

基于扩散模型的皮肤病变分割方法

徐 行, 杨瑜蓉, 于春蕾, 暴 爽

燕山大学理学院, 河北 秦皇岛

收稿日期: 2025年2月28日; 录用日期: 2025年3月27日; 发布日期: 2025年4月3日

摘要

皮肤病变分割在临床诊断中扮演着重要角色。为了加强对多尺度上下文信息的学习, 提高模型在皮肤病变边缘处的分割准确度, 本研究提出一种新的基于扩散模型的皮肤病变分割框架DCA-SegDiff。该框架将双交叉注意模块集成到网络的跳跃连接部分, 能够同时接收来自不同层编码器的多尺度特征信息, 同时与解码器部分的挤压激励机制结合使用, 以充分利用上下文特征信息。同时, 采用基于Haar小波的下采样模块, 以保留更多边缘细节信息。此外, 在编码器深层采用高效增强多尺度注意力模块, 通过跨空间学习的方法来融合多尺度特征信息。为验证模型的优越性, 在ISIC2018、PH2、HAM10000三个皮肤病变分割数据集上进行实验, 主要评价指标Dice分别达到了0.8944、0.9446、0.9445。结果表明, DCA-SegDiff模型在三个数据集上表现优于其他现有模型, 而且相比基线模型的参数量显著减少, 证明了其在皮肤病变分割任务中的有效性和泛化能力。

关键词

扩散模型, 皮肤病变分割, Haar小波下采样, 高效增强多尺度注意, 双交叉注意

Skin Lesions Segmentation Method Based on Diffusion Model

Xing Xu, Yurong Yang, Chunlei Yu, Shuang Bao

School of Science, Yanshan University, Qinhuangdao Hebei

Received: Feb. 28th, 2025; accepted: Mar. 27th, 2025; published: Apr. 3rd, 2025

Abstract

Segmentation of skin lesions plays an important role in clinical diagnosis. In this paper, a new skin lesion segmentation framework based on diffusion model, DCA-SegDiff, is proposed to enhance the learning of multi-scale context information to improve the segmentation accuracy of the model at the edge of skin lesions. The dual cross-attention module is integrated into the skip connection part

of the denoising network in this framework, which can receive multi-scale feature information from different encoders at the same time. Meanwhile, the down-sampling module based on Haar wavelet is adopted to retain more edge detail information. In addition, an efficient enhanced multi-scale attention module is used in the deep layer of the encoder to integrate multi-scale feature information by cross-space learning method. In order to verify the superiority of the model, experiments are conducted on three skin lesion segmentation datasets, ISIC2018, PH2, and HAM10000, and Dice reaches 0.8944, 0.9446, and 0.9445, respectively. The results show that the DCA-SegDiff model outperforms other existing models on the three datasets, and the number of parameters is significantly reduced compared to the Baseline, demonstrating its effectiveness and generalization ability in the skin lesion segmentation task.

Keywords

Diffusion Model, Skin Lesion Segmentation, Haar Wavelet Downsampling, Efficient Enhanced Multi-Scale Attention, Dual Cross-Attention

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

皮肤病的发病率和死亡率不断增长，皮肤癌已跻身最常见癌症之列[1]。皮肤病变区域在形状、颜色、大小上呈现多样性，且病变与正常区域之间边界模糊、对比度低，这使得对病变区域进行准确的分割成为一项具有挑战性的任务[2]。

目前，皮肤病变图像分割方法分为传统图像分割方法和基于深度学习的分割方法，其中深度学习方法主要基于卷积神经网络(Convolutional Neural Network, CNN)、Transformer 和扩散模型。

传统分割方法主要包括阈值法、边缘检测法和区域生长法等。胡越黎等[3]提出的自适应阈值方法，能够根据图像的 S 域标准差选择不同方法确定分割阈值，以适应不同症状的皮肤图像分割。Yasmin 等[4]使用 Canny 边缘检测器进行迭代分割，通过不断迭代对边缘检测图像进行中值滤波。徐国雄等[5]结合分水岭算法和区域合并算法，有效克服了过分割的不良效果。但这些传统分割方法比较依赖人工，不仅费时费力、效率低，分割效果也差强人意。

随着深度学习的发展，CNN 广泛应用于各类分割任务。2015 年，Ronneberger 等[6]提出具有对称编解码器结构的 U-Net，并引入跳跃连接，在各类医学图像分割任务中表现十分优异。之后在此基础上，Diakogiannis 等[7]提出 ResUNet，加入残差思想，以缓解梯度消失问题。Oktay 等[8]提出 AttU-Net，在 U-Net 的跳跃连接部分加入注意力门控机制。然而，CNN 结构由于局部感受野的限制，很难捕获全局上下文信息，这可能会对分割精度产生影响，尤其是在病变区域的形状复杂或边缘不清晰的情况下。

为解决这一问题，Dosovitskiy 等[9]提出了一种基于 Transformer 的网络，通过自注意力机制捕获长距离依赖关系，能够同时获取局部细节和全局上下文信息。之后的 TransUNet[10]，首次采用混合 CNN-Transformer 架构，有效结合了 CNN 在局部特征提取的优势和 Transformer 在全局关系建模的能力。Valanarasu 等[11]提出通用医学图像分割模型 MedT，引入一种门控轴向注意力模块，使用局部 - 全局训练策略，在多个医学图像数据集上都取得了不错的分割效果。杨澜[12]设计了双分支并行结构，能够有效提取多尺度特征信息。但 Transformer 架构通常需要大量的数据用于训练，而医学图像数据集通常规模较小且获取成本高，导致其在医学图像任务上的应用往往会受到更多限制。

去噪扩散概率模型(Denoising Diffusion Probability Model, DDPM) [13]在图像处理领域已经展现出了独特的优势。其训练过程具有更好的可解释性和稳定性,减少了训练中出现的模式崩溃现象,提高了模型的泛化能力。Julia 等[14]提出 EnsDiff, 使用原始图像作为引导图像进行训练,并且在采样过程中使用隐式分割集成。Wu 等[15]提出首个通用医学图像分割模型 MedSegDiff, 添加动态条件编码,同时通过特征频率解析器 FF-Parser 来消除高频噪声的影响。之后 Wu 等[16]又提出改进的 MedSegDiff-V2, 在条件UNet 中使用一种新的频谱空间变压器 SS-Former 来实现噪声和语义特征之间的交互。尽管这些方法已经展现出很好的性能,但对于皮肤分割任务,应该加强对病变边界处的关注。Afshin 等[17]在 DermoSegDiff 中引入了一个新的强调边界信息的加权损失函数。但这种方法需要反复调整系数来平衡边缘和非边缘区域之间的学习,而这依赖于费力的试错,临床应用难度高。因此,本研究对 DermoSegDiff 加以改进,强调对多尺度上下文信息的捕获,以提高模型在皮肤病变边缘处的分割精度。首先,在跳跃连接部分融入双交叉注意模块,捕获多尺度编码器之间的长距离依赖关系。其次,考虑到扩散模型参数量大的问题,在编码器部分引入更轻量级的高效增强多尺度注意机制,增强模型对病变区域的聚焦能力。另外,使用基于 Haar 小波下采样方法,保留更多的空间细节信息。

