

面向深度感知的空间层次学习用于稳健的医学图像分割

周世钦

五邑大学电子与信息工程学院智能制造部, 广东 江门

收稿日期: 2026年4月27日; 录用日期: 2026年6月23日; 发布日期: 2026年6月30日

摘要

皮肤镜图像分割是计算机辅助诊断皮肤癌的关键步骤, 但现有方法多依赖二维纹理信息, 忽略了单目图像中隐含的视觉远近线索(伪深度)。本文提出一种融合伪深度注意力机制与边界监督的多任务U-Net框架——DASH-Net (Depth-Aware Spatial Hierarchy Network)。其核心创新在于通过深度引导的空间-通道双注意力模块(DSCA), 建立二维图像特征通道与视觉远近层次之间的自适应关联, 使分割网络能够根据深度先验选择性地增强来自相关深度范围的特征通道。此外, 针对临床图像提出轻量变体DASH-Light, 以亮度映射替代伪深度图, 实现无外部深度估计模型的高效分割。在ISIC 2018皮肤病变基准数据集上的实验表明, DASH-Net的Dice系数达88.99%, Jaccard指数达82.02%, 参数量仅7.85 M; 替换为PVTv2骨干后Dice进一步提升至90.61%。在私有静脉畸形临床数据集上, DASH-Light取得66.80% Dice和30.21的HD95, 较当前最优方法提升1.1%和下降12.4。消融实验与可视化分析揭示了DSCA模块的工作机制: 部分通道专门化编码深层病灶信息, 部分编码背景表层信息, 形成类似人类视觉系统的深度语义分工。

关键词

皮肤病变分割, 多任务学习, 伪深度估计, 空间-通道注意力, 边界监督, 静脉畸形

Depth-Aware Spatial Hierarchy Learning for Robust Medical Image Segmentation

Shiqin Zhou

Department of Intelligent Manufacturing, School of Electronics and Information Engineering, Wuyi University, Jiangmen Guangdong

Received: April 27, 2026; accepted: June 23, 2026; published: June 30, 2026

Abstract

Skin lesion segmentation is a critical step in computer-aided diagnosis of skin cancer. Existing methods mostly rely on two-dimensional texture information and ignore the implicit visual depth cues in monocular images. This study proposes DASH-Net, a multi-task U-Net framework integrating pseudo-depth attention with boundary supervision. Its core innovation is the Depth-guided Spatial-Channel Attention (DSCA) module that establishes an adaptive association between image feature channels and visual depth hierarchy. A lightweight variant DASH-Light is also proposed for clinical images where brightness serves as a depth proxy. On ISIC 2018, DASH-Net achieves 88.99% Dice with only 7.85 M parameters, and reaches 90.61% Dice with PVTv2 backbone. On a private venous malformation clinical dataset, DASH-Light achieved a 66.80% Dice score and a 30.21 HD95, improving by 1.1% and decreasing by 12.4 compared to the current best method. Ablation studies reveal that DSCA channels spontaneously specialize into deep-preferring and background-preferring groups, forming a depth semantic division.

Keywords

Skin Lesion Segmentation, Multi-Task Learning, Pseudo-Depth Estimation, Spatial-Channel Attention, Boundary Supervision, Venous Malformation

Copyright © 2026 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

皮肤癌是全球最常见的恶性肿瘤之一，其中黑色素瘤的死亡率占皮肤癌相关死亡的绝大部分[1]。早期筛查与诊断对于改善患者预后具有至关重要的临床意义。皮肤镜(Dermoscopy)作为一种非侵入性皮肤成像技术，能够放大皮损区域并消除表面反射，已成为皮肤科医生进行黑色素瘤早期检测的首选工具[2]。在计算机辅助诊断(CAD)系统中，准确分割皮肤镜图像中的病灶区域是所有后续定量分析(如不对称性、边界不规则性、颜色多样性和直径测量，即 ABCD 法则)的基础与前提[2] [3]。

