# FRS-Nets: Fourier Parameterized Rotation and Scale Equivariant Networks for Retinal Vessel Segmentation

Zihong Sun, Qi Xie, and Deyu Meng

Abstract—With translation equivariance, convolution neural networks (CNNs) have achieved great success in retinal vessel segmentation. However, some other symmetries of the vascular morphology are not characterized by CNNs, such as rotation and scale symmetries. To embed more equivariance into CNNs and achieve the accuracy requirement for retinal vessel segmentation, we construct a novel convolution operator (FRS-Conv), which is Fourier parameterized and equivariant to rotation and scaling. Specifically, we first adopt a new parameterization scheme, which enables convolutional filters to arbitrarily perform transformations with high accuracy. Secondly, we derive the formulations for the rotation and scale equivariant convolution mapping. Finally, we construct FRS-Conv following the proposed formulations and replace the traditional convolution filters in U-Net and Iter-Net with FRS-Conv (FRS-Nets). We faithfully reproduce all compared methods and conduct comprehensive experiments on three public datasets under both indataset and cross-dataset settings. With merely 13.9% parameters of corresponding baselines, FRS-Nets have achieved state-of-the-art performance and significantly outperform all compared methods. It demonstrates the remarkable accuracy, generalization, and clinical application potential of FRS-Nets.

**Index Terms**—Convolution neural networks, equivariance, group, retinal vessel segmentation

# 1 Introduction

The morphological changes of the retinal vasculature are substantially relevant to certain diseases, such as diabetes, glaucoma, and hypertension [1]. With retinal vessel segmentation, clinicians can observe abnormal symptoms and diagnose them in an early stage. However, in clinical practice, it's subjective and laborious for professional ophthalmologists to manually annotate fundus images. Therefore, it's of great clinical significance to develop an automatic algorithm for retinal vessel segmentation [2].

Yet vessel segmentation is still a challenging task for the following reasons. First, the low contrast of fundus images can make it hard to distinguish vessels from the background. Second, pathological exudates and hemorrhage can be easily misclassified as vessels. Third, the great variation of orientations and scales in the complex morphology of

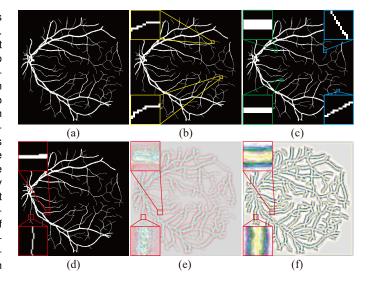


Figure 1. Illustration of the symmetries existed in retinal vessels and the performance of the rotation and scale equivariance. (a) A typical retinal vessel image. (b) Local patterns with the translation symmetry. (c) Local patterns with the rotation symmetry and the scale symmetry, respectively. (d) Local patterns with the rotation and scale symmetry. (e)-(f) Outputs of randomly initialized CNN and the proposed Fourier parameterized rotation and scale equivariant convolution network.

the retinal vasculature can severely affect the segmentation results [3].

Traditional methods are mainly based on unsupervised image processing techniques, which heavily rely on hand-crafted features and domain knowledge [4]–[6]. Recently, deep learning has achieved remarkable performance in medical image segmentation, especially the convolution neural networks-based methods (CNNs) [7].

One of the most important reasons why CNNs can achieve such great success is attributed to the skillful implementation of the convolution operator. With a dynamic shift window performing the convolution operation, CNNs rationally save parameters by weight sharing, which facilitates the generalization, while more significantly, embed the translation equivariance into neural networks [8]. Translation equivariance means when we shift the input images, all intermediate feature maps and output images of CNNs will perform the same shift operation accordingly. That

Z. Sun, Q. Xie, and D. Meng are with the School of Mathematics and Statistics and Ministry of Education Key Laboratory of Intelligent Networks and Network Security, Xi'an Jiaotong University, Shaanxi 710049, China. Email: szhc0gk@stu.xjtu.edu.cn, xie.qi@mail.xjtu.edu.cn, and dymeng@mail.xjtu.edu.cn

is, when similar patterns appear at different positions in an image, CNNs will always give similar responses. It's a distinct advantage for image segmentation, since this kind of translation symmetry is a universal image prior and widely exists in fundus images, as shown in Fig. 1(a).

Besides the translation symmetry, there're still other symmetries extensively existed in fundus images, like rotation and scale symmetries, as shown in Fig. 1(b). Just like translation symmetry can bring great advantages for CNNs, as compared with fully connected networks, these rotation and scale symmetries should also be very helpful for segmentation tasks. However, for rotation and scaling transforms, CNNs are not able to directly make use of these transformation symmetries for parameter saving or performance improvement. Data augmentation is a widely used method to cope with the issue [9]. In retinal vessel segmentation, [10] demonstrates that excessive data augmentation can drive vanilla U-Net [11] to near state-of-the-art performance. In spite of the great advantages, data augmentation is an indirect way to characterize the symmetries of the retinal vessels, which will cause additional training costs and the trained models are still without the equivariance guarantee. Previous works attempt to pile up convolution filters of different scales as the spatial pyramid to capture multi-scale features, or design constraints to guide filters to learn more symmetries [12]–[14]. By contrast, these methods have partially characterized local symmetries, however, in a heuristic approach.

Very recently, equivariant convolution neural networks (E-CNNs) have been proposed to internally embed more kinds of transformation equivariance [15]. By constructing a map from the transformation group domain to the real number field, the group convolution-based CNNs sufficiently incorporate the convolution operator with the corresponding equivariance. Besides, the weight sharing between channels further prompts E-CNNs to achieve stronger generalization.

Nonetheless, for retinal vessel segmentation, two main issues should be considered for deep network design. On one hand, as shown in Fig. 1(c), the rotation and scale symmetry usually appear in the retinal vasculature simultaneously. It indicates the need for the group convolution framework to satisfy the rotation and scale equivariance simultaneously. Similar to translation equivariance, the rotation and scale equivariance implies, when similar local patterns are rotated and/or rescaled, the network will give similar responses. On the other hand, due to the complexity of the retinal vascular morphology, the fitting capability of the deep network is fairly important for the accuracy of segmentation, while most of the equivariant CNNs exploit filter parameterization with insufficient representation accuracy. Very recently, the Fourier series expansion-based filter parameterization (FSE-FP) has been shown to be capable to achieve high representation accuracy [16]. However, the current FSE-FP-based convolution is only designed for rotation equivariance. Therefore, there is still room for improvement in current equivariant CNNs in retinal vessel segmentation tasks.