2. 扩散模型介绍

本文采用的核心架构基于 DDPM。扩散模型包括两个过程:固定的正向扩散过程 q 和可训练的反向去噪过程 p_θ 。在正向扩散过程中,首先将标签掩码 $\mathbf{x}_0 \in \mathbf{R}^{H \times W \times 1}$ 输入模型,随着时间步 t 逐步加入高斯噪声,将其转化为噪声掩码。这个过程是一个可推导的马尔可夫过程,可表示为:

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = N\left(\sqrt{1-\beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}\right), \forall t \in \{1, 2, \dots, T\} \quad (1)$$

$$q(\mathbf{x}_{1:T} | \mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1}) \quad (2)$$

其中, $\beta_1, \beta_2, \dots, \beta_T$ 表示扩散步骤的噪声方差序列, \mathbf{I} 为单位矩阵。定义 $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$ 和 $\varepsilon \sim N(0, 1)$,

随机采样 \mathbf{x}_t 及其概率分布如下:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + (1 - \bar{\alpha}_t) \varepsilon \quad (3)$$

$$q(\mathbf{x}_t | \mathbf{x}_0) = N\left(\sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}\right) \quad (4)$$

扩散模型的性质表明,从 $q(\mathbf{X}_T) \sim N(0, \mathbf{I})$ 到 $q(\mathbf{X}_0)$ 的逆过程是马尔可夫过程。但条件高斯分布 $q(\mathbf{X}_t | \mathbf{X}_{t-1})$ 的均值和方差参数未知,需要利用神经网络去学习一个模型 p_θ 生成预测噪声。

给定原始图像 $\mathbf{g} \in \mathbf{R}^{H \times W \times 3}$ 作为指导,和标签掩码 $\mathbf{x} \in \mathbf{R}^{H \times W \times 1}$ 连接起来作为去噪网络的输入,表示为 $\mathbf{I}_t \in \mathbf{R}^{H \times W \times (3+1)}$ 。该过程可定义为:

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = N\left(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{I}_t, t), \sum_\theta(\mathbf{I}_t, t)\right) \quad (5)$$

通过优化负对数似然函数的变分下界得到最终的目标函数如下:

$$L = \|\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_\theta(\mathbf{I}_t, t)\|^2 \quad (6)$$

其中, $\boldsymbol{\varepsilon}$ 和 $\boldsymbol{\varepsilon}_\theta$ 表示真实噪声和预测噪声。此目标函数作为网络训练过程中的损失函数,指导训练过程的更新学习。而采样过程 $t-1$ 步的图片由 t 步的图片减去一个预测噪声 $\boldsymbol{\varepsilon}_\theta(\mathbf{I}_t, t)$ 再加上一个高斯噪声 $\boldsymbol{\varepsilon}$ 得到,最终得到预测掩码。

3. DCA-SegDiff 模型

3.1. 去噪网络模型 DCA-SegDiff

本文的去噪网络采用的是一个改进的 ResUnet 架构, 如下图 1 所示。模型输入包括噪声掩码 x 、引导图像 g 以及时间嵌入 t 。

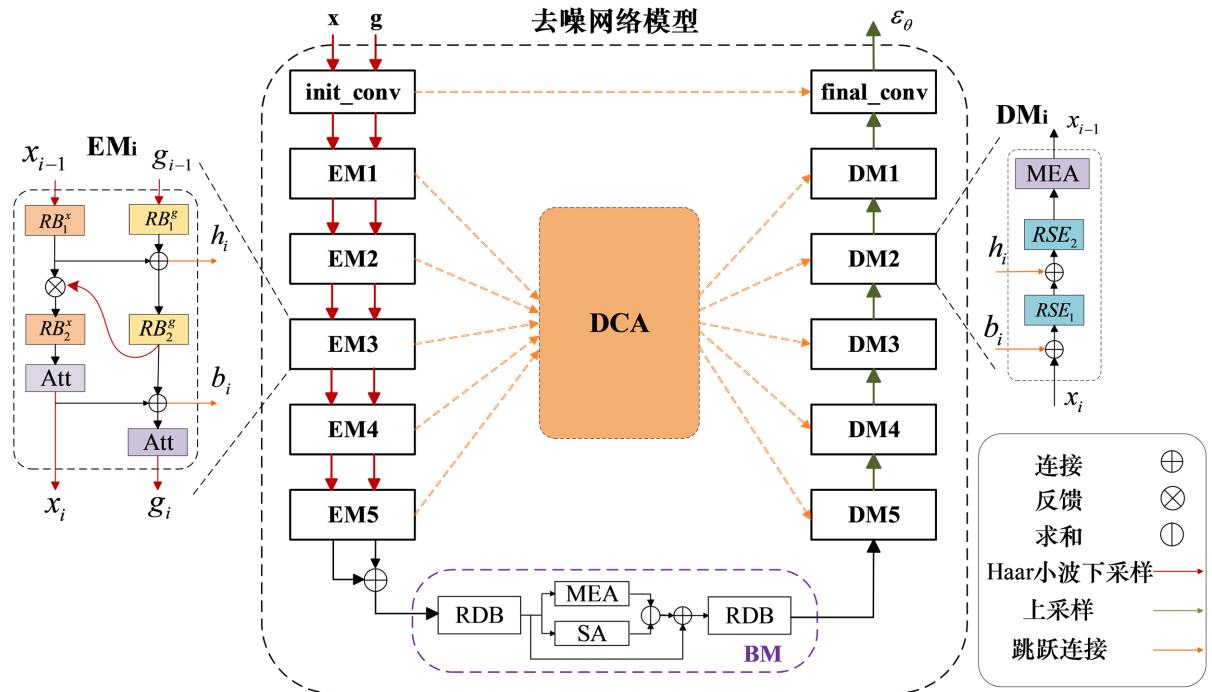


Figure 1. Denoising network model

图 1. 去噪网络模型

编码器部分 EM: 每个编码器模块包括 x 和 g 双路径输入, 每条路径包含两个连续的残差块 RB 和一个注意力模块 Att。相比单路径来说, 双路径编码器可以学习到更加多样化的特征表示, 以更好地实现特征提取[17]。为融合时间步信息, 在每个残差块中都引入了时间嵌入, 这是一个包括正弦位置嵌入、线性层和 GELU 激活函数的处理序列。将 RB_1^x 和 RB_1^g 的输出进行特征拼接, 得到融合了噪声和图像信息的 h_i 。接着, 对 RB_2^g 的输出进行卷积操作反馈连接到 RB_2^x 的输入上, 这种反馈和特征映射相乘充当了一种注意力机制, 可以突出重要特征的同时衰减一些不太重要的特征。之后 RB_2^x 的输出经过注意力模块之后与 RB_2^g 的输出进行特征拼接, 得到另一个输出 b_i 。每个 EM_i 模块最终会得到四个输出, 其中 (h_i, b_i) 用于编解码器之间的跳跃连接; (x_i, g_i) 用于传递给下一个编码器模块继续进行特征提取。