近年来，基于深度学习的语义分割方法在皮肤病变分割领域取得了显著进展。Sultana 等[4]对卷积神经网络在图像分类中的发展历程进行了系统综述，为后续医学图像分割的快速发展奠定了认知基础。Shelhamer 等[5]提出的全卷积网络(FCN)首次实现了端到端的像素级分类，Ronneberger 等[6]设计的 U-Net 通过编码器 - 解码器结构与跳跃连接在医学图像分割中展现了强大的特征提取与细节恢复能力。在此基础上，研究者从多个方向对 U-Net 进行了改进：Zhou 等[7]通过嵌套稠密跳跃连接充分利用多尺度特征(UNet++); Oktay 等[8]引入门控注意力机制以抑制不相关背景响应(Attention U-Net); Hu 等[9]提出的 SENet 通过通道关系建模自适应重标定特征响应，Woo 等[10]进一步将通道注意力与空间注意力相结合(CBAM)。在皮肤病变分割的具体应用中，Dai 等[11]提出了多尺度残差编解码网络(Ms RED)，Yang 和 Zhang [12]设计了金字塔傅里叶可变形网络(PFD-Net)，Huang 等[13]提出了通道先验卷积注意力机制(CPCA)，Ruan 等[14]构建了多注意力轻量级 U-Net (MALUNet)。Al-Masni 等[15]则将分割与分类集成到统一的深度卷积网络框架中，实现了对皮肤病变的同步诊断与定位。Sarker 等[16]采用空洞残差与金字塔池化网络(SLSDeep)扩大感受野以捕获多尺度上下文，Gu 等[17]设计的 CE-Net 通过上下文

编码器模块有效聚合全局语义信息, Kaul 和 Zhang [18]提出的 FocusNet 将注意力机制融入 FCN 以聚焦病灶区域, Feng 等[19]的 CPFNet 通过上下文金字塔融合实现多尺度特征的高效整合。此外, 基于纯 Transformer 的医学分割网络如 Chen 等[20]的 TransUNet、Wang 等[21]的 PVTv2、Cao 等[22]的 Swin-Unet、Valanarasu 等[23]的 MedT 以及 Isensee 等[24]的 nnU-Net 也在不同程度上推动了该领域的发展。近期, WinGraphUNet [25]通过窗口化图建模捕获皮肤病变的长程依赖关系, DFAFormer [26]采用双解码器 Transformer 结构显式建模边界信息, SkinAttn-Net [27]通过多级注意力机制增强特征表示, BA-RA-UNet [28]结合逆向注意力实现精细的边界恢复, AD-Net [29]利用空洞卷积残差网络扩大感受野, SkinFormer [30]学习统计纹理表示以提升分割效率, MRCNN [31]在掩膜分类与分割任务上展现了良好的多任务学习能力, CRFNet [32]通过交叉注意力和精细化特征融合提升病灶定位精度, HHBSNet [33]引入全局通道 - 空间注意力以增强深层语义表达, MSCB-UNet [34]构建多尺度空间 - 通道桥接模块促进信息流传播。上述方法共同构成了皮肤病变分割领域的技术全景, 为本研究的对比实验提供了全面而可靠的参照体系。

尽管上述方法在特定场景下取得了良好效果, 但它们均依赖二维 RGB 纹理特征。当面临病灶边界模糊、毛发伪影遮挡、对比度低下等挑战时, 纯粹基于纹理的模型难以区分病灶与背景。人类视觉系统借助纹理梯度、相对大小、遮挡关系等单目深度线索推断物体相对距离[35], 受此启发, 本文提出核心假设: 即使从单目皮肤镜图像估计的伪深度图并非真实物理深度, 其蕴含的空间层次结构仍可为分割提供几何先验。基于此, 本文提出 DASH-Net, 一种融合伪深度注意力与边界监督的多任务 U-Net 框架。其核心创新在于通过 DSCA 模块建立特征通道与视觉远近层次的自适应关联, 使模型通过调节通道权重选择性地获取特定远近范围内的病灶信息。此外, 针对静脉畸形临床图像提出轻量化的 DASH-Light 变体, 以亮度映射替代伪深度图, 验证该范式在不同医学成像模态下的有效性。

