Highlights

用于眼底血管分割的具有全局学习相对偏移的可变形卷积模块

Lexuan Zhu, Yuxuan Li, Yuning Ren

- We propose a plug-and-play deformable convolution with globally learned relative offsets, a module well suited to handle complex zigzag edge features. Compared with other existing deformable convolution, the proposed convolution has the following different characteristics:
  - 偏移量是通过多头注意力和前馈网络学习的，能够有效捕捉全局语义依赖关系。
  - 偏移作用于卷积特征图，形成亚像素级的变形场，从而导致卷积核发生“相对变形”。
  - 该模块可以解耦卷积核的大小和偏移学习网络，并且可以适应具有任何参数的常规卷积。
- We add the proposed convolutional module to Unet and propose a deep learning model for fundus vascular segmentation, called GDCUnet.
- We construct a unified comparison framework with the same data processing methods and loss functions and other configurations for the mainstream existing state-of-the-art medical image segmentation models in our project. Experiments on the public dataset have verified that GDCUnet has reached state-of-the-art performance.

用于眼底血管分割的具有全局学习相对偏移的可变形卷积模块

Lexuan Zhu[a], Yuxuan Li[b], Yuning Ren[c]

[a]*New York University, New York City, 10003, New York State, United States*
[b]*The University of Sydney, Sydney, NSW 2006, New South Wales, Australia*
[c]*China University of Political Science and Law, Beijing, 100088, China*

Abstract

可变形卷积通过学习偏移来适应性地改变卷积核的形状，以应对复杂的形状特征。我们提出了一种新颖的即插即用可变形卷积模块，该模块通过使用注意力机制和前馈网络来学习偏移，使可变形模式可以捕捉长距离的全局特征。与现有的可变形卷积相比，所提出的模块学习亚像素位移场并适应性地扭曲所有通道的特征图（而不是直接变形卷积核），这等同于内核采样网格的相对变形，实现全局特征变形和（核大小）-（学习网络）的解耦。考虑到眼底血管具有全球自相似的复杂边缘，我们设计了一种用于眼底血管分割的深度学习模型 GDCUnet，该模型基于所提出的卷积模块。在相同的配置和统一的框架下的实证评估显示，GDCUnet 在公共数据集上已达到最新的性能。进一步的消融实验表明，所提出的可变形卷积模块可以更显著地学习眼底血管的复杂特征，增强模型的表现和泛化能力。所提出的模块类似于传统卷积的接口，我们建议将其应用于更多具有复杂全球自相似特征的机器视觉任务。我们的源代码将在 `https://github.com/LexyZhu/GDCUnet.git` 提供。

Keywords: deformable convolution, plug-and-play module, medical image segmentation, attention, relative deformation.

1. 介绍

医学图像分割技术将感兴趣区域提取为二值掩码，并在分析和诊断中有重要应用 [1] 。这个任务可以由专业的医疗工作者手动完成。然而，准确的计算机

*Corresponding author: Y. R. yuning.ren.cupl.cn@gmail.com

辅助分割可以提高效率和客观性 [2] 。Unet 是一个基于卷积和残差特征重用的 U 形结构，已成为生物图像分割等任务的重要解决方案 [3] 。一些改进版本的 Unet 可以在某些任务中取得更好的性能 [4, 5] 。卷积的全局特征学习能力存在局限性。Vision Transformer (ViT) 使用视觉模态的注意力机制来增强远距离特征学习 [6] 。医学图像通常具有自相似的全局特征，因此类似 ViT 的结构在分析和理解能力上表现出色 [7] 。然而，Transformer 结构在局部特征表示和多尺度特征融合方面略显弱势。近年来，机器视觉模型的改进方向一直是局部和全局特征的融合理解 [8] 。医学图像分割任务也需要整合全局和局部特征以实现更好的性能 [9] 。

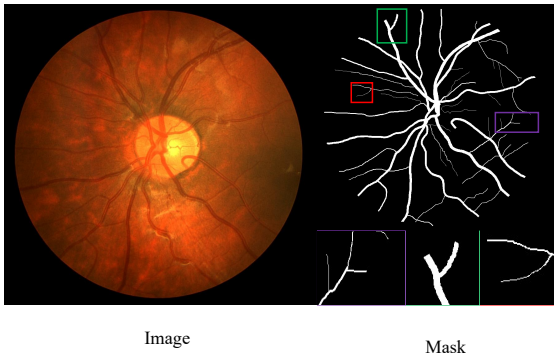


Figure 1: 视网膜血管的局部和全局特征。

如图 1 所示，血管具有微妙的局部特征，同时在整体上是自相似的。本文着重于如何增强深度学习模型表示局部复杂但整体自相似的血管特征的能力，以提高分割性能。我们将可变形卷积偏移的学习视为解决此问题的方法。

可变形卷积是基于网络学习的偏移量来自适应地改变卷积核的形状，从而使卷积能够关注具有复杂形状的物体的特征 [10] 。可变形卷积非常适合处理医学图像的复杂局部边缘 [11] 。然而，传统的可变形卷积依赖于另一个卷积来学习偏移，这限制了其对全局自相似特征形状的表示。在本文中，我们使用空间注意和前馈网络来学习偏移，并提出了一种新颖的可变形卷积，我们将其命名为 SAFD 卷积。SAFD 卷积即插即用，接口与常规卷积相似，非常方便替换卷积神经网络中的常规卷积。基于此，我们设计了一种类似于 Unet 的眼底血管图像分割模型，使用常规卷积和所提议的卷积，即全局可变形卷积 Unet（GDCUnet）。

鉴于当前医学图像分割模型缺乏统一的比较框架，限制了比较实验的说服力，我们构建了一个统一的配置框架，并对现有的一些最先进模型进行了统一配置。公共数据集上的实验表明，GDCUnet 在参数数量最少的情况下实现了最佳性能。我们的主要贡献总结在上述亮点中，本文其余部分组织如下：第 **??** 节回顾了与局部全球模型和可变形卷积相关的工作。在第 2 节中，我们提出了 SAFD 卷积和 GDCUnet。第 3 节介绍了实验的主要结果。第 4 节通过消融实验展示了所提出方法的特征学习能力和泛化能力，第 5 节总结了全文。

医疗图像的全局特征逐渐受到关注，因此提出了类似于 ViT 的医学图像分析模型。由于纯 Transformer 缺乏捕捉局部细节的能力，一些研究集中于使用局部-全局多尺度特征融合的理解模型。处理方法包括类似于 Swin Transformer 的局部窗口注意结构，以及一些其他结合卷积和注意力的方法。考虑到计算复杂度和其他问题，一些模型放弃了传统注意力机制来处理卷积特征图的长距离语义依赖。局部-全局概念也被应用于处理更高维度和更多模态的医学图像。

血管具有更多独特的特征，表现出全局自相似的局部复杂边缘，这激发我们找到使用卷积和注意机制的新解决方案。

许多研究已经开展，以增强卷积神经网络的空间全局特征表示能力。近年来，针对不同视觉任务的空间注意力模型被提出，有些需要结合通道注意力来进一步增强学习能力。对于连续的视频帧，空间注意力可以与时间注意力结合，以有效处理时间连续的全局特征。

空间注意力的成功为我们处理复杂形状的全球自相似特征的方法提供了启发。

1.1. 具有多样采样模式的卷积

可变形卷积 *V1 [10]*．J. Dai 等人首次提出了可变形卷积，这种方法利用另一个专门的卷积层来学习偏移，实现卷积采样点形状的自适应变化，并增强卷积层对复杂形状的适应性。