同时, 编码器部分使用 Haar 小波下采样模块(Haar Wavelet Downsampling Module, HWD), 能够保留更多的语义信息。另外需要注意的是, 不同位置的编码器采用了不同的注意力机制, 编码器浅层(前 3 层)采用了改进的线性注意力模块(Modified Linear Attention Module, MLA), 编码器深层(后 2 层)采用了高效增强多尺度注意力模块(Efficient Enhanced Multi-Scale Attention Module, EEMA), 能够更加有效地处理图像信息, 提高分割精度。

跳跃连接部分: 跳跃连接部分融入了双交叉注意模块(Dual Cross-Attention Module, DCA), 能够有效提取多尺度编码器特征之间的全局通道和空间依赖。

底部部分 BM: 连接最深层的编解码器模块。将编码器最后一层的两个输出($\mathbf{x}_5, \mathbf{g}_5$)拼接到一起传入 BM 部分, 依次通过一个残差空洞卷积模块(Residual Dilated Convolution Block, RDB)、一个线性自注意力模块(Linear Self-Attention Module, LSA)和另一个 RDB 模块。RDB 模块中使用空洞率为 2 的空洞卷积[18]替换了普通二维卷积, 能够在不增加参数量的情况下扩大感受野, 适用于底部特征图分辨率较小的情况, 可以有效减少语义信息的损失。LSA 包含一个线性注意力模块[19]和一个自注意力模块[9], 自注意力模块能够捕捉序列各元素间的全局依赖关系, 使网络更好地理解和处理上下文信息。与线性注意力模块结合使用, 进一步增强了模型的特征表达能力。

解码器部分 DM: 与编码器块 EM_i 的双路径不同, 解码器块 DM_i 是单路径。它包括两个连续的 RSE 模块和一个线性注意力模块 EA。同时, RSE 在 RB 模块的基础上添加了挤压激励机制(Squeeze Excitation, SE) [20], 根据通道关系来调整各权重大小, 自动校准每个特征通道的重要性。与跳跃连接部分的 DCA 模块结合使用, 进一步丰富了网络对特征的处理和解析能力。将应用 RSE_i^d 模块之前和之后的特征图分别与编码器块的输出($\mathbf{h}_i, \mathbf{b}_i$)连接起来, 完成编解码器之间的跳跃连接。最后经过一个 MLA 模块, 得到最终的输出。

3.2. Haar 小波下采样模块

传统的下采样方法如最大池化或跨步卷积, 可能会丢失一些重要的空间细节信息, 如边界和纹理信息。为了减少网络下采样过程中的信息损失, 这里采用了一种 Haar 小波下采样模块 HWD, 利用 Haar 小波变换来降低特征图的空间分辨率, 以保留尽可能多的信息[21], 如图 2 所示。

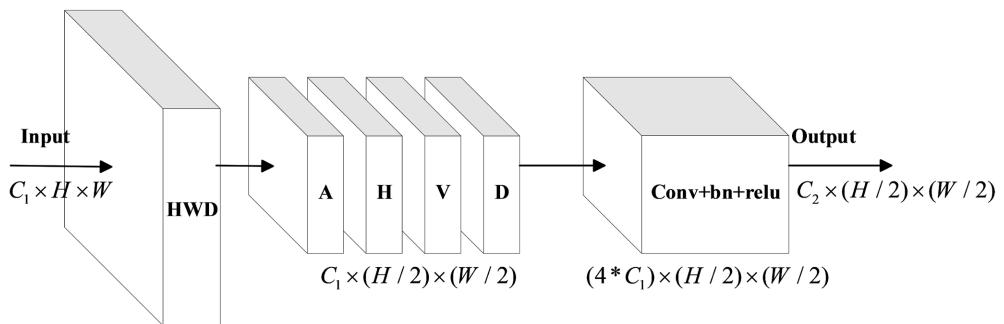


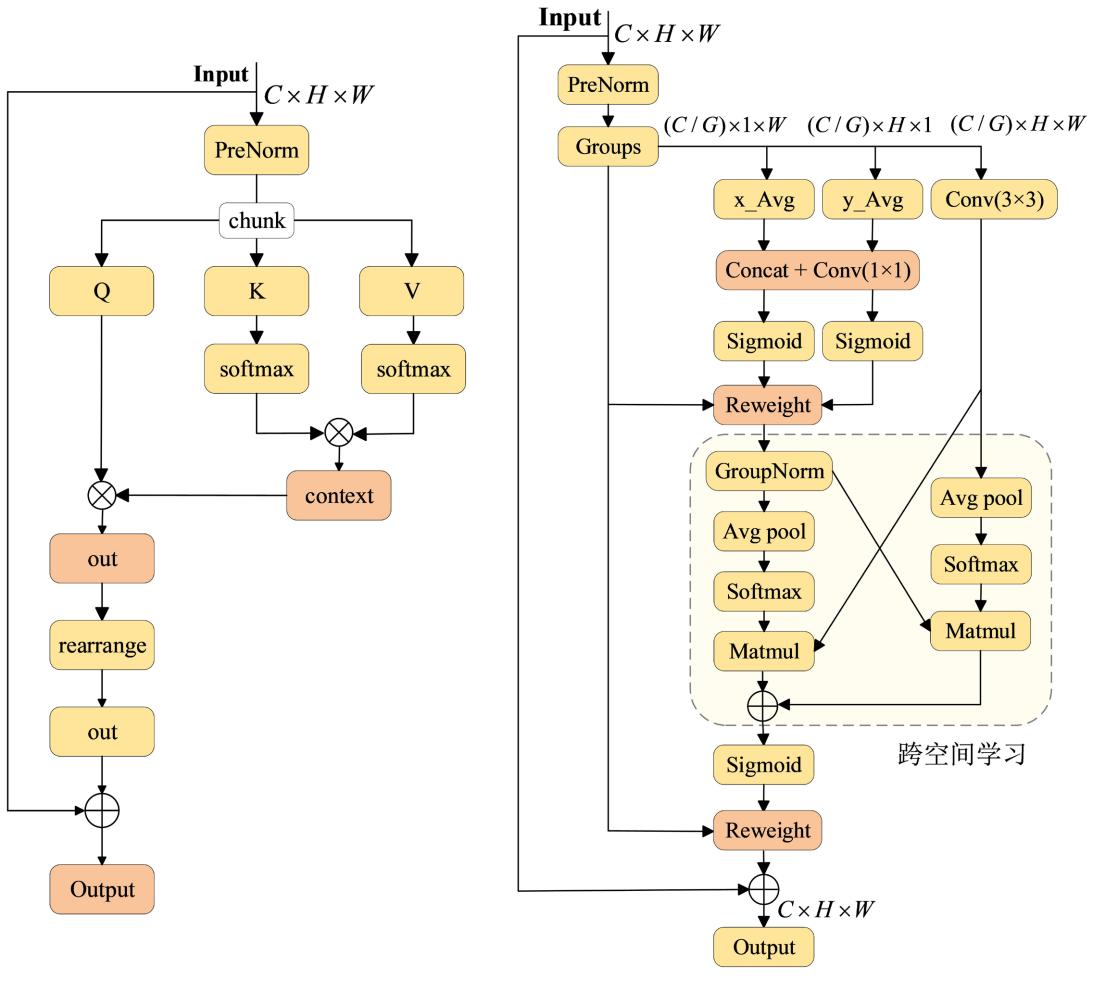
Figure 2. Haar-wavelet downsampling module

