from the mask of synthetic image and the prediction mask respectively. We partition pixels of I into two groups: curvilinear object locations I^+ and background locations I^-. For synthetic images, we perform the partition directly based on the mask G_{syn}, i.e., $I_{syn}^+ = \{i \in I_{syn} \mid g_i = 1\}$ and $I_{syn}^- = \{i \in I_{syn} \mid g_i = 0\}$. For target images, as the ground-truth mask is not available, we alternatively perform the partition based on the prediction probability mask Y_{target}, i.e., $I_{target}^+ = \{i \in I_{target} \mid y_i \geq 1 - \alpha\}$ and $I_{target}^- = \{i \in I_{target} \mid y_i \leq \alpha\}$, where α=0.1 is a small threshold and is fixed in our method.

We let $q_{syn}^+ = \left\{\mathbf{z}_i \mid i \in \S\left(I_{syn}^+, \sigma\right)\right\}$ and $k_{target}^+ = \left\{\mathbf{z}_i \mid i \in \S\left(I_{target}^+, \sigma\right)\right\}$ denote the curvilinear keys of synthetic and target images respectively. Similarly, we define $k_{syn}^- = \left\{\mathbf{z}_i \mid i \in \S\left(I_{syn}^-, \sigma\right)\right\}$ and $k_{target}^- = \left\{\mathbf{z}_i \mid i \in \S\left(I_{target}^-, \sigma\right)\right\}$ as the background keys of synthetic and target images, where $\S(\bullet, \sigma)$ is a random sampling operator which samples a subset from a set randomly with a proportion ratio σ. We combine N negative queries of the features of synthetic and target images to form a negative set $k^- = \left(k_{syn}^-, k_{target}^-\right)$. The CMCL is defined as:

$$\begin{aligned} \mathcal{L}_{CMCL} = &-\log\left(\exp\left(q_{syn}^+ \cdot k_{target}^+ / \tau\right)\right) + \\ &\log\left(\exp\left(q_{syn}^+ \cdot k_{target}^+ / \tau\right) + \sum_{i=0}^{N} \exp\left(q_{syn}^+ \cdot k_i^- / \tau\right)\right) \end{aligned} \tag{9}$$

where τ is a temperature hyper-parameter.

3) **Final Loss**. The final loss is a combination of the segmentation loss, the PSAL and CMCL as:

$$\mathcal{L}_{seg} = \mathbb{E}\left[G_{syn} \cdot \log\left(Y_{syn}\right)\right] \tag{10}$$

$$\mathcal{L} = \mathcal{L}_{seg} + \mathcal{L}_{PSAL} + \lambda \mathcal{L}_{CMCL} \tag{11}$$

4. Experiments

XCAD dataset. The X-ray angiography coronary artery disease (XCAD) dataset [24] is obtained during stent placement using a General Electric Innova IGS 520 system. Each image has a resolution of 512 × 512 pixels with one channel. The training set contains 1621 coronary angiograms without annotations as target images. The testing set contains 126 independent coronary angiograms with vessel segmentation maps annotated by experienced radiologists.

Retinal dataset. We also employ two public retinal datasets to validate the effectiveness of the proposed method. The DRIVE dataset [35] consists of 40 color retinal images of size 565 × 584 pixels. We use 20 images as target images and 20 remaining as test images. The STARE

dataset [17] contains 20 color retinal images of size 700 × 605 pixels with annotations as test images. There are 377 images without annotation which are used as target images.

CrackTree dataset. The CrackTree dataset [49] contains 206 800×600 pavement images with different kinds of cracks with curvilinear structures. The whole dataset is split into 160 target images and 46 test images by [33] setting. Following [33], we dilate the annotated centerlines by 4 pixels to form the ground-truth segmentation.

4.1. Evaluation Metrics

For XCAD and CrackTree, we follow [24, 33] to use the following widely-used metrics in our evaluation, i.e., Jaccard Index (Jaccard), Dice Coefficient (Dice), accuracy (Acc.), sensitivity (Sn.) and specificity (Sp.). For the DRIVE and STARE datasets, we follow the state-of-the-art works for retina vessel segmentation [24] to report accuracy (Acc.), sensitivity (Sn.) specificity (Sp.) and area under curve (AUC) in our evaluation.