可变形卷积 *V2 [30]*．V2 引入了调制幅度控制，并采用特征模仿的概念来限制感兴趣特征点的变形行为。

可变形卷积 *V3 [31]*．V3 稀疏化了采样核并采用了空间聚合的概念，增强了可变形卷积的长距离感知能力，并在与 ViT 结构的竞争中取得了有竞争力的表

现。

可变形卷积 *V4 [32]*. V4 移除了 V3 中空间聚合的 softmax 归一化，从而进一步增强了动态性和表现力。V4 的主要关注点是计算效率，减少多余操作并提升推理效率。

其他. 为了在不增加参数数量的情况下扩大感受野，F. Yu 等人将膨胀引入卷积核，从而提出了膨胀卷积 [33] 。此外，针对特殊管状拓扑结构的专用卷积也进行了研究，这种卷积被称为动态蛇卷积 [34] 。

然而，目前所有可变形卷积的概念都是每个采样网格的偏移，这无法解决卷积核大小与偏移学习网络的解耦，并且在全局学习模块（如注意力）的集成友好性方面存在局限性。

2. 方法

2.1. SAFD 卷积

假设卷积核大小 $\boldsymbol{K_s}$ 是一个奇数，使得半径为 $r = \frac{K_s-1}{2}$ ，然后具有膨胀率 $\boldsymbol{D_s}$ 的二维传统采样网格 $\mathcal{R}$ 被定义为：

例如，

给定输入特征图 $\boldsymbol{x}(\boldsymbol{p}_0)$ ，其空间坐标为 $\boldsymbol{p}_0 \in \mathbb{Z}^2$ ，膨胀率为 $\boldsymbol{D_s}$ 的标准卷积表示为

$$\boldsymbol{y}(\boldsymbol{p}_0) = \sum_{\boldsymbol{p}_n \in \mathcal{R}_{\boldsymbol{K_s},\boldsymbol{D_s}}} \boldsymbol{w}(\boldsymbol{p}_n)\, \boldsymbol{x}(\boldsymbol{p}_0 + \boldsymbol{p}_n), \tag{1}$$

，其中 $\boldsymbol{w}(\boldsymbol{p}_n)$ 表示卷积核权重。为了使卷积核能够自适应地进行几何形变，现有的可变形卷积引入了一个可学习的偏移量 $\{\Delta\boldsymbol{p}_n\}$ ：

$$\boldsymbol{y}(\boldsymbol{p}_0) = \sum_{\boldsymbol{p}_n \in \mathcal{R}_{\boldsymbol{K_s},\boldsymbol{D_s}}} \boldsymbol{w}(\boldsymbol{p}_n)\, \boldsymbol{x}(\boldsymbol{p}_0 + \boldsymbol{p}_n + \Delta\boldsymbol{p}_n). \tag{2}$$

。我们学习一个连续的子像素位移场

$$\Delta\boldsymbol{p} : \ \mathbb{Z}^2 \longrightarrow \mathbb{R}^2, \qquad \boldsymbol{p} \mapsto \Delta\boldsymbol{p}(\boldsymbol{p}) \tag{3}$$

，其值由相同空间坐标处的所有通道共享，而不是如方程 (2) 那样为每个核位置 $\boldsymbol{p_n}$ 学习一组离散偏移 $\{\Delta \boldsymbol{p}_n\}_{n=1}^{K_s^2}$ 。特征图首先由该场变形：

$$\tilde{\boldsymbol{x}}(\boldsymbol{p}) \;=\; \boldsymbol{x}(\boldsymbol{p} + \Delta \boldsymbol{p}(\boldsymbol{p})), \tag{4}$$

，然后对 $\tilde{\boldsymbol{x}}$ 应用常规卷积：

$$\begin{aligned} \boldsymbol{y}(\boldsymbol{p}_0) &= \sum_{\boldsymbol{p}_n \in \mathcal{R}_{K_s, D_s}} \boldsymbol{w}(\boldsymbol{p}_n)\, \tilde{\boldsymbol{x}}(\boldsymbol{p}_0 + \boldsymbol{p}_n) \\ &= \sum_{\boldsymbol{p}_n \in \mathcal{R}_{K_s, D_s}} \boldsymbol{w}(\boldsymbol{p}_n)\, \boldsymbol{x}(\boldsymbol{p}_0 + \boldsymbol{p}_n + \Delta \boldsymbol{p}(\boldsymbol{p}_0 + \boldsymbol{p}_n)). \end{aligned} \tag{5}$$

。

为了通过注意力机制获取偏差 $\Delta \boldsymbol{p}(\boldsymbol{p})$ ，对于输入张量 $X \in \mathbb{R}^{B \times H \times W \times C}$ ：

$$X_u = \text{reshape}(X,\; B, HW, C) \in \mathbb{R}^{B \times N \times C} \tag{6}$$

，其中 $N = HW$ ，然后我们引入一个可选的超参数 $\mathcal{E}$ 来增强特征维度。基于这个超参数，特征维度通过一个线性层进行嵌入：

$$X_e = X_u\, W^E \qquad\qquad \in \mathbb{R}^{B \times N \times \mathcal{E}C} \tag{7}$$

，其中 W^E 是嵌入层的权重矩阵，标记 $\mathcal{E}C$ 为 D 。注意力机制通过线性变换将 X_e 划分为三个张量：查询、键和值：

$$Q = X_e\, W^Q, \quad K = X_e\, W^K, \quad V = X_e\, W^V \qquad \in \mathbb{R}^{B \times N \times D} \tag{8}$$

。学习来自 ViT，3 个张量 Q 、K 和 V 被分块成 h 个头：

$$Q_{h_i} = \text{chunk}_{\text{h}_\text{i}}(Q,\; B, N, d_h) \qquad \in \mathbb{R}^{B \times N \times d_h} \tag{9}$$

$$[2pt] K_{h_i} = \text{chunk}_{\text{h}_\text{i}}(K,\; B, N, d_h) \qquad \in \mathbb{R}^{B \times N \times d_h} \tag{10}$$

$$[2pt] V_{h_i} = \text{chunk}_{\text{h}_\text{i}}(V,\; B, N, d_h) \qquad \in \mathbb{R}^{B \times N \times d_h} \tag{11}$$

在这里 $d_h = \frac{D}{h}$ ，$h_i \in 1, 2, \ldots, h$ 。然后执行常规的注意力乘法：

$$\mathcal{A}_{h_i} = \text{SoftMax}(\tfrac{Q_{h_i} K_{h_i}^{\top}}{\sqrt{d_h}}) V_{h_i} \qquad \in \mathbb{R}^{B \times N \times d_h} \tag{12}$$

这里 $\mathcal{A}_{h_i}$ 代表第 h_i 个头部的输出，所有头部的输出连接在一起以获得注意力的输出 $\mathcal{A}$：

$$\mathcal{A} = \text{Concat}[\mathcal{A}_1, \cdots, \mathcal{A}_h]W^O \qquad \in \mathbb{R}^{B\times N\times D} \tag{13}$$