To address the problems as aforementioned, this study explores high-accuracy rotation and scale equivariant convolutions for retinal vessel segmentation. The main contribution of this work can be summarized as follows:

- 1) We proposed a rotation and scale equivariant convolution framework for retinal vessel segmentation, named Fourier parameterized rotation and scale equivariant convolution (FRS-Conv). The key idea is to exploit the carefully designed FSE-FP for adopting rotation and scale transformation on convolution filters simultaneously. The proposed equivariant convolution achieves pixel-level accuracy to characterize local symmetries of blood vessels. As shown in Fig. 1(d) and (e), the rotation and scale equivariant convolution filters tend to characterize the retinal vessels in different orientations and scales with similar structured patterns, while the output of traditional convolutions presents chaotic and unstructured local patterns.
- 2) We further construct high-accuracy rotation and scale equivariant CNNs for vessel segmentation, based on the proposed FRS-Convs. Specifically, by adopting FRS-Convs instead of the traditional convolution filters in U-Net [11] and Iter-Net [17], without changing any architectures of backbone networks, we obtain the FRS-Conv-based neural networks (FRS-Nets), named FRS U-Net and FRS Iter-Net respectively. By rational weight sharing between channels, our FRS-Nets have successfully achieved fewer parameters, faster convergence, and stronger generalization.
- 3) We have faithfully reproduced multiple state-of-theart methods and conducted comprehensive experiments under identical conditions. The experiment results indicate FRS-Nets evidently outperform all contrast methods with merely 13.9% parameters of corresponding baselines. It illustrates the potential of our method, as a basic tool for deep network design, in future clinical applications.

# 2 RELATED WORK

# 2.1 Retinal Vessel Segmentation

Early studies about retinal vessel segmentation are mostly based on the traditional image processing [4]–[6], such as handcraft filters and morphological operations, which generally have an unsatisfactory performance in extreme cases.

Recently, deep learning-based methods have achieved remarkable performance, among which U-Net [11] is one of the most widely used methods in medical imaging segmentation. Despite the excellent accuracy compared with the traditional methods, U-Net is still not effective enough to handle the segmentation for complex retinal vasculature.

To improve the segmentation connectivity of retinal vessels, Iter-Net [17] cascades a U-Net with several mini U-Nets and shrinks the number of channels, which makes the network deep enough and lightweight meanwhile. U-Net++ [18] redesigned the skip connections in U-Net, which enables more sufficient information fusion between multiple scales. To alleviate the loss of some spatial information caused by consecutive pooling operations in U-Net, CE-Net [12] adopted a multi-scale branches structure with dilated dense blocks and residual poolings. In CS-Net [19], spatial attention and channel attention were used to advance the fusion of local and global information, which benefits the capture of the vascular morphology. SCS-Net [20] proposed a feature aggregation module to adjust the receptive fields adaptively, and replaced the skip connection in U-Net with a feature fusion module, which flexibly fuses the spatial

and semantic information, and suppresses the background noise.

Very recently, in order to reduce the time and the required experience in the network design, Genetic U-Net [21] first applied the evolutionary neural architecture search (NAS) to retinal vessel segmentation and achieved an excellent improvement with a compact network structure. Besides, DE-DCGCN-EE [22] constructed a dynamic-channel graph CNN with dual encoders and edge enhancement, which alleviates the loss of the edge information and utilizes topological relations in feature maps. LIOT [23] further improved the generalization of Iter-Net with a novel imaging preprocess, which yet is certainly sensitive to curvilinear structures and invariant to contrast perturbation.

In this study, different from the previous methods, we don't change network architectures, but merely replace the traditional convolution in U-Net and Iter-Net with our proposed FRS-Conv. With the most basic network architecture and training strategy, FRS-Nets have achieved state-of-theart performance in both accuracy and generalization.

# 2.2 Equivarient CNNs

Compared with multi-layer perceptrons (MLPs) [24], CNNs successfully embed translation equivariance into networks and have achieved great improvement in image processing. It arouses widespread interest in how to equip networks with more equivariance. Data augmentation [7] is the most widely used approach, which enables networks to learn symmetries by enriching datasets with multiple transformations. In retinal vessel segmentation, [10] indicates that, with sufficient data augmentation, vanilla U-Net still can achieve near state-of-the-art performance. The method is straightforward and embeds networks with global symmetries, but accordingly, it's time-consuming, and moreover, lacks the characterization of the symmetries in local patterns.

One of the first CNN-based networks that focus on the local scale symmetry is SiCNN [14]. It interpolated convolution kernels in multi-columns to force filters to have identical patterns in different scales. For retinal vessel segmentation, DRIS-GP [13] optimized the convolution filters under designed constraints to learn the rotation and scale symmetries in the vessel morphology. This category of methods characterizes local symmetries in a heuristic approach, which more or less limits the expressiveness of equivariance.

A series of recent works attempt to incorporate equivariance into networks by utilizing the symmetry of groups. G-CNN [15] firstly constructed the group equivariant framework, which achieves the equivariance to the discrete  $\pi/2$  rotation. HexaConv [25] further extended the discrete rotation group to  $\pi/3$  by changing the image representation into hexagonal lattices. Based on the scale-spaces theory, DSS [26] constructed the scale equivariant framework under the formulation of semi-groups, and used dilated convolutions to represent multi-scales. It restricts the method only to integer rescaling factors. By rearranging the convolution structure, these methods embed the equivariance in an explicit expression. However, bounded by the traditional discrete convolution operation, it's still difficult for the methods to represent groups adequately.

Currently, the filter parameterization strategy is proposed to address the aforementioned problem. By utilizing harmonic functions as bases, SFCNN [27] and E2-CNN [28] enabled the parameterized convolution filters to arbitrarily rotate to any angle. PDO-eConv [29] utilized the partial differential operators to impose the rotation equivariance and firstly derived the boundary of the approximation error for the rotation discretization process. With the bases of Hermite polynomials, SESN [30] expanded the factors of scale groups from integers to the continuous domain. The parameterization approach breaks the limitation of the traditional convolution, and makes it possible to transform filters continuously. However, these methods still suffer from problems in the expression accuracy of the parameterized bases, which results in an inferior performance for tasks that require high accuracy. To address the problem, the F-Conv [16] proposed the aforementioned FSE-FP, which enables the parameterized filters to achieve high accuracy in both static and transformed cases.

Considering the geometric symmetries existed in the retinal vasculature and the requirement for high accuracy, we exploit FSE-FP and extend the current equivariant convolution framework to satisfy the rotation and scale equivariance simultaneously. Then, we obtain FRS-Conv, which achieves a fine performance in retinal vessel segmentation.

#### 3 METHODOLOGY

#### 3.1 Rotatable and Scalable Convolution Filter

The key issue for constructing the rotation and scale equivariant convolution is to transform convolution filters into different orientations and sizes while keeping filters learnable in CNNs. However, traditional convolution filters are usually discrete 2D arrays, which are hard to be rotated or rescaled. Thus, it's necessary to represent traditional convolution filters with 2D continuous functions instead of the original discrete 2D arrays, while the latent 2D functions are supposed to be learnable. Formally, as shown in Fig. 2(c) and (d), the element of a traditional convolution filter  $\Psi \in \mathbb{R}^{p \times p}$  is represented as:

$$\Psi_{ij} = \psi(x_{ij}), \forall i, j = 1, 2, \cdots, p, \tag{1}$$

where  $x_{ij} = \left[(i-\frac{p+1}{2})h,(j-\frac{p+1}{2})h\right]^T \in \mathbb{R}^2$  denotes the  $p \times p$  mesh grid on 2D spatial coordinates, which is origin centered. p and h are the filter size and the mesh size, respectively.  $\psi$  is the latent 2D continuous function, which is designed to be learnable. We call  $\Psi$  a discretization of  $\psi(x)$ .