图 2. Haar 小波下采样模块

在 HWD 模块中, 首先对输入的特征图进行二维离散 Haar 小波变换, 得到低频分量 A 、水平方向的高频分量 H 、垂直方向的高频分量 V 和对角线方向的高频分量 D 。低频分量包含了图像的基本轮廓信息, 而高频分量则包含了边缘、纹理等细节信息。接着, 从高频分量中提取出不同方向的高频分量进行重组, 构建新的特征图。最后, 将处理后的低频信息和重组后的高频信息拼接起来, 并通过一个包含卷积、批归一化和 RELU 激活函数的处理序列, 生成最终的下采样特征图。

3.3. 改进的注意力模块

高效多尺度注意力模块通过其中的并行子结构, 能够同时处理多尺度的特征信息, 同时使用了更少的参数[22]。为了让网络能够学习到更丰富的多尺度特征信息, 解决皮肤病变区域边界模糊等问题, 本研究在网络编码器深层采用了一种高效增强多尺度注意模块 EEMA, 以满足深层网络对复杂特征提取的需求, 且参数量更少。而在编码器浅层, 由于特征相对简单且保留了较多原始信息, 采用了改进的线性注意力模块 MLA。下图 3 给出了 MLA 模块和 EEMA 模块结构。



(a) 改进的高效线性注意力模块MEA

(b) 改进的高效多尺度注意力模块MEMA

Figure 3. Modified attention module

图 3. 改进的注意力模块

在 MLA 模块中，使用线性层将输入特征映射到查询(\mathbf{Q})、键(\mathbf{K})和值(\mathbf{V})。经过 Softmax 函数运算后通过一个卷积层输出，可表示为：

$$LA(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = Conv\left(\left(\rho_q(\mathbf{Q})\rho_k(\mathbf{K})^T\right)\mathbf{V}\right) \quad (7)$$

其中， ρ_q 和 ρ_k 分别表示 \mathbf{Q} 和 \mathbf{K} 的规范化函数。

另外，MLA 模块在普通线性注意力模块中增加了预归一化操作和残差连接，表示为如下形式：

$$MEA(\mathbf{x}) = \mathbf{x} \oplus PerNorm(LA(\mathbf{x})) \quad (8)$$

预归一化操作有助于保持变量稳定，减少训练过程中的梯度消失或爆炸问题；残差连接可以加速模型收敛，使模型更加高效地学习皮肤病变区域的多尺度特征信息[7]。

在 EEMA 模块中，将输入 $\mathbf{X} \in \mathbf{R}^{C \times H \times W}$ 在通道维度方向上划分为 G 个子特征，使模型能够同时处理多尺度的信息。接着，引入多尺度并行子网络，使用三个并行子路径来提取不同尺度的特征权重。在两个 1×1 分支上，沿着宽度和高度两个方向分别进行一维全局平均池化，以分别在两个空间方向上编码通

道信息。接着，将池化后的特征图拼接在一起，通过 1×1 卷积和激活函数得到两个新的特征图，分别表示高度和宽度方向上的注意力权重。最后，将这两个注意力权重与子特征图相乘进行加权，再经过归一化操作得到特征图 x_i 。在 3×3 分支上，通过 3×3 卷积捕获多尺度特征表示，得到特征图 x_2 。

使用跨空间学习的方法对并行子网络的输出特征图 x_1, x_2 进行聚合，得到如下四个注意力权重矩阵：

$$x_{11} = \text{Soft max}(\text{Avgpool}(\text{Reshape}(x_1))) \quad (9)$$

$$x_{12} = \text{Reshape}(x_2) \quad (10)$$

$$x_{21} = \text{Soft max}(\text{Avgpool}(\text{Reshape}(x_2))) \quad (11)$$

$$x_{22} = \text{Reshape}(x_1) \quad (12)$$

通过矩阵点积操作计算两组注意力权重的加权和，以捕获像素级配对关系，突出全局上下文信息。得到最终的注意力权重矩阵 **weights**：

$$\text{weights} = x_{11}x_{12} + x_{21}x_{22} \quad (13)$$

再将各组的子特征图与注意力权重矩阵相乘得到最终加权后的输出特征图。另外，EEMA 模块在高效多尺度注意力模块中增加了预归一化操作和残差连接，表示为如下形式：

$$\text{EEMA}(x) = x \oplus \text{PreNorm}(\text{EMA}(x)) \quad (14)$$

浅层 MLA 模块利用线性映射处理简单特征，加以预归一化和残差连接来稳定训练和加速收敛；深层 EEMA 模块通过多尺度并行子结构处理复杂的多尺度信息，且参数量更少。二者结合有效利用了网络不同位置的特征处理优势，提升分割精度。

3.4. 双交叉注意模块

双交叉模块可以同时接受来自各层编码器的多尺度特征嵌入，按顺序捕获多尺度编码器特征之间的全局空间和通道依赖关系，来解决编解码器特征之间的语义差距[23]。为有效提取多尺度的上下文信息，提高模型在皮肤病变边缘处的分割准确度，本研究在网络跳跃连接部分融入双交叉注意模块 DCA。整体结构如图 4 所示。

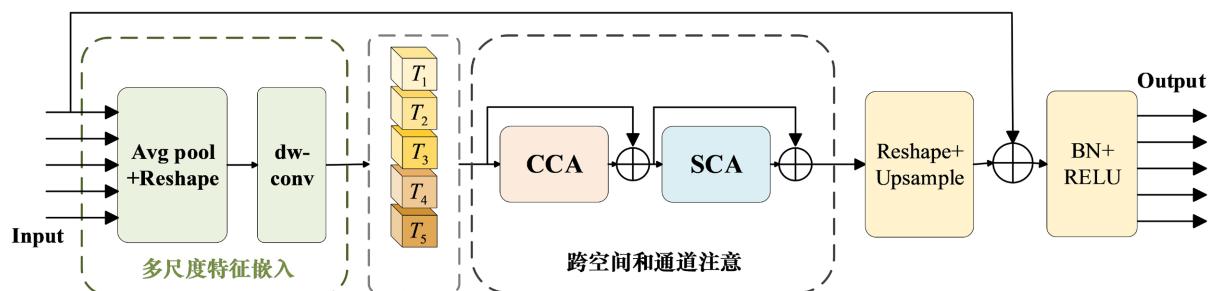


Figure 4. Dual cross-attention module

图 4. 双交叉注意模块

在 DCA 模块中，首先经过一个多尺度特征嵌入模块，从 5 个编码器阶段提取多尺度特征表示。使用 2D 平均池化进行特征提取，再利用 1×1 深度可分离卷积进行映射，对应得到的特征表示为：

$$T_i = \text{dwconv}(\text{AvgPool2D}(P_i)) \quad (15)$$

其中, $\mathbf{T}_i \in \mathbf{R}^{p \times C_i}, i = 1, 2, 3, 4, 5$ 。接着, 对每层得到的 \mathbf{T}_i 依次通过跨空间和通道注意模块, 并加以残差连接。