这里 W^O 是另一个线性权重矩阵。在输入前馈网络之前，我们执行残差操作：

$$X_A = \mathcal{A} + X_e \qquad \in \mathbb{R}^{B\times N\times D} \tag{14}$$

然后我们提供另一个可选的超参数 D_{hidden} 作为前馈的隐藏维度，让 X_A 通过两个线性层：

$$X_{hidden} = X_A W^{(1)}_{hidden} \qquad \in \mathbb{R}^{B\times N\times D_{hidden}} \tag{15}$$

$$X_F = X_{hidden} W^{(2)} \qquad \in \mathbb{R}^{B\times N\times D} \tag{16}$$

这里 $W^{(1)}_{hidden} \in \mathbb{R}^{B\times D\times D_{hidden}}$ ， $W^{(2)} \in \mathbb{R}^{B\times D_{hidden}\times D}$ 。之后，我们进行另一个残差操作和反向嵌入：

$$\hat{X} = (X_F + X_A)W^E_{reverse} \qquad \in \mathbb{R}^{B\times N\times C} \tag{17}$$

这里 $W^E_{reverse} \in \mathbb{R}^{B\times D\times C}$ 。需要将 $\hat{X}$ 重塑回卷积特征图的形式：

$$X_{offset} = \text{reshape}(\hat{X},\ B, H, W, C) \in \mathbb{R}^{B\times H\times W\times C} \tag{18}$$

通道数量 C 通常可以被 2 整除，我们表示为 $M_c = C/2$ 并分割最后一个维度：

然后我们对 M_c 张量取平均值以获得最终的 $\Delta\boldsymbol{p}(\boldsymbol{p}_0 + \boldsymbol{p}_n)$ ，这由所有通道共享：

最后，子像素位移场被添加到公式 (5) 中的每个采样点 $\boldsymbol{p}_0 + \boldsymbol{p}_n$ 。图 2 展示了 SAFDConvolution 和现有可变形卷积之间的差异，这种结构实现了（卷积核大小）-（学习网络）的解耦，从而允许偏移学习网络的多样性。所有通道的相对偏移共享实现了参数数量的轻量化。

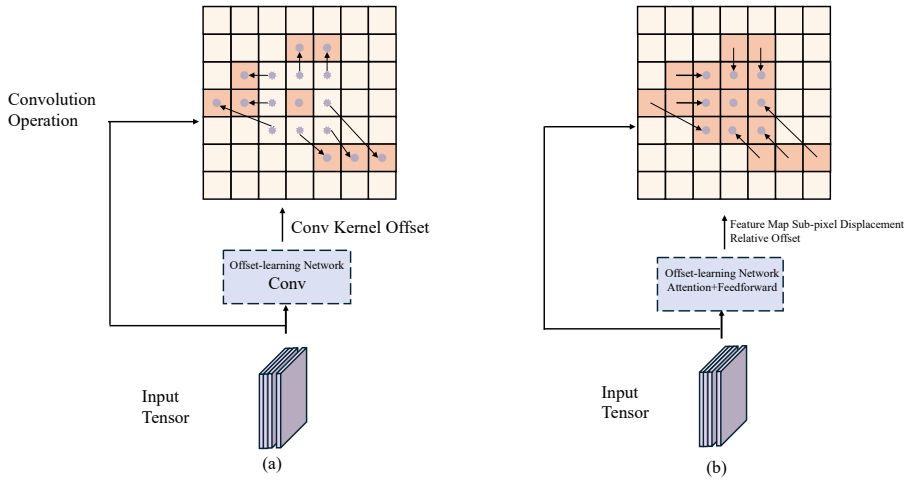


Figure 2: SAFDConvolution 与现有可变形卷积之间的概念差异，(a) 现有可变形卷积，(b) 我们的 SAFDConvolution。SAFDConvolution 并不直接改变卷积核的形状，而是变形特征图以实现卷积核的相对变形。特征图的位移场对所有通道共享。相对偏移的学习通过注意力和前馈网络（如 ViT）实现了对全局自相似复杂局部细节的捕捉。

我们在表格 1 中总结了 SAFDConvolution 可选择的超参数，并为后续实验（在部分 3.4 中）提供了一些超参数设置。

Table 1: SAFDConvolution 在实验中的超参数设置。$\boldsymbol{K_s}$ -核大小，$\boldsymbol{D_s}$ -膨胀率，$\mathcal{E}$ -嵌入的倍数，h -注意力头的数量，D_{hidden} -前馈的隐藏维度。

	Setting 1	Setting 2	Setting 3	Setting 4	Setting 5	Setting 6
$\boldsymbol{K_s}$	5	7	3	5	5	5
$\boldsymbol{D_s}$	1	1	2	1	1	1
$\mathcal{E}$	1	1	1	1	2	4
h	4	4	4	4	4	4
D_{hidden}	32	32	32	64	64	64

2.2. GDCUnet

在编码阶段，Conv 模块包含两个 3×3 传统卷积。SAFDConv 模块使用两个连续的提议 SAFD 卷积，其中超参数可以根据表 1 进行设置。此外，特征激励模块使用一个 7×7 传统卷积。

解码阶段．解码阶段使用的结构块（D-Block）与编码阶段采用的模块相同。

总体流程．假设输入张量的形状为 $B \times H_s \times W_s \times 3$ ，GDCUnet 的流程如图 3 所示，最终输出是一个 1 通道的分割掩码 $B \times H_s \times W_s \times 1$ 。

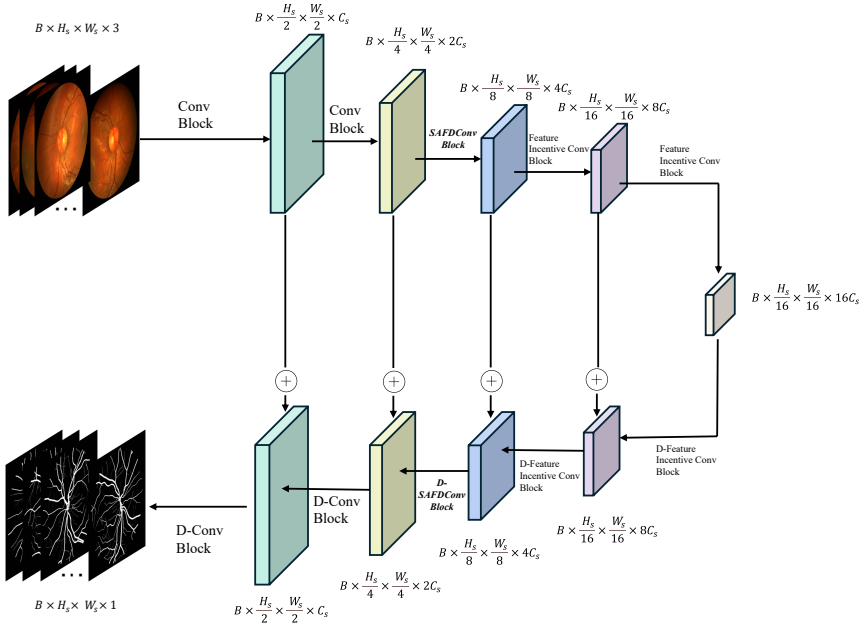


Figure 3: GDCUnet 的整体流程，其中 C_s 是 16 。我们的网络采用了类似于 Unet 的结构。在编码阶段，使用最大池化来减小特征图的尺寸，而在解码阶段，使用双线性插值来恢复特征图的尺寸。我们使用张量加法来实现前一阶段特征图的特征融合。使用我们提出的 SAFDConvolution 的模块以斜体加粗形式显示。据我们所知，提出的可变形卷积更适合于特征提取和解码的中间阶段。

