Medical Image Analysis 91 (2024) 103022

Contents lists available at ScienceDirect

Medical Image Analysis

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

C-DARL: Contrastive diffusion adversarial representation learning for label-free blood vessel segmentation

Boah Kim [a,1], Yujin Oh [b,1], Bradford J. Wood [a], Ronald M. Summers [a,*], Jong Chul Ye [b,*]

[a] Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, MD, USA
[b] Kim Jaechul Graduate School of AI, Korea Advanced Institute of Science & Technology (KAIST), Daejeon, Republic of Korea

ARTICLE INFO

Dataset link: https://github.com/boahK/MEDIA_CDARL

MSC:
68U10
68T45
92C55

Keywords:
Image segmentation
Vascular structures
Diffusion model
Self-supervised learning

ABSTRACT

Blood vessel segmentation in medical imaging is one of the essential steps for vascular disease diagnosis and interventional planning in a broad spectrum of clinical scenarios in image-based medicine and interventional medicine. Unfortunately, manual annotation of the vessel masks is challenging and resource-intensive due to subtle branches and complex structures. To overcome this issue, this paper presents a self-supervised vessel segmentation method, dubbed the contrastive diffusion adversarial representation learning (C-DARL) model. Our model is composed of a diffusion module and a generation module that learns the distribution of multi-domain blood vessel data by generating synthetic vessel images from diffusion latent. Moreover, we employ contrastive learning through a mask-based contrastive loss so that the model can learn more realistic vessel representations. To validate the efficacy, C-DARL is trained using various vessel datasets, including coronary angiograms, abdominal digital subtraction angiograms, and retinal imaging. Experimental results confirm that our model achieves performance improvement over baseline methods with noise robustness, suggesting the effectiveness of C-DARL for vessel segmentation.Our source code is available at https://github.com/boahK/MEDIA_CDARL.[2]

1. Introduction

Angiography is an invasive or non-invasive exam to visualize blood vessels towards diagnosis or treatment of a wide variety of diseases that impact vascular structures, or where vascular maps provide the roadmap to delivery of therapeutics. For example, to plan therapy and accurately deliver drugs and devices in minimally invasive image-guided therapies, identification, characterization, and quantification of the blood vessels and their branches are foundational elements towards blood flow, endothelial pathology, landmarks, reference points, and roadmaps towards tumors or target anatomy (Dehkordi et al., 2011). As manual annotation of vessel masks is time-consuming due to tiny and low-contrast vessel branches (see Fig. 1), automatic vessel segmentation methods have been extensively studied to enhance efficiencies and to facilitate large data for training (Delibasis et al., 2010; Jiang et al., 2019; Wu et al., 2019).

Classical rule-based vessel segmentation methods utilize various features of vessel images such as geometric models, ordered region growing, and vessel intensity distributions (Meijering et al., 2004; Lesage et al., 2009; Taghizadeh Dehkordi et al., 2014; Zhao et al., 2019). However, these approaches require complicated preprocessing and may need user interaction in inference, particularly for image-specific parameters optimization, posing resource barriers and challenges to practical clinical deployment. On the other hand, recent learning-based techniques (Nasr-Esfahani et al., 2016; Fan et al., 2018; Wu et al., 2019), which segment blood vessels through neural networks, can generate outputs in real-time, but they require the supervision of large amounts of annotated data.

Recently, self-supervised learning methods, which do not require ground-truth vessel masks when training networks, have been extensively studied. For example, Ma et al. (2021) presents a fractal synthetic module and an adversarial vessel segmentation method, in which the fractal module generates fractal masks that look similar to vessels by drawing distorted rectangles with various thicknesses.

In our previous work (Kim et al., 2023), we propose a diffusion adversarial representation learning model (DARL) that combines the diffusion model and the adversarial model. Specifically, the DARL model learns the distribution of background images using the diffusion denoising probabilistic model (DDPM) (Ho et al., 2020) so that the

* Corresponding author.
E-mail addresses: rms@nih.gov (R.M. Summers), jong.ye@kaist.ac.kr (J.C. Ye).
[1] Co-first authors.
[2] This paper extends the work (Kim et al., 2023) presented at the Eleventh International Conference on Learning Representations (ICLR) 2023.

https://doi.org/10.1016/j.media.2023.103022
Received 31 July 2023; Received in revised form 6 October 2023; Accepted 31 October 2023
Available online 11 November 2023
1361-8415/

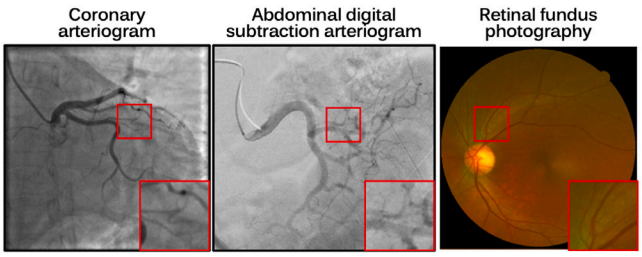


Fig. 1. Examples of various blood vessel image domains. Red boxes show magnified vessel structures.

vessel structures can be easily discerned in the latent. Accordingly, a subsequent generation module can extract foreground vessel regions. Since the DARL model performs the generation and segmentation using a single generator, the DARL alleviates the network training complexity compared to Ma et al. (2021) as shown in Appendix C.8 of Kim et al. (2023). While the DARL method provides a high-quality vessel segmentation map through single-step inference, one of the main limitations of DARL is that it uses both the pre-contrast background and angiography images for network training. This may limit its use in various clinical applications that do not normally provide similar background images (eg., retinal fundus images). Future training with digital subtraction and raw images might reduce this training variability.

This paper is to extend and overcome the aforementioned weaknesses in our DARL (Kim et al., 2023). We present a label-free vessel segmentation method that can utilize a variety of blood vessel images in training the model, enabling its generalizability as a vessel segmentator in a variety of clinical applications. Specifically, we design a model that basically follows the overall structure of DARL consisting of the diffusion and generation modules, with customizations. One of the key improvements over DARL comes from our observation that we can still obtain a semantically meaningful segmentation map by omitting the background image input path.

In addition, to effectively learn vessel representation, we employ contrastive learning, namely contrastive-DARL (C-DARL), for further improvement (Wu et al., 2018; Chen et al., 2020; He et al., 2020). While DARL is trained to generate vessel masks via adversarial learning using the fractal masks, the input fractal masks and the real blood vessels have intrinsically different shapes and sizes. Accordingly, the contrastive loss function is designed to dissociate the estimated masks of real vessel images and the fractal masks while maximizing the similarity between the estimated masks of fractal-based synthetic vessel images and the fractal. This is achieved by leveraging contrastive unpaired translation (CUT) (Park et al., 2020) that computes the similarity of source and target in a patch-wise manner.

Thanks to this simplification of the data preparation without requiring background images, our framework can be trained using multiple domains of two-dimensional (2D) blood vessel representations, such as X-ray angiography or retinal imaging. Experimental results show that our C-DARL model outperforms the comparative methods in vessel segmentation of various datasets. Also, when comparing our model to the DARL, our method achieves consistent improvement both on internal and external test data with respect to the training data. As the C-DARL provides vessel segmentation maps in real-time (0.176 s per frame), this holds great promise as a platform in clinical practices, upon validation of the integrity of the resulting representations. In summary, the contributions of this paper are as follows:

- We introduce a label-free vessel segmentation method that can leverage multi-domain blood vessel images without requiring background images and provide vessel masks for those various datasets.

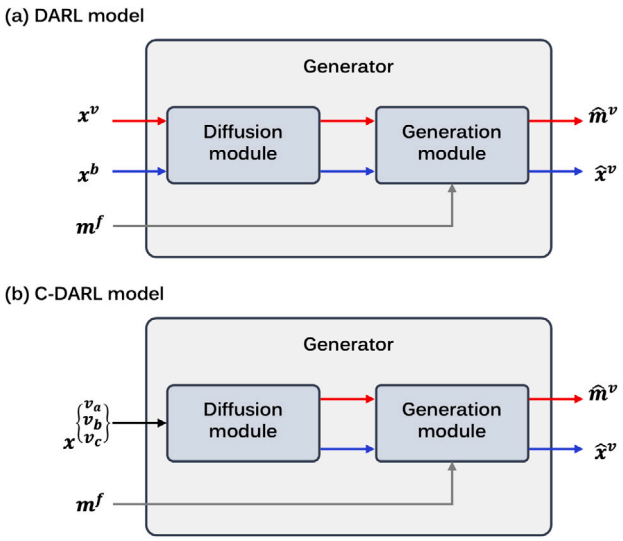


Fig. 2. Comparison of the training pipelines between (a) the DARL model proposed in our previous work (Kim et al., 2023) and the C-DARL model proposed in this paper. x^v and x^b are a real vessel image and a background image, respectively, and m^f denotes a fractal mask. $\hat{m}^v$ is an estimated vessel segmentation mask, and $\hat{x}^v$ is a synthetic vessel image. For the C-DARL, the real vessel image in one of the various domains (e.g. $\{v_a, v_b, v_c\}$) can be fed into the model.

- In contrast to the DARL, our proposed model applies contrastive learning in generating vessel segmentation maps, allowing the network to intensively learn vessel representations.
- Extensive experimental results demonstrate that the proposed C-DARL is robust across diverse blood vessel data in a variety of clinical applications, and has superior performance with more efficiency than the existing self-supervised methods.

2. Related works

2.1. Diffusion model

The DDPM (Ho et al., 2020) generates images by converting the Gaussian noise distribution into the data distribution through the Markov chain process. Specifically, for the forward diffusion, the clean data x_0 is corrupted with the noise level at t by the following distribution.:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I), \tag{1}$$

where $\mathcal{N}(a, b^2)$ denotes the normal distribution with a mean $a \in \mathbb{R}$ and a variance b^2, I is the identity matrix, and β_t is a scalar variance in the range of $[0, 1]$. Then, through Markov chain, a noisy image x_t for the data x_0 can be computed by:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_0, (1-\alpha_t)I), \tag{2}$$

where $\alpha_t = \Pi_{s=1}^{t}(1-\beta_s)$. For this corrupted image, the DDPM learns the reverse diffusion:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \boldsymbol{\mu}_\theta(x_t, t), \sigma_t^2 I), \tag{3}$$

where σ_t is a scalar variance and $\boldsymbol{\mu}_\theta$ is a learnt mean computed by the network G_ϵ:

$$\boldsymbol{\mu}_\theta(x_t, t) = \frac{1}{\sqrt{1-\beta_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\alpha_t}}G_\epsilon(x_t, t)\right). \tag{4}$$

Accordingly, one can obtain images from the Gaussian noise using the DDPM through the reverse diffusion process as follows:

$$x_{t-1} = \boldsymbol{\mu}_\theta(x_t, t) + \sigma_t z, \tag{5}$$

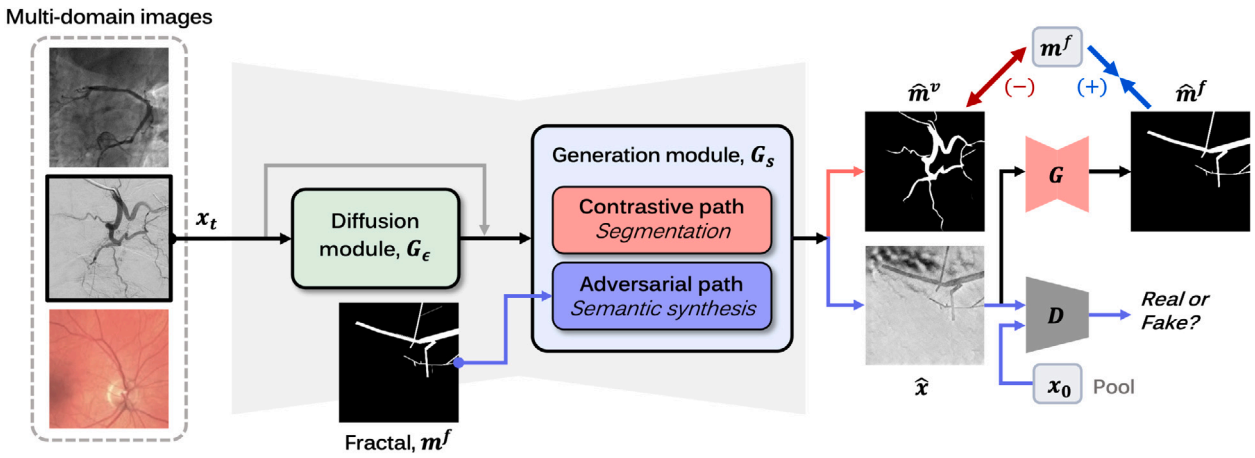


Fig. 3. Overall training framework of the proposed contrastive diffusion adversarial representation learning model (C-DARL).

where $z \sim \mathcal{N}(0, I)$.

This DDPM has been successfully adapted to various computer vision tasks including semantic image synthesis (Wang et al., 2022; Huang et al., 2023). Moreover, the potentials of learned representations from DDPM have been revealed through semantic segmentation, which effectively captures semantic features to improve segmentation performance (Baranchuk et al., 2022; Brempong et al., 2022; Rahman et al., 2023).