In this paper, we exploited the linear combination-based filter parameterization technique [16], [27]–[29] for representing  $\psi(x)$ . Specifically, as shown in Fig. 2(a) and (b), we have

$$\psi(x) = \sum_{n=1}^{N} w_n \phi_n(x), \tag{2}$$

where  $\phi_n(x)$ ,  $n=1,2,\cdots,N$  are fixed basis functions, N is the number of basis functions for representing  $\psi(x)$ , and  $w_n$  are learnable coefficient parameters.

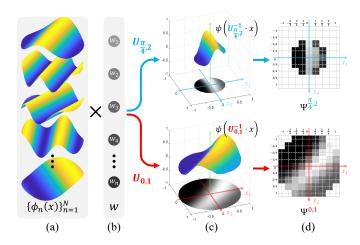


Figure 2. Illustration of the generating filters  $\Psi$  at different orientations and sizes. (a) The basis function set. (b) The shared coefficient parameters. (c) Latent functions  $\psi(x)$  at different orientations and sizes. (d) Generated filters at different orientations and sizes.

As shown in Fig. 2(b)(c) and (d), by adopting coordinate transformation on  $\psi(x)$  (or  $\phi_n(x)$ ), we can obtain the rotated and rescaled filter  $\Psi^{\theta,s}$ , whose elements are defined as:

$$\Psi_{ij}^{\theta,s} = \psi\left(U_{\theta,s}^{-1} \cdot x_{ij}\right) = \sum_{n=1}^{N} w_n \phi_n\left(U_{\theta,s}^{-1} \cdot x_{ij}\right), \quad (3)$$

where  $\forall i, j = 1, 2, \dots, p$  and  $x_{ij}$  are points on the mesh grid, and  $U_{\theta,s}$  is the rotation and scale transformation matrix, i.e.,

$$U_{\theta,s} = s \cdot \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}. \tag{4}$$

It's straightforward to see that, as convolution filters in CNNs,  $\Psi^{\theta,s}$  are not only rotatable and scalable with different  $\theta$  and s, but also learnable with shared parameters  $w_n$ .

Meanwhile, due to the complexity of the retinal vascular morphology, there is a high requirement for representation accuracy in retinal vessel segmentation. Therefore, the choice of the basis function set aforementioned in (1) and (2) is quite crucial, which is highly relevant to the representation capability of parameterized filters [16], [27]–[29].

Very recently, [16] has shown that the enhanced Fourier basis function set can represent any  $\Psi \in \mathbb{R}^{p \times p}$  without representation error, and first time achieve the high representation accuracy that satisfies the requirements for pixellevel vision tasks. Therefore, we adopt the enhanced Fourier basis function set [16] to construct the proposed rotation and scale equivariant convolution. As well as we know, this is the first rotation and scale equivariant convolution framework with high accuracy that is suitable for retinal vessel segmentation.

# 3.2 Rotation and Scale Equivariant Convolution

To make the formulation easier to understand, we first introduce the proposed Fourier parameterized rotation and scale equivariant convolution (FRS-Conv) in the continuous domain and then introduce its discretization.

# 3.2.1 FRS-Conv in continuous domain

By adopting the proposed parameterized convolution filters into the framework of group convolutions [15], [16], [27]–[29], [31], we can construct novel rotation and scale equivariant convolutions. In particular, the initial equivariant convolution for the first layer of the network and the intermediate equivariant convolutions for other layers will be needed.

Similar to the relationship between  $\Psi$  and  $\psi(x)$  in (1), we model input images as 2D continuous functions r(x), whose discretization is the commonly used 2D array type of images I. We model feature maps as 2D mappings  $f_{(\theta,s)}(x)$ , which are indexed by  $\theta$  and s, two extra dimensions in equivariant CNNs. The discretization of  $f_{(\theta,s)}(x)$  is  $F^{\theta,s}$ , which would be a 4D tensor with 2 spatial dimensions and 2 index dimensions.

Specifically, in the continuous domain, the initial equivariant convolution  $\Psi^R$  maps the input images r to the feature maps f, i.e.,  $f_{(\theta,s)}(x) = [\Psi^R \circ r]_{(\theta,s)}(x)$ . Formally, it is defined as:

$$[\Psi^R \circ r]_{(\theta,s)}(x) = \int_{\mathbb{R}^2} \mu^{-2s} \psi\left(U_{\theta,\mu^s}^{-1}\tilde{x}\right) \cdot r(x+\tilde{x}) d\sigma(\tilde{x}), \tag{5}$$

where  $\mu$  is the step size for scaling the filters and  $\sigma$  denotes the Haar measure on  $\mathbb{R}^2$ . By substituting filter function  $\psi(x)$  with (2), we can find that it's easy to perform gradient feedback on the parameters  $w_n$ , when we apply this convolution in constructing CNNs.

The intermediate equivariant convolution  $\Psi^H$  maps the input feature maps  $\hat{f}$  to the output feature maps  $\hat{f}$ , i.e.,  $\hat{f}_{(\theta,s)}(x) = [\Psi^H \circ f]_{(\theta,s)}(x)$ . Formally, it is defined as:

$$[\Psi^{H} \circ f]_{(\theta,s)}(x) = \int_{R} \int_{S} \int_{\mathbb{R}^{2}} \mu^{-2s} \psi_{(\tilde{\theta}-\theta,\tilde{s}-s)} \left( U_{\theta,\mu^{s}}^{-1} \tilde{x} \right)$$

$$\cdot f_{(\tilde{\theta},\tilde{s})}(x+\tilde{x}) \sigma(\tilde{x}) \sigma(\tilde{s}) \sigma(\tilde{\theta}),$$
(6)

where R is the rotation transformation group, whose elements are defined by  $\theta$ ; S is the scale transformation group, whose elements are defined by s.  $\psi_{(\theta,s)}(x)$  defines the parameterized filter with index  $\theta$  and s in the 2 extra dimensions of filters in equivariant convolution. Similar to (5), by substituting the filter functions  $\psi_{(\theta,s)}(x)$  in (2), the coefficient parameters  $w_n$  are also learnable.

**Equivariance Analysis.** Based on the theorem analysis in previous works [15], [16], [27]–[29], [31], we have the following results for the proposed FRS-Conv in (5) and (6).

**Lemma 1.** (5) and (6) satisfy following equations:

$$\Psi^{R} \circ \pi_{\hat{\theta},\hat{s}}^{R}[r] = \pi_{\hat{\theta},\hat{s}}^{H}[\Psi^{R} \circ r],$$

$$\Psi^{H} \circ \pi_{\hat{\theta},\hat{s}}^{H}[f] = \pi_{\hat{\theta},\hat{s}}^{H}[\Psi^{H} \circ f].$$
(7)

where  $\pi^R_{\hat{\theta},\hat{s}}$  and  $\pi^H_{\hat{\theta},\hat{s}}$  are the transformations<sup>1</sup> (i.e., group elements in R and S, indexed by  $\hat{\theta}$  and  $\hat{s}$ ) acting on input images and feature maps respectively.