4.2. Implementation Details

For FFS, we set the $ruleset$ as $(rule_1, \cdots, rule_4)$, such as:

$rule_1$:$F_{random} \rightarrow F_{random}\,[+F_{random} - F_{random}]$.
$rule_2$:$F_{random} \rightarrow F_{random}\,[-F_{random} - F_{random}]$.
$rule_3$:$F_{random} \rightarrow F_{random} - F_{random} - F_{random}$.
$rule_4$:$F_{random} \rightarrow F_{random} + F_{random} + F_{random}$.

We set the initial parameters of the Fractal system as follows. For all four datasets, the angle is randomly selected from $(20°, 120°)$, the initial length is randomly selected from (120px, 200px), the decreased parameter γ is randomly selected from (0.7, 1). The initial width w_{init} is ranging from (8px, 14px) for XCAD, DRIVE and STARE. For CrackTree, the w_{init} is selected from (2px, 6px). The kernel size of Gaussian blur for FFS is 13. We generate 150, 150, 600 synthetic fractal images X_{frac} for XCAD, CrackTree, and retinal datasets, respectively.

We apply data augmentation including horizontal flipping, random brightness and contrast changes ranging from 1.0 to 2.1, random saturation ranging from 0.5 to 1.5, and random rotation with 90°, 180°, and 270°. The standard deviation for Gaussian noise is set to a random value within (-5, +5). All images are cropped to 256×256 pixels for training. All the data-augmented operations are applied before the GIA module. The segmentation network is trained using the SGD with a momentum of 0.9 for optimization and the initial learning rate is 0.01. The discriminator network is trained using an Adam optimizer with an initial learning rate of 10^{-3}. We employ a batch size of 8 to train the network for 600 epochs. The number of negative queries q^+_{syn}, k^+_{target} and k^- per batch is taken up to 500, 500 and 1000. The amplitude map center region selection parameter of β is set to 0.3. the sampling ratio is 0.3, the hyper-parameter λ is 0.4 and the temperature hyper-parameter τ is 0.1.

	Methods	Jaccard	Dice	Acc.	Sn.	Sp.
Upper bound	U-Net [32]	0.571	0.724	0.981	0.868	0.996
Domain Adaptation	U-Net [32]	0.228	0.365	0.831	0.444	0.906
	MMD [8]	0.262	0.416	0.873	0.553	0.920
	YNet [31]	0.287	0.434	0.891	0.523	0.935
Traditional	Hessian [13]	0.307	0.465	0.948	0.406	0.981
	OOF [22]	0.241	0.386	0.899	0.566	0.920
Unsupervised	IIC [19]	0.124	0.178	0.738	0.487	0.754
	ReDO [9]	0.151	0.261	0.753	0.392	0.923
Self-supervised	SSVS [24]	0.389	0.557	0.945	0.583	0.972
	DARL [21]	0.471	0.636	**0.962**	0.597	**0.985**
	Ours	**0.499**	**0.661**	0.960	**0.687**	0.977

Table 1. Quantitative evaluation of FreeCOS compared with different methods on the XCAD dataset.

	Methods	DRIVE				STARE			
		Acc.	Sn.	Sp.	AUC	Acc.	Sn.	Sp.	AUC
Traditional	Hessian [13]	0.941	0.644	0.97	0.847	0.938	0.690	0.957	0.858
	OOF [22]	0.936	0.688	0.959	0.920	0.920	0.770	0.932	0.955
	Memari [25]	**0.961**	0.761	0.981	0.871	0.951	0.782	0.965	0.783
	Khan [20]	0.958	0.797	0.973	0.885	**0.996**	0.792	**0.998**	0.895
Unsupervised	IIC [19]	0.738	0.632	0.840	0.736	0.710	0.586	0.832	0.709
	ReDO [9]	0.761	0.593	0.927	0.760	0.756	0.567	0.899	0.733
Self-supervised	SSVS [24]	0.913	0.794	**0.982**	0.888	0.910	0.774	0.980	0.877
	DARL [21]	–	0.456	–	–	–	0.480	–	–
	Ours	0.921	**0.810**	0.932	**0.941**	0.952	**0.797**	0.964	**0.971**

Table 2. Quantitative evaluation of FreeCOS compared with different methods on the retinal dataset.