2.2. Self-supervised vessel segmentation

Semantic segmentation problems have been traditionally addressed using supervised learning. Unfortunately, the performance of supervised learning methods (Fan et al., 2019; Yang et al., 2019) largely depends on the huge amount of labels, which requires time-demanding manual segmentation performed by medical experts.

Recently, self-supervised learning (SSL) methods have been actively investigated to mitigate this issue. Mahmood et al. (2019) presents a deep adversarial (DA) training model to segment cell geometries in histopathology images without ground-truth labels. Also, Caron et al. (2021) and Oquab et al. (2023) propose large-scale models pre-trained on natural images, which extract meaningful information from their attention map, and have been used for medical image analysis (Ye et al., 2022; Yeganeh et al., 2023). Nonetheless, a naive application of these SSL methods designed for natural images or non-vessel-like images to medical blood vessel images is challenging to accurately extract blood vessel structures, which contain tiny branches within highly interfering background signals. To address this, SSL methods tailored for the vessel segmentation task have been recently developed (Ma et al., 2021; Kim et al., 2023), which learn vessel semantic information using synthetic fractal masks as pseudo vessel labels.

Previously, we propose a diffusion-based adversarial vessel segmentation method (DARL) (Kim et al., 2023) that learns the background signal using the diffusion module, which effectively improves vessel segmentation performance with noise robustness. Specifically, the DARL model estimates vessel segmentation maps through the guidance of semantic image synthesis that incorporates the given pre-contrast background image and fractal vessel masks, as shown in Fig. 2(a). This is motivated by Ma et al. (2021) which synthesizes vessel images by adding the fractal masks to the background images and lets the network estimate vessel masks using the information of fractal-based synthetic vessel images. However, the DARL model has limitations in that (1) it still utilizes the background images as input, and (2) the vessel segmentation is learned through the adversarial loss by regarding the fractal masks as real and the network output as fake even though the fractals are different from the ground-truth vessel masks.

3. Method

3.1. Motivation

To deal with the aforementioned issues, firstly, we propose a model that eliminates the need for background images, alleviating the constraint of using the angiography dataset, as shown in Fig. 2(b). This is based on the empirical observation that the diffusion module in DARL is effective in extracting the sparsity of blood vessel structure by intensively learning the background structure but also in estimating the noise that captures the information of the given data distribution and enables diverse image synthesis. Here, the diffusion model can estimate the noise by nulling out the learned image distribution regardless of the existence of vessels in the training data. Accordingly, as long as the vessel structures are sparsely distributed, the vessels can be regarded as outliers and represented in the diffusion module output when generating vessel masks. Therefore, the diffusion module can be trained using vessel images in various domains where vessel-free backgrounds are difficult to obtain.

Moreover, to further refine the segmentation accuracy, in the vessel segmentation path, we replace an adversarial loss of DARL with a contrastive loss (Chen et al., 2020; Zhong et al., 2021; Hu et al., 2021; Oh et al., 2022). Specifically, by reflecting the fact that the fractal masks and real vessel masks have different features, we present a mask-based contrastive loss that utilizes the fractals and the estimated vessel masks as negative pairs, while using the fractals and the cyclically estimated segmentation masks as positive pairs. In particular, inspired by contrastive unsupervised translation (CUT) (Park et al., 2020), which maximizes the mutual information of the patches at the same locations in the source and the target images, we employ the CUT-loss to maximize the structural patch similarity between the fractal mask (as a query) and the cyclically synthesized vessel mask (as a positive), and to disassociate the query signals from the segmented mask of the real vessel image (as a negative), allowing the model to learn vessel structure more effectively with no use of any labeled dataset.

3.2. Overall architecture

The overall learning flow of C-DARL is illustrated in Fig. 3. The C-DARL model has a generator G consisting of a diffusion module G_ϵ and a generation module G_s. When the diffusion module based on the DDPM takes an input image, the module estimates a latent feature by learning data distribution for various noisy levels, in which the latent feature is the output of the last layer of the module. Then, the generation module generates a vessel mask $\hat{m}^v$ and a synthetic

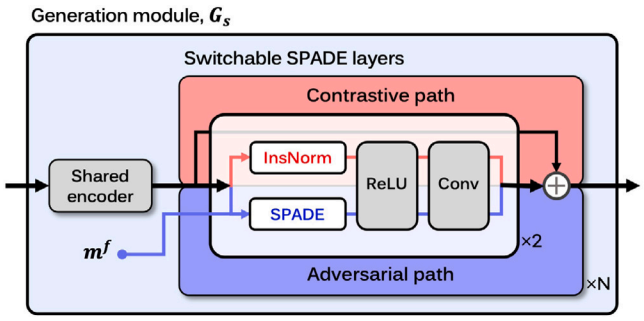


Fig. 4. Detailed architecture of the generation module G_s that has a shared encoder and N residual blocks composed of switchable SPADE layers, ReLU, and convolution (Conv) layers. If a semantic layout m^f is given to the generation module, the SPADE layer is activated. Otherwise, the instance normalization (InsNorm) is activated.

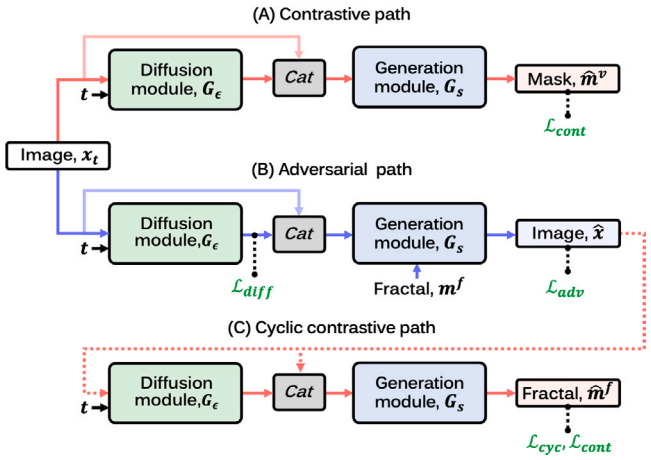


Fig. 5. Diagram of loss formulation to train our proposed C-DARL. t is the time of the noisy level used in the perturbed image x_t and *Cat* denotes the channel-wise concatenation of two inputs. $\mathcal{L}_{diff}$, $\mathcal{L}_{cont}$, $\mathcal{L}_{adv}$, and $\mathcal{L}_{cyc}$ are the diffusion loss, the mask-based contrastive loss, the adversarial loss, and the cycle loss, respectively.

vessel image $\hat{x}$. Here, the vessel image is generated using the latent of the diffusion module and a fractal mask m^f, where the fractal mask is synthesized by the fractal synthetic module proposed in Ma et al. (2021).

Then, in the path of vessel segmentation, contrastive learning is applied by using the fractal mask m^f and the estimated vessel masks $(\hat{m}^v, \hat{m}^f)$, where $\hat{m}^f$ is generated by feeding the synthetic vessel image $\hat{x}$ into the generator through the cycle pathway. On the other hand, in the path of semantic image synthesis, the synthetic vessel image $\hat{x}$ and real vessel image x are fed into a discriminator D for adversarial learning. As described in Fig. 3, we denote the segmentation path as a *contrastive* path and the image synthesis path as an *adversarial* path. In the following, we describe the proposed method in detail.

3.2.1. Model input

Let $\mathbf{X} = \bigcup_{k=1}^{k=K} X^k$ be a group of given blood vessel image datasets with K different domains where X^k has one or more images of the kth domain, i.e. $X^k = \{x^{k_1}, x^{k_2}, \dots, x^{k_N}\}$ with $k_N \geq 1$. For each iteration, our model randomly samples one of the multi-domain images $x_0^{k_n}$ among the given datasets. Then, for the model input, the image is corrupted by the forward diffusion (2):

$$x_t = \sqrt{\alpha_t} x_0^{k_n} + \sqrt{1-\alpha_t}\epsilon, \tag{6}$$

where t is the noisy level in range of $[0, T]$ and $\epsilon \sim \mathcal{N}(0, I)$. Using this perturbed image, the diffusion module G_ϵ learns the vessel image distribution and provides latent information to the generation module G_s.

3.2.2. Switchable SPADE layers

When the diffusion module estimates the latent feature, the feature map is concatenated with the perturbed image x_t in channel dimension. Then, the generation module generates vessel segmentation masks and synthetic vessel images simultaneously. These simultaneous tasks can be performed by the switchable SPADE (S-SPADE) layers proposed in the DARL model. Specifically, as shown in Fig. 4, the generation module consists of N residual blocks with S-SPADE layers that perform different normalization operations depending on whether or not the semantic layout is given to the model. For the feature maps e from a shared encoder in the generation module, the residual blocks synthesize semantic images through the spatially adaptive normalization (SPADE) (Park et al., 2019) if the semantic mask m is given, whereas they generate segmentation masks through the instance normalization (InsNorm) (Ulyanov et al., 2016) otherwise:

$$e = \begin{cases} \text{SPADE}(e, m), & \text{if semantic mask } m \text{ is given,} \\ \text{InsNorm}(e), & \text{otherwise.} \end{cases} \tag{7}$$

Thus, the contrastive path activates the InsNorm and estimates the vessel masks, and at the same time, the adversarial path, which takes the fractal masks m^f, activates the SPADE layer and generates the synthetic images. Also, by sharing all network parameters except for the S-SPADE layers in the two paths of the generation module, our model can synergistically learn vessel representation through semantic image synthesis. This helps the model to estimate real vessel masks that become a semantically meaningful negative pair with respect to the synthetic fractal masks.

Therefore, although the contrastive path does not need to learn the reverse process from the pure Gaussian noise in that this path estimates the vessel masks of the input image, the corrupted input image is given also to the contrastive path to share the same generation module with SPADE layer. Instead, by setting the maximum noisy level T to be smaller than that for the adversarial path, we design the model to estimate vessel masks for noisy input images.

3.3. Loss formulation

Fig. 5 shows our loss function with training flow. The proposed C-DARL model has three paths in the training phase: (A) the contrastive path estimates the vessel segmentation mask $\hat{m}^v$ that is used to compute a mask-based contrastive loss $\mathcal{L}_{cont}$, (B) the adversarial path learns data distribution through a diffusion loss $\mathcal{L}_{diff}$ and generates the vessel image $\hat{x}$ based on the fractal layout m^f via an adversarial loss $\mathcal{L}_{adv}$, and (C) the cyclic contrastive path feeds the fractal-based synthetic image into the model and segments the fractal region $\hat{m}^f$ that is utilized in a cycle loss $\mathcal{L}_{cyc}$ and the contrastive loss $\mathcal{L}_{cont}$. A detailed description of each loss function is as follows.

3.3.1. Diffusion loss

The diffusion loss aims to learn input data distribution through the diffusion module G_ϵ, yielding the latent feature including meaningful information about the input. For a given vessel image x_0, a noise $\epsilon \sim \mathcal{N}(0, I)$, and a time step t uniformly sampled from $[0, T]$, by following the DDPM training scheme (Ho et al., 2020), the loss can be represented as:

$$\mathcal{L}_{diff} = \mathbb{E}_{x_0,\epsilon,t}\left[\|G_\epsilon(\sqrt{\alpha_t}x_0 + \sqrt{(1-\alpha_t)}\epsilon, t) - \epsilon\|^2\right]. \tag{8}$$

As aforementioned, since the input image is perturbed within the full range of noisy levels in the adversarial path compared to the contrastive path, the diffusion loss is calculated only in the adversarial path.

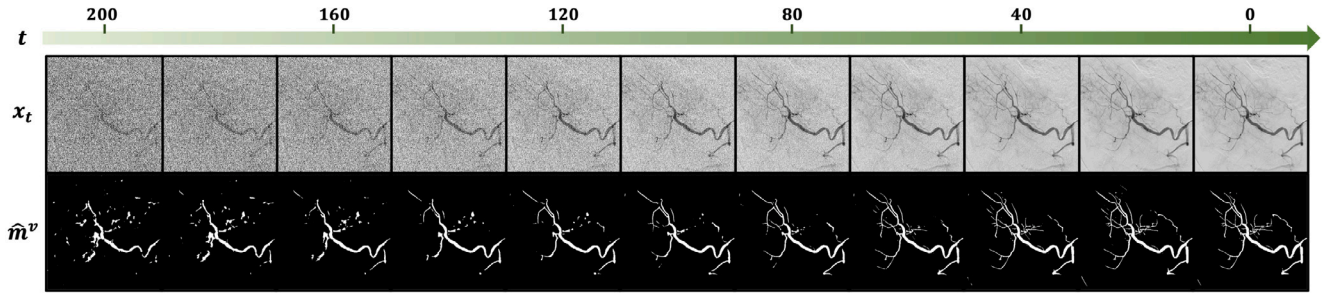


Fig. 6. Inference of the blood vessel segmentation according to the noisy level t. For given perturbed input images x_t, the proposed C-DARL model generates vessel masks $\hat{m}^v$ in a single step. We show an example of the vessel segmentation of the hepatic angiogram.

3.3.2. Mask-based contrastive loss

To address the limitation of not having real vessel masks in our label-free training scheme, we employ a patch-based contrastive learning objective (Park et al., 2020) that maximizes the mutual information of corresponding patches between two images, which utilizes a noise contrastive estimation method (Oord et al., 2018). Here, instead of using the image features from the network encoder, we design a *mask-based* contrastive loss that leverages the segmentation masks.