3. 实验

3.1. 统一基准和配置

据我们所知，与多维时间序列分析（例如 [35] ）不同，目前的医学图像分割模型在比较表示和泛化能力时缺乏统一的框架，导致在正则化方法、损失函数和学习率等配置上存在差异。这限制了对算法性能的全面解释。我们建立了一个统一的基准，并为一些经典模型和现有的最先进模型进行了兼容性处理。更多详情请访问我们的项目网站。我们希望这对于未来其他工作的比较实验有所帮助。

3.2. 实现细节

我们的实验基于 CHASEDB1 [36] ，这是一个公共的视网膜血管参考数据集。该数据集按 0.86 : 0.14 的比例分为训练集和测试集。实验配置概要见表 2 。所有的实验都是使用 Pytorch 框架 [37] 进行的，学习率通过余弦退火进行统一优化。我们在每个时期结束后对测试集上的模型进行评估，并报告所达到的最高分数。

Table 2: 实验的配置设置

Settings	Details
Batch Size	4
Initial Learning Rate	1×10^{-4}
Minimum Learning Rate	1×10^{-5}
Optimizer	Adam [38] (Default $\beta_1 = 0.0$, $\beta_2 = 0.99$)
Input Resolution	256×256
Output Resolution	256×256
Total Epochs	4,000
GPU	Nvidia Tesla V100 32G

本文采用的损失函数 BCEDiceLoss $\mathcal{L}$ 是交叉熵和 DiceLoss [39] 的混合，对于网络输出的对数值 p 和二进制目标 $y \in \{0, 1\}$ ，该损失由两个部分组成：

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log \sigma(p_i) + (1 - y_i) \log(1 - \sigma(p_i))], \tag{19}$$

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_i \sigma(p_i)\, y_i + \varepsilon}{\sum_i \sigma(p_i) + \sum_i y_i + \varepsilon}, \qquad \varepsilon = 10^{-5}, \tag{20}$$

$$\mathcal{L} = 0.5\,\mathcal{L}_{\text{BCE}} + \mathcal{L}_{\text{Dice}}, \tag{21}$$

，其中 $\sigma(p) = \frac{1}{1+e^{-p}}$ 、$i \in 0, 1, 2, \ldots, N$ 、$N = H_s \times W_s$ 是所有像素的总和。

3.3. 度量

我们采用公认的评价指标，总结在表 3 中，其中 P 表示预测的前景，G 表示真实前景，TP ，FP ，TN ，FN 表示真正/假正/假负。

Table 3: 评估指标总结

Metric	Formula	Interpretation
IoU	$\frac{\lvert P \cap G \rvert}{\lvert P \cup G \rvert}$	Overlap between prediction and ground truth.
Dice	$\frac{2\lvert P \cap G \rvert}{\lvert P \rvert + \lvert G \rvert}$	Harmonic mean of precision and recall.
HD	$max\{\sup_{p \in P} \inf_{g \in G} d(p,g), \sup_{g \in G} \inf_{p \in P} d(p,g)\}$	Largest boundary error; lower is better.
HD_{95}	95 % quantile of HD	Less sensitive to outliers than HD.
Recall	$\frac{TP}{TP+FN}$	Ability to find all positives.
Specificity	$\frac{TN}{TN+FP}$	Ability to reject negatives.
Precision	$\frac{TP}{TP+FP}$	Proportion of predicted positives that are true.

3.4. 主要结果

表 4 展示了我们 GDCUNet 在分割性能方面的定量评估，对应于表 1 的各种不同设置。对比的现有最新方法包括模型表示能力的改进和医学图像分割先验知识的建模（主要考虑边缘的拓扑先验）。我们用绿色标记明显突出优点的指标。

Table 4: 与现有方法的定量比较。

Type	Model(With # Parameters)	Venue	IoU	Dice	HD	HD $_{95}$	Re-call	Speci-ficity	Pre-cision
Model Enhancement	Unet # 496M [3]	MICCAI'2015	0.5978	0.7483	15.62	2.000	0.7498	0.9817	0.7469
	Unet++ # 36.1M [5]	ML-CDS'2018 [1]	0.6134	0.7604	15.62	1.414	0.7620	0.9826	0.7588
	Att-Unet # 127M [40]	Arxiv'2018	0.6039	0.7531	16.00	2.236	0.7146	0.9868	0.7959
	Unext # 1.47M [42]	MICCAI'2022	0.5007	0.6673	17.26	3.000	0.6594	0.9772	0.6754
	Uctransnet # 66.2M [15]	AAAI'2022	0.5982	0.7486	15.00	2.000	0.7282	0.9844	0.7702
	Rolling-Unet # 1.78M [9]	AAAI'2024	0.6008	0.7506	15.00	1.414	0.7745	0.9828	0.7568
Prior Knowledge [2]	DconnNet # 25.4M [41]	CVPR'2023	0.5863	0.7444	17.76	3.000	0.7929	0.9752	0.7407
	DSCNet # - -M [34]	ICCV'2023	0.6073	0.7521	15.00	2.0000	0.7534	0.9832	0.7638
Our GDCUnet	Setting 1 # 1.15M	—	0.6160	0.7623	15.56	2.000	0.7495	0.9844	0.7757
	Setting 2 # 1.45M	—	0.6201	0.7655	15.00	1.732	0.7511	0.9848	0.7805
	Setting 3 # 0.954M	—	0.6139	0.7608	15.00	2.000	0.7470	0.9844	0.7750
	Setting 4 # 1.17M	—	0.6161	0.7625	17.23	2.000	0.7433	0.9852	0.7827
	Setting 5 # 1.40M	—	0.6304	0.7733	15.36	1.732	0.7596	0.9853	0.7875
	Setting 6 # 2.16M	—	0.6132	0.7602	15.00	2.000	0.7394	0.9852	0.7823

所有模型的结果表明，在相同配置下，这是一个具有挑战性的项目。几乎所有的 GDCUNet 设置都取得了具有竞争力的表现。设置 5 在 IoU 和 Dice 等多个指标上实现了最佳结果，仅有 1.40M 的参数量。同时，还有一种更轻量的选择（设置 3），在参数数量和分割性能之间达到了平衡。DconnNet 具有最高的召回率，但整体上有限。得益于 SAFDConvolution 对全局自相似复杂局部特征的表征能力，GDCUNet 设置 5 达到了世界领先的综合水平。

图 4 展示了视觉效果的定性比较。DconnNet 建立了先验知识模型，具有相对清晰且完整的分割掩码，但出现了更多的假阳性例子。GDCUnet 拥有显著更好的分割效果，能够分析被现有最先进模型忽略的微小血管。即使是最轻量的设置 3 也能达到具有竞争力的性能。

[1]Held in conjunction with MICCAI 2018.

[2]We emphasize conducting experiments under the unified benchmark and configuration. However, the model based on prior knowledge characterizing the vascular topological structure involves separate settings for data processing, loss functions, etc., and experiments need to be conducted under separate configurations.