	Methods	AUC	Dice	Acc.	Sn.	Sp.
Traditional	Hessian [13]	0.780	0.122	0.935	0.310	0.945
	OOF [22]	0.482	0.031	0.770	0.244	0.778
Unsupervised	ReDO [9]	0.422	0.035	0.632	0.450	0.635
Self-supervised	SSVS [24]	0.477	0.078	0.299	0.042	0.912
	DARL [21]	0.888	0.395	**0.974**	0.542	**0.981**
	Ours	**0.920**	**0.525**	**0.974**	**0.576**	0.979

Table 3. Quantitative evaluation of FreeCOS compared with different methods on the CrackTree dataset.

4.3. Experimental Results

4.3.1 Comparison with State-of-the-art

Table. 1 compares the performance of vessel segmentation on XCAD between FreeCOS and the state-of-the-art methods, including the unsupervised methods [9, 19], the self-supervised methods [24, 21], domain adaptation methods [8, 31] and the traditional methods [13, 22]. The results in the 1st row are based on supervised U-Net as [24] (i.e., identical segmentation network trained using real images with manual labels) which are the upper bound.

For domain adaption methods, we pretrain a vessel segmentation model based on U-Net using training images of DRIVE. Then we adapt the pre-trained model to XCAD using MMD [8] and Ynet [31]. Even with supervised information in the annotated source domain, the performance of MMD and YNet is still inferior to ours. Specifically, our method achieves 21.2% improvement in Jaccard, 22.7% im-

	Jaccard	Dice	Acc.	Sn.	Sp.
FFS	0.450	0.615	0.955	0.664	0.974
FFS+PSAL	0.468	0.633	0.957	0.680	0.974
FFS+PSAL+IA	0.485	0.647	0.958	0.667	0.976
FFS+GIA	**0.499**	**0.661**	**0.960**	**0.687**	**0.977**

Table 4. Ablation study for modules.

	Jaccard	Dice	Acc.	Sn.	Sp.
FFS w/o Gaussian blur	0.383	0.547	0.940	0.654	0.959
FFS w/o FDA	0.302	0.459	0.920	0.609	0.940
FFS w/o various intensities	0.405	0.569	**0.957**	0.525	**0.984**
FFS w/o various angles	0.267	0.415	0.939	0.409	0.972
FFS w/o various lengths	0.224	0.358	0.943	0.300	0.983
FFS w/o various widths	0.224	0.354	0.935	0.348	0.972
FFS	**0.450**	**0.615**	0.955	**0.664**	0.974

Table 5. Ablation study for FFS.

Numbers of images	Jaccard	Dice	Acc.	Sn.	Sp.
Synthetic-15	0.464	0.629	0.958	0.630	0.979
Synthetic-45	0.478	0.642	0.958	0.674	0.976
Synthetic-75	0.488	0.652	**0.960**	0.668	**0.978**
Synthetic-150	**0.499**	**0.661**	**0.960**	**0.687**	0.977

Table 6. Ablation study for different numbers of synthetic images.

provement in Dice, 6.9% improvement in Acc, 16.4% improvement in Sn, and 4.2% improvement in Sp compared with YNet.

Compared with the unsupervised methods IIC [19] and ReDO [9], our method achieves significantly better performance for all metrics on XCAD. The results show that unsupervised methods cannot achieve satisfactory performance on the gray-scale X-ray images where the segmentation objects can be hardly distinguished from the background.

Self-supervised method SSVS [24] is specifically designed for XCAD, yet our method still achieves much better performance for all metrics, i.e., 11% improvement in Jaccard, 10.4% improvement in Dice and 10.4% improvement in Sn.

Table. 2 and 3 further compare our method with the existing methods on the retinal and crack datasets and a similar trend can be observed in these three datasets. Figure. 3 shows the visualization results of images from various kinds of curvilinear datasets.

4.3.2 Ablation Study

We first conduct ablation studies to evaluate the impact of different modules. To this end, we build the following variants based on our method. For all the variant models and our final model, we use the same U-Net model as our backbone.

Numbers of images	Jaccard	Dice	Acc.	Sn.	Sp.
Target-162	0.455	0.620	0.952	0.678	0.969
Target-486	0.476	0.637	0.955	0.683	0.972
Target-810	0.466	0.630	0.958	0.635	**0.978**
Target-1620	**0.499**	**0.661**	**0.960**	**0.687**	0.977

Table 7. Ablation study for different numbers of target images.