Specifically, based on the observation that the real blood vessel and the fractal masks have different features of shapes and sizes, for the fractal m^f as a query, we refer to the estimated vessel mask $\hat{m}^v$ in the contrastive path as negatives (See Fig. 3). In contrast, since the cyclic contrastive path estimates the fractal embedded in the synthetic vessel image, we regard the segmented fractal mask $\hat{m}^f$ as a positive of the query. Then, the model is trained for corresponding patches of m^f and $\hat{m}^f$ to be more strongly associated than those of m^f and $\hat{m}^v$ using our contrastive loss.

More specifically, for each segmentation mask $m \in \mathbb{R}^{1\times HW}$, we obtain R different-sized tensors, in which each tensor is obtained by folding the mask as $m_r \in \mathbb{R}^{p_r{}^2\times \frac{H}{p_r}\frac{W}{p_r}}$ where p_r is a division factor for $r \in \{1,2,\dots,R\}$. Then, the tensor is fed into a light-weight network H_r composed of two fully-connected layers, generating a stack of the features $\{h_r(m)\}_R = \{H_1(m_1), H_2(m_2), \dots, H_R(m_R)\}$ where $h_r(m) \in \mathbb{R}^{C_r\times \frac{H}{p_r}\frac{W}{p_r}}$ is the feature with C_r channels for the rth tensor. Through this process, we can get the feature stacks of $\{h_r(m^f)\}_R$, $\{h_r(\hat{m}^f)\}_R$, and $\{h_r(\hat{m}^v)\}_R$ for the masks of m^f, $\hat{m}^f$, and $\hat{m}^f$, respectively. By randomly selecting Q_r spatial locations in the range of $[0, \frac{H}{p_r}\frac{W}{p_r}]$, the contrastive loss is computed by:

$$\mathcal{L}_{cont} = \mathbb{E}_{m^f,\hat{m}^f,\hat{m}^v} \sum_{r=1}^{R}\sum_{q=1}^{Q_r} \ell_{MI}\left(h_r^q(m^f), h_r^q(\hat{m}^f), h_r^q(\hat{m}^v)\right), \tag{9}$$

where ℓ_{MI} is the mutual information using the cross-entropy loss as:

$$\ell_{MI}(u,u^+,u^-) = -log\left[\frac{\exp(u\cdot u^+/\tau)}{\exp(u\cdot u^+/\tau)+\sum_i \exp(u\cdot u_i^-/\tau)}\right], \tag{10}$$

where τ is a temperature scaling the distances between the query and other positives/negatives. This calculates the probability that the positive patches are selected over the negative patches at specific locations, enabling the model to generate vessel masks not very similar to the fractal masks while learning the mask features. In our experiments, we set $R = 3$, and $(p_1, p_2, p_3) = (4, 8, 16)$.

3.3.3. Adversarial loss

In the adversarial path, our model generates the synthetic vessel image using the semantic fractal mask and the noisy features generated by the diffusion module. Since real vessel images exist in the training phase, we apply adversarial learning to the model and enable the generator G to output realistic images while fooling the discriminator D that distinguishes real and fake images. By employing LSGAN (Mao et al., 2017), the adversarial loss of the generator can be represented as:

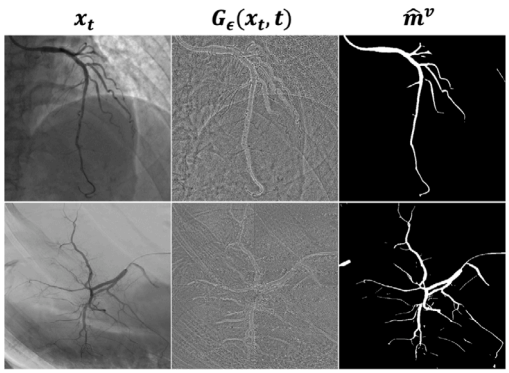


Fig. 7. Example images of the latent feature and the segmentation mask of the coronary angiogram (top row) and the hepatic angiogram (bottom row) when testing our proposed C-DARL model. x_t is an input, $G_\epsilon(x_t, t)$ is the latent feature, and $\hat{m}^v$ is the segmentation mask. In this figure, we use the input image with $t = 0$.

$$\mathcal{L}_{adv}^G = \mathbb{E}_{x_t,m^f}\left[(D(G(x_t,m^f))-1)^2\right], \tag{11}$$

and the adversarial loss of the discriminator can be written as:

$$\mathcal{L}_{adv}^D = \frac{1}{2}\mathbb{E}_{x_0}\left[(D(x_0)-1)^2\right] + \frac{1}{2}\mathbb{E}_{x_t,m^f}\left[(D(G(x_t,m^f)))^2\right]. \tag{12}$$

Through this loss function, the model is trained to generate the synthetic vessel image $\hat{x}$ that is indistinguishable from real data x by the discriminator.

3.3.4. Cycle loss

Recall that when the fractal-based synthetic vessel image is generated, in the cyclic contrastive path, the proposed model takes this synthetic image and generates the fractal segmentation mask $\hat{m}^f$. Here, in addition to using the mask $\hat{m}^f$ in the contrastive loss, to learn semantic information about blood vessels in the vessel medical data, we utilize the estimated fractal mask in the cycle loss that computes a distance between the real and fake images. As we handle the mask as an image, the cycle loss can be formulated using l_1 norm:

$$\mathcal{L}_{cyc} = \mathbb{E}_{x_t,m^f}\left[\|G(G(x_t,m^f),0) - m^f\|_1\right]. \tag{13}$$

Accordingly, the model can capture vessel information and estimate the mask for vessel-like regions even though there is no ground-truth vessel mask of real data.

3.3.5. Full loss formulation

Using the diffusion loss, the contrastive loss, the adversarial loss, and the cycle loss, our C-DARL model is trained in an end-to-end learning manner. Hence, the complete loss function of the generator can be defined by:

$$loss = \mathcal{L}_{diff} + \lambda_\alpha \mathcal{L}_{cont} + \lambda_\beta \mathcal{L}_{adv}^G + \lambda_\gamma \mathcal{L}_{cyc}, \tag{14}$$

Table 1
The number of data used in each training, validation (val), and test phase.

	Input	CA			AA	RI			
		XCAD	134XCA	30XCA	AAIR	UWF	FP	DRIVE	STARE
Train	Image $\boldsymbol{x}$	1621	–	–	327	451	745	–	–
	Fractal mask $\boldsymbol{m}^f$	1621	–	–	–	–	–	–	
Val	Image $\boldsymbol{x}$	12	–	–	–	–	–	–	–
	Vessel label $\boldsymbol{m}^v$	12	–	–	–	–	–	–	–
Test	Image $\boldsymbol{x}$	114	134	30	–	–	–	20	20
	Vessel label $\boldsymbol{m}^v$	114	134	30	–	–	–	20	20

CA: Coronary angiography, AA: Abdomen angiography, RI: Retinal imaging.

where λ_α, λ_β, and λ_γ are hyper-parameters that control each loss function.

3.4. Inference

Compared to conventional diffusion models, which generate images from pure Gaussian noise through the iterative reverse process, the DARL model provides a segmentation mask in one step in that the model is trained to estimate the mask for a given input image. Similarly, the proposed C-DARL model also generates the mask for a given vessel medical image in a single step. In other words, for a perturbed vessel image x_t, the vessel segmentation mask $\hat{m}^v$ is provided through the contrastive path of our model. Here, as shown in Fig. 6, while the model can estimate the mask for any corrupted image with a noise level $t \in [0, T]$, we evaluate our method using the clean image x_0 which can be considered as a target image for the diffusion model. Fig. 7 shows examples of the latent feature from the diffusion module and the final segmentation mask from the generation module when the image at $t = 0$ is given to our model. We further discuss the segmentation performance according to the noise level in Section 5.3.

4. Experiments and results

4.1. Experimental setup

4.1.1. Datasets

To train our C-DARL model for blood vessel segmentation, we utilize a variety of vessel images from different medical domains, including coronary arteriograms, abdominal pancreatic and hepatic arteriograms, and retinal fundus photography images. Also, we test the proposed model using several benchmark datasets of blood vessel segmentation. The number of data used in the training, validation, and test phases is summarized in Table 1, and detailed descriptions and preprocessing methods of each dataset are as follows.

XCAD dataset. The X-ray angiography coronary artery disease (XCAD) dataset (Ma et al., 2021) contains 1621 coronary angiography images taken during stent placement. For an additional 126 angiograms and the corresponding vessel segmentation masks labeled by experienced radiologists, we use 12 pairs of images and masks for the validation set, and the remaining 114 pairs for the test set. For each image with a size of 512 × 512, in training, we randomly subsample the image by factor 2 for data augmentation, while in testing, we extract the vessel masks for 4 subsampled 256 × 256 sized patches and get the full vessel masks by un-subsampling the 4 masks into one.

Also, using the fractal synthetic module presented by Ma et al. (2021), we synthesize 1621 fractal masks for model training. We generate 1621 synthetic fractal masks with 512 × 512 size, in which the fractals have various shapes and thicknesses. In Appendix A, we provide the details on how to synthesize the fractal masks used in our experiment.

134XCA dataset. To test the model for vessel segmentation of coronary angiography images using externally distributed data over the training dataset, we use the 134XCA dataset (Cervantes-Sanchez et al., 2019) that provides 134 coronary angiograms with ground-truth vessel masks, where the masks are obtained by an expert cardiologist. We resize the images to 512 × 512. We extract vessel masks using 4 subsampled 256 × 256 images. The full vessel masks are then obtained by un-subsampling the 4 masks into one. The segmentation performance is evaluated for the vessel masks in the region of interest of the angiography images.

30XCA dataset. We also utilize the 30XCA dataset (Hao et al., 2020) in the evaluation of coronary angiography segmentation. This dataset has 30 X-ray coronary angiography series, and one image for each series and its corresponding vessel segmentation masks annotated by experts are used for the test. We resize the image to 512 × 512 and subsample it with 4 patches with 256 × 256 size. We infer the segmentation maps for those 4 patches per image and un-subsample them to get the whole-sized mask.

AAIR dataset. As one of the multi-domain angiography images, we use 327 abdominal angiography images that were obtained from 42 subjects at the National Institutes of Health Clinical Center. The angiography series were obtained with 2 frames per second (fps) during arteriography, embolization, or calcium stimulation and showed arteries in the abdomen such as the celiac, pancreatic, and hepatic vasculatures. For each scan, we select one frame in which the blood vessels were most visible and manually cropped the region that includes the vessels. We call these collected data the AAIR dataset. In the training phase, we randomly extract image patches with a size of 256 × 256.

UWF and FP datasets. In addition to the coronary and abdominal angiograms, our model is trained using retinal images of the ultra-widefield (UWF) data and the fundus photograph (FP) data which are provided by Yoo (2020). The UWF dataset has 451 normal or pathologic retinal images with artifacts, and the FP dataset has 745 retinal images without artifacts. We use these retinal data for training the model. For data processing, we resize each image to 512 × 512 and crop the center area to 256 × 256. Also, we convert the RGB scale images to grayscale for the data to have one channel.

DRIVE dataset. To evaluate the model for retinal imaging data, we use the DRIVE dataset (Staal et al., 2004) that provides 20 retinal images and their vessel segmentation labels annotated by experts. We transform the images from RGB scale to grayscale and resize them to 768 × 768. Then, we extract patches with a size of 256 × 256 that are not overlapped, thus getting the estimated vessel masks for a total of 9 patches. The final full vessel mask with the original resolution is obtained by stitching all patches of the segmentation map. We evaluate the segmentation performance for nonzero region-of-interest areas of the vessel images.

Table 2
Quantitative evaluation results of vessel segmentation performance for the comparison study with the self-supervised methods. The baseline methods are the Meijering filter (Meijering et al., 2004), DINO (Caron et al., 2021), SSVS (Ma et al., 2021), DARL (Kim et al., 2023), and DA (Mahmood et al., 2019). "w/o training" denotes the rule-based model, "w/natural" means the model trained using natural images, "w/CA+background" denotes the model trained using only coronary angiography (CA) data with their background images, and "w/multi-domain" means the model trained using multi-domain datasets.