It implies that the convolution filters constructed under (5) and (6) are equivariant with respect to the group elements  $\hat{\theta}$  and  $\hat{s}$ . In other words, for similar patterns at different orientations and sizes in images, the proposed FRS-Conv will always give similar responses. The proof of

$$1. \ \pi^R_{\hat{\theta},\hat{s}}[r](x) = r\left(U^{-1}_{\hat{\theta},\mu^{\hat{s}}}x\right), \pi^H_{\hat{\theta},\hat{s}}[f]_{(\theta,s)}(x) = f_{(\theta-\hat{\theta},s-\hat{s})}\left(U^{-1}_{\hat{\theta},\mu^{\hat{s}}}x\right).$$

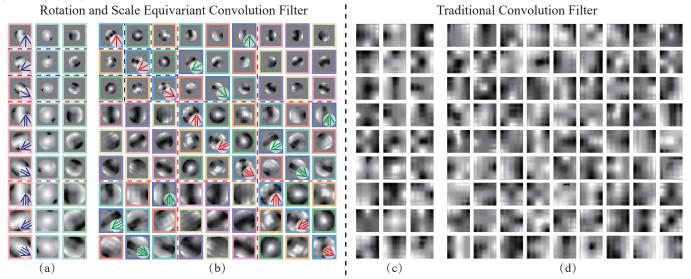


Figure 3. Illustration of the FRS-Conv, compared with the traditional convolution. (a)(b) The initial and intermediate equivariant convolution filters of FRS-Conv, with 3/9 in/out channels and 9/9 in/out channels respectively. The kernels with the same border color share the same weights. The discrete rotation group R is  $\{2\pi/3, i=1,2,3\}$ . The arrow orientation indicates the rotation angle of the filter. The discrete scale group R has three scale levels, which are indexed by the number of arrow shafts. (c)(d) The corresponding traditional convolution filters for comparison, which have no pre-designed structures.

Lemma 1 follows the same framework as the proof of the rotation equivariance in [16], and more details can be found in [16].

#### 3.2.2 FRS-Conv in discrete domain

When applying FRS-Conv to digital images, the discrete versions of (5) and (6) are necessary. By replacing integral with summation, 2D continuous functions with 2D arrays (such as (1)), and continuous transformation groups with discrete subgroups<sup>2</sup> respectively, we can easily obtain the discrete FRS-Conv.

Due to the space limitation, instead of the formal definitions, we provide the illustration of the discrete FRS-Conv filters in Fig. 3 in comparison with the traditional convolution filters, and one can refer to [16] for more details. From Fig. 3(a) and (b), we can observe that, in FRS-Conv, a series of convolution filers at different orientations and sizes share the same pattern, while the filters in the traditional convolution are independent of each other. Besides, the arrangement of filters in FRS-Conv is carefully designed, e.g., from Fig 3(b), we can see that the FRS-Conv filters with the same pattern are implemented by cyclically shifting along the rotation and scale dimensions, which is consistent with (6).

### 3.3 Network Architecture and Loss Function

The main contribution of this work is to develop a novel convolution filter for retinal vessel segmentation, rather than a network architecture. Therefore, we simply replace the traditional convolution filters in typical methods, U-Net and Iter-Net, with our proposed FRS-Conv and don't change any other network architectures for a fair comparison. Then we obtain two FRS-Nets, i.e., FRS U-Net and FRS Iter-Net.

2. A certain range of scales needs to be set for the subgroup of T.

As for the loss function, we only apply the binary crossentropy loss, as shown:

$$L = \sum_{i} (-y_i \log(p_i) - (1 - y_i) \log(1 - p_i))$$
 (8)

where y denotes the binary ground truth and p is the predicted probability.

#### 4 EXPERIMENTS

#### 4.1 Datasets and Evaluation Metrics

We conduct experiments on the following three widely used public datasets: DRIVE [32], STARE [33], CHASE\_DB1 [34].

DRIVE consists of 40 fundus images with a resolution of  $584 \times 565$ , which are divided into 20 training and 20 testing images. STARE is composed of 20  $700 \times 605$  retinal images divided into 16 training images and 4 testing images. CHASE\_DB1 contains 28  $999 \times 960$  images and are split into 20 for training and 8 for testing. All datasets have two expert annotations and only the first annotation is used as the ground truth according to previous works.

Field of view masks (FOVs) are offered in DRIVE, but not in STARE and CHASE\_DB1. Therefore, we generate the corresponding FOVs following [35]. In all experiments, we only calculate evaluation metrics inside FOVs for both comparison methods and ours.

For the assessment of segmentation, we choose the following five commonly used evaluation metrics: sensitivity (Se), specificity (Sp), F1-Score (F1), accuracy (Acc), and area under the receiver operating characteristic curve (AUC).

#### 4.2 Implementation Details and Comparison Methods

We performed all experiments on the Pytorch framework with an NVIDIA 3090 GPU. The Adam optimizer [36] is used with a learning rate of 0.0002.

For the hyper-parameters of FRS-Nets, we adopt p as 6, h as 0.5, the discrete rotation group R as  $\{\frac{i\pi}{4}, i=0,1,\cdots,7\}$ ,

Table 1
Numerical Results under In-Dataset Validation. the Best and Second Best are highlighted in **Bold** and *Italic* respectively.