2. 材料与方法

2.1. 数据集与评价指标

为验证所提方法的有效性, 本研究在 ISIC 2018 数据集[3]及一个私有静脉畸形临床图像数据集上进行了大量实验(ISIC 2018 数据集公开获取于 <https://challenge.isic-archive.com/>)。ISIC 2018 数据集[3]是目前评估皮肤病变分割算法的主流基准, 包含 2594 张训练图像、100 张验证图像和 1000 张测试图像, 图像分辨率为 1920×1080 至 4K 不等。此外, 本研究还收集了一个私有静脉畸形临床图像数据集, 包含静脉畸形的临床照片及专家标注的分割掩膜。

输入图像统一缩放至 256×1256 并归一化至 $[0, 1]$ 。数据增强包括随机翻转、旋转($\pm 30^\circ$)、颜色抖动和仿射变换。训练集、验证集与测试集按 70%:15%:15%划分。评价指标包括 Jaccard 指数(JI)、Dice 系数(DC)、准确率(Acc)、敏感性(Sen)、特异性(Spe)、95%分位 Hausdorff 距离(HD95)及平均对称表面距离(ASSD)。

2.2. 整体网络架构

DASH-Net 采用多任务编码器 - 解码器架构, 整体结构如图 1 所示, 包含三个核心组件: 共享编码器、深度估计解码器和分割解码器(含边界监督分支)。共享编码器可采用原生 U-Net 编码器[6] (4 级下采样, 输出通道依次为 64-128-256-512)或预训练骨干网络如 He 等[36]的 ResNet-50、Huang 等[37]的 DenseNet 以及 Wang 等[21]的 PVTv2。Dosovitskiy 等[38]提出的 Vision Transformer (ViT)通过将图像划分为 Patch 序列并利用自注意力机制建模全局关系, 开创了 Transformer 在视觉任务中的应用先河; Liu 等[39]在此基础上提出 Swin Transformer, 通过层次化窗口移位注意力机制在降低计算复杂度的同时实现了跨窗口信

息交互。PVTv2 [21]融合了金字塔结构与 Transformer 的优势，通过渐进式收缩的空间注意力有效捕获多尺度特征，与 DSCA 模块形成良好协同。深度估计解码器通过 4 级上采样恢复空间分辨率。分割解码器在每一级上采样后插入 DSCA 模块，融合深度与分割特征，并行输出病灶分割图与边界图。