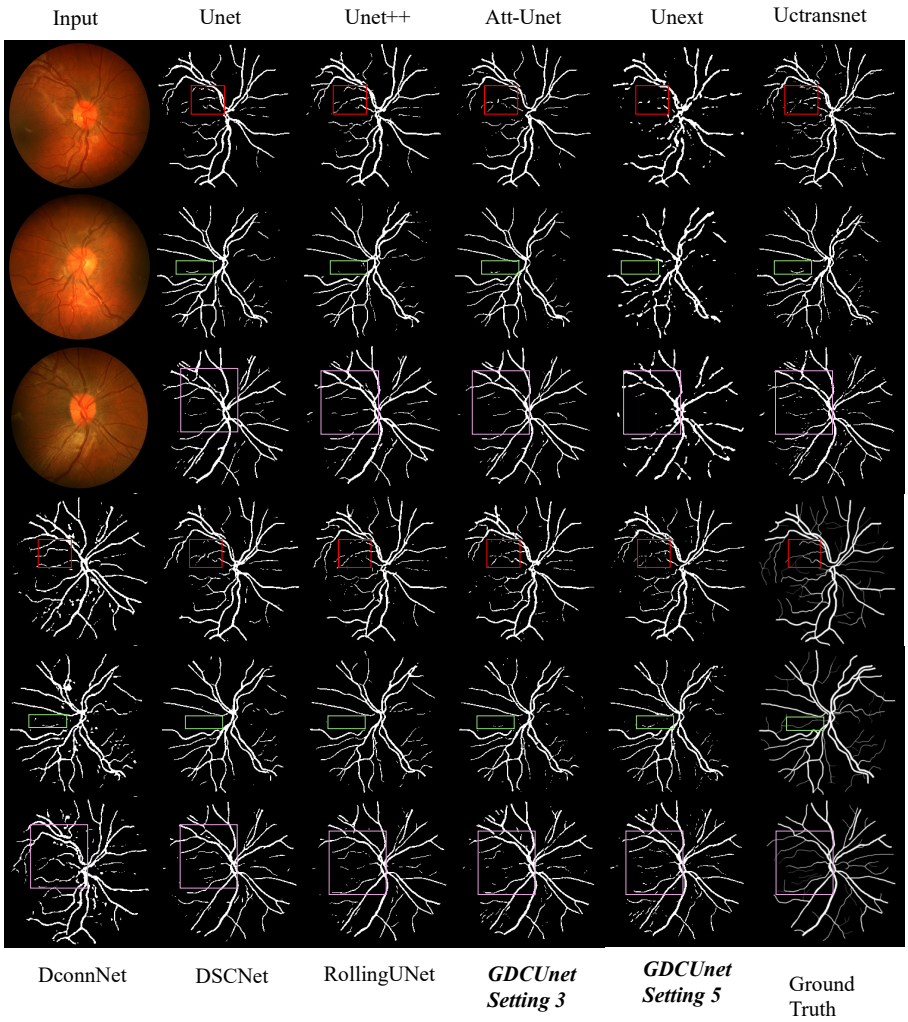


Figure 4: 分割结果的视觉比较。请放大以获得更好的展示效果。

4. 消融研究

4.1. SAFD 卷积的效果

GDCUNet 的主要改进是在特征处理过程中添加了 SAFDConvolution，从而更有效地捕捉具有全局自相似性的复杂局部细节。与其他现有的可变形卷积方法相比，SAFDConvolution 可以更有效地处理全局特征。我们用其他卷积

（统一核大小 5×5 ）替换了 SAFDConvolution Block 中的卷积，定量评估结果如表 5 所示。

Table 5: 卷积变体插入 SAFD 卷积块的比较。

Conv Variant	IoU	Dice
Conventional	0.5833	0.7368
Deformable V1	0.6053	0.7472
Deformable V2	0.5992	0.7438
Deformable V3	0.6130	0.7601
SAFDConvolution	0.6304	0.7733

与现有工作相比，SAFDConvolution 具有显著更高的指标。视觉比较如图 5 所示。即使与 V3 相比，所提出的模块在视觉效果上仍具有显著的优势 [1]。

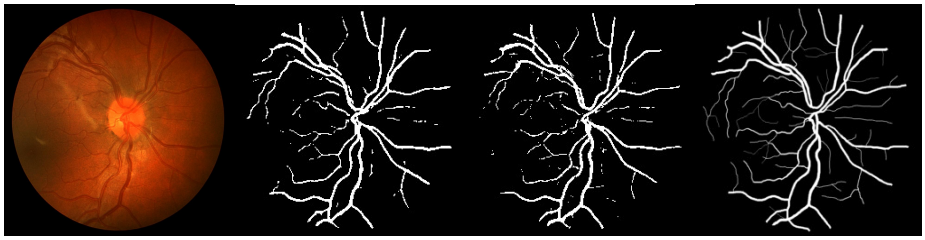


Figure 5: 卷积变体的比较（可变形 V3 和我们的 SAFD 卷积）。

4.2. 特征学习的重要性

SAFDConvolution 对全局自相似的复杂局部细节非常敏感。为了验证这一点，我们在 GDCUnet 编码阶段的 SAFDConvolution 中得到的特征图如图 **??** 所示。

在其他相同的情况下，仅将插入到 SAFDConv 块中的卷积更改为常规卷积，在相同位置获得的特征图如图 6 所示。

[1]The main focus of V4 is on the efficient computing of CUDA and the optimization of memory access. Its essential structure is similar to that of V3. For dynamic snake convolution, we have presented the results in DSCNet in Table 4.

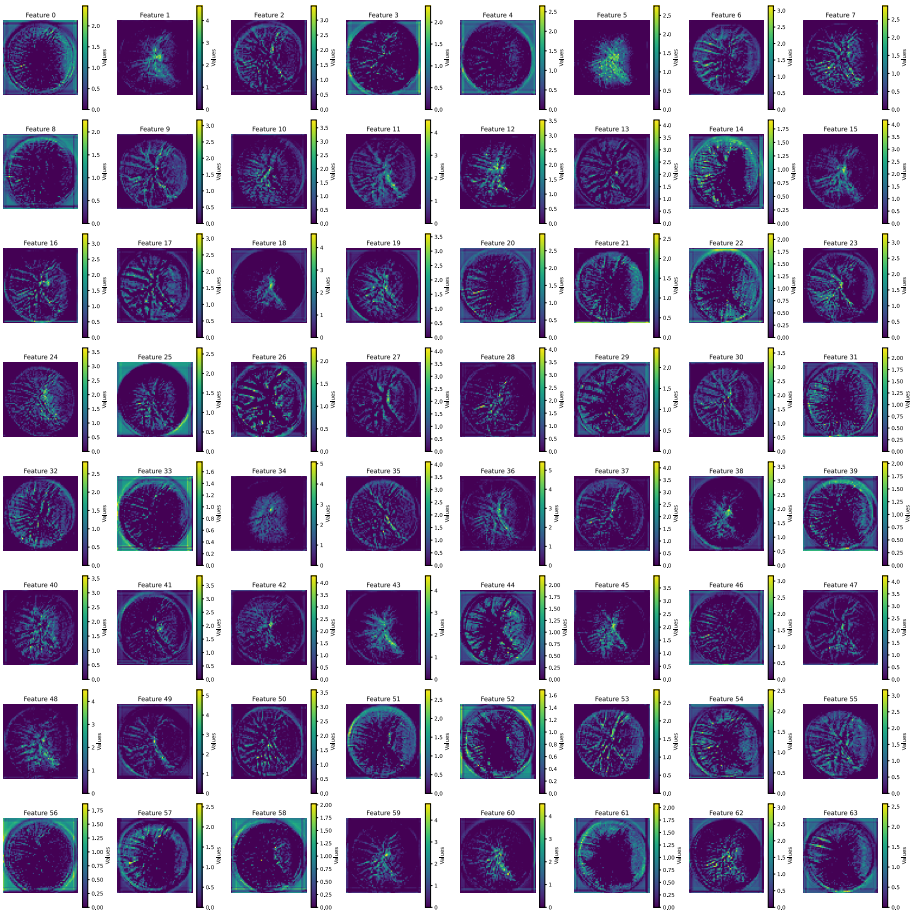


Figure 6: 常规卷积获得的特征图可视化。请放大以获得更好的说明。

SAFDConvolution 利用全局特征变形特征图，从而实现卷积核的相对变形。因此，得到的特征图在人眼中似乎是扭曲的。与图 6 相比，图 **??** 中的特征图显示了更多突出的特征，表明学习到的特征更加显著。