通道交叉注意 CCA: 首先对 \mathbf{T}_i 进行层归一化, 沿通道维度进行拼接创建键(\mathbf{K})和值(\mathbf{V}), 同时使用 \mathbf{T}_i 进行查询(\mathbf{Q}), 计算点积注意力, 然后通过 *Softmax* 函数将结果投影回原始维度。最后, 使用深度可分离卷积得到 CCA 部分的输出, 并将其作为 SCA 模块的输入。

空间交叉注意 SCA: 类似的, 首先进行层归一化操作, 沿通道维度进行拼接作为查询(\mathbf{Q})和键(\mathbf{K}), 使用 \mathbf{T}_i 作为值(\mathbf{V})。接着, 计算点积注意力, 通过 *Softmax* 函数将结果投影回原始维度。

最后, 对 SCCA 部分的输出特征图进行上采样, 与原始输入相加形成残差连接。最后, 通过 *BN* 层归一化和 *RELU* 激活处理, 连接到解码器部分, 完成跳跃连接。另外, 在解码器中加入挤压激励机制, 实现了对特征的动态加权。与 DCA 模块结合使用, 进一步丰富了网络对特征的处理和解析能力, 能够在保持计算效率的同时进一步提升分割性能。

4. 实验与结果分析

4.1. 数据集与数据预处理

为验证算法性能, 本文使用了三个公开的皮肤病变分割数据集, ISIC2018、PH2 和 HAM10000, 使得实验结果更具可比性和普遍性。

ISIC2018 数据集[24]包含 2594 张图像, 其中 1815 张用作训练, 259 张用作验证, 520 张用于测试。PH2 数据集[25]包含 200 张图像, 其中 80 张用作训练, 20 张用作验证, 100 张用作测试。HAM10000 数据集[26]包含 10,015 张图像, 其中 7200 张用作训练, 1800 张用作验证, 1015 张用作测试。

本文对这三个数据集进行了与文献[27]相同的预处理操作。在训练过程中, 使用随机噪声、随机翻转等方法进行数据增强, 提高模型的泛化能力。实验中所有图像分辨率统一调整为 128×128 像素。

4.2. 实验设置

4.2.1. 实验环境

本文所有实验基于 Python3.8 和 PyTorch 框架, 使用 NVIDIA GeForce RTX 3090 的 GPU 和 24 GB 的内存。输入标签掩码为 1 通道, 引导图像为 3 通道。扩散过程时间步设置为 $T=250$, 定义正向扩散过程加入的噪声标准差 $\beta_1 = 0.0004$, $\beta_T = 0.08$, 噪声强度随时间步呈线性增加。训练过程中使用的优化器是 Adam, 超参数设置为 $[0.7, 0.99]$ 。设置初始学习率为 0.0001, 如果在连续 5 个轮数性能没有提升则降低为之前的 0.5 倍。

仿照文献[17]的训练方法, 对每个数据集进行 320,000 次的迭代训练。训练过程中按照验证集损失最小的方式保存模型文件。用于训练的批量大小设置为 8, 验证和测试过程的批量大小设置为 32。测试过程集成 5 次模型的运行结果, 以产生更稳定和高质量的输出。训练过程使用指数移动平均(Exponential Moving Average, EMA) [28]来计算模型参数的指数加权平均值, 将这个平均值用于下一步的参数更新, 这样可以避免由于参数更新过大而导致的训练不稳定问题。设定在 200 个训练步之后开始更新, 每步更新一次, 指数移动平均因子为 0.9999, 倒数乘法因子为 1.0, 指数因子为 0.95。

4.2.2. 评价指标

为了评估所提出模型的有效性, 本文选取了 Dice 相似系数(Dice)、Jaccard 相似系数(也叫交并比, IoU)、准确度(Accuracy, Acc)、灵敏度(Sensitivity, SE)、特异度(Specificity, SP)以及豪斯多夫距离(Hausdorff Distance, HD)这 6 个常用的医学图像分割评价指标来对模型的性能进行评估, 其中以 Dice 相似系数为主要评价指标。

4.3. 实验结果

4.3.1. 对比试验

Table 1. Experimental results of different models on ISIC2018 dataset

表 1. 不同模型在 ISIC2018 数据集上的实验结果

模型	ISIC2018					
	Dice	IoU	Acc	SE	SP	HD
U-Net	0.8573	0.7503	0.9214	0.8618	0.9439	26.4572
AttU-Net	0.8527	0.7432	0.9188	0.8572	0.9421	29.0955
MedT	0.8751	<u>0.7987</u>	0.9359	0.8979	0.9391	<u>11.5169</u>
MedSegDiff	0.8548	0.7718	0.9267	<u>0.8851</u>	0.9577	15.0613
EnsDiff	0.8654	0.7900	0.9324	0.8582	<u>0.9746</u>	11.8954
DermoSegDiff	<u>0.8851</u>	0.7940	0.9398	0.8468	0.9749	11.6461
DCA-SegDiff (Ours)	0.8944	0.8089	0.9432	0.8771	0.9682	10.8902

Table 2. Experimental results of different models on PH2 dataset

表 2. 不同模型在 PH2 数据集上的实验结果

模型	PH2					
	Dice	IoU	Acc	SE	SP	HD
U-Net	0.8905	0.8026	0.9264	0.8357	0.9770	35.9472
AttU-Net	0.9178	0.8481	0.9434	0.8818	0.9778	36.5759
MedT	0.8990	0.8292	0.9232	0.8470	0.9903	12.5720
MedSegDiff	0.8710	0.7856	0.9034	0.8626	0.9692	18.9947
EnsDiff	<u>0.9332</u>	<u>0.8791</u>	0.9480	<u>0.9161</u>	0.9783	<u>10.4538</u>
DermoSegDiff	0.9238	0.8584	<u>0.9484</u>	0.8745	<u>0.9896</u>	11.5919
DCA-SegDiff (Ours)	0.9446	0.8950	0.9604	0.9431	0.9700	9.8265

Table 3. Experimental results of different models on HAM10000 dataset

表 3. 不同模型在 HAM10000 数据集上的实验结果

模型	HAM10000					
	Dice	IoU	Acc	SE	SP	HD
U-Net	0.8918	0.8048	0.9480	0.8993	0.9632	16.9833
AttU-Net	0.9108	0.8362	0.9572	0.9168	0.9698	14.2049
MedT	0.9247	0.8723	0.9635	0.9511	0.9640	8.6837
MedSegDiff	0.9138	0.8525	0.9595	0.9196	0.9740	9.9713
EnsDiff	0.9136	0.8571	0.9627	0.9147	0.9797	8.1425
DermoSegDiff	<u>0.9386</u>	0.8977	<u>0.9681</u>	0.9308	<u>0.9814</u>	<u>6.3815</u>
DCA-SegDiff (Ours)	0.9445	<u>0.8949</u>	0.9738	<u>0.9374</u>	0.9851	6.3609