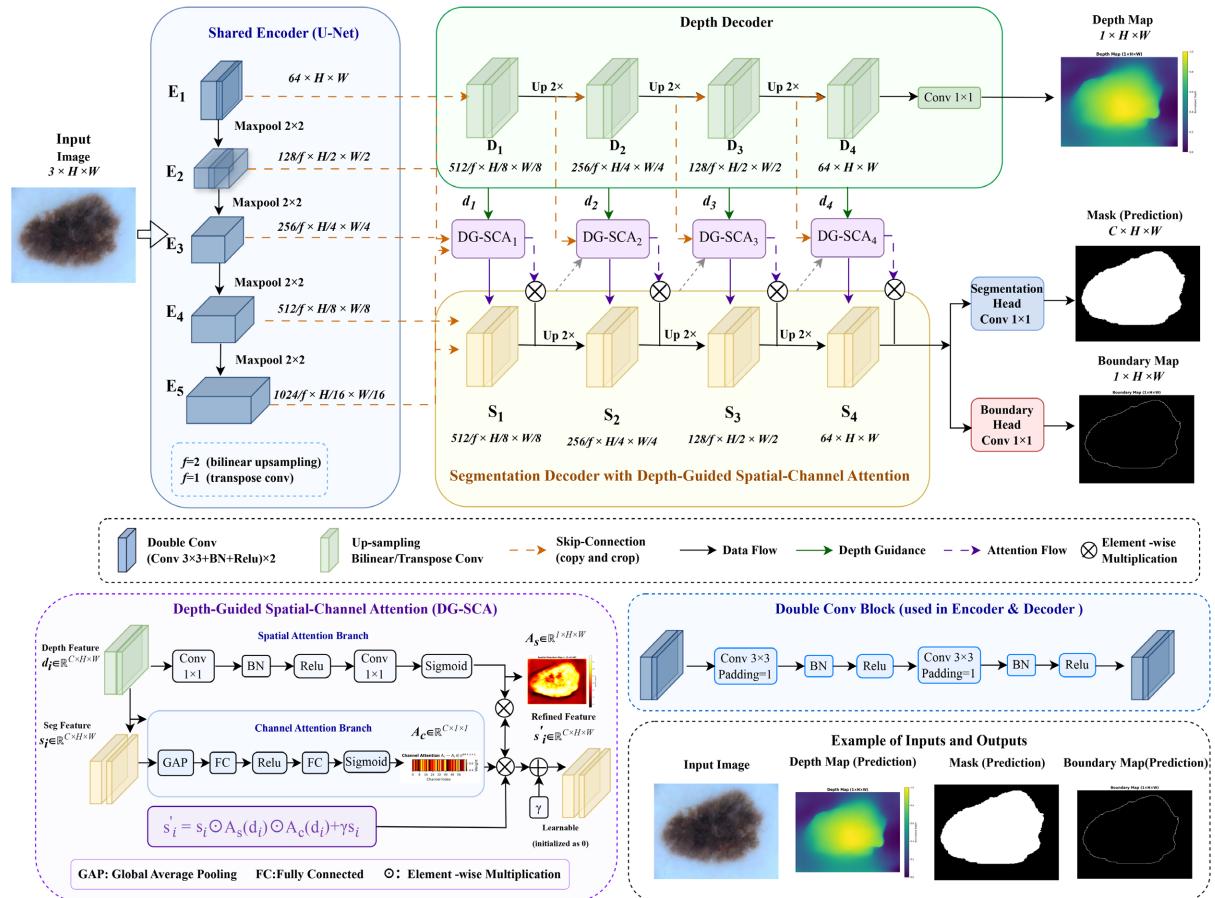


Figure 1. Overall architecture of DASH-Net. The shared encoder extracts multi-scale features, the depth decoder generates pseudo-depth prediction maps for auxiliary supervision, the segmentation decoder fuses depth and segmentation features through the DSCA module, and outputs lesion segmentation maps and boundary maps in parallel

图 1. DASH-Net 整体架构示意图。共享编码器提取多尺度特征，深度解码器生成伪深度预测图进行辅助监督，分割解码器通过 DSCA 模块融合深度与分割特征，并行输出病灶分割图与边界图

2.3. 伪深度图生成与预处理

伪深度图通过预训练的 DPT-SwinV2-Large-384 单目深度估计模型离线生成。Ranftl 等 [40] 提出的 DPT (Dense Prediction Transformer) 将视觉 Transformer 与卷积解码器相结合，能够基于单张 RGB 图像生成高质量的相对深度图。图 2、图 3 分别为皮肤镜图像及静脉畸形临床图像。本研究采用在大量自然场景图像上预训练的 DPT 模型，对于输入皮肤镜图像 I ，生成相对深度估计图 $D_{\text{pseudo}} \in \mathbb{R}^{H \times W}$ ：

$$D_{\text{pseudo}} = \text{DPT-SwinV2}(I) \in [0, 1]^{H \times W} \quad (1)$$

生成的伪深度图经 Min-Max 归一化至 $[0, 1]$ 后作为辅助输入。此处深度图并非真实物理深度，而是反映场景相对远近关系的“伪深度”。病灶与周围皮肤在表面高度和组织密度上的差异会形成可辨识的深度梯度，为分割提供几何结构先验 [41] [42]。