| Method        |           | DRI       | VE ⇒ Dl   | RIVE       |            |           | STA       | $RE \Rightarrow ST$ | ARE        |            | CHASE_DB1 ⇒ CHASE_DB1 |           |           |            |            |
|---------------|-----------|-----------|-----------|------------|------------|-----------|-----------|---------------------|------------|------------|-----------------------|-----------|-----------|------------|------------|
|               | Se<br>(%) | Sp<br>(%) | F1<br>(%) | Acc<br>(%) | AUC<br>(%) | Se<br>(%) | Sp<br>(%) | F1<br>(%)           | Acc<br>(%) | AUC<br>(%) | Se<br>(%)             | Sp<br>(%) | F1<br>(%) | Acc<br>(%) | AUC<br>(%) |
| U-Net         | 79.05     | 98.14     | 82.43     | 95.71      | 98.08      | 77.81     | 98.88     | 82.06               | 97.05      | 98.90      | 78.29                 | 98.43     | 80.69     | 96.61      | 98.61      |
| Iter-Net      | 79.15     | 98.12     | 82.44     | 95.71      | 98.11      | 77.62     | 98.92     | 82.15               | 97.07      | 98.92      | 79.57                 | 98.23     | 80.63     | 96.54      | 98.57      |
| U-Net++       | 79.54     | 98.08     | 82.54     | 95.72      | 98.14      | 79.47     | 98.82     | 82.81               | 97.14      | 98.92      | 79.64                 | 98.31     | 81.01     | 96.62      | 98.57      |
| CE-Net        | 76.67     | 98.15     | 80.99     | 95.42      | 97.75      | 79.85     | 98.52     | 81.70               | 96.90      | 98.78      | 77.90                 | 98.27     | 79.80     | 96.43      | 98.30      |
| CS-Net        | 78.13     | 98.09     | 81.71     | 95.55      | 97.85      | 77.70     | 98.77     | 81.53               | 96.95      | 98.76      | 78.63                 | 98.31     | 80.39     | 96.52      | 98.44      |
| SCS-Net       | 77.79     | 98.10     | 81.53     | 95.51      | 97.72      | 76.71     | 98.84     | 81.19               | 96.92      | 98.48      | 77.15                 | 98.20     | 79.04     | 96.29      | 98.06      |
| Genetic U-Net | 78.64     | 98.20     | 82.34     | 95.71      | 98.09      | 79.94     | 98.83     | 83.15               | 97.19      | 99.07      | 79.85                 | 98.25     | 80.88     | 96.58      | 98.51      |
| DE-DCGCN-EE   | 78.40     | 98.16     | 82.08     | 95.64      | 97.99      | 73.98     | 98.96     | 80.01               | 96.79      | 98.57      | 76.25                 | 98.35     | 79.09     | 96.35      | 98.17      |
| LIOT U-Net    | 78.74     | 97.85     | 81.39     | 95.42      | 97.59      | 77.62     | 98.76     | 81.41               | 96.93      | 98.66      | 78.57                 | 98.22     | 80.01     | 96.44      | 98.33      |
| LIOT Iter-Net | 77.35     | 98.12     | 81.32     | 95.48      | 97.73      | 78.53     | 98.69     | 81.68               | 96.94      | 98.78      | 75.66                 | 98.43     | 79.07     | 96.37      | 98.26      |
| FRS U-Net     | 79.95     | 98.05     | 82.70     | 95.74      | 98.16      | 80.02     | 98.82     | 83.18               | 97.19      | 98.99      | 81.26                 | 98.26     | 81.79     | 96.72      | 98.74      |
| FRS Iter-Net  | 79.97     | 98.04     | 82.69     | 95.74      | 98.16      | 80.13     | 98.93     | 83.71               | 97.30      | 99.14      | 81.55                 | 98.22     | 81.79     | 96.71      | 98.75      |

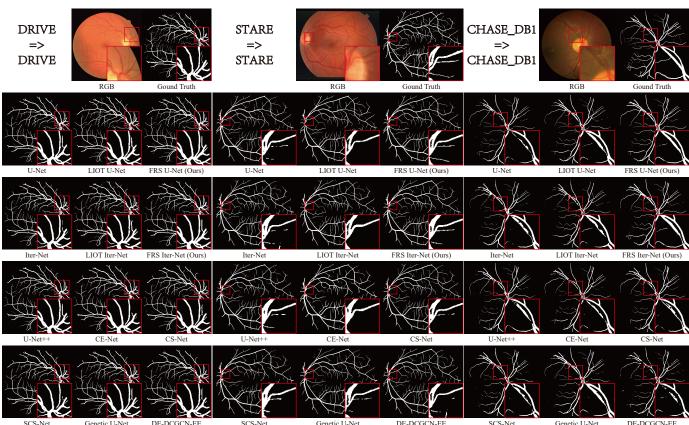


Figure 4. Some typical visualized segmentation results of in-dataset validation.

and the discrete scale group S as  $\{(\frac{5}{4})^i, i=0,\cdots,3\}$ , which means  $\mu=1.25$ .

During training, to avoid overfitting on small datasets, we apply various random data augmentations, including rotation, rescaling, flip, shearing, brightness, saturation, and contrast. We randomly extract  $256\times256$  patches from images with a batch size of 2 to train the network for 200 epochs. During testing, overlapping  $256\times256$  patches are extracted with the stride of 128, which alleviates border effects. And the final segmentation results are obtained by binarizing the predicted probability maps with the threshold of 0.5.

In order to make a fair comparison, we faithfully repro-

duce all comparison methods, including U-Net (2015) [11], Iter-Net (2019) [17], U-Net++ (2019) [18], CE-Net (2019) [12], CS-Net (2019) [19], SCS-Net (2021) [20], and some recent state-of-the-art methods: Genetic U-Net (2022) [21], DE-DCGCN-EE (2022) [22], LIOT (2022) [23]. Especially, besides the original LIOT using Iter-Net as the backbone, we also implemented a U-Net version and these two are named LIOT U-Net and LIOT Iter-Net respectively for distinction. Considering fairness, all experiments are conducted under the same experimental conditions.

6

Table 2
Numerical Results under Cross-Dataset Validation. the Best and Second Best are highlighted in **Bold** and *Italic* respectively.