Dataset	Metric	Non-learning	Self-supervised learning					
		w/o training	w/natural	w/CA+background			w/multi-domain	
		Meijering	DINO	SSVS	DARL	C-DARL	DA	C-DARL
Coronary angiography (CA)								
XCAD	IoU	$0.223_{\pm 0.077}$	$0.279_{\pm 0.047}$	$0.421_{\pm 0.073}$	$0.443_{\pm 0.074}$	$0.476_{\pm 0.082}$	$0.429_{\pm 0.059}$	$0.505_{\pm 0.086}$
	Dice	$0.358_{\pm 0.102}$	$0.434_{\pm 0.058}$	$0.588_{\pm 0.075}$	$0.609_{\pm 0.073}$	$0.641_{\pm 0.079}$	$0.597_{\pm 0.060}$	$0.666_{\pm 0.078}$
	Precision	$0.619_{\pm 0.180}$	$0.387_{\pm 0.082}$	$0.571_{\pm 0.104}$	$0.638_{\pm 0.123}$	$0.727_{\pm 0.114}$	$0.573_{\pm 0.096}$	$0.751_{\pm 0.109}$
	Recall	$0.281_{\pm 0.116}$	$0.528_{\pm 0.112}$	$0.631_{\pm 0.111}$	$0.603_{\pm 0.084}$	$0.585_{\pm 0.094}$	$0.648_{\pm 0.095}$	$0.611_{\pm 0.094}$
	95-HD ▼	$0.393_{\pm 0.117}$	$0.294_{\pm 0.089}$	$0.412_{\pm 0.111}$	$0.338_{\pm 0.116}$	$0.328_{\pm 0.102}$	$0.351_{\pm 0.107}$	$0.307_{\pm 0.093}$
134XCA	IoU	$0.399_{\pm 0.127}$	$0.248_{\pm 0.053}$	$0.370_{\pm 0.122}$	$0.461_{\pm 0.100}$	$0.491_{\pm 0.119}$	$0.419_{\pm 0.133}$	$0.534_{\pm 0.097}$
	Dice	$0.558_{\pm 0.134}$	$0.394_{\pm 0.070}$	$0.529_{\pm 0.136}$	$0.623_{\pm 0.100}$	$0.648_{\pm 0.117}$	$0.578_{\pm 0.141}$	$0.690_{\pm 0.091}$
	Precision	$0.783_{\pm 0.118}$	$0.313_{\pm 0.071}$	$0.484_{\pm 0.126}$	$0.624_{\pm 0.131}$	$0.626_{\pm 0.121}$	$0.567_{\pm 0.115}$	$0.664_{\pm 0.132}$
	Recall	$0.463_{\pm 0.161}$	$0.554_{\pm 0.110}$	$0.600_{\pm 0.183}$	$0.653_{\pm 0.118}$	$0.703_{\pm 0.165}$	$0.615_{\pm 0.193}$	$0.752_{\pm 0.121}$
	95-HD ▼	$0.364_{\pm 0.116}$	$0.342_{\pm 0.119}$	$0.360_{\pm 0.118}$	$0.300_{\pm 0.130}$	$0.305_{\pm 0.128}$	$0.318_{\pm 0.113}$	$0.283_{\pm 0.122}$
30XCA	IoU	$0.311_{\pm 0.088}$	$0.314_{\pm 0.054}$	$0.347_{\pm 0.160}$	$0.388_{\pm 0.057}$	$0.443_{\pm 0.066}$	$0.379_{\pm 0.090}$	$0.461_{\pm 0.061}$
	Dice	$0.468_{\pm 0.101}$	$0.476_{\pm 0.064}$	$0.490_{\pm 0.208}$	$0.556_{\pm 0.060}$	$0.610_{\pm 0.065}$	$0.542_{\pm 0.103}$	$0.629_{\pm 0.059}$
	Precision	$0.737_{\pm 0.144}$	$0.402_{\pm 0.087}$	$0.630_{\pm 0.189}$	$0.710_{\pm 0.134}$	$0.801_{\pm 0.097}$	$0.664_{\pm 0.106}$	$0.834_{\pm 0.087}$
	Recall	$0.360_{\pm 0.106}$	$0.611_{\pm 0.086}$	$0.430_{\pm 0.206}$	$0.467_{\pm 0.052}$	$0.496_{\pm 0.060}$	$0.471_{\pm 0.113}$	$0.508_{\pm 0.058}$
	95-HD ▼	$0.359_{\pm 0.130}$	$0.294_{\pm 0.104}$	$0.335_{\pm 0.108}$	$0.318_{\pm 0.093}$	$0.310_{\pm 0.113}$	$0.312_{\pm 0.116}$	$0.299_{\pm 0.125}$
Retinal Imaging (RI)								
DRIVE	IoU	$0.381_{\pm 0.066}$	$0.223_{\pm 0.017}$	$0.229_{\pm 0.050}$	$0.265_{\pm 0.047}$	$0.237_{\pm 0.052}$	$0.229_{\pm 0.041}$	$0.340_{\pm 0.061}$
	Dice	$0.548_{\pm 0.071}$	$0.365_{\pm 0.022}$	$0.370_{\pm 0.065}$	$0.415_{\pm 0.060}$	$0.379_{\pm 0.066}$	$0.371_{\pm 0.053}$	$0.505_{\pm 0.068}$
	Precision	$0.687_{\pm 0.088}$	$0.314_{\pm 0.033}$	$0.502_{\pm 0.083}$	$0.768_{\pm 0.097}$	$0.786_{\pm 0.096}$	$0.493_{\pm 0.085}$	$0.905_{\pm 0.107}$
	Recall	$0.469_{\pm 0.091}$	$0.443_{\pm 0.042}$	$0.295_{\pm 0.058}$	$0.289_{\pm 0.055}$	$0.253_{\pm 0.056}$	$0.300_{\pm 0.042}$	$0.354_{\pm 0.063}$
	95-HD ▼	$0.134_{\pm 0.063}$	$0.106_{\pm 0.017}$	$0.130_{\pm 0.022}$	$0.131_{\pm 0.019}$	$0.140_{\pm 0.032}$	$0.171_{\pm 0.035}$	$0.121_{\pm 0.021}$
STARE	IoU	$0.124_{\pm 0.076}$	$0.207_{\pm 0.016}$	$0.284_{\pm 0.100}$	$0.347_{\pm 0.119}$	$0.289_{\pm 0.117}$	$0.282_{\pm 0.091}$	$0.364_{\pm 0.124}$
	Dice	$0.212_{\pm 0.118}$	$0.342_{\pm 0.022}$	$0.433_{\pm 0.128}$	$0.501_{\pm 0.140}$	$0.433_{\pm 0.149}$	$0.431_{\pm 0.116}$	$0.520_{\pm 0.144}$
	Precision	$0.703_{\pm 0.291}$	$0.281_{\pm 0.028}$	$0.499_{\pm 0.137}$	$0.673_{\pm 0.169}$	$0.633_{\pm 0.200}$	$0.503_{\pm 0.137}$	$0.714_{\pm 0.192}$
	Recall	$0.130_{\pm 0.078}$	$0.446_{\pm 0.053}$	$0.386_{\pm 0.124}$	$0.408_{\pm 0.126}$	$0.336_{\pm 0.126}$	$0.381_{\pm 0.105}$	$0.427_{\pm 0.132}$
	95-HD ▼	$0.325_{\pm 0.165}$	$0.122_{\pm 0.018}$	$0.129_{\pm 0.034}$	$0.139_{\pm 0.038}$	$0.142_{\pm 0.046}$	$0.125_{\pm 0.022}$	$0.150_{\pm 0.055}$

Note: If ▼, lower is better, otherwise, higher is better.

STARE dataset. For the retinal image domain, the model is also tested using the STARE dataset (Hoover et al., 2000). This is composed of 20 retinal images and human-labeled blood vessel segmentation masks. As with the DRIVE dataset, each image is converted to grayscale and resized into 768×768. Then, the vessel masks are estimated for 9 non-overlapped patches with a size of 256×256. We stitch the patches to get full vessel masks and evaluate the performance for the region of interest of the retinal images.

4.1.2. Implementation details

Our proposed model is implemented using the PyTorch (Paszke et al., 2019) in Python. For the network architectures of the diffusion module and the generation module, we employ the DDPM (Ho et al., 2020) and the SPADE (Park et al., 2019), respectively. The DDPM network has a U-Net-like structure and takes an embedding vector of the time as well as the noise-corrupted image. Also, the SPADE model has a shallow encoder and a decoder with SPADE blocks, in which the SPADE layer is replaced with the S-SPADE layer for our C-DARL model. Also, for the discriminator, we utilize PatchGAN (Isola et al., 2017) to distinguish real and generated fake images in patch levels.

For the inputs of multi-domain images, as shown in Table 1, we use all images from the XCAD, AAIR, UWF, and FP datasets. For each iteration, the input image is randomly sampled among the given training datasets, rescaled into $[-1, 1]$, and augmented using random horizontal or vertical flips and rotation at 90 degrees. We set the batch size to 1.

To optimize our model, we set the range of noisy time steps as $[0, 2000]$ to linearly schedule the noise levels from 10^{-6} to 10^{-2}. In the contrastive path, we set the maximum noisy level to 200, allowing the model to perform robust segmentation even on noisy images. The hyperparameters of our loss function are set as $\lambda_\alpha = 4 \times 10^{-4}$, $\lambda_\beta = 0.2$, and $\lambda_\gamma = 2$. The study of how the hyperparameters are set can be found in Section 4.2.4. We adopt the Adam algorithm (Kingma and Ba, 2015) with a learning rate of 1×10^{-5}. The model is trained for 150 epochs using a single GPU card of Nvidia A100-SXM4-40 GB, where the generator G and the discriminator D are optimized with the two-player setting for adversarial learning that the generator competes against the discriminator in generating realistic images. Then, we test the proposed method using the network weights that achieve the best performance on the validation dataset.

4.1.3. Evaluation

For baseline methods, we adopt rule-based and self-supervised-learning based approaches: Meijering (Meijering et al., 2004), DINO (Caron et al., 2021), Self-Supervised Vessel Segmentation (SSVS) (Ma et al., 2021), Diffusion Adversarial Representation Learning (DARL) (Kim et al., 2023), and Deep Adversarial (DA) (Mahmood et al., 2019). The Meijering filter is a vessel function that has been developed to extract vessel structure in a rule-based manner. The DINO learns class-specific features of natural images in an unsupervised manner so that it can be used in self-supervised segmentation. We use the pre-trained ViT-B/16 of DINO as a backbone and extract self-attention maps following the original paper. For both the Meijering and the DINO methods that require heuristic thresholds for each dataset, the optimal segmentation performance is selected by varying data-specific thresholds within the range from 0.1 to 0.9 in increments of 0.1.

In addition, the SSVS, DARL, DA method learns semantic segmentation using pseudo masks via adversarial learning. For the models of SSVS and DARL that require background images acquired before contrast agent injection, we use only coronary angiography (CA) data from the XCAD dataset. For the DA method, we train the model using the same multi-domain data as our method. We implement the adversarial

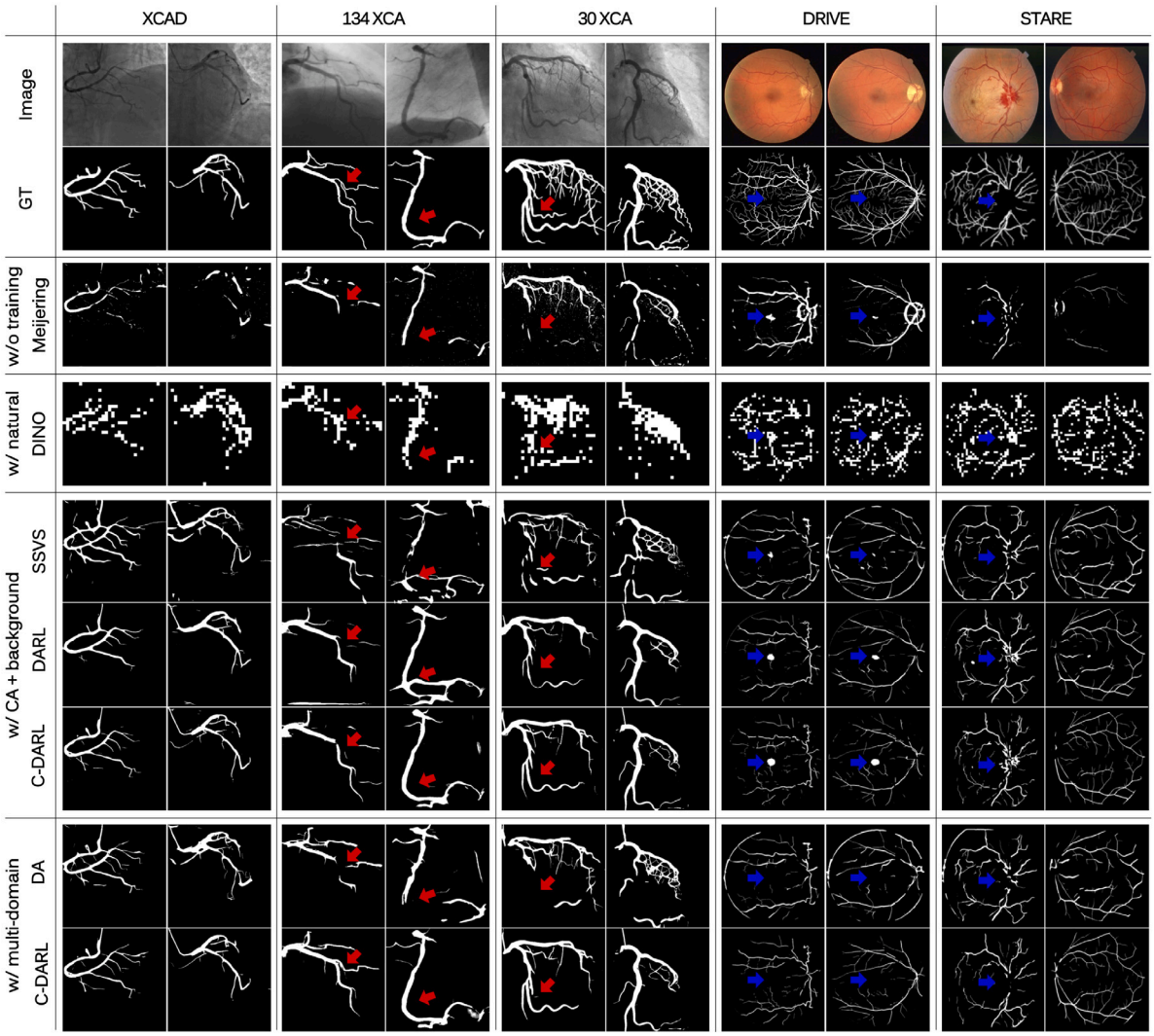


Fig. 8. Visual comparisons of the vessel segmentation results on various vessel datasets. The baseline methods are the Meijering filter (Meijering et al., 2004), DINO (Caron et al., 2021), SSVS (Ma et al., 2021), DARL (Kim et al., 2023), and DA (Mahmood et al., 2019). "w/o training" denotes the rule-based model, "w/natural" means the model trained using natural images, "w/CA+background" denotes the model trained using only coronary angiography (CA) data with their background images, and "w/multi-domain" means the model trained using multi-domain datasets. Red and blue arrows indicate remarkable parts of tiny vessel structures and false positive artifacts, respectively.

learning approaches using the same learning rate and hyperparameters weight ratio of the loss function as our C-DARL model. Also, when comparing ours to these methods, we train each model two times and report the ensembled results to mitigate variance across multiple experiments.