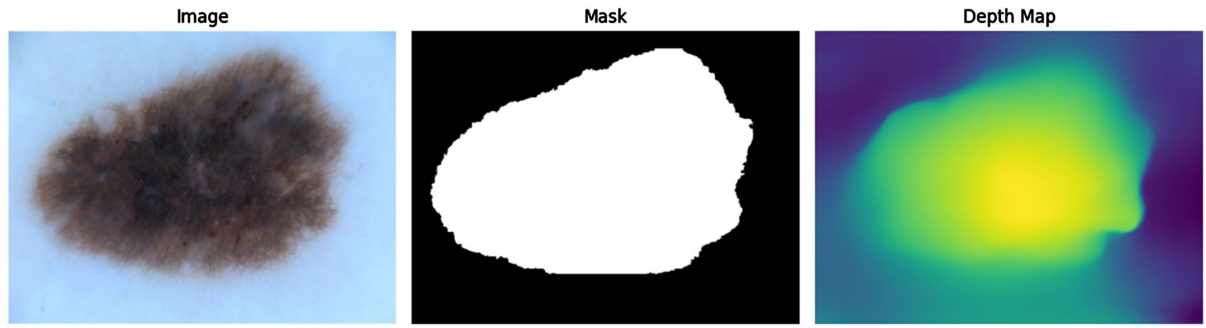


Figure 2. Examples of ISIC dermoscopy images, corresponding segmentation masks, and pseudo-depth maps generated by DPT
图 2. ISIC 皮肤镜图像、对应分割掩膜及 DPT 生成的伪深度图示例

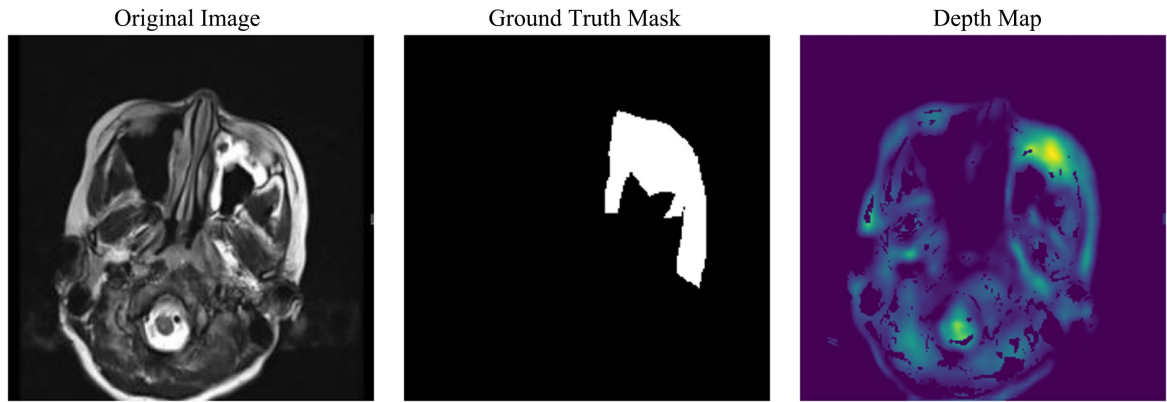


Figure 3. Examples of venous malformation clinical images, corresponding segmentation masks, and brightness proxy depth maps
图 3. 静脉畸形临床图像、对应分割掩膜及亮度代理深度图示例

2.4. 深度引导的空间 - 通道双注意力模块(DSCA)

DSCA 模块是 DASH-Net 的核心创新，其设计动机源于本文的核心思想：让二维图像的特征通道与视觉上的远近层次建立关联。具体而言，DSCA 通过并行计算空间注意力图和通道注意力向量，使分割网络能够自适应地关注特定深度层次对应的特征通道和空间区域，从而在“通道维度”上实现对“远近维度”的选择性增强。

空间注意力分支的处理流程如下。给定深度解码器输出的深度特征 $F_{\text{depth}} \in R^{C_d \times H \times W}$ 和分割解码器特征 $F_{\text{seg}} \in R^{C_s \times H \times W}$ ，空间分支首先通过 1×1 卷积将深度特征投影到与分割特征相同的通道空间：

$$F_{\text{depth_proj}} = \text{BN} \left(\text{ReLU} \left(\text{Conv}_{1 \times 1} \left(F_{\text{depth}} \right) \right) \right) \in R^{C_s \times H \times W} \quad (2)$$

然后对投影后的深度特征应用 1×1 卷积和 Sigmoid 激活生成空间注意力图：

$$\alpha_s = \sigma \left(\text{Conv}_{1 \times 1} \left(F_{\text{depth_proj}} \right) \right) \in R^{1 \times H \times W} \quad (3)$$

其中 $\sigma(\cdot)$ 为 Sigmoid 激活函数， $\alpha_s \in R^{1 \times H \times W}$ 即为空间注意力图。高响应区域对应深度图中具有显著远近梯度变化的边界区域，低响应区域对应深度相对均匀的内部区域。