| M d 1         |                   | DRIVE     | ⇒ CHA     | SE_DB1     |            |           | DRI       | VE ⇒ ST   | ARE        |            | STARE ⇒ DRIVE     |           |           |            |            |  |
|---------------|-------------------|-----------|-----------|------------|------------|-----------|-----------|-----------|------------|------------|-------------------|-----------|-----------|------------|------------|--|
| Method        | Se<br>(%)         | Sp<br>(%) | F1<br>(%) | Acc<br>(%) | AUC<br>(%) | Se<br>(%) | Sp<br>(%) | F1<br>(%) | Acc<br>(%) | AUC<br>(%) | Se<br>(%)         | Sp<br>(%) | F1<br>(%) | Acc<br>(%) | AUC<br>(%) |  |
| U-Net         | 60.29             | 98.00     | 66.85     | 94.58      | 95.64      | 76.21     | 98.55     | 79.59     | 96.61      | 97.62      | 66.76             | 99.05     | 77.05     | 94.94      | 96.56      |  |
| Iter-Net      | 58.57             | 97.73     | 64.58     | 94.18      | 93.77      | 74.61     | 98.71     | 79.29     | 96.62      | 97.53      | 67.21             | 98.84     | 76.75     | 94.82      | 96.49      |  |
| U-Net++       | 57.40             | 98.00     | 64.68     | 94.32      | 93.94      | 76.60     | 98.63     | 80.19     | 96.72      | 97.90      | 68.70             | 98.75     | 77.52     | 94.93      | 96.66      |  |
| CE-Net        | 73.23             | 96.84     | 71.46     | 94.70      | 96.38      | 76.42     | 98.47     | 79.40     | 96.56      | 98.03      | 69.17             | 98.35     | 76.65     | 94.63      | 95.98      |  |
| CS-Net        | 61.40             | 97.77     | 66.83     | 94.48      | 94.54      | 73.87     | 98.72     | 78.85     | 96.56      | 97.50      | 69.15             | 98.18     | 76.13     | 94.48      | 96.03      |  |
| SCS-Net       | 44.91             | 98.38     | 55.73     | 93.54      | 92.21      | 54.64     | 99.20     | 67.02     | 95.34      | 93.72      | 68.40             | 98.62     | 76.91     | 94.77      | 95.99      |  |
| Genetic U-Net | 59.55             | 97.69     | 65.17     | 94.24      | 95.12      | 77.18     | 98.72     | 80.96     | 96.85      | 98.47      | 67.31             | 98.96     | 77.17     | 94.93      | 97.02      |  |
| DE-DCGCN-EE   | 59.19             | 97.45     | 64.06     | 93.98      | 94.10      | 74.98     | 98.64     | 79.24     | 96.59      | 97.65      | 63.50             | 98.75     | 73.80     | 94.26      | 95.98      |  |
| LIOT U-Net    | 59.51             | 97.29     | 63.75     | 93.87      | 94.40      | 77.43     | 98.45     | 79.94     | 96.63      | 98.07      | 68.23             | 97.63     | 73.96     | 93.89      | 96.12      |  |
| LIOT Iter-Net | 62.09             | 96.79     | 63.92     | 93.65      | 94.35      | 79.93     | 98.21     | 80.42     | 96.62      | 98.11      | 63.48             | 98.36     | 72.66     | 93.92      | 96.26      |  |
| FRS U-Net     | 76.79             | 96.87     | 73.76     | 95.05      | 96.44      | 81.63     | 98.48     | 82.63     | 97.02      | 98.66      | 75.03             | 98.09     | 79.76     | 95.15      | 97.09      |  |
| FRS Iter-Net  | 76.95             | 96.53     | 72.68     | 94.76      | 96.76      | 82.38     | 98.43     | 82.85     | 97.04      | 98.85      | 73.62             | 98.40     | 79.77     | 95.25      | 97.12      |  |
| 36.1.1        | STARE ⇒ CHASE_DB1 |           |           |            |            |           | CHASE     | _DB1 ⇒    | STARE      |            | CHASE_DB1 ⇒ DRIVE |           |           |            |            |  |
| Method        | Se<br>(%)         | Sp<br>(%) | F1<br>(%) | Acc<br>(%) | AUC<br>(%) | Se<br>(%) | Sp<br>(%) | F1<br>(%) | Acc<br>(%) | AUC<br>(%) | Se<br>(%)         | Sp<br>(%) | F1<br>(%) | Acc<br>(%) | AUC<br>(%) |  |
| U-Net         | 67.87             | 97.32     | 69.70     | 94.66      | 95.81      | 56.04     | 99.61     | 70.01     | 95.84      | 97.84      | 61.83             | 98.67     | 72.33     | 93.98      | 95.81      |  |
| Iter-Net      | 58.36             | 98.08     | 65.72     | 94.49      | 96.03      | 54.19     | 99.59     | 68.35     | 95.65      | 97.45      | 61.67             | 98.69     | 72.27     | 93.97      | 95.30      |  |
| U-Net++       | 60.63             | 97.72     | 66.07     | 94.36      | 95.59      | 55.16     | 99.55     | 69.01     | 95.70      | 97.42      | 60.35             | 98.82     | 71.66     | 93.92      | 95.79      |  |
| CE-Net        | 69.43             | 97.14     | 70.09     | 94.63      | 95.95      | 63.18     | 99.19     | 73.59     | 96.07      | 97.51      | 62.89             | 98.39     | 72.31     | 93.87      | 94.68      |  |
| CS-Net        | 68.46             | 96.99     | 68.90     | 94.40      | 95.50      | 51.77     | 99.58     | 66.31     | 95.44      | 96.54      | 58.26             | 98.74     | 69.82     | 93.59      | 94.24      |  |
| SCS-Net       | 64.24             | 97.01     | 66.14     | 94.04      | 95.32      | 50.24     | 99.69     | 65.47     | 95.40      | 96.70      | 54.73             | 99.08     | 67.97     | 93.44      | 93.84      |  |
| Genetic U-Net | 59.20             | 97.78     | 65.24     | 94.29      | 95.18      | 70.83     | 99.04     | 78.29     | 96.59      | 98.46      | 61.15             | 98.58     | 71.58     | 93.82      | 95.65      |  |
| DE-DCGCN-EE   | 46.97             | 98.09     | 56.54     | 93.46      | 93.45      | 62.84     | 98.92     | 72.13     | 95.79      | 96.25      | 60.13             | 98.80     | 71.42     | 93.88      | 94.50      |  |
| LIOT U-Net    | 66.20             | 96.95     | 67.25     | 94.16      | 94.76      | 70.23     | 98.94     | 77.42     | 96.45      | 98.08      | 54.69             | 99.05     | 67.85     | 93.4       | 94.89      |  |
| LIOT Iter-Net | 64.32             | 97.35     | 67.38     | 94.36      | 94.69      | 69.37     | 99.01     | 77.18     | 96.44      | 98.29      | 56.85             | 98.84     | 69.00     | 93.5       | 94.71      |  |
| FRS U-Net     | 74.55             | 97.19     | 73.55     | 95.14      | 96.93      | 72.73     | 99.02     | 79.45     | 96.74      | 97.95      | 62.06             | 99.08     | 73.72     | 94.37      | 96.39      |  |
| FRS Iter-Net  | 72.66             | 97.24     | 72.51     | 95.01      | 96.85      | 71.15     | 99.21     | 79.29     | 96.78      | 98.58      | 64.03             | 98.76     | 74.23     | 94.34      | 95.96      |  |

## 4.3 In-Dataset Evaluation

In-dataset evaluation implies the training and the testing process are performed on the same dataset. Given the complex morphology of retinal vasculature, it has a high requirement for the accuracy of methods. We conduct indataset evaluations on DRIVE, STARE, and CHASE\_DB1 with identical experimental conditions and calculate metrics inside FOVs. The experiment results are summarized in Table 1 and some visualization results are illustrated in Fig. 4.

As shown in Table 1, our proposed methods FRS U-Net and FRS Iter-Net, obtain almost all the top two best scores in Se, F1, Acc, and AUC. It indicates that FRS-Nets not only achieve the best overall performance (F1, Acc, AUC), but also are more capable of capturing vessels (Se). Although the Sp of FRS-Nets is slightly inferior to the best, it's negligible compared with the improvement in Se. As shown in Fig. 4, FRS-Nets have better identification of small blood vessels and better connectivity and smoothness of large vessels.

These demonstrate our FRS-Nets are optimal in both numerical and visualized results. Moreover, such evident improvements are achieved simply by replacing the traditional convolution filters of the backbone methods, U-Net and Iter-Net, with our proposed FRS-Conv.

Interestingly, it's should be noticed that the data augmentations mentioned in Section 4.2 are applied for all methods, which include random rotation in 360 degrees and random rescaling from 0.8 to 1.4. The results in Table 1

denote that, even with sufficient global data augmentations, especially rotation and scaling, FRS U-Net and FRS Iter-Net still outperform the corresponding backbone methods, U-Net and Iter-Net. It strongly verifies the necessity to characterize the local symmetries of the vascular morphology, and demonstrates the effectiveness of FRS-Nets to embed equivariance compared with data augmentation.