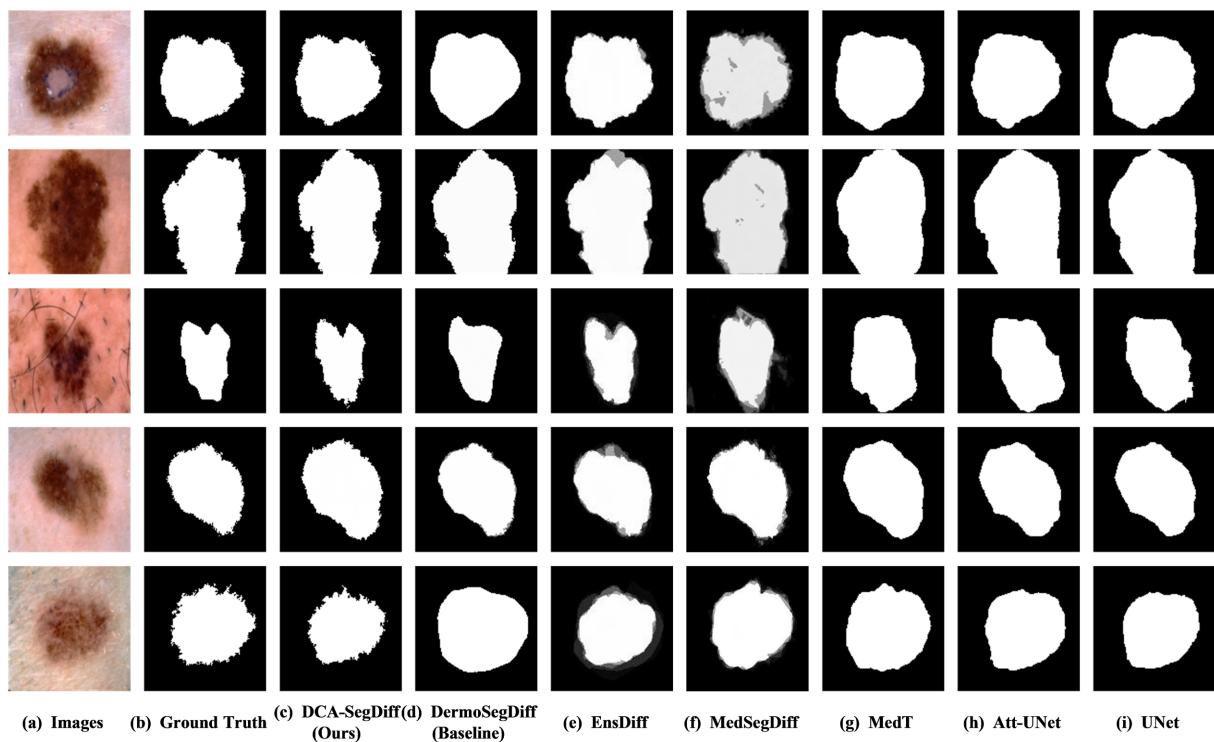


Figure 5. Comparison of experimental results on ISIC2018 dataset

图 5. ISIC2018 数据集上的实验结果对比

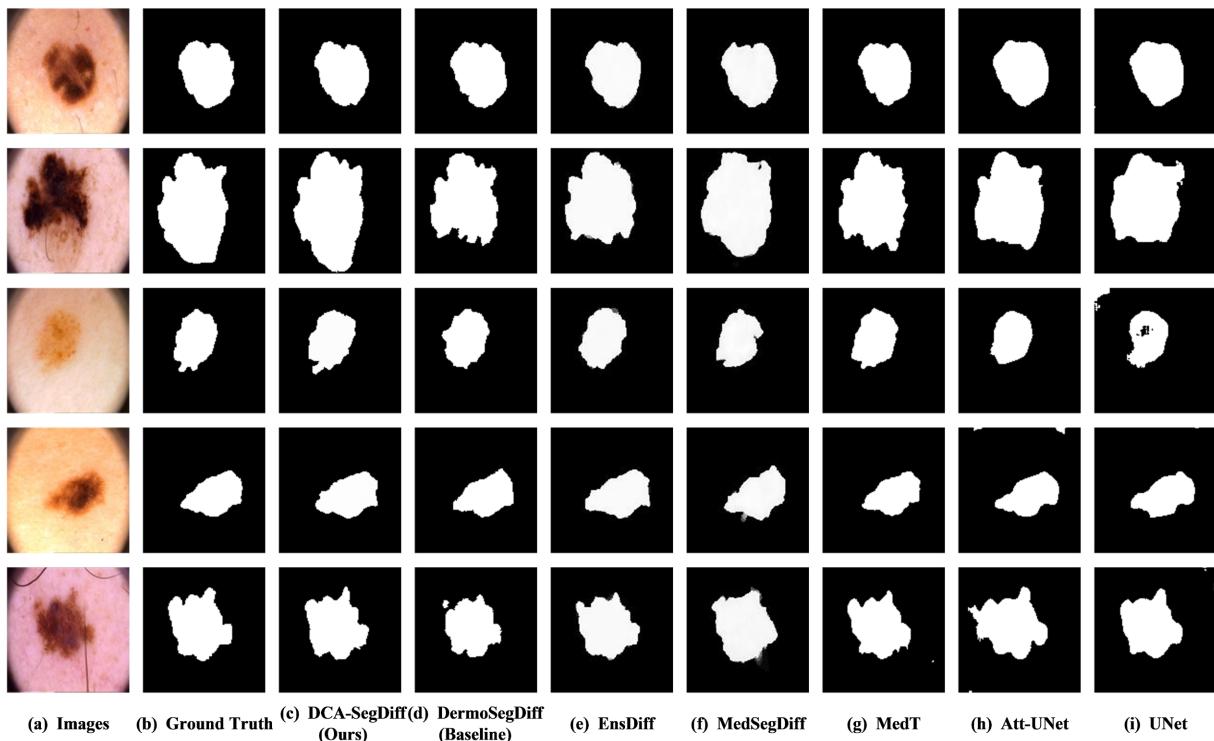


Figure 6. Comparison of experimental results on PH2 dataset

图 6. PH2 数据集上的实验结果对比

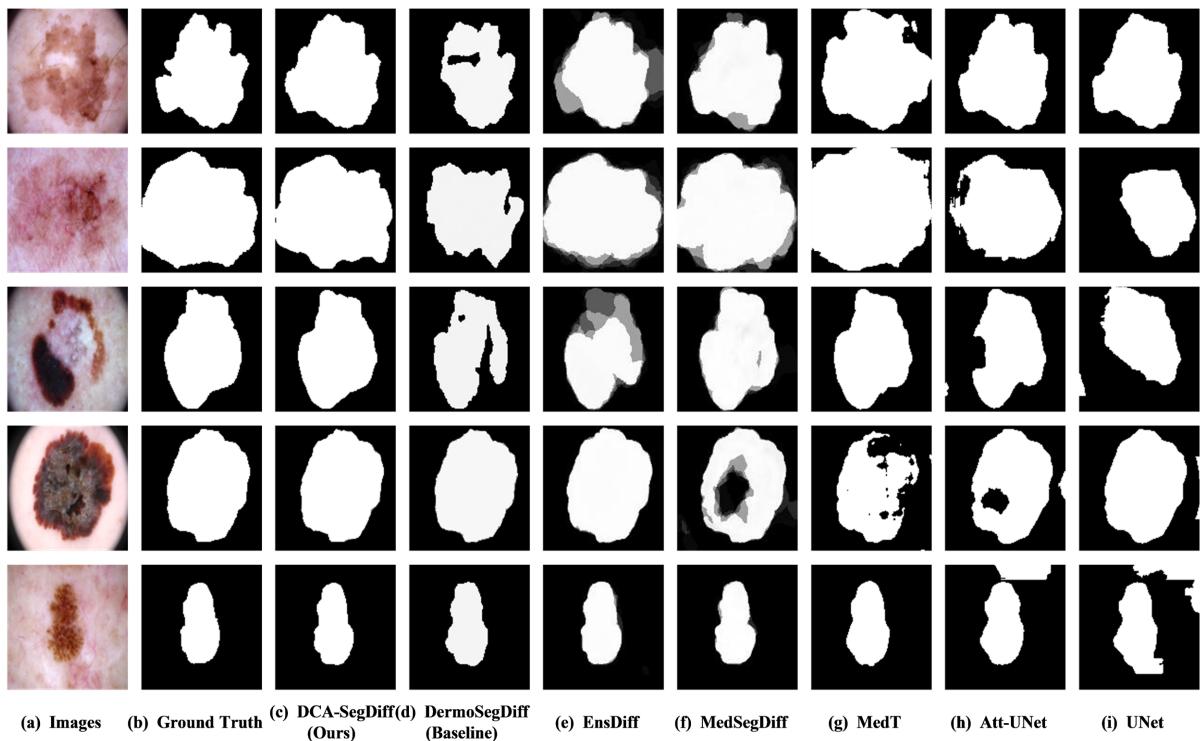


Figure 7. Comparison of experimental results on HAM10000 dataset

图 7. HAM10000 数据集上的实验结果对比

为了进一步分析所提出方法的有效性，将其与现有的六种模型方法进行了比较。为保证实验的公平性，采取相同的预处理操作，训练策略与原论文保持一致。实验结果如表 1~3 所示，表中用加粗加黑表示最好的结果，加下划线表示第二好的结果。

从对比实验结果来看，所提出模型在三个数据集上的分割性能基本优于其他参与比较的模型方法，主要指标 Dice 系数分别达到了 0.8944、0.9446、0.9445，且前三个指标 Dice、IoU、Acc 值都是最好或次好的。模型结果在 ISIC2018、PH2 和 HAM10000 三个公开的皮肤病变分割数据集上进行测试验证，进一步说明了模型的泛化能力较好。

为了更加直观地对比各个模型算法在分割效果上的优劣，给出 ISIC2018、PH2 和 HAM10000 三个数据集上的实验结果对比，如图 5~7 所示。从对比结果来看，所提出的模型能够有效地分割出皮肤病变区域。即使在面对病变区域边界复杂、对比不清晰或是有毛发遮挡的情况下，所提出模型也能得到不错的分割结果，而其他模型可能会出现有伪影或是边界分割不完全的问题。

4.3.2. 消融实验

该部分以 ISIC2018 数据集为例，对各模块的有效性进行了检验。在相同参数设置的情况下，以 DermoSegDiff 为基线模型，首先将网络减少一层，同时将 BM 部分的卷积替换为空洞卷积，然后依次对 HWD 模块、EEMA 模块以及 DCA 模块进行消融实验，并分别给出评价指标值以及模型参数量的变化情况，实验结果如下表 4 所示。

在消融实验中，首先将基线模型的网络层数减少一层，同时为保证充足的感受野，将 BM 部分的卷积替换为空洞卷积。这一改动不仅提高了模型的分割性能，使得 Dice 系数提升了 0.13%，同时模型的参数量显著减少至原本的一半。在此基础上，进一步引入 Haar 小波下采样模块得到模型“+HWD”，可以