通道注意力分支旨在建立“特征通道”与“深度语义”之间的关联。首先对深度特征进行全局平均池化(GAP)，将空间信息压缩为通道描述符：

$$z = \text{GAP} \left(F_{\text{depth}} \right) = \left(1 / (H \cdot W) \right) \sum_{i,j} F_{\text{depth}} (:, i, j) \in R^{C_d} \quad (4)$$

然后通过两层的共享 MLP (压缩比 $r = 16$) 生成通道注意力权重:

$$\alpha_c = \sigma(W_2(\text{ReLU}(W_1 \cdot z))) \in R^{C_s} \quad (5)$$

其中 $W_1 \in R^{C_d/r \times C_d}$ 、 $W_2 \in R^{C_s \times C_d/r}$ 为全连接层权重, $\alpha_c \in R^{C_s}$ 即为通道注意力向量。 α_c 的每个元素 $\alpha_c(k)$ 表示第 k 个特征通道与深度语义的关联强度——数值越大, 说明该通道编码的特征越与特定远近层次的深度信息相关。通过这种方式, 模型实现了“调节通道即选择远近”的机制: 当 α_c 在某些通道上赋予高权重时, 相当于选择了对应深度层次的特征响应。这种设计在思想上与 Hu 等[9]的 SENet 和 Woo 等[10]的 CBAM 一脉相承, 但区别在于 DSCA 的注意力权重由深度特征驱动而非分割特征自身, 从而实现了跨模态的通道选择。

融合输出的最终表达式为:

$$F_{\text{out}} = F_{\text{seg}} \odot \alpha_s \odot \alpha_c + \gamma \cdot F_{\text{seg}} \quad (6)$$

其中 \odot 表示广播乘法(空间注意力沿通道广播, 通道注意力沿空间广播), $\gamma \in R$ 为可学习的残差缩放参数, 初始化为 0, 使网络在训练初期保持原始特征不变, 随着训练深入逐步学习深度注意力的贡献。

2.5. 边界监督分支与动态边界生成

病灶边界的精确勾勒对皮肤癌诊断具有直接的临床意义——ABCD 法则中的“B” (Border irregularity, 边界不规则性) 是判断黑色素瘤良恶性的关键指标之一[2]。为显式提升边界质量, DASH-Net 在分割解码器末端并行输出一个边界概率图 $\hat{B} \in R^{H \times W}$, 并通过动态 Canny 边缘检测算法[43]从 ground truth 分割掩膜 $Y \in \{0, 1\}^{H \times W}$ 生成自适应边界标签 B_{canny} 。

具体流程如下: 首先对掩膜图像应用高斯平滑($\sigma = 1.0$)抑制噪声; 然后使用 Sobel 算子计算梯度幅值 $G = \sqrt{G_x^2 + G_y^2}$ 和梯度方向 $\theta = \arctan(G_y/G_x)$; 接着沿梯度方向进行非极大值抑制(NMS), 细化边缘; 最后通过双阈值滞后处理生成二值边界图, 其中低阈值 T_{low} 和高阈值 T_{high} 根据当前掩膜的梯度幅值分布自适应计算:

$$T_{\text{low}} = \mu_G - 0.5 \cdot \sigma_G, T_{\text{high}} = \mu_G + 0.5 \cdot \sigma_G \quad (7)$$

其中 μ_G 和 σ_G 分别为掩膜梯度幅值的均值和标准差。自适应阈值策略的优势在于: 不同病灶具有不同的最佳边缘检测参数, 动态调整可确保生成的边界标签始终紧贴病灶真实轮廓。在损失计算方面, 边界分支采用二元交叉熵损失 $L_{\text{BCE}}(\hat{B}, B_{\text{canny}})$ 。此外, 为进一步直接惩罚边界定位误差, 引入 Karimi 和 Salcudean [44]提出的可微分 95%分位 Hausdorff 距离损失:

$$L_{\text{HD95}} = \max \left\{ d_{95}(\partial S_{\text{pred}} \rightarrow \partial S_{\text{gt}}), d_{95}(\partial S_{\text{gt}} \rightarrow \partial S_{\text{pred}}) \right\} \quad (8)$$