7

#### 4.4 Cross-Dataset Evaluation

Cross-dataset evaluation implies that models are trained on one dataset and tested on another, which is more consistent with clinical applications. Accordingly, it's more challenging to the generalization and robustness of models, compared with in-dataset evaluation.

With the same conditions, we assess all methods under cross-dataset evaluation on three datasets inside FOVs. It should be noted that, since different datasets are different in resolutions, we test models directly with the original resolutions of testing datasets, rather than rescale them to fit the resolutions of training sets. It's fairly applied to all methods under cross-dataset evaluations. The numerical results of the six cross-dataset experiments are listed in Table 2 and some visualization results are shown in Fig. 5.

It's obvious that FRS-Nets still achieve almost all the top two best performances in Se, F1, Acc, and AUC, and significantly outperform other methods by a more significant margin. It indicates the fine generalization and robustness of FRS-Nets and the promising potential for clinical applications. Note that the Sp of FRS-Nets is still unsubstantially

Figure 5. Some typical visualized segmentation results under cross-dataset validation.

lower than the best in some cases. Considering evident improvements in other metrics, especially in Se, it's still rational to say that the proposed method is superior. As

shown in Fig. 5, FRS-Nets not only achieve the best performance for the segmentation of capillaries and large vessels, but more importantly, have a significant advantage over

Table 3
Ablation Studies about FRS U-Net. the Best is highlighted in **Bold**.

| Method    | Rotation<br>Equi. | Scale<br>Equi. | Fourier<br>Param. | Params<br>(M) |           | DRI       | VE ⇒ DI   | RIVE       |            | $DRIVE \Rightarrow STARE$ |           |           |            |            |  |
|-----------|-------------------|----------------|-------------------|---------------|-----------|-----------|-----------|------------|------------|---------------------------|-----------|-----------|------------|------------|--|
|           |                   |                |                   |               | Se<br>(%) | Sp<br>(%) | F1<br>(%) | Acc<br>(%) | AUC<br>(%) | Se<br>(%)                 | Sp<br>(%) | F1<br>(%) | Acc<br>(%) | AUC<br>(%) |  |
| U-Net     | X                 | X              | X                 | 31.04         | 77.35     | 97.95     | 80.82     | 95.33      | 97.37      | 51.02                     | 98.93     | 62.88     | 94.77      | 82.69      |  |
| F U-Net   | ×                 | X              | <b>✓</b>          | 138.04        | 76.59     | 98.12     | 80.85     | 95.38      | 97.33      | 58.83                     | 98.99     | 69.42     | 95.50      | 90.20      |  |
| FS U-Net  | ×                 | /              | <b>✓</b>          | 34.52         | 76.24     | 98.26     | 81.05     | 95.46      | 97.65      | 61.53                     | 98.71     | 70.28     | 95.49      | 95.08      |  |
| FR U-Net  | <b>✓</b>          | X              | <b>✓</b>          | 17.26         | 76.69     | 98.17     | 81.05     | 95.44      | 97.63      | 64.81                     | 98.94     | 73.66     | 95.98      | 89.01      |  |
| FRS U-Net | <b>✓</b>          | <b>✓</b>       | ✓                 | 4.32          | 81.18     | 97.66     | 82.31     | 95.56      | 97.89      | 69.29                     | 98.86     | 76.42     | 96.29      | 97.65      |  |

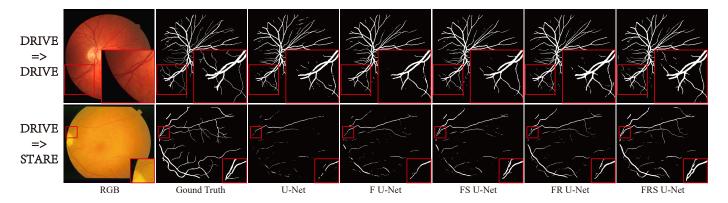


Figure 6. Some typical visualized segmentation results in ablation study.

other methods in the segmentation of the overall vascular structure.

Moreover, we still use the same data augmentations as in-dataset evaluation, including random rotation and rescaling. Once again, without any changes to network architectures, FRS-Nets outperform the backbone methods, U-Net and Iter-Net, by an even more evident margin. It further demonstrates the effectiveness of FRS-Nets to embed the symmetries of the vascular morphology into networks themselves, compared with external influence strategies, especially data augmentations.

# 4.5 Ablation Study

In order to verify the effectiveness of the Fourier parameterization, the scale equivariance and the rotation equivariance in FRS-Nets, we construct the Fourier parameterized convolution (F-Conv), the Fourier parameterized scale equivariant convolution (FS-Conv), and the Fourier parameterized rotation equivariant convolution (FR-Conv) by adjusting the hyper-parameters of the discrete rotation group R and the discrete scale group S. We respectively replace the traditional convolution filters in U-Net with the aforementioned variants of FRS-Conv, and obtain the corresponding variants of FRS U-Net, as shown in Table 3. To exclude other external influences to the ablation study, we don't apply any data augmentations to the experiments of this section.

We compare the five networks under two evaluation strategies. The in-dataset evaluation is on DRIVE, and the cross-dataset is from DRIVE to STARE. Additionally, we summarize the model parameter numbers. The numerical results are listed in Table 3, and some visual results are shown in Fig. 6.

#### 4.5.1 In-Dataset Ablation

For in-dataset evaluation, as shown in Table 3, F U-Net performs very close to U-Net. It implies that the Fourier parameterization scheme can approximate the traditional convolution filters with high accuracy. FS U-Net and FR U-Net have almost identical performances, both of which are superior to U-Net and F U-Net but inferior to FRS U-Net. It indicates the rotation symmetry and the scale symmetry in the retinal vasculature, and more significantly, demonstrates the necessity to simultaneously embed the rotation and scale equivariance into networks.

# 4.5.2 Cross-Dataset Ablation

For cross-dataset evaluation, F U-Net has an overall improvement compared with U-Net. It implies that the Fourier parameterized convolution filters have stronger generalization capability besides accuracy. FS U-Net is superior to FR U-Net in AUC but inferior in F1, Acc, Se, and Sp, both of which are still better than U-Net and F U-Net, but worse than FRS U-Net. It denotes the powerful generalization of FRS-Nets and the potential for clinical applications.

# 4.5.3 Comparison of Parameters

As aforementioned in Section 4.2, we choose 4 scale levels as the scale group and 8 angles as the rotation group. Since the weights are shared between channels as shown in Fig. 3, the number of parameters of FS U-Net, FR U-Net, and FRS U-Net are 1/4, 1/8, and 1/32 of F U-Net respectively. Surprisingly, FRS U-Net achieves such an overall superior performance with merely 13.9% parameters of U-Net, which is also suitable for FRS Iter-Net. It further illustrates the promising potential of FRS-Nets for clinical application deployment.

As the visualization shown in Fig. 6, U-Net and F U-Net tend to have similar visual effects. The segmentation of FS U-Net has better connectivity and smoothness for blood vessels, and less noise. FR U-Net is prone to explore more vessel structures. FRS U-Net has a favorable combination of both advantages and achieves the best visual effects among all networks.