在保持模型参数量不明显增加的情况下使 Dice 再次提高了 0.24%。接着，在编码器深层使用 EEMA 模块替换掉原本的 MLA 模块，得到模型“+HWD + EEMA”，Dice 系数再次提高了 0.37%，同时降低了模型的参数量。最后，在跳跃连接部分融入 DCA 模块，并在解码器部分添加了挤压激励机制，得到最终模型“+HWD + EEMA + DCA (DCA-SegDiff)”，Dice 系数又提高了 0.19%，最终达到 89.44%，同时并没有带来模型参数量的显著增加。相较基线模型，DCA-SegDiff 不仅显著减少了模型参数量，并且在分割指标上得到了一定程度的提高，证实了所作改进的有效性。

Table 4. Comparison of ablation experimental results
表 4. 消融实验结果对比

模型	Dice	IoU	Acc	SE	SP	HD	参数量(M)
基线模型(DermoSegDiff)	0.8851	0.7940	0.9398	0.8468	0.9749	11.6461	321.18
网络减少一层，BM 替换为空洞卷积	0.8864	0.7960	0.9403	0.8507	0.9741	11.3981	156.03
+HWD	0.8888	0.7999	0.9416	0.8523	0.9753	11.3626	156.04
+HWD + EEMA	0.8925	0.8059	0.9405	0.9017	0.9551	10.8842	154.74
+HWD + EEMA + DCA (DCA-SegDiff)	0.8944	0.8089	0.9432	0.8771	0.9682	10.8902	156.79

4.4. 进一步讨论

另外发现，数据集中个别标签并不完全准确，但这种医生手工分割导致的误差是难以避免的。下图 8 中蓝色区域表示所提出模型预测结果，绿色区域表示注释标签。可以看出，在某些边缘复杂的情况下，所提出方法在一定程度上可以弥补标签本身的不足，表现出更优的分割效果。

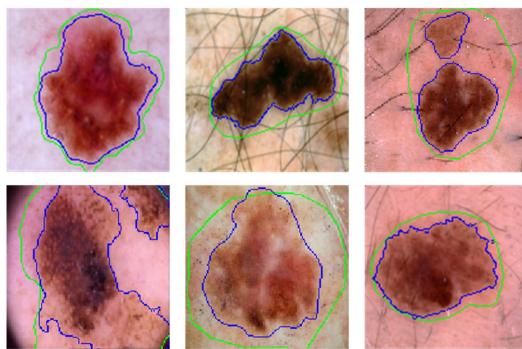


Figure 8. Label restriction
图 8. 标签限制

5. 结论

本文提出了一种新的基于扩散模型的皮肤病变分割模型 DCA-SegDiff，并在 ISIC2018、PH2、HAM10000 三个公开的皮肤病变分割数据集上进行实验验证，主要分割指标 Dice 分别达到了 0.8944、0.9446、0.9445，优于参与比较的其他模型，证实了本模型的优势和泛化能力。在可视化结果对比图中可以看出，所提出模型即使在边缘复杂等情况下依然可以得到不错的分割结果，并进一步通过消融实验证实了各模块的有效性。相比基线模型 DermoSegDiff，本文方法不仅在分割性能上得到提升，参数量也显著减少。但扩散模型的通病是模型参数量大，运行时间比较长，下一步将继续尝试对模型进行轻量化，减少模型参数量，提高模型运行效率。