1) **FFS**. We utilize the FFS to generate synthetic training images and the corresponding labels to train the U-Net segmentation model. 2) **FFS+PSAL**. We further apply PSAL to align features of synthetic and real target images on top of FFS. 3) **FFS+PSAL+IA**. We apply image-level alignment via relative intensity order transformation (i.e., IA) on top of **FFS+PSAL**. 4) **FFS+GIA**. We apply our GIA module (including IA, PSAL and CMCL) on top of FFS. This model is also our final curvilinear segmentation model. The ablation studies are conducted on XCAD. The trend on other datasets are similar and thus the results on the other datasets are omitted due to the space limit.

The results in Table 4 show that 1) by training a U-Net model using synthetic images from FFS, we can already achieve better performance than the SOTA self-supervised method SSVS which is based on adversarial learning [24]. Such results reveal a very interesting phenomenon although adversarial learning can synthesize visually similar images as real target images, it cannot explicitly control the generated visual patterns and thus fail to enforce the segmentation-oriented patterns and properties to be encoded in synthetic images. In comparison, FFS can explicitly control the synthesis of both curvilinear structures and background patterns and hence can achieve better performance. 2) Aligning features via prediction space adaptation loss (PSAL) can help reduce domain shifts between synthetic images and real target images and thus **FFS+PSAL** achieves performance improvements compared with FFS. 3) Aligning synthetic images and real target images via IA method can also reduce the domain shifts and thus **FFS+PSAL+IA** can provide complementary improvements to **FFS+PSAL**. 4) Finally, the best performance is obtained when combining both FFS and GIA.

We further explore the importance of different parameters in FFS on the final performance and discuss how to generate images of curvilinear structures to encode sufficient and comprehensive visual cues for learning robust and distinctive features. To this end, we build 6 variant models based on FFS. 1) **FFS w/o Gaussian blur**. We do not perform Gaussian blur to synthetic images from FFS. 2) **FFS w/o FDA**. We remove FDA-based synthesis from FFS and only generate curvilinear structures without backgrounds. 3) **FFS w/o various intensities**. We remove intensity variations c_i in the Fractal system and set the intensity to a fixed value 60. 4) **FFS w/o various angles**. We utilize a small