To evaluate the vessel segmentation performance quantitatively, we use several metrics: Intersection over Union (IoU), Dice similarity coefficient, Precision, and Recall. In the comparison study of ours to the baseline methods, we also compute the 95th percentile of Hausdorff Distance (95-HD) (Crum et al., 2006) that can measure overall spatial distances between the labels and the corresponding predictions. When calculating the 95-HD, all the measured distances in the pixel unit are normalized with respect to the input image dimension.

4.2. Results

4.2.1. Comparison study

For the comparison study with the existing models, we evaluate the blood vessel segmentation performance on coronary angiography (CA) datasets and retinal imaging (RI) datasets. Here, for a fair comparison of ours to the SSVS and DARL methods which require background images in training, we also implement our C-DARL model in the same scenario that uses only CA data and their background images, by providing the background images to the adversarial path. Table 2 and Fig. 8 show the quantitative and qualitative comparison results, respectively.

As reported in Table 2, our C-DARL model trained using multiple domains of vessel images achieves state-of-the-art (SOTA) performance on most of the datasets compared to the baseline methods of non-learning and self-supervised vessel segmentation. Specifically, when comparing the results using all metrics, we can observe that our model accurately segments true-positive vessel structures by minimizing false-negative and false-positive predictions. For the DRIVE dataset, the Meijering shows higher performance of IoU, Dice, Recall, and 95-HD than ours, however, it shows inferior performance on the other datasets compared to most of the baseline methods. This suggests that the rule-based method requires image-specific parameter optimization, which poses barriers to practical deployment. Also, although the DINO method performs better than our C-DARL model in terms of 95-HD, we can see that the DINO omits details for tiny vessels and estimates rough vessel structure in Fig. 8, leading to lower performance on the other metrics. In contrast, our model is superior in segmenting vessel structures including subtle branches.

Moreover, when comparing the models trained using only CA data with the background images, we can observe that the proposed C-DARL model performs the best on the CA datasets of XCAD, 134XCA,

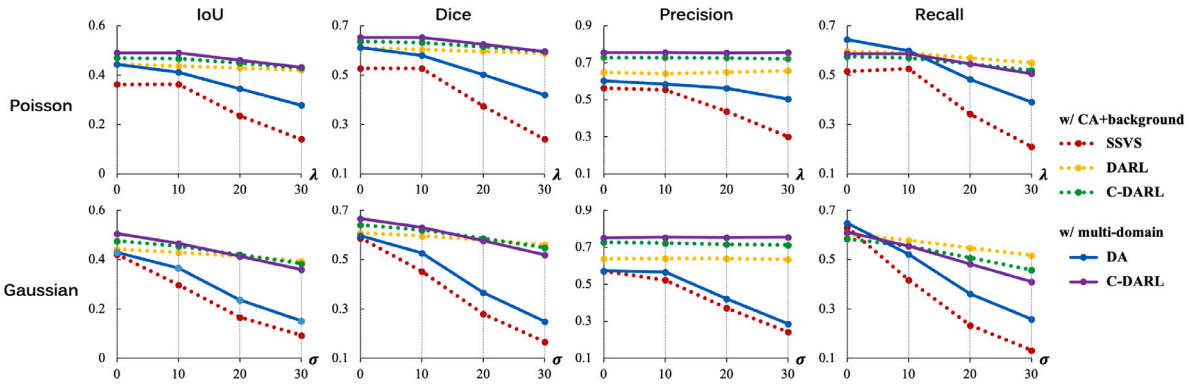


Fig. 9. Quantitative results of vessel segmentation on various noise-corrupted image scenarios. The top row shows the evaluation results according to λ level for Poisson noisy images, and the bottom row shows the results according to σ level for Gaussian noisy images. The models of SSVS (Ma et al., 2021), DARL (Kim et al., 2023), and DA (Mahmood et al., 2019) are used for the baseline methods. "w/CA+background" denotes the model trained using only coronary angiography (CA) data with their background images, and "w/multi-domain" means the model trained using multi-domain datasets.

and 30XCA, while the DARL model shows the best in the RI datasets of DRIVE and STARE. This may come from the proposed mask-based contrastive loss (9) that maximizes the similarity between the fractal mask and the estimated fractal mask that is embedded in the synthetic vessel image generated using the training data, whereas the DARL model simply distinguishes the fractal mask from the estimated vessel mask. Accordingly, our C-DARL improves the vessel segmentation performance over the DARL model for the seen domain images in training. This verifies that the proposed contrastive loss allows our model not only to capture the semantic information of blood vessels but also to learn domain-specific image segmentation even though the performance on the unseen datasets may decrease. Also, it is noteworthy that our C-DARL achieves the best generalization performance by enabling the model to use unlabeled multiple-domain datasets while eliminating the necessity of background images in training.

On the other hand, in Fig. 8, we can observe that the segmentation performance on tiny vessel regions is significantly improved (see red arrows), which demonstrates the advantages of our contrastive learning that effectively differentiates real vessel structure from the pseudo fractal mask signal. Moreover, our C-DARL mitigates false positive artifacts compared to the existing models trained with only the CA dataset (see blue arrows) and shows the most promising precision metrics. This shows the training efficiency of the proposed label-free framework which is capable of incorporating various data distributions with no use of paired background images.

4.2.2. Noise corruption study

In clinical practice, X-ray angiography images can be degraded by noise due to various factors such as low radiation dose, body mass, organ or patient motion, breathing, or X-ray energy parameters in the data acquisition procedure. Accordingly, we further evaluate the segmentation performance of our C-DARL under noise degradation scenarios. To simulate the noisy images, we add Poisson and Gaussian noises with different levels of λ and σ, respectively, to the clean XCAD test data. Then, we extract vessel structures within the noise-degraded images.

Figs. 9 and 10 show the quantitative and qualitative results of the XCAD test data between our model and the baseline methods. The detailed values of evaluation metrics are reported in Table B.1 of Appendix B. We can observe that our model is robust to noise compared to the existing models even though the performance decreases according to the noise level being stronger. Also, the DARL and C-DARL are the only methods that endure harsh noise corruption with reasonable performance. These results suggest that our proposed diffusion adversarial learning framework is superior in unseen noise distributions since the model is trained with highly perturbed input images through the diffusion module.

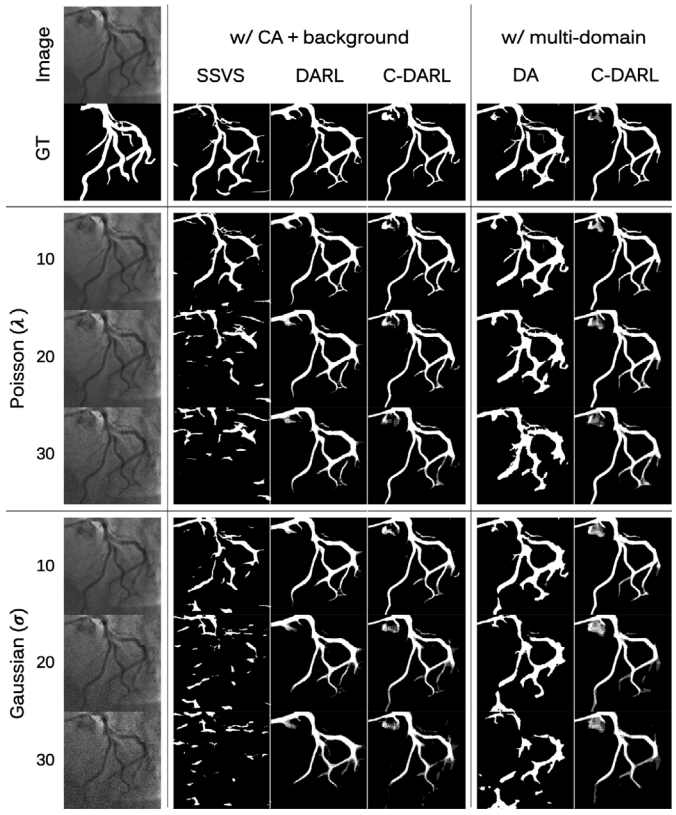


Fig. 10. Visual vessel segmentation results of the noisy corruption study. The models of SSVS (Ma et al., 2021), DARL (Kim et al., 2023), and the DA (Mahmood et al., 2019) are used for the baseline methods. "w/CA+background" denotes the model trained using only coronary angiography (CA) data with their background images, and "w/multi-domain" means the model trained using multi-domain datasets.

4.2.3. Ablation study

Our proposed method is trained by minimizing the objective function (14) which is composed of diffusion loss, contrastive loss, adversarial loss, and cycle loss. To verify the effectiveness of each loss, we conduct an ablation study on the loss function and report the results in Table 3.

When our C-DARL model is trained without the diffusion loss (w/o $\mathcal{L}_{diff}$), the segmentation performance slightly decreases than the proposed method. This suggests that the DDPM network architecture is

Table 3
Blood vessel segmentation results for the XCAD test dataset in the ablation study.

Ablation method	Metric			
	IoU	Dice	Precision	Recall
w/o $\mathcal{L}_{diff}$	$0.483_{\pm 0.094}$	$0.646_{\pm 0.088}$	$0.741_{\pm 0.114}$	$0.584_{\pm 0.105}$
w/o $\mathcal{L}_{cont}$	$0.361_{\pm 0.067}$	$0.527_{\pm 0.073}$	$0.593_{\pm 0.101}$	$0.486_{\pm 0.086}$
w/o $\mathcal{L}_{adv}$	$0.004_{\pm 0.005}$	$0.008_{\pm 0.010}$	$0.006_{\pm 0.008}$	$0.014_{\pm 0.017}$
w/o $\mathcal{L}_{cyc}$	$0.091_{\pm 0.038}$	$0.164_{\pm 0.063}$	$0.107_{\pm 0.045}$	$0.374_{\pm 0.121}$
Replace $\mathcal{L}_{cont} \rightarrow \mathcal{L}_{adv}$	$0.456_{\pm 0.098}$	$0.620_{\pm 0.095}$	$0.669_{\pm 0.134}$	$0.594_{\pm 0.107}$
w/o *Cat* before G_s	$0.489_{\pm 0.090}$	$0.652_{\pm 0.084}$	$0.717_{\pm 0.115}$	$0.612_{\pm 0.104}$
Proposed	$0.498_{\pm 0.086}$	$0.661_{\pm 0.079}$	$0.750_{\pm 0.108}$	$0.602_{\pm 0.095}$

effective in extracting features for given noisy input images, but learning data distribution through the diffusion loss, which enhances the performance of semantic image synthesis required to capture vessel structures, gives more improvement in vessel segmentation performance.

Also, the ablation method that excludes the contrastive loss (w/o $\mathcal{L}_{cont}$) indicates that the proposed mask-based contrastive loss makes the model effectively learn vessel representations without ground-truth labels. In particular, although self-supervised learning can be possible via adversarial learning on the real and fake vessel masks, the experiment in which the contrastive loss is replaced with the adversarial loss (Replace $\mathcal{L}_{cont} \rightarrow \mathcal{L}_{adv}$), such as the loss function of the DARL, have 4% lower IoU and Dice scores and 8% lower Precision than the proposed method. This shows that our contrastive loss considering that the input fractal masks have different features from real blood vessels outperforms the adversarial loss regarding the synthetic fractal masks as real masks.

For the C-DARL model trained without the adversarial loss (w/o $\mathcal{L}_{adv}$), the vessel structures are hardly extracted due to no guidance to learn vessel representations. Since the fake vessel images synthesized using the fractal mask provide vessel-like structure information to the model in the cyclic contrastive path, the adversarial loss is required to generate realistic vessel images. In addition, the model trained without the cycle loss (w/o $\mathcal{L}_{cyc}$) shows much inferior vessel segmentation performance compared to the proposed method, implying that the cycle loss in the cyclic contrastive path enables the model to capture vessel structures of input images.