为了进一步验证这一点，我们展示了特征图值的分布。图 7 对应图 **??** ，图 8 对应图 6 。

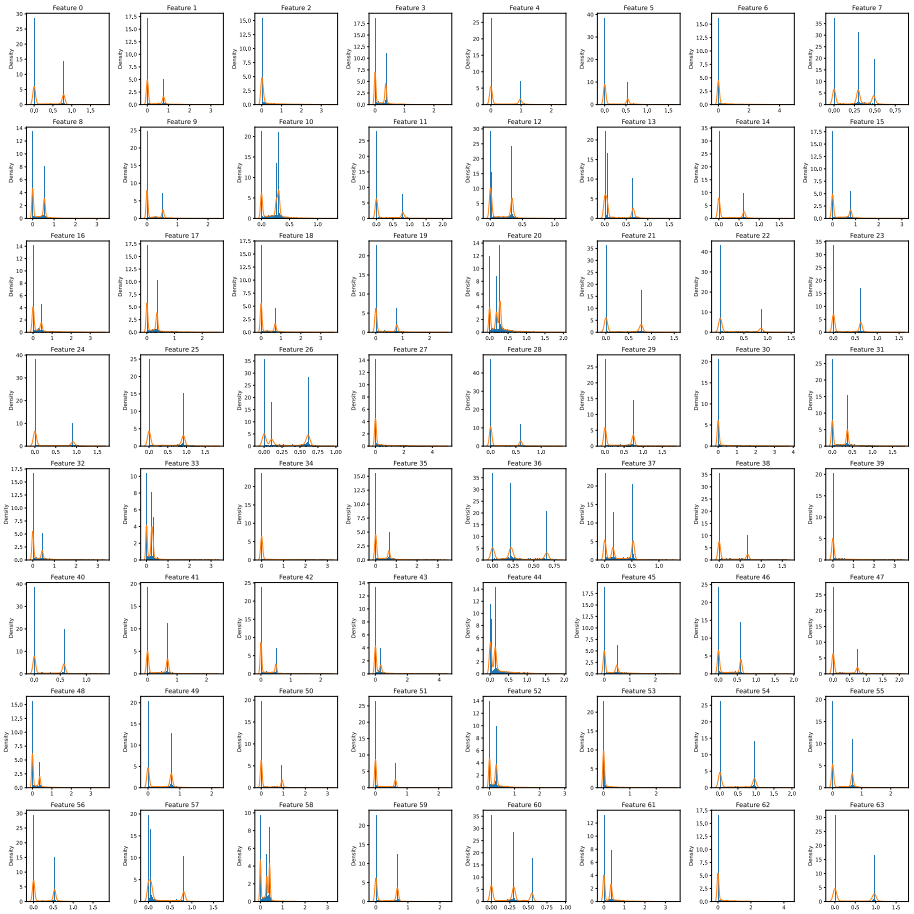


Figure 7: SAFD 卷积的特征图网格值的分布直方图。请放大查看以获得更好的示例说明。

SAFDConvolution 的分布直方图相对尖锐，这表明具有很高的特征分离度、显著的特征学习和一个强目标。

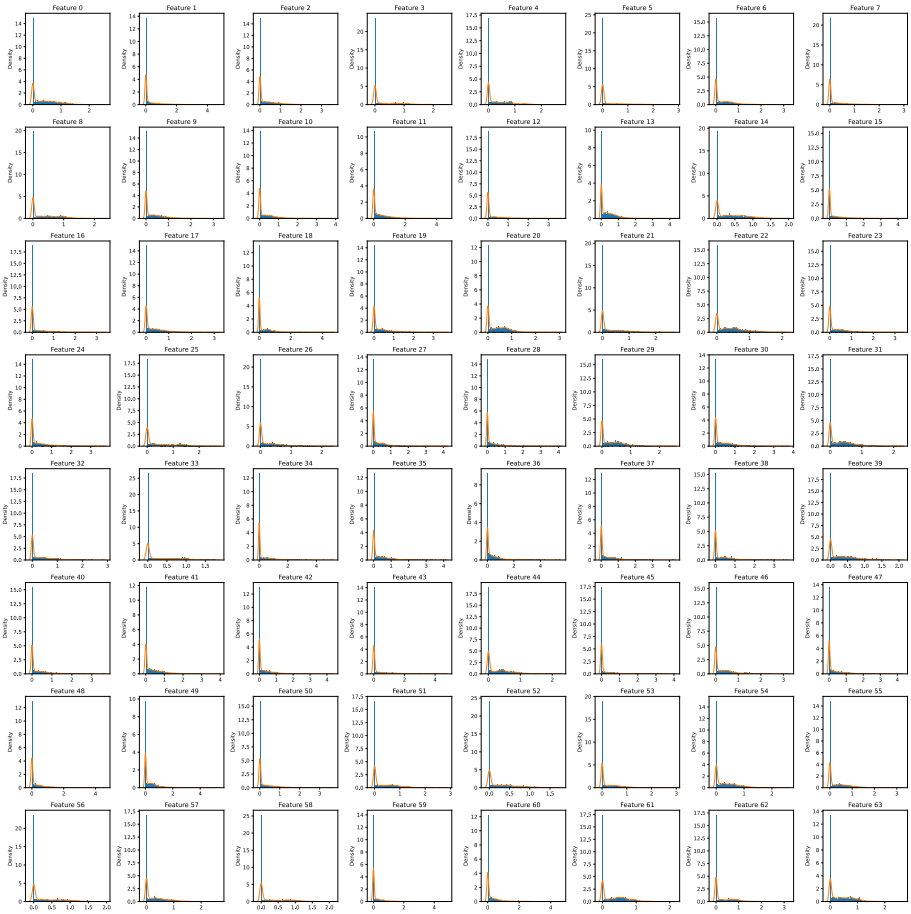


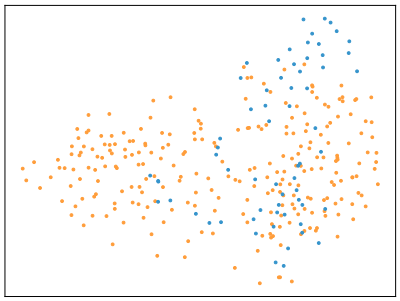
Figure 8: SAFDConvolution 的特征图网格值的分布直方图。请放大以获得更清晰的说明。

与 SAFDConvolution 相比，在相同其他条件下，通过传统卷积获得的特征图中网格点值的分布更加平滑和均匀。传统卷积核在整个视野范围内的响应相对均匀，这导致了特征缠结和辨别能力的下降。