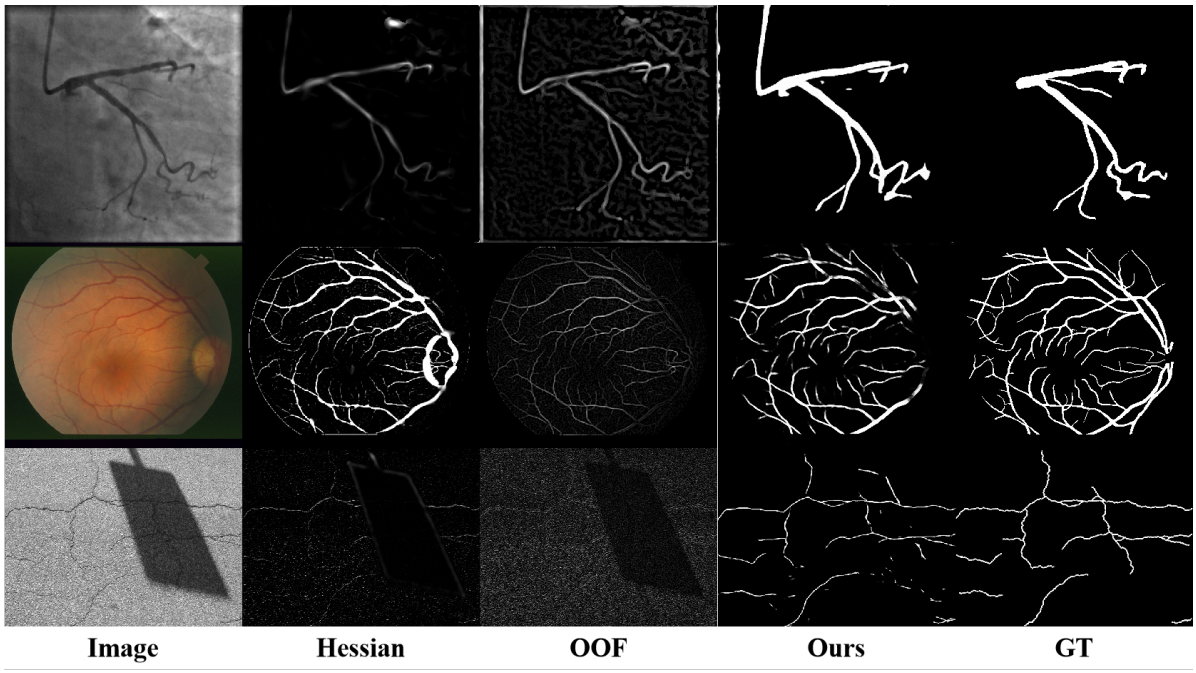

Figure 3. The visualization results of images from the various kinds of curvilinear datasets.

branching angle range from 1° to 5° to reduce the branching angle variations. 5) **FFS w/o various lengths**. We reduce the length variation range to (10px, 20px). 6) **FFS w/o various widths**. We reduce the width variation to (1px, 5px).

The results in Table 5 show that 1) Gaussian blur can smooth the transition region between curvilinear objects and backgrounds, better mimicking the target images and providing more challenging samples for training than those with sharp boundaries. As a result, excluding Gaussian blur decreases the performance of FFS by 6.8% in Dice. 2) FDA could provide background patterns of target images which are essential for learning features of negative samples. Thus, **FFS w/o FDA** is 15.6% worse than FFS in Dice. 3) Among the four parameters of the Fractal system, i.e., intensity, angle, length and width, width plays the most important role in the final performance, i.e., **FFS w/o various widths** decreases the performance by 26.1% compared with FFS in Dice. To summarize, we identify that the appearance patterns in the curvilinear object, object-background transition regions and background regions are important for self-supervised segmentation. Meanwhile, proper parameter settings which can provide similar geometric characteristics with the real target images are key to the success of our self-supervised method.

4.3.3 Self-supervised training with more data

We also examine the segmentation performance when varying the number of synthetic images on the XCAD dataset. Results in Tables. 6 and 7 provide the results when using different numbers of synthetic structures and real target images for generating synthetic images for self-supervised training respectively. Synthetic-15 $\sim$ Synthetic-150 denote using the number of synthetic fractal object images X_{frac} from 15 to 150, and Target-162 $\sim$ Target-1620 denote using the number of target images X_t from 162 to 1620. Increasing both synthetic structures and real target images can accordingly improve the final segmentation performance.

5. Conclusion

In this paper, we propose a novel self-supervised curvilinear object segmentation method that learns robust and distinctive features from fractals and unlabeled images (FreeCOS). Different from existing methods, FreeCOS applies the proposed FFS and GIA approach to effectively guide learning distinctive features to distinguish curvilinear objects and backgrounds and aligns information of synthetic and target images at both image and feature levels. One limitation of FreeCOS is that may generate false positives in other curvilinear objects (e.g., catheters in XCAD) and requires proper selection for width range. However, we have successfully utilized this self-supervised learning method for coronary vessel segmentation, retinal vessel segmentation and crack segmentation by reasonable parameter selection. To the best of our knowledge, FreeOCS is the first self-supervised learning method for various curvilinear object segmentation applications.

References

[1] Rameen Abdal, Peihao Zhu, Niloy J Mitra, and Peter Wonka. Labels4free: Unsupervised segmentation using stylegan. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13970–13979, 2021. 1, 3

[2] Michael D Abràmoff, Mona K Garvin, and Milan Sonka. Retinal imaging and image analysis. *IEEE Reviews in Biomedical Engineering*, 3:169–208, 2010. 1

[3] Inigo Alonso, Alberto Sabater, David Ferstl, Luis Montesano, and Ana C Murillo. Semi-supervised semantic segmentation with pixel-level contrastive learning from a class-wise memory bank. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8219–8228, 2021. 3

[4] Shekoofeh Azizi, Basil Mustafa, Fiona Ryan, Zachary Beaver, Jan Freyberg, Jonathan Deaton, Aaron Loh, Alan Karthikesalingam, Simon Kornblith, Ting Chen, et al. Big self-supervised models advance medical image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3478–3488, 2021. 3