On the other hand, we also study the necessity of input image concatenation before the generation module. We implement the ablation method that excludes the concatenation (w/o *Cat* before G_s), by setting the contrastive path of the generation module to take only noisy vessel images while the adversarial path to take only latent images estimated from the diffusion module. As a result, the segmentation performance is similar to the proposed C-DARL model, but this ablation method shows relatively higher false positives. This indicates that although the two paths of the generation module may require different image information, the latent feature from the diffusion module and the noisy vessel images synergistically improve the vessel segmentation performance and their concatenation further simplifies the model flows.

4.2.4. Hyperparameter setting

To investigate the optimal setting of hyper-parameters in our loss function (14), we implement our model by adjusting the values for each λ_α, λ_β, and λ_γ. When one of the parameters is adjusted, we set the other parameters to fixed values. Fig. 11 shows quantitative evaluation results on the validation dataset according to the hyperparameters.

When the parameter λ_α weighting the contrastive loss increases, the IoU and Dice scores increase and then converge, and the Precision value continues to increase slightly, leading to the achievement of the highest performance when λ_α is 4×10^{-4}. In the study of λ_β that controls the adversarial loss, we can observe that our model hardly captures vessel information if the model does not learn semantic image synthesis with $\lambda_\beta = 0$, but λ_β is between 0.1 and 0.3, the model learns vessel representation with high performance. In particular, when λ_β is set to 0.2, the model shows the best results in terms of IoU and Precision. Lastly, in the study of λ_γ for the cycle loss, the model shows the best performance at $\lambda_\gamma = 2.0$, and the performance gradually decreases as the parameter increases. By considering these results, we report the performance of our model that is trained by setting the hyperparameters as $\lambda_\alpha = 4 \times 10^{-4}$, $\lambda_\beta = 0.2$, and $\lambda_\gamma = 2.0$.

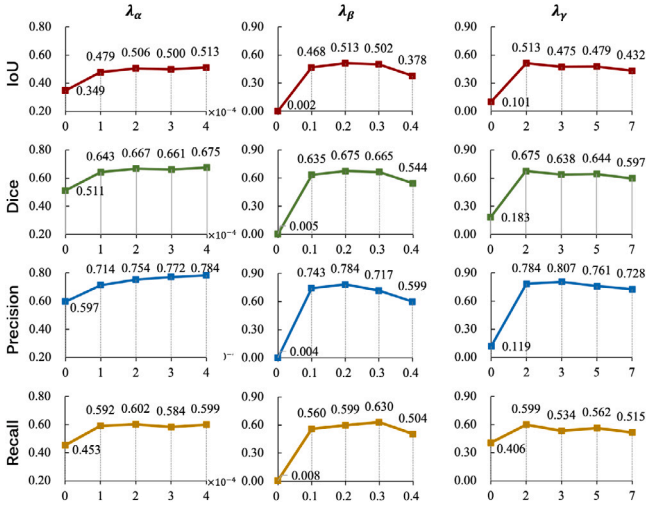


Fig. 11. Study of hyperparameters in the loss function of C-DARL model. We compare the evaluation results on the validation dataset according to the value of hyperparameters for λ_α (first column), λ_β (second column), and λ_γ (third column).

Table 4
Quantitative evaluation results of vessel segmentation performance for the comparison study with the self-supervised methods.

Dataset	Metric	Training dataset		
		CA	CA+RI	CA+AA+RI
Coronary angiography (CA)				
XCAD	IoU	$0.466_{\pm 0.073}$	$0.491_{\pm 0.088}$	$0.498_{\pm 0.086}$
	Dice	$0.633_{\pm 0.070}$	$0.654_{\pm 0.081}$	$0.661_{\pm 0.079}$
	Precision	$0.712_{\pm 0.114}$	$0.745_{\pm 0.111}$	$0.750_{\pm 0.108}$
	Recall	$0.583_{\pm 0.084}$	$0.594_{\pm 0.097}$	$0.602_{\pm 0.095}$
134 XCA	IoU	$0.508_{\pm 0.103}$	$0.521_{\pm 0.106}$	$0.545_{\pm 0.094}$
	Dice	$0.668_{\pm 0.097}$	$0.678_{\pm 0.101}$	$0.700_{\pm 0.087}$
	Precision	$0.665_{\pm 0.119}$	$0.649_{\pm 0.120}$	$0.673_{\pm 0.128}$
	Recall	$0.695_{\pm 0.135}$	$0.741_{\pm 0.144}$	$0.758_{\pm 0.119}$
30 XCA	IoU	$0.428_{\pm 0.054}$	$0.464_{\pm 0.057}$	$0.453_{\pm 0.061}$
	Dice	$0.598_{\pm 0.054}$	$0.632_{\pm 0.053}$	$0.621_{\pm 0.059}$
	Precision	$0.806_{\pm 0.105}$	$0.817_{\pm 0.093}$	$0.826_{\pm 0.083}$
	Recall	$0.479_{\pm 0.047}$	$0.520_{\pm 0.054}$	$0.500_{\pm 0.058}$
Retinal imaging (RI)				
DRIVE	IoU	$0.302_{\pm 0.049}$	$0.314_{\pm 0.065}$	$0.345_{\pm 0.061}$
	Dice	$0.462_{\pm 0.058}$	$0.475_{\pm 0.073}$	$0.510_{\pm 0.067}$
	Precision	$0.838_{\pm 0.102}$	$0.888_{\pm 0.103}$	$0.904_{\pm 0.090}$
	Recall	$0.323_{\pm 0.054}$	$0.328_{\pm 0.067}$	$0.359_{\pm 0.063}$
STARE	IoU	$0.333_{\pm 0.131}$	$0.342_{\pm 0.117}$	$0.367_{\pm 0.133}$
	Dice	$0.484_{\pm 0.157}$	$0.498_{\pm 0.137}$	$0.522_{\pm 0.154}$
	Precision	$0.685_{\pm 0.190}$	$0.691_{\pm 0.179}$	$0.736_{\pm 0.179}$
	Recall	$0.380_{\pm 0.133}$	$0.399_{\pm 0.125}$	$0.420_{\pm 0.140}$

CA: Coronary angiography, AA: Abdoman angiography, RI: Retinal imaging.

5. Discussion

5.1. Training using multi-domain data

One of the strengths of our framework is that only vessel images are taken as input without background images before the contrast agent injection so that the model can utilize various vessel medical data, which

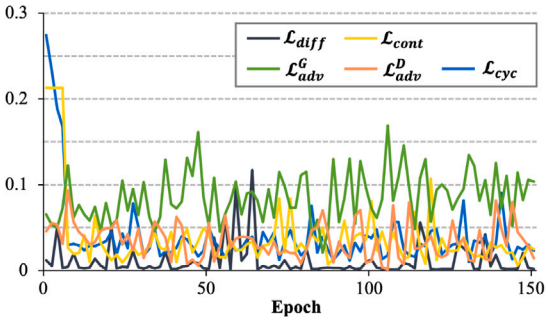
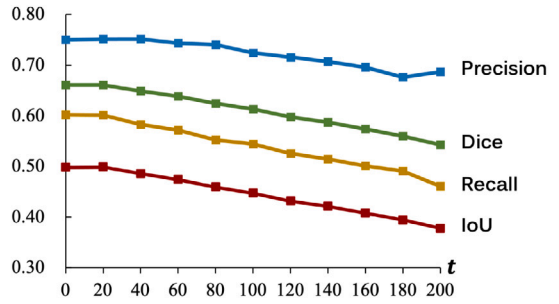


Fig. 12. Convergence curves of our losses during training the C-DARL model. Each loss of the diffusion loss $\mathcal{L}_{diff}$, the contrastive loss $\mathcal{L}_{cont}$, the adversarial loss ($\mathcal{L}^{G}_{adv}$, $\mathcal{L}^{D}_{adv}$), and the cycle loss $\mathcal{L}_{cyc}$ is weighted by 1, 4×10^{-4}, 0.2, and 2.0, respectively, according to the hyperparameter setting described in Section 4.1.2.



13: The quantitative results of our model on the validation set acc

Fig. 13. The quantitative results of our model on the validation set according to the noise level in the inference phase. The graph shows the value of each evaluation metric (y-axis) with respect to the noise level t (x-axis).

improves the generalization performance. In order to demonstrate the effect of using multi-domain data in the training phase, we train our C-DARL model with respect to the number of vessel image domains.

Table 4 shows the vessel segmentation performance according to the input image datasets. We can observe that the results of the proposed C-DARL using only CA are slightly higher than the DARL results reported in Table 2, even though the C-DARL does not use the background images while the DARL uses the background images in addition to the angiography images. Also, the C-DARL leveraging additional retinal images improves the segmentation performance with a gain of up to 5% for both internal and external test datasets, over the C-DARL using only CA. In addition, the model trained using various datasets in three domains, including coronary images from the XCAD dataset, abdomen angiography images from the AAIR dataset, and retinal images from the UWF and FP datasets, achieves the highest performance in most of the test data.

These results show that the segmentation performance increases even though the training data have different domain information, which suggests that our model outperforms to incorporate various domains and achieves better generalization. Although the ablation study of replacing the proposed contrastive loss with the adversarial loss in Section 4.2.3 also demonstrates that our performance improvement over the DARL model comes from the contrastive loss, this study on training using multi-domain data shows that the ability of our proposed model that uses diverse images in different domains is more attributable to enhancing the generalization performance in the segmentation of various blood vessel image datasets. Future work will study performance for different clinical tasks in scenarios where the underlying vessel pathway is desirable (navigation tasks), as well as settings where diagnostic surrogates for vessel or endothelial pathology are valuable.

5.2. Convergence of loss function

In training the proposed C-DARL, the learnable parameters are optimized by minimizing our loss function (14). To show how each loss is minimized during the network training, here we illustrate the convergence curve of our losses in Fig. 12. We can see that the diffusion loss $\mathcal{L}_{diff}$ is decreased initially, and then the proposed contrastive loss $\mathcal{L}_{cont}$ and cycle loss $\mathcal{L}_{cyc}$ are minimized. These losses then converge as the training continues. Also, it is noticeable that the adversarial loss for the generator $\mathcal{L}^{G}_{adv}$ and that for the discriminator $\mathcal{L}^{D}_{adv}$ are relatively stable during the training, even though the generator is trained with the other loss functions as well as the adversarial loss. This shows that our proposed losses are jointly optimized through stable learning.

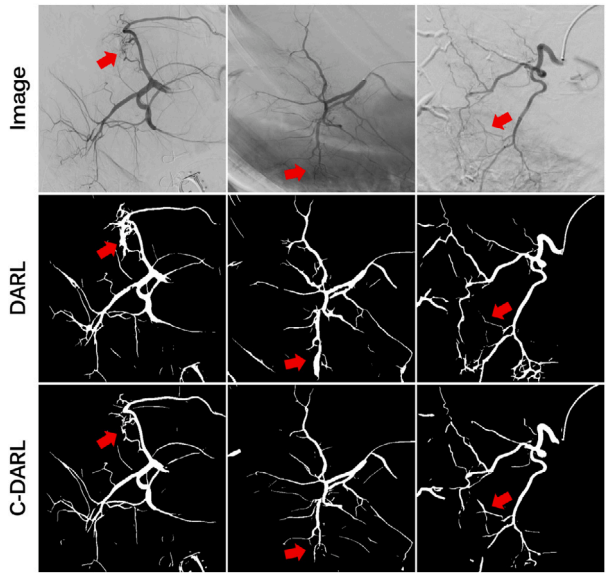


Fig. 14. Visual vessel segmentation results of abdomen angiography dataset. Red arrows point to remarkable regions.

5.3. Noise level for inference

Our framework has another strong point in providing robust performance under noise corruption scenarios thanks to the proposed diffusion adversarial learning strategy, which leverages both the perturbed real vessel image through the forward diffusion process and the synthesized fake vessel image for learning blood vessel segmentation. We analyze the segmentation performance according to the noise level at a time step $t \in [0, 200]$ that is uniformly sampled with an interval of 20.

In Fig. 13, our model achieves the reliable performance of vessel segmentation until t reaches 200, which relates to Section 4.2.2 which shows the additional noise corruption study with different types of noise on external datasets. Although the model reveals the best segmentation performance at $t = 0$, this inference study with different levels of noise demonstrates the great potential of our model in generalizing to out-of-distribution datasets.

5.4. Abdomen angiography segmentation

We also verify the segmentation performance of our model for abdominal angiography to further study the generalization of our model

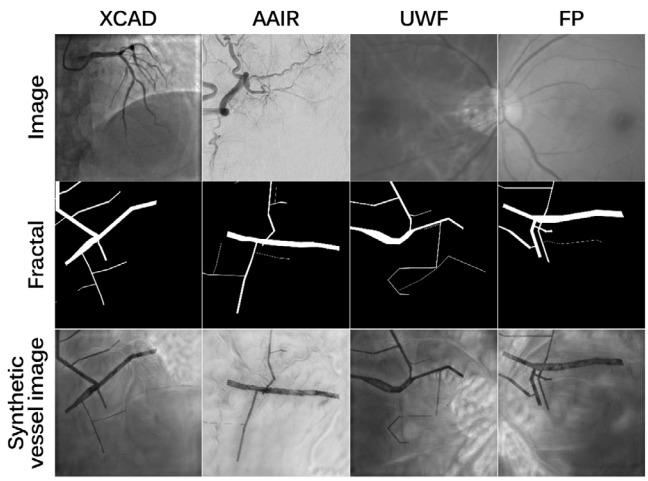


Fig. 15. Synthetic blood vessel images from our C-DARL model according to the datasets of different domains. The vessel images (bottom row) are generated using the input image (top row) and the fractal mask (middle row) in the adversarial path.