参考文献

- [1] Feigelson, H.S., Powers, J.D., Kumar, M., Carroll, N.M., Pathy, A. and Ritzwoller, D.P. (2019) Melanoma Incidence, Recurrence, and Mortality in an Integrated Healthcare System: A Retrospective Cohort Study. *Cancer Medicine*, **8**, 4508-4516. <https://doi.org/10.1002/cam4.2252>
- [2] 陈旋. 基于深度学习的皮肤病图像分割算法研究[D]: [硕士学位论文]. 贵阳: 贵州大学, 2024.
- [3] 胡越黎, 曹家麟. 皮肤显微图像的自适应分割[J]. 计算机工程, 2006, 32(10): 175-177, 182.
- [4] Yasmin, J. and Mohamed Sadiq, M. (2012) An Improved Iterative Segmentation Algorithm Using Canny Edge Detector with Iterative Median Filter for Skin Lesion Border Detection. *International Journal of Computer Applications*, **50**, 37-42. <https://doi.org/10.5120/7779-0865>
- [5] 徐国雄, 王海娜, 胡进贤, 等. 一种基于混合方法的彩色图像分割算法——结合分水岭算法、种子区域生长和区域合并的混合方法[J]. 计算机科学与应用, 2013, 3(2): 134-138.
- [6] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, Berlin, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [7] Diakogiannis, F.I., Waldner, F., Caccetta, P. and Wu, C. (2020) ResUNet-A: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **162**, 94-114. <https://doi.org/10.1016/j.isprsjprs.2020.01.013>
- [8] Oktay, O., Schlemper, J., Folgoc, L., et al. (2024) Attention U-Net: Learning Where to Look for the Pancreas. <https://arxiv.org/abs/1804.03999>
- [9] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021) An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale. *Proceedings of International Conference on Learning Representations*, Virtual, 3-7 May 2021. <https://openreview.net/forum?id=YicbFdNTTy>
- [10] Chen, J., Lu, Y., Yu, Q., et al. (2024) TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. <https://arxiv.org/abs/2102.04306>
- [11] Valanarasu, J.M.J., Oza, P., Hacihaliloglu, I., et al. (2020) Medical Transformer: Gated Axial-Attention for Medical Image Segmentation. *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, Strasbourg, 27 September-1 October 2020, 6840-6851.
- [12] 杨澜. 一种基于 TransUNet 的双分支并行医学图像分割模型[J]. 计算机科学与应用, 2024, 14(10): 74-84.
- [13] Ho, J., Jain, A. and Abbeel, P. (2022) Denoising Diffusion Probabilistic Models. *Proceedings of Advances in Neural Information Processing Systems*, Vancouver, 8-11 December 2022, 6840-6851.
- [14] Wolleb, J., Sandkühler, R., Bieder, F., et al. (2022) Diffusion Models for Implicit Image Segmentation Ensembles. *Proceedings of International Conference on Medical Imaging with Deep Learning*, Zurich, 6-8 July 2022, 1336-1348.
- [15] Wu, J.D., Fu, R., Fang, H.H., et al. (2024) MedSegDiff: Medical Image Segmentation with Diffusion Probabilistic Model. *Proceedings of International Conference on Medical Imaging with Deep Learning*, Paris, 3-5 July 2024, 1623-1639.
- [16] Wu, J., Ji, W., Fu, H., Xu, M., Jin, Y. and Xu, Y. (2024) Medsegdiff-v2: Diffusion-Based Medical Image Segmentation with Transformer. *Proceedings of the AAAI Conference on Artificial Intelligence*, **38**, 6030-6038. <https://doi.org/10.1609/aaai.v38i6.28418>
- [17] Bozorgpour, A., Sadegheih, Y., Kazerouni, A., Azad, R. and Merhof, D. (2023) DermoSegDiff: A Boundary-Aware Segmentation Diffusion Model for Skin Lesion Delineation. *Proceedings of International Workshop on Predictive Intelligence in Medicine*, Vancouver, 8 October 2023, 146-158. https://doi.org/10.1007/978-3-031-46005-0_13
- [18] Yu, F., Koltun, V. and Funkhouser, T. (2017) Dilated Residual Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 472-480. <https://doi.org/10.1109/cvpr.2017.75>
- [19] Shen, Z., Zhang, M., Zhao, H., et al. (2021) Efficient Attention: Attention with Linear Complexities. *Proceedings of Winter Conference on Applications of Computer Vision*, 3-8 January 2021, 3531-3539.
- [20] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 7132-7141. <https://doi.org/10.1109/cvpr.2018.00745>
- [21] Xu, G., Liao, W., Zhang, X., Li, C., He, X. and Wu, X. (2023) Haar Wavelet Downsampling: A Simple but Effective Downsampling Module for Semantic Segmentation. *Pattern Recognition*, **143**, Article ID: 109819. <https://doi.org/10.1016/j.patcog.2023.109819>
- [22] Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., et al. (2023) Efficient Multi-Scale Attention Module with Cross-Spatial Learning. *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, 4-10 June 2023, 1-5. <https://doi.org/10.1109/icassp49357.2023.10096516>

-
- [23] Ates, G.C., Mohan, P. and Celik, E. (2023) Dual Cross-Attention for Medical Image Segmentation. *Engineering Applications of Artificial Intelligence*, **126**, Article ID: 107139. <https://doi.org/10.1016/j.engappai.2023.107139>
 - [24] Codella, N.C.F., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., *et al.* (2018) Skin Lesion Analysis toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, 4-7 April 2018, 168-172. <https://doi.org/10.1109/isbi.2018.8363547>
 - [25] Mendonca, T., Ferreira, P.M., Marques, J.S., Marcal, A.R.S. and Rozeira, J. (2013) PH²—A Dermoscopic Image Database for Research and Benchmarking. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, 3-7 July 2013, 5437-5440. <https://doi.org/10.1109/embc.2013.6610779>
 - [26] Tschandl, P., Rosendahl, C. and Kittler, H. (2018) The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions. *Scientific Data*, **5**, 1-9. <https://doi.org/10.1038/sdata.2018.161>
 - [27] Alom, M.Z., Hasan, M., Yakopcic, C., *et al.* (2024) Recurrent Residual Convolutional Neural Network Based on U-Net (R2u-Net) for Medical Image Segmentation. <http://arxiv.org/abs/1802.06955>
 - [28] Karras, T., Aittala, M., Lehtinen, J., Hellsten, J., Aila, T. and Laine, S. (2024) Analyzing and Improving the Training Dynamics of Diffusion Models. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, 17-21 June 2024, 24174-24184. <https://doi.org/10.1109/cvpr52733.2024.02282>