[5] Amir Bar, Xin Wang, Vadim Kantorov, Colorado J Reed, Roei Herzig, Gal Chechik, Anna Rohrbach, Trevor Darrell, and Amir Globerson. Detreg: Unsupervised pretraining with region priors for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14605–14615, 2022. 3

[6] Favyen Bastani, Songtao He, Sofiane Abbar, Mohammad Alizadeh, Hari Balakrishnan, Sanjay Chawla, Sam Madden, and David DeWitt. Roadtracer: Automatic extraction of road networks from aerial images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4720–4728, 2018. 1

[7] Anil Batra, Suriya Singh, Guan Pang, Saikat Basu, CV Jawahar, and Manohar Paluri. Improved road connectivity by joint learning of orientation and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10385–10393, 2019. 1

[8] Róger Bermúdez-Chacón, Pablo Márquez-Neila, Mathieu Salzmann, and Pascal Fua. A domain-adaptive two-stream u-net for electron microscopy image segmentation. In *IEEE International Symposium on Biomedical Imaging*, pages 400–404. IEEE, 2018. 1, 6

[9] Mickaël Chen, Thierry Artières, and Ludovic Denoyer. Unsupervised object segmentation by redrawing. *Advances in Neural Information Processing systems*, 32, 2019. 1, 2, 3, 6, 7

[10] Mingfei Cheng, Kaili Zhao, Xuhong Guo, Yajing Xu, and Jun Guo. Joint topology-preserving and feature-refinement network for curvilinear structure segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7147–7156, 2021. 1

[11] Kien Do, Truyen Tran, and Svetha Venkatesh. Clustering by maximizing mutual information across views. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9928–9938, 2021. 1, 2

[12] Carl Doersch, Abhinav Gupta, and Alexei A Efros. Unsupervised visual representation learning by context prediction. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1422–1430, 2015. 3

[13] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 130–137. Springer, 1998. 2, 6

[14] Muhammad Moazam Fraz, Paolo Remagnino, Andreas Hoppe, Bunyarit Uyyanonvara, Alicja R Rudnicka, Christopher G Owen, and Sarah A Barman. Blood vessel segmentation methodologies in retinal images–a survey. *Computer Methods and Programs in Biomedicine*, 108(1):407–433, 2012. 1

[15] Matteo Frigo and Steven G Johnson. Fftw: An adaptive software architecture for the fft. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 1381–1384. IEEE, 1998. 4

[16] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. In *International Conference on Learning Representations*, 2018. 3

[17] AD Hoover, Valentina Kouznetsova, and Michael Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical imaging*, 19(3):203–210, 2000. 1, 2, 6

[18] Xiaoling Hu, Fuxin Li, Dimitris Samaras, and Chao Chen. Topology-preserving deep image segmentation. *Advances in Neural Information Processing systems*, 32, 2019. 1

[19] Xu Ji, Joao F Henriques, and Andrea Vedaldi. Invariant information clustering for unsupervised image classification and segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9865–9874, 2019. 1, 2, 3, 6, 7

[20] Khan Bahadar Khan, Muhammad Shahbaz Siddique, Muhammad Ahmad, and Manuel Mazzara. A hybrid unsupervised approach for retinal vessel segmentation. *BioMed Research International*, 2020, 2020. 2, 6

[21] Boah Kim, Yujin Oh, and Jong Chul Ye. Diffusion adversarial representation learning for self-supervised vessel segmentation. *arXiv preprint arXiv:2209.14566*, 2022. 2, 3, 6

[22] Max WK Law and Albert Chung. Three dimensional curvilinear structure detection using optimally oriented flux. In *European Conference on Computer Vision*, pages 368–382. Springer, 2008. 2, 6

[23] Yunfan Li, Peng Hu, Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng. Contrastive clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 8547–8555, 2021. 1, 2

[24] Yuxin Ma, Yang Hua, Hanming Deng, Tao Song, Hao Wang, Zhengui Xue, Heng Cao, Ruhui Ma, and Haibing Guan. Self-supervised vessel segmentation via adversarial learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7536–7545, 2021. 1, 2, 3, 5, 6, 7