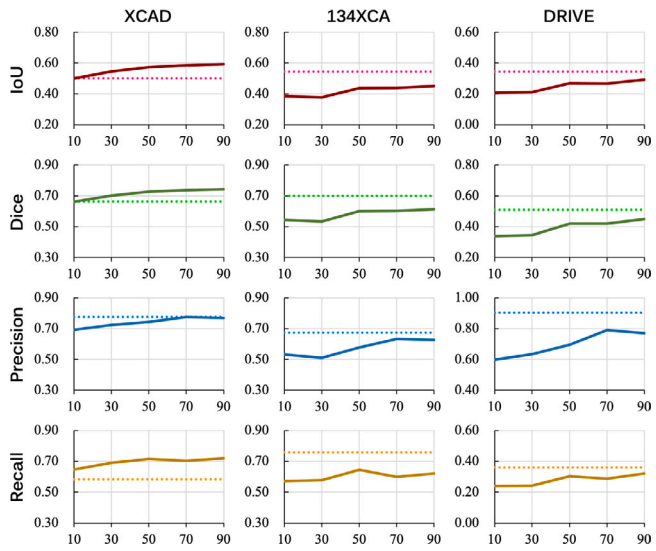


Fig. 16. Quantitative comparison results of ours to supervised learning models for the XCAD, 134XCA, and DRIVE datasets. The dashed line indicates the results from our C-DARL model, while the solid line indicates the results from the supervised learning model with respect to the percentage of labeled data usage in training.

to various domain datasets. As there is no benchmark dataset containing ground-truth vessel labels for the abdominal vessel images, we instead compare the qualitative results of the vessel segmentation using the AAIR data. For a baseline model, we adopt the DARL model which achieves the second-best performance in the comparison study of Section 4.2.1 (Table 2).

Fig. 14 shows the results of the blood vessel segmentation for several abdomen angiography images. Compared with the DARL model, the results show that the proposed C-DARL model outperforms the segmentation of vascular regions, including tiny and low-contrast branches. Also, our model shows superior segmentation performance of even blood vessels that are difficult to distinguish from the background structures. This suggests that our model can be applied to a variety of vascular images by improving the capability of learning vessel representations through contrastive learning under the condition of no ground-truth vessel masks.

5.5. Synthetic vessel image generation

In our experiments, we show that the C-DARL model segments blood vessel structures with high performance on multi-domain datasets. In fact, our proposed model learns semantic information about the blood vessels with the guidance of synthetic image generation in the adversarial path. To investigate how our model synthesizes the vessel images, we visualize the synthetic images generated during training according to the datasets of different domains.

Fig. 15 shows several synthetic vessel images from our C-DARL model. We can observe that the model generates diverse realistic vessel images using the given input vessel image and fractal masks. In particular, the generated image maintains the information of background images and performs semantic image synthesis with the fractal masks while nulling out the vessel structures of the inputted real vessel images. This suggests that our model learns blood vessel segmentation efficiently by estimating realistic vessel images.

5.6. Comparison to supervised learning

The most significant contribution of our work is that C-DARL is fully unsupervised and no ground truth labels are required to yield vessel segmentation performance with generalizability across different datasets. In order to demonstrate the superiority of our label-free framework, we train a supervised version of C-DARL using only the contrastive path. Here, we replace the contrastive loss with the cross-entropy loss to predict segmentation maps using labels. For training and testing the model, we use a total of 114 pairs of labeled XCAD data used for the test data in our main experiments. In particular, we vary the amount of training data within the range from 10% to 90% in increments of 20% for each trial and use the remaining data as a test set.

In Fig. 16, we can observe that the supervised C-DARL tends to overfit the XCAD dataset in that it becomes on par with our C-DARL even when using only 10%–30% of the labeled data. On the other hand, for the external 134 XCA and DRIVE datasets, even the supervised C-DARL model trained using 90% of labeled data cannot beat the performance of C-DARL. The detailed metric scores can be found in Table C.1. These results demonstrate that our self-supervised C-DARL shows promising cross-data generalization capabilities compared to the supervised learning method.

5.7. Limitation and future works

Our proposed self-supervised blood vessel segmentation method achieves the Dice score of around 63%–69% on coronary angiography data and around 50%–52% on retinal imaging data. While our C-DARL model outperforms the existing self-supervised learning methods, the performances are still far from those of fully-supervised learning methods, which limits its use in clinical practice. Specifically, compared to ours that achieves 66% of Dice on the XCAD dataset as shown in Table 2, Ma et al. (2021) reports the performance of the supervised learning method with 72% of Dice, where the result comes from the UNet (Ronneberger et al., 2015) by adapting 3-fold cross-validation over all labeled images in the XCAD dataset. Also, in the supplementary material of Kim et al. (2023), we can see that the DARL model trained in a supervised manner using a few labels of the XCAD shows 70% of Dice.

Nevertheless, it is remarkable that our method achieves high generalization performance on unseen data. To get the performance closer to the supervised methods and be used for presumptive clinical guidance, developing self-supervised methods for the blood vessel segmentation task needs to be continued. Also, since our method can utilize various image domain data and capture the semantic information related to pseudo masks, if label-like pseudo masks can be synthesized, future

work in this area could be performed to apply our model to other medical imaging modalities or 3D image segmentation tasks. To the best of our knowledge, this paper is the first self-supervised method that uses multiple vessel image domains, so we expect that our method can be a baseline platform for developing self-supervised learning algorithms.

6. Conclusion

We propose a novel label-free C-DARL model that achieves reliable blood vessel segmentation performance on multi-domain vessel images. Our model addresses the limitations of previously introduced DARL by effectively simplifying two different vessel image generation and vessel segmentation tasks, enabling self-supervised learning of blood vessel features using vessel images in a variety of domains. We further introduce mask-based contrastive learning to the diffusion adversarial learning framework, by leveraging both the synthetic fractal and the estimated segmentation masks to intensely extract vessel representation under the label-free condition. Extensive experiments conducted on various vessel segmentation benchmarks demonstrate that our proposed C-DARL outperforms existing label-free vessel segmentation methods, offering noise robustness and promising generalization performance across diverse vessel datasets. This performance with a variety of clinical settings for vessel segmentation suggests that our model can be a platform for future studies to provide clinicians with presumptive guidance during the diagnostic process for a variety of organs and diseases, as well as aid in detection, treatment planning, navigation, or reference data augmentation for various clinical interventional and diagnostic tasks without requiring any labeling process.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jong Chul Ye reports financial support was provided by National research Foundation of Korea. Yujin Oh reports financial support was provided by National research Foundation of Korea. Jong Chul Ye has patent #10-2023-0086537 issued to Diffusion Adversarial Representation Learning for Self-supervised Vessel Segmentation. Boah Kim has patent #10-2023-0086537 issued to Diffusion Adversarial Representation Learning for Self-supervised Vessel Segmentation. Yujin Oh has patent #10-2023-0086537 issued to Diffusion Adversarial Representation Learning for Self-supervised Vessel Segmentation. Boah Kim previously studied at Korea Advanced Institute of Science and Technology (KAIST) for Ph.D. Ronald M. Summers received royalties for patents or software licenses from iCAD, Philips, ScanMed, PingAn, Translation Holdings, and MGB, and his lab received research funding through a Cooperative Research and Development Agreement from PingAn.

Data availability

We used the public datasets except for the AAIR dataset which is confidential. The source code is available at https://github.com/boahK/MEDIA_CDARL.

Acknowledgments

This research was supported in part by the Intramural Research Program of the National Institutes of Health, Clinical Center, United States, and in part by the National Research Foundation of Korea under Grant NRF-2020R1A2B5B03001980.

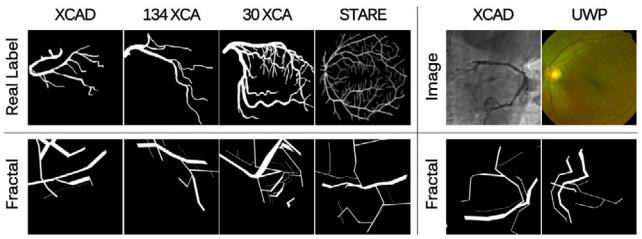


Fig. A.1. Visualization of the randomly synthesized fractal masks, which resemble real labels and vessel images from various datasets.

Table B.1
Quantitative results of the noise corruption study on the XCAD test dataset. For each Poisson and Gaussian noise, we use different levels of λ and σ, respectively. The models of SSVS (Ma et al., 2021), DARL (Kim et al., 2023), and DA (Mahmood et al., 2019) are used for the baseline methods. "w/CA+background" denotes the model trained using only coronary angiography (CA) data with their background images, and "w/multi-domain" means the model trained using multi-domain datasets.

Level	Metric	Self-supervised learning				
		w/CA+background			w/multi-domain	
		SSVS	DARL	C-DARL	DA	C-DARL
Poisson noise (λ)						
10	IoU	$0.363_{\pm 0.073}$	$0.437_{\pm 0.074}$	$0.467_{\pm 0.082}$	$0.412_{\pm 0.064}$	$0.490_{\pm 0.089}$
	Dice	$0.527_{\pm 0.079}$	$0.604_{\pm 0.073}$	$0.632_{\pm 0.079}$	$0.579_{\pm 0.067}$	$0.653_{\pm 0.082}$
	Precision	$0.554_{\pm 0.098}$	$0.642_{\pm 0.123}$	$0.728_{\pm 0.112}$	$0.584_{\pm 0.100}$	$0.755_{\pm 0.107}$
	Recall	$0.526_{\pm 0.124}$	$0.590_{\pm 0.086}$	$0.570_{\pm 0.095}$	$0.599_{\pm 0.107}$	$0.587_{\pm 0.102}$
20	IoU	$0.235_{\pm 0.065}$	$0.429_{\pm 0.074}$	$0.449_{\pm 0.082}$	$0.344_{\pm 0.071}$	$0.461_{\pm 0.096}$
	Dice	$0.374_{\pm 0.083}$	$0.596_{\pm 0.074}$	$0.615_{\pm 0.079}$	$0.502_{\pm 0.083}$	$0.625_{\pm 0.091}$
	Precision	$0.437_{\pm 0.101}$	$0.649_{\pm 0.120}$	$0.726_{\pm 0.109}$	$0.563_{\pm 0.105}$	$0.754_{\pm 0.108}$
	Recall	$0.342_{\pm 0.100}$	$0.570_{\pm 0.090}$	$0.545_{\pm 0.096}$	$0.483_{\pm 0.109}$	$0.546_{\pm 0.112}$
30	IoU	$0.140_{\pm 0.047}$	$0.421_{\pm 0.074}$	$0.429_{\pm 0.083}$	$0.277_{\pm 0.070}$	$0.432_{\pm 0.101}$
	Dice	$0.240_{\pm 0.070}$	$0.588_{\pm 0.075}$	$0.596_{\pm 0.082}$	$0.419_{\pm 0.088}$	$0.595_{\pm 0.100}$
	Precision	$0.300_{\pm 0.096}$	$0.656_{\pm 0.117}$	$0.721_{\pm 0.108}$	$0.503_{\pm 0.112}$	$0.755_{\pm 0.112}$
	Recall	$0.209_{\pm 0.069}$	$0.550_{\pm 0.092}$	$0.519_{\pm 0.098}$	$0.391_{\pm 0.099}$	$0.506_{\pm 0.119}$
Gaussian noise (σ)						
10	IoU	$0.298_{\pm 0.077}$	$0.429_{\pm 0.075}$	$0.455_{\pm 0.082}$	$0.366_{\pm 0.072}$	$0.466_{\pm 0.095}$
	Dice	$0.453_{\pm 0.090}$	$0.596_{\pm 0.075}$	$0.620_{\pm 0.080}$	$0.527_{\pm 0.082}$	$0.630_{\pm 0.089}$
	Precision	$0.523_{\pm 0.109}$	$0.640_{\pm 0.122}$	$0.723_{\pm 0.112}$	$0.566_{\pm 0.104}$	$0.754_{\pm 0.109}$
	Recall	$0.418_{\pm 0.119}$	$0.578_{\pm 0.090}$	$0.555_{\pm 0.096}$	$0.522_{\pm 0.111}$	$0.555_{\pm 0.111}$
20	IoU	$0.166_{\pm 0.055}$	$0.413_{\pm 0.075}$	$0.420_{\pm 0.084}$	$0.236_{\pm 0.077}$	$0.413_{\pm 0.104}$
	Dice	$0.280_{\pm 0.079}$	$0.580_{\pm 0.077}$	$0.586_{\pm 0.085}$	$0.366_{\pm 0.102}$	$0.577_{\pm 0.106}$
	Precision	$0.372_{\pm 0.116}$	$0.639_{\pm 0.119}$	$0.716_{\pm 0.111}$	$0.422_{\pm 0.131}$	$0.753_{\pm 0.114}$
	Recall	$0.234_{\pm 0.076}$	$0.548_{\pm 0.094}$	$0.508_{\pm 0.100}$	$0.362_{\pm 0.105}$	$0.482_{\pm 0.122}$
30	IoU	$0.093_{\pm 0.037}$	$0.394_{\pm 0.072}$	$0.384_{\pm 0.087}$	$0.151_{\pm 0.061}$	$0.360_{\pm 0.109}$
	Dice	$0.167_{\pm 0.061}$	$0.561_{\pm 0.076}$	$0.549_{\pm 0.091}$	$0.249_{\pm 0.089}$	$0.519_{\pm 0.122}$
	Precision	$0.243_{\pm 0.099}$	$0.635_{\pm 0.118}$	$0.712_{\pm 0.114}$	$0.286_{\pm 0.115}$	$0.755_{\pm 0.127}$
	Recall	$0.132_{\pm 0.050}$	$0.517_{\pm 0.091}$	$0.458_{\pm 0.102}$	$0.259_{\pm 0.092}$	$0.411_{\pm 0.126}$

Appendix A. Fractal mask synthesis

We generate fractal masks following the method introduced in Ma et al. (2021) as follows:

$$m^f = \{(Draw, s_d), (Affine, s_a), (Rotate, s_r)\}, \tag{15}$$

where hyperparameters of $s_d \in [1/4, 3/4]$, $s_a \in [0.09, 0.13]$, $s_r \in [-30, 30]$ are randomly selected for generating each fractal mask to control stochasticity. Sample fractal masks are visualized in Fig. A.1.