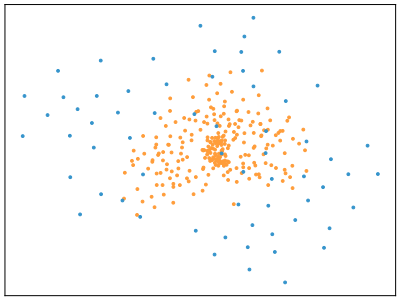
因此，对于局部复杂而整体自相似的血管特征，SAFDConvolution 可以更有效且显著地处理这些特征。

为了进一步验证 SAFDConvolution 的泛化性能，按照 J. Feng 等人的方法，我们使用 t-SNE 来降低学习到的特征张量的维度。首先，使用 SAFDConvolutions 在 GDCUnet 中提取的域不变特征进行可视化。然后，用传统卷积替换它们，同时保持其他结构不变，以可视化传统卷积提取的域不变特征。结果如图

10 所示，其中降维后的源域特征表示为橙色点，目标域特征表示为蓝色点。



(a)



(b)

Figure 9: 学习特征张量的 T-SNE 可视化。(a) SAFD 卷积。(b) 常规卷积。

传统卷积在目标域中的特征分布相对分散，并且源域和目标域之间的重叠较低，表现出明显的域偏移。SAFD 卷积所学习的特征在相同聚类内的源域和目标域样本之间表现出高程度的重叠，表明这种卷积能够有效提取域不变且高度判别的表示，降低域偏移，并具有强的泛化能力。

5. 结论

为了增强深度学习模型表示局部复杂但整体自相似血管特征的能力，我们提出了插拔式 SAFD 卷积，这是一种新型的可变形卷积。与现有的可变形卷积

相比，SAFD 卷积能够实现相对变形，并支持基于空间注意力的偏移学习网络。SAFD 卷积用于实现所提出的眼底血管分割模型 GDCUNet。我们建立了一个统一的实验基准。本基准下的实验表明，GDCUNet 达到了最新水平。消融研究验证了 SAFD 卷积具有更强的表示和泛化能力，可以显著学习关键特征，并显著提高模型性能。未来的工作可以关注 SAFD 卷积在具有类似特征的其他机器视觉任务中的应用。

局限性讨论。引入空间注意力会在矩阵乘法中带来计算开销，这是所有类似 Transformer 结构的常见问题，并限制了在较低性能 GPU 上的训练和推理。我们希望未来通过使用稀疏注意力等方法来改进 SAFDConvolution，以进一步增强其适用性。

6.

CRediT 作者贡献声明

朱乐轩：数据整理、正式分析、调查、方法、资源、软件、可视化、撰写初稿。李宇轩：资源、软件、验证、撰写初稿。任宇宁：方法、概念化、资金获取、调查、监督、验证、撰写—评审和编辑。

7.

利益声明 作者声明他们没有已知的可影响本文工作的竞争性财务利益或个人关系。

8.

数据可用性 本研究使用的数据集是公开可用的。所有代码在我们在摘要中提到的项目网站上可获得。

9.

致谢 本工作得到了中华人民共和国教育部的资助，项目编号为 2024GH-ZDA-GJ-Y-09。

References

[1] A. Shaker, M. Maaz, H. Rasheed, S. Khan, M. -H. Yang, F. Shahbaz Khan, UNETR++: Delving Into Efficient and Accurate 3D Medical Image Segmentation, IEEE Trans. Med. Imaging , 43 (9) (2024) 3377–3390. https://doi.org/10.1109/TMI.2024.3398728.

[2] J. Wang, Y. Tang, Y. Xiao, J. T. Zhou, Z. Fang, F. Yang, GREnet: Gradually REcurrent Network With Curriculum Learning for 2-D Medical Image Segmentation, IEEE Trans. Neural Netw. Learn. Syst. , 35 (7) (2024) 10018–10032. https://doi.org/10.1109/TNNLS.2023.3238381.

[3] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in: N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 , Springer International Publishing, Cham, 2015, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

[4] A. DiSpirito, D. Li, T. Vu, M. Chen, D. Zhang, J. Luo, R. Horstmeyer, J. Yao, Reconstructing Undersampled Photoacoustic Microscopy Images Using Deep Learning, IEEE Trans. Med. Imaging , 40 (2) (2021) 562–570. https://doi.org/10.1109/TMI.2020.3031541.

[5] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, UNet++: A Nested U-Net Architecture for Medical Image Segmentation, in: D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. S. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, A. Madabhushi (Eds.), Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support , Springer International Publishing, Cham, 2018, pp. 3–11. https://doi.org/10.1007/978-3-030-00889-5_1.

[6] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J.

Uszkoreit, N. Houlsby, An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, arXiv preprint arXiv:2010.11929, 2021. https://arxiv.org/abs/2010.11929.

[7] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, V. M. Patel, Medical Transformer: Gated Axial-Attention for Medical Image Segmentation, in: M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, C. Essert (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2021 , Springer International Publishing, Cham, 2021, pp. 36–46. https://doi.org/10.1007/978-3-030-87193-2_4.

[8] Y. Cui, A. Knoll, Enhancing Local–Global Representation Learning for Image Restoration, IEEE Trans. Ind. Inform. , 20 (4) (2024) 6522–6530. https://doi.org/10.1109/TII.2023.3345464.

[9] Y. Liu, H. Zhu, M. Liu, H. Yu, Z. Chen, J. Gao, Rolling-Unet: Revitalizing MLP' s Ability to Efficiently Extract Long-Distance Dependencies for Medical Image Segmentation, Proc. AAAI Conf. Artif. Intell. , 38 (4) (2024) 3819–3827. https://doi.org/10.1609/aaai.v38i4.28173.

[10] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, Deformable Convolutional Networks, in: Proc. IEEE Int. Conf. Comput. Vis. (ICCV) , 2017, pp. 764–773. https://doi.org/10.1109/ICCV.2017.89.

[11] Q. Ming, X. Xiao, Towards Accurate Medical Image Segmentation With Gradient-Optimized Dice Loss, IEEE Signal Process. Lett. , 31 (2024) 191–195. https://doi.org/10.1109/LSP.2023.3329437.

[12] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation, in: L. Karlinsky, T. Michaeli, K. Nishino (Eds.), Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops , Springer Nature Switzerland, Cham, 2023, pp. 205–218. https://doi.org/10.1007/978-3-031-25066-8_9.

[13] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows, in: Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV) , 2021, pp. 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986.

[14] S.-I. Jang, T. Pan, Y. Li, P. Heidari, J. Chen, Q. Li, K. Gong, Spach Transformer: Spatial and Channel-Wise Transformer Based on Local and Global Self-Attentions for PET Image Denoising, IEEE Trans. Med. Imaging , 43 (6) (2024) 2036–2049. https://doi.org/10.1109/TMI.2023.3336237.

[15] H. Wang, P. Cao, J. Wang, O. R. Zaiane, UCTransNet: Rethinking the Skip Connections in U-Net from a Channel-Wise Perspective with Transformer, Proc. AAAI Conf. Artif. Intell. , 36 (3) (2022) 2441–2449. https://doi.org/10.1609/aaai.v36i3.20144.

[16] S. He, P. E. Grant, Y. Ou, Global-Local Transformer for Brain Age Estimation, IEEE Trans. Med. Imaging , 41 (1) (2022) 213–224. https://doi.org/10.1109/TMI.2021.3108910.