#### 4.6 Future Work

Although FRS-Nets achieve an overall remarkable performance compared with existing methods, they still have several limitations. One of the problems is the increase in FLOPs due to the scale expansion of convolution filters, even though we have significantly reduced the number of parameters. Another one is the truncation of the infinite scale group due to limited practical resources, which affects the accuracy of equivariance. In future work, we will focus on alleviating the two issues to further improve the performance of FRS-Nets and make them more effective for clinical applications.

# 5 CONCLUSION

In this work, we propose a novel convolution filter, FRS-Conv, for retinal vessel segmentation. By utilizing the theory of group and Fourier series expansion, it's Fourier parameterized and equivariant to rotation and scaling. These properties enable FRS-Conv to successfully characterize the local symmetries existed in the retinal vascular morphology. Without changing network architectures, we replace the traditional convolution filters in U-Net and Iter-Net, and conduct comprehensive experiments on three public datasets with FRS-Nets, including FRS U-Net and FRS Iter-Net. The numerical and visualized results demonstrate that FRS-Nets achieve state-of-the-art performance in both accuracy and generalization, with merely 13.9% parameters of corresponding baselines. It illustrates the promising potential of FRS-Nets for clinical applications. Future work will focus on further refining FRS-Nets for better applications.

# REFERENCES

- [1] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE reviews in biomedical engineering*, vol. 3, pp. 169–208, 2010.
- pp. 169–208, 2010.
  [2] Z. Li, X. Zhang, H. Müller, and S. Zhang, "Large-scale retrieval for medical image analytics: A comprehensive review," *Medical image analysis*, vol. 43, pp. 66–84, 2018.
- [3] J. Mo and L. Zhang, "Multi-level deep supervised networks for retinal vessel segmentation," *International journal of computer as*sisted radiology and surgery, vol. 12, no. 12, pp. 2181–2193, 2017.
- [4] A. M. Mendonca and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," *IEEE transactions on medical imaging*, vol. 25, no. 9, pp. 1200–1213, 2006.
- [5] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Transactions on medical imaging*, vol. 8, no. 3, pp. 263–269, 1989.
- [6] F. Zana and J.-C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," *IEEE transactions on image processing*, vol. 10, no. 7, pp. 1010–1019, 2001.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications* of the ACM, vol. 60, no. 6, pp. 84–90, 2017.

[8] Y. LeCun, Y. Bengio et al., "Convolutional networks for images, speech, and time series," The handbook of brain theory and neural networks, vol. 3361, no. 10, p. 1995, 1995.

- [9] E. Barnard and D. Casasent, "Invariance and neural nets," *IEEE Transactions on neural networks*, vol. 2, no. 5, pp. 498–508, 1991.
- [10] E. S. Uysal, M. Ş. Bilici, B. S. Zaza, M. Y. Özgenç, and O. Boyar, "Exploring the limits of data augmentation for retinal vessel segmentation," arXiv preprint arXiv:2105.09365, 2021.
- [11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, 2015, pp. 234–241.
- [12] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.
- [13] V. Cherukuri, V. K. Bg, R. Bala, and V. Monga, "Deep retinal image segmentation with regularization under geometric priors," *IEEE Transactions on Image Processing*, vol. 29, pp. 2552–2567, 2019.
- [14] Y. Xu, T. Xiao, J. Zhang, K. Yang, and Z. Zhang, "Scale-invariant convolutional neural networks," arXiv preprint arXiv:1411.6369, 2014
- [15] T. Cohen and M. Welling, "Group equivariant convolutional networks," in *International conference on machine learning*. PMLR, 2016, pp. 2990–2999.
- [16] Q. Xie, Q. Zhao, Z. Xu, and D. Meng, "Fourier series expansion based filter parametrization for equivariant convolutions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [17] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks," in *Proceedings of the IEEE/CVF winter* conference on applications of computer vision, 2020, pp. 3656–3665.
- [18] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE transactions on medical imaging*, vol. 39, no. 6, pp. 1856–1867, 2019.
- [19] L. Mou, Y. Zhao, L. Chen, J. Cheng, Z. Gu, H. Hao, H. Qi, Y. Zheng, A. Frangi, and J. Liu, "Cs-net: channel and spatial attention network for curvilinear structure segmentation," in Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22. Springer, 2019, pp. 721–730.
- [20] H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, and J. Qin, "Scs-net: A scale and context sensitive network for retinal vessel segmentation," *Medical Image Analysis*, vol. 70, p. 102025, 2021.
- [21] J. Wei, G. Zhu, Z. Fan, J. Liu, Y. Rong, J. Mo, W. Li, and X. Chen, "Genetic u-net: Automatically designed deep networks for retinal vessel segmentation using a genetic algorithm," *IEEE Transactions* on *Medical Imaging*, vol. 41, no. 2, pp. 292–307, 2022.
- [22] Y. Li, Y. Zhang, W. Cui, B. Lei, X. Kuang, and T. Zhang, "Dual encoder-based dynamic-channel graph convolutional network with edge enhancement for retinal vessel segmentation," *IEEE Transactions on Medical Imaging*, vol. 41, no. 8, pp. 1975–1989, 2022.
- [23] T. Shi, N. Boutry, Y. Xu, and T. Géraud, "Local intensity order transformation for robust curvilinear object segmentation," *IEEE Transactions on Image Processing*, vol. 31, pp. 2557–2569, 2022.
- [24] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," California Univ San Diego La Jolla Inst for Cognitive Science, Tech. Rep., 1985.
- [25] E. Hoogeboom, J. W. Peters, T. S. Cohen, and M. Welling, "Hexa-conv," arXiv preprint arXiv:1803.02108, 2018.
- [26] D. Worrall and M. Welling, "Deep scale-spaces: Equivariance over scale," Advances in Neural Information Processing Systems, vol. 32, 2019
- [27] M. Weiler, F. A. Hamprecht, and M. Storath, "Learning steerable filters for rotation equivariant cnns," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 849–858.
- [28] M. Weiler and G. Cesa, "General e (2)-equivariant steerable cnns," Advances in Neural Information Processing Systems, vol. 32, 2019.
- [29] Z. Shen, L. He, Z. Lin, and J. Ma, "Pdo-econvs: Partial differential operator based equivariant convolutions," in *International Confer*ence on Machine Learning. PMLR, 2020, pp. 8697–8706.
- [30] I. Sosnovik, M. Szmaja, and A. Smeulders, "Scale-equivariant steerable networks," arXiv preprint arXiv:1910.11093, 2019.

[31] L. Gao, G. Lin, and W. Zhu, "Deformation robust roto-scale-translation equivariant cnns," *Transactions on Machine Learning Research*.

- [32] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE transactions on medical imaging*, vol. 23, no. 4, pp. 501–509, 2004.
- [33] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical imaging*, vol. 19, no. 3, pp. 203–210, 2000.
- [34] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–2548, 2012.
- [35] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE transactions on medical imaging*, vol. 35, no. 11, pp. 2369–2380, 2016.
- [36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.