Appendix B. Noise corruption study

Table B.1 reports the detailed quantitative evaluation results of Fig. 9, which is the vessel segmentation performance according to the different Gaussian and Poisson noise levels.

Table C.1
Quantitative comparison results of ours to supervised learning.

Metric	C-DARL	The amount of training data for supervised learning				
		10%	30%	50%	70%	90%
XCAD dataset						
IoU	$0.502_{\pm 0.100}$	$0.500_{\pm 0.084}$	$0.545_{\pm 0.075}$	$0.573_{\pm 0.072}$	$0.586_{\pm 0.068}$	$0.593_{\pm 0.068}$
Dice	$0.662_{\pm 0.090}$	$0.662_{\pm 0.079}$	$0.702_{\pm 0.066}$	$0.726_{\pm 0.060}$	$0.737_{\pm 0.056}$	$0.742_{\pm 0.055}$
Precision	$0.775_{\pm 0.092}$	$0.691_{\pm 0.116}$	$0.723_{\pm 0.105}$	$0.742_{\pm 0.086}$	$0.774_{\pm 0.070}$	$0.769_{\pm 0.068}$
Recall	$0.583_{\pm 0.102}$	$0.646_{\pm 0.064}$	$0.689_{\pm 0.046}$	$0.715_{\pm 0.057}$	$0.704_{\pm 0.052}$	$0.719_{\pm 0.048}$
134XCA dataset						
IoU	$0.545_{\pm 0.094}$	$0.385_{\pm 0.127}$	$0.377_{\pm 0.132}$	$0.437_{\pm 0.103}$	$0.439_{\pm 0.096}$	$0.450_{\pm 0.098}$
Dice	$0.700_{\pm 0.087}$	$0.543_{\pm 0.143}$	$0.533_{\pm 0.151}$	$0.600_{\pm 0.105}$	$0.603_{\pm 0.100}$	$0.613_{\pm 0.104}$
Precision	$0.673_{\pm 0.128}$	$0.532_{\pm 0.164}$	$0.509_{\pm 0.162}$	$0.578_{\pm 0.113}$	$0.633_{\pm 0.101}$	$0.627_{\pm 0.098}$
Recall	$0.758_{\pm 0.119}$	$0.572_{\pm 0.153}$	$0.577_{\pm 0.168}$	$0.644_{\pm 0.135}$	$0.600_{\pm 0.142}$	$0.621_{\pm 0.143}$
DRIVE dataset						
IoU	$0.345_{\pm 0.061}$	$0.207_{\pm 0.060}$	$0.211_{\pm 0.058}$	$0.268_{\pm 0.051}$	$0.267_{\pm 0.049}$	$0.292_{\pm 0.052}$
Dice	$0.510_{\pm 0.067}$	$0.339_{\pm 0.080}$	$0.345_{\pm 0.077}$	$0.420_{\pm 0.063}$	$0.419_{\pm 0.062}$	$0.449_{\pm 0.063}$
Precision	$0.904_{\pm 0.090}$	$0.599_{\pm 0.077}$	$0.633_{\pm 0.068}$	$0.695_{\pm 0.071}$	$0.790_{\pm 0.065}$	$0.769_{\pm 0.078}$
Recall	$0.359_{\pm 0.063}$	$0.240_{\pm 0.072}$	$0.242_{\pm 0.069}$	$0.304_{\pm 0.060}$	$0.288_{\pm 0.054}$	$0.320_{\pm 0.057}$

Appendix C. Comparison to supervised learning

Table C.1 reports the detailed quantitative evaluation results of Fig. 16. This shows the comparison of our C-DARL model to the supervised learning models that are trained using different amounts of labeled data.

References

Baranchuk, D., Voynov, A., Rubachev, I., Khrulkov, V., Babenko, A., 2022. Label-efficient semantic segmentation with diffusion models. In: International Conference on Learning Representations. URL: https://openreview.net/forum?id=SlxSY2UZQT.

Brempong, E.A., Kornblith, S., Chen, T., Parmar, N., Minderer, M., Norouzi, M., 2022. Decoder denoising pretraining for semantic segmentation. Trans. Mach. Learn. Res. URL: https://openreview.net/forum?id=D3WI0QG7dC.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A., 2021. Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 9650–9660.

Cervantes-Sanchez, F., Cruz-Aceves, I., Hernandez-Aguirre, A., Hernandez-Gonzalez, M.A., Solorio-Meza, S.E., 2019. Automatic segmentation of coronary arteries in X-ray angiograms using multiscale analysis and artificial neural networks. Appl. Sci. 9 (24), http://dx.doi.org/10.3390/app9245507, URL: https://www.mdpi.com/2076-3417/9/24/5507.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning. PMLR, pp. 1597–1607.

Crum, W.R., Camara, O., Hill, D.L., 2006. Generalized overlap measures for evaluation and validation in medical image analysis. IEEE Trans. Med. Imaging 25 (11), 1451–1461.

Dehkordi, M.T., Sadri, S., Doosthoseini, A., 2011. A review of coronary vessel segmentation algorithms. J. Med. Signals Sens. 1 (1), 49.

Delibasis, K.K., Kechriniotis, A.I., Tsonos, C., Assimakis, N., 2010. Automatic model-based tracing algorithm for vessel segmentation and diameter estimation. Comput. Methods Programs Biomed. 100 (2), 108–122.

Fan, Z., Mo, J., Qiu, B., Li, W., Zhu, G., Li, C., Hu, J., Rong, Y., Chen, X., 2019. Accurate retinal vessel segmentation via octave convolution neural network. arXiv preprint arXiv:1906.12193.

Fan, J., Yang, J., Wang, Y., Yang, S., Ai, D., Huang, Y., Song, H., Hao, A., Wang, Y., 2018. Multichannel fully convolutional network for coronary artery segmentation in X-ray angiograms. IEEE Access 6, 44635–44643.

Hao, D., Ding, S., Qiu, L., Lv, Y., Fei, B., Zhu, Y., Qin, B., 2020. Sequential vessel segmentation via deep channel attention network. Neural Netw. 128, 172–187. http://dx.doi.org/10.1016/j.neunet.2020.05.005, URL: https://www.sciencedirect.com/science/article/pii/S0893608020301672.

He, K., Fan, H., Wu, Y., Xie, S., Girshick, R., 2020. Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9729–9738.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Adv. Neural Inf. Process. Syst. 33, 6840–6851.

Hoover, A., Kouznetsova, V., Goldbaum, M., 2000. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans. Med. Imaging 19 (3), 203–210.

Hu, H., Cui, J., Wang, L., 2021. Region-aware contrastive learning for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 16291–16301.

Huang, Z., Chan, K.C., Jiang, Y., Liu, Z., 2023. Collaborative diffusion for multi-modal face generation and editing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6080–6090.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1125–1134.

Jiang, Y., Zhang, H., Tan, N., Chen, L., 2019. Automatic retinal blood vessel segmentation based on fully convolutional neural networks. Symmetry 11 (9), 1112.

Kim, B., Oh, Y., Ye, J.C., 2023. Diffusion adversarial representation learning for self-supervised vessel segmentation. In: The Eleventh International Conference on Learning Representations. URL: https://openreview.net/forum?id=H0gdPxSwkPb.

Kingma, D., Ba, J., 2015. Adam: A method for stochastic optimization. In: International Conference on Learning Representations (ICLR). San Diega, CA, USA.

Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G., 2009. A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. Med. Image Anal. 13 (6), 819–845.

Ma, Y., Hua, Y., Deng, H., Song, T., Wang, H., Xue, Z., Cao, H., Ma, R., Guan, H., 2021. Self-supervised vessel segmentation via adversarial learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7536–7545.

Mahmood, F., Borders, D., Chen, R.J., McKay, G.N., Salimian, K.J., Baras, A., Durr, N.J., 2019. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. IEEE Trans. Med. Imaging 39 (11), 3257–3267.

Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S., 2017. Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2794–2802.

Meijering, E., Jacob, M., Sarria, J.-C., Steiner, P., Hirling, H., Unser, e.M., 2004. Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. Cytom. A: J. Int. Soc. Anal. Cytol. 58 (2), 167–176.

Nasr-Esfahani, E., Samavi, S., Karimi, N., Soroushmehr, S.R., Ward, K., Jafari, M.H., Felfeliyan, B., Nallamothu, B., Najarian, K., 2016. Vessel extraction in X-ray angiograms using deep learning. In: The 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 643–646.

Oh, Y., Ko, E.S., Park, H., 2022. Semi-supervised breast lesion segmentation using local cross triplet loss for ultrafast dynamic contrast-enhanced MRI. In: Proceedings of the Asian Conference on Computer Vision. pp. 2713–2728.

Oord, A.v.d., Li, Y., Vinyals, O., 2018. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.-Y., Li, S.-W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P., 2023. DINOv2: Learning robust visual features without supervision. arXiv:2304.07193.

Park, T., Efros, A.A., Zhang, R., Zhu, J.-Y., 2020. Contrastive learning for unpaired image-to-image translation. In: European Conference on Computer Vision. Springer, pp. 319–345.

Park, T., Liu, M.-Y., Wang, T.-C., Zhu, J.-Y., 2019. Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2337–2346.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems, Vol. 32.

Rahman, A., Valanarasu, J.M.J., Hacihaliloglu, I., Patel, V.M., 2023. Ambiguous medical image segmentation using diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11536–11546.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M.A., Van Ginneken, B., 2004. Ridge-based vessel segmentation in color images of the retina. IEEE Trans. Med. Imaging 23 (4), 501–509.

Taghizadeh Dehkordi, M., Doost Hoseini, A.M., Sadri, S., Soltanianzadeh, H., 2014. Local feature fitting active contour for segmenting vessels in angiograms. IET Comput. Vis. 8 (3), 161–170.

Ulyanov, D., Vedaldi, A., Lempitsky, V., 2016. Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022.

Wang, W., Bao, J., Zhou, W., Chen, D., Chen, D., Yuan, L., Li, H., 2022. Semantic image synthesis via diffusion models. arXiv:2207.00050.

Wu, Z., Xiong, Y., Yu, S.X., Lin, D., 2018. Unsupervised feature learning via non-parametric instance discrimination. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3733–3742.

Wu, C., Zou, Y., Yang, Z., 2019. U-GAN: Generative adversarial networks with U-net for retinal vessel segmentation. In: The 14th International Conference on Computer Science & Education. IEEE, pp. 642–646.

Yang, S., Kweon, J., Roh, J.-H., Lee, J.-H., Kang, H., Park, L.-J., Kim, D.J., Yang, H., Hur, J., Kang, D.-Y., et al., 2019. Deep learning segmentation of major vessels in X-ray coronary angiography. Sci. Rep. 9 (1), 1–11.

Ye, Y., Zhang, J., Chen, Z., Xia, Y., 2022. DeSD: Self-supervised learning with deep self-distillation for 3D medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 545–555.

Yeganeh, Y., Farshad, A., Weinberger, P., Ahmadi, S.-A., Adeli, E., Navab, N., 2023. DIAMANT: Dual image-attention map encoders for medical image segmentation. arXiv preprint arXiv:2304.14571.

Yoo, T.K., 2020. Deep learning-based style transfer from ultra-widefield to traditional fundus photography. URL: https://data.mendeley.com/datasets/m3kg8p8cxf/2.

Zhao, F., Chen, Y., Hou, Y., He, X., 2019. Segmentation of blood vessels using rule-based and machine-learning-based methods: a review. Multimedia Syst. 25, 109–118.

Zhong, Y., Yuan, B., Wu, H., Yuan, Z., Peng, J., Wang, Y.-X., 2021. Pixel contrastive-consistent semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7273–7282.