[17] J. Li, MCTE: Marrying Convolution and Transformer Efficiently for End-to-End Medical Image Segmentation, in: Proc. IEEE Int. Conf. Image Process. (ICIP) , 2023, pp. 1100–1104. https://doi.org/10.1109/ICIP49359.2023.10222041.

[18] J. Song, X. Chen, Q. Zhu, F. Shi, D. Xiang, Z. Chen, Y. Fan, L. Pan, W. Zhu, Global and Local Feature Reconstruction for Medical Image Segmentation, IEEE Trans. Med. Imaging , 41 (9) (2022) 2273–2284. https://doi.org/10.1109/TMI.2022.3162111.

[19] X. Zhao, J. Zhang, Q. Li, T. Zhao, Y. Li, Z. Wu, Global and local multi-modal feature mutual learning for retinal vessel segmentation, Pattern Recognit. , 151 (2024) 110376. https://doi.org/10.1016/j.patcog.2024.110376.

[20] X. Wang, R. Girshick, A. Gupta, K. He, Non-local Neural Networks, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) , 2018, pp. 7794–7803. https://doi.org/10.1109/CVPR.2018.00813.

[21] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, CBAM: Convolutional Block Attention Module, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), Proc. Eur. Conf. Comput. Vis. (ECCV) , Springer International Publishing, Cham, 2018, pp. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.

[22] Q. Hou, D. Zhou, J. Feng, Coordinate Attention for Efficient Mobile Network Design, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) , 2021, pp. 13708–13717. https://doi.org/10.1109/CVPR46437.2021.01350.

[23] L. Qi, X. Qin, F. Gao, J. Dong, X. Gao, SAWU-Net: Spatial Attention Weighted Unmixing Network for Hyperspectral Images, IEEE Geosci. Remote Sens. Lett. , 20 (2023) 1–5. https://doi.org/10.1109/LGRS.2023.3270183.

[24] G. Sun, Z. Pan, A. Zhang, X. Jia, J. Ren, H. Fu, K. Yan, Large Kernel Spectral and Spatial Attention Networks for Hyperspectral Image Classification, IEEE Trans. Geosci. Remote Sens. , 61 (2023) 1–15. https://doi.org/10.1109/TGRS.2023.3292065.

[25] B. Zhang, Y. Chen, S. Xiong, X. Lu, Hyperspectral Image Classification via Cascaded Spatial Cross-Attention Network, IEEE Trans. Image Process. , 34 (2025) 899–913. https://doi.org/10.1109/TIP.2025.3533205.

[26] Y. Xu, Z. Xu, W. Mei, Q. Yang, Y. Zhang, H. Chen, Y. Xu, SLCAM: A Lightweight Spatial Location Channel Attention Module for Image Classification, Expert Syst. Appl. , 287 (2025) 128100. https://doi.org/10.1016/j.eswa.2025.128100.

[27] Y. Bai, Q. Zou, X. Chen, L. Li, Z. Ding, L. Chen, Extreme Low-Resolution Action Recognition with Confident Spatial-Temporal At-

tention Transfer, Int. J. Comput. Vis. , 131 (6) (2023) 1550–1565. https://doi.org/10.1007/s11263-023-01771-4.

[28] L. Wang, M. Cao, Y. Zhong, X. Yuan, Spatial-Temporal Transformer for Video Snapshot Compressive Imaging, IEEE Trans. Pattern Anal. Mach. Intell. , 45 (7) (2023) 9072–9089. https://doi.org/10.1109/TPAMI.2022.3225382.

[29] T. Pu, T. Chen, H. Wu, Y. Lu, L. Lin, Spatial–Temporal Knowledge-Embedded Transformer for Video Scene Graph Generation, IEEE Trans. Image Process. , 33 (2024) 556–568. https://doi.org/10.1109/TIP.2023.3345652.

[30] X. Zhu, H. Hu, S. Lin, J. Dai, Deformable ConvNets V2: More Deformable, Better Results, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) , 2019, pp. 9300–9308. https://doi.org/10.1109/CVPR.2019.00953.

[31] W. Wang, J. Dai, Z. Chen, Z. Huang, Z. Li, X. Zhu, X. Hu, T. Lu, L. Lu, H. Li, X. Wang, Y. Qiao, InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) , 2023, pp. 14408–14419. https://doi.org/10.1109/CVPR52729.2023.01385.

[32] Y. Xiong, Z. Li, Y. Chen, F. Wang, X. Zhu, J. Luo, W. Wang, T. Lu, H. Li, Y. Qiao, L. Lu, J. Zhou, J. Dai, Efficient Deformable ConvNets: Rethinking Dynamic and Sparse Operator for Vision Applications, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) , 2024, pp. 5652–5661. https://doi.org/10.1109/CVPR52733.2024.00540.

[33] F. Yu, V. Koltun, Multi-Scale Context Aggregation by Dilated Convolutions, arXiv preprint arXiv:1511.07122, 2016. https://arxiv.org/abs/1511.07122.

[34] Y. Qi, Y. He, X. Qi, Y. Zhang, G. Yang, Dynamic Snake Convolution Based on Topological Geometric Constraints for Tubular Structure Segmentation, in: Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV) , 2023, pp. 6047–6056. https://doi.org/10.1109/ICCV51070.2023.00558.

[35] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, M. Long, iTransformer: Inverted Transformers Are Effective for Time Series Forecasting, arXiv preprint arXiv:2310.06625, 2024. https://arxiv.org/abs/2310.06625.

[36] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, S. A. Barman, An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation, IEEE Trans. Biomed. Eng. , 59 (9) (2012) 2538–2548. https://doi.org/10.1109/TBME.2012.2205687.

[37] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An Imperative Style, High-Performance Deep Learning Library, arXiv preprint arXiv:1912.01703, 2019. https://arxiv.org/abs/1912.01703.

[38] D. P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, arXiv preprint arXiv:1412.6980, 2017. https://arxiv.org/abs/1412.6980.

[39] S. A. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, G. Hamarneh, Combo loss: Handling input and output imbalance in multi-organ segmentation, Comput. Med. Imaging Graph. , 75 (2019) 24–33. https://doi.org/10.1016/j.compmedimag.2019.04.005.

[40] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, D. Rueckert, Attention U-Net: Learning Where to Look for the Pancreas, arXiv preprint arXiv:1804.03999, 2018. https://arxiv.org/abs/1804.03999.

[41] Z. Yang, S. Farsiu, Directional Connectivity-Based Segmentation of Medical Images, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) , 2023, pp. 11525–11535. https://doi.org/10.1109/CVPR52729.2023.01109.

[42] J. M. J. Valanarasu, V. M. Patel, UNeXt: MLP-Based Rapid Medical Image Segmentation Network, in: L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, S. Li (Eds.), Proc. Med. Image Comput. Comput. Assist. Interv. (MICCAI) , Springer Nature Switzerland, Cham, 2022, pp. 23–33. https://doi.org/10.1007/978-3-031-16443-9_3.

[43] J. Feng, T. Zhang, J. Zhang, R. Shang, W. Dong, G. Shi, L. Jiao, S4DL: Shift-Sensitive Spatial–Spectral Disentangling Learning for Hyperspectral Image Unsupervised Domain Adaptation, IEEE Trans. Neural Netw. Learn. Syst. , 2025, pp. 1–15. https://doi.org/10.1109/TNNLS.2025.3556386.