[25] Nogol Memari, Abd Rahman Ramli, M Saripan, Syamsiah Mashohor, and Mehrdad Moghbel. Retinal blood vessel segmentation by using matched filtering and fuzzy c-means clustering with integrated level set method for dia-

betic retinopathy assessment. *Journal of Medical and Biological Engineering*, 39(5):713–731, 2019. 2, 6

[26] Ishan Misra and Laurens van der Maaten. Self-supervised learning of pretext-invariant representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6707–6717, 2020. 3

[27] Agata Mosinska, Mateusz Koziński, and Pascal Fua. Joint segmentation and path classification of curvilinear structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(6):1515–1521, 2019. 1

[28] Agata Mosinska, Pablo Marquez-Neila, Mateusz Koziński, and Pascal Fua. Beyond the pixel-wise loss for topology-aware delineation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3136–3145, 2018. 1

[29] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *European Conference on Computer Vision*, pages 319–345. Springer, 2020. 3

[30] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2536–2544, 2016. 3

[31] Joris Roels, Julian Hennies, Yvan Saeys, Wilfried Philips, and Anna Kreshuk. Domain adaptive segmentation in volume electron microscopy imaging. In *IEEE International Symposium on Biomedical Imaging*, pages 1519–1522. IEEE, 2019. 1, 6

[32] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015. 5, 6

[33] Tianyi Shi, Nicolas Boutry, Yongchao Xu, and Thierry Géraud. Local intensity order transformation for robust curvilinear object segmentation. *IEEE Transactions on Image Processing*, 31:2557–2569, 2022. 1, 5, 6

[34] Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien PW Pluim, Ulrich Bauer, and Bjoern H Menze. cldice-a novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16560–16569, 2021. 1

[35] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 2004. 1, 2, 5

[36] Joseph L Thomas, Simon Winther, Robert F Wilson, and Morten Bøttcher. A novel approach to diagnosing coronary artery disease: acoustic detection of coronary turbulence. *The International Journal of Cardiovascular Imaging*, 33(1):129–136, 2017. 1

[37] Feigege Wang, Yue Gu, Wenxi Liu, Yuanlong Yu, Shengfeng He, and Jia Pan. Context-aware spatio-recurrent curvilinear structure segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12648–12657, 2019. 1

[38] Jierong Wang and Albert Chung. Higher-order flux with spherical harmonics transform for vascular analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 55–65. Springer, 2020. 2

[39] Xuehui Wang, Kai Zhao, Ruixin Zhang, Shouhong Ding, Yan Wang, and Wei Shen. Contrastmask: Contrastive learning to segment every thing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11604–11613, 2022. 3, 5

[40] Yan Wang, Xu Wei, Fengze Liu, Jieneng Chen, Yuyin Zhou, Wei Shen, Elliot K Fishman, and Alan L Yuille. Deep distance transform for tubular structure segmentation in ct scans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3833–3842, 2020. 1

[41] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021. 3

[42] Enze Xie, Jian Ding, Wenhai Wang, Xiaohang Zhan, Hang Xu, Peize Sun, Zhenguo Li, and Ping Luo. Detco: Unsupervised contrastive learning for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8392–8401, 2021. 3

[43] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4085–4095, 2020. 2, 4

[44] M Zamir. On fractal properties of arterial trees. *Journal of Theoretical Biology*, 197(4):517–526, 1999. 1

[45] Mair Zamir. Arterial branching within the confines of fractal l-system formalism. *The Journal of General Physiology*, 118(3):267–276, 2001. 1, 2, 4

[46] M Zamir. Fractal dimensions and multifractility in vascular branching. *Journal of Theoretical Biology*, 212(2):183–190, 2001. 1

[47] Yuanyi Zhong, Bodi Yuan, Hong Wu, Zhiqiang Yuan, Jian Peng, and Yu-Xiong Wang. Pixel contrastive-consistent semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7273–7282, 2021. 3

[48] Wei Zhou, Yukang Wang, Jiajia Chu, Jiehua Yang, Xiang Bai, and Yongchao Xu. Affinity space adaptation for semantic segmentation across domains. *IEEE Transactions on Image Processing*, 30:2549–2561, 2020. 5

[49] Qin Zou, Yu Cao, Qingquan Li, Qingzhou Mao, and Song Wang. Cracktree: Automatic crack detection from pavement images. *Pattern Recognition Letters*, 33(3):227–238, 2012. 1, 2, 6