其中 $d_{95}(A \rightarrow B)$ 表示从集合 A 中所有点到集合 B 最近点距离的第 95%分位数, 采用排序可微近似实现端到端训练。边界监督分支的总损失为 $L_{\text{boundary}} = L_{\text{BCE}} + \lambda_{\text{hd}} \cdot L_{\text{HD95}}$, 其中 $\lambda_{\text{hd}} = 0.1$ 为权重系数。

2.6. 损失函数

DASH-Net 的总训练损失为三项损失的加权求和:

$$L_{\text{total}} = L_{\text{seg}} + \lambda_d \cdot L_{\text{depth}} + \lambda_b \cdot L_{\text{boundary}} \quad (9)$$

分割损失 L_{seg} 结合 Dice 损失与二元交叉熵损失:

$$L_{\text{seg}} = \left[1 - \frac{2 \cdot |Y \cap \hat{Y}| + \varepsilon}{(|Y| + |\hat{Y}| + \varepsilon)} \right] + \left[-\sum_i (Y_i \cdot \log \hat{Y}_i + (1 - Y_i) \cdot \log(1 - \hat{Y}_i)) / N \right] \quad (10)$$

其中 $Y \in \{0, 1\}^{H \times W}$ 为真实标签, $\hat{Y} \in [0, 1]^{H \times W}$ 为预测概率图, $\varepsilon = 1e-6$ 为平滑因子, N 为像素总数。Dice 损失关注区域重叠度, 对类别不平衡更鲁棒; BCE 损失保证像素级分类精度。深度损失采用 L1 损失:

$$L_{\text{depth}} = \|\hat{D} - D_{\text{pseudo}}\|_1 = (1/(H \cdot W)) \sum_{i,j} |\hat{D}(i,j) - D_{\text{pseudo}}(i,j)| \quad (11)$$

通过消融实验确定最优权重组合为 $\lambda_d = 0.5$ 、 $\lambda_b = 0.1$ 。深度损失权重 λ_d 的取值至关重要：过小时深度分支的辅助监督作用不足，过大时则分割主任务受深度估计误差干扰。

2.7. DASH-Light: 基于亮度代理的轻量变体

皮肤镜图像与静脉畸形临床图像在成像特性上存在本质差异：前者病灶与周围皮肤的颜色/纹理差异较为复杂，而后者中静脉畸形区域往往呈现显著的亮度暗化特征(血红蛋白富集导致对可见光的吸收增强)。针对这一特性，本研究提出 DASH-Light，一种无需外部深度估计模型的轻量变体。

DASH-Light 的网络结构与 DASH-Net 基本一致，唯一区别在于深度代理的生成方式。给定输入图像 $x \in R^{3 \times H \times W}$ ，亮度映射计算如下：

$$D_{\text{brightness}}(x) = (x_R + x_G + x_B)/3 \in R^{1 \times H \times W} \quad (12)$$

对于静脉畸形图像，由于病灶区域通常呈现紫红色(红蓝通道高、绿色通道低)，还采用了红 - 蓝通道均值作为辅助深度代理：

$$D_{RB}(x) = (x_R + x_B)/2 \in R^{1 \times H \times W} \quad (13)$$

上述亮度图作为“伪深度图”输入深度估计分支进行辅助监督训练。从物理意义上理解：亮度值较低的暗区对应于视觉上的“凹陷”或“深层”组织，而亮度值较高的亮区对应于“表层”皮肤。这种“以亮度代深度”的策略不仅避免了 DPT 推理的计算开销，还使深度代理与图像内容紧密耦合，形成一种“自监督”式的深度线索。

3. 实验结果

3.1. 实施细节

所有实验在 PyTorch 框架下实现，训练环境为配备 10 GB 显存的 NVIDIA GeForce RTX 3080 GPU。优化器配置：AdamW 优化器(学习率 1×10^{-4} ，权重衰减 5×10^{-4} ，余弦退火)和 SGD 优化器(学习率 1×10^{-2} ，动量 0.9，阶梯式衰减，step_size = 30 epoch)，批量大小设为 8。所有模型训练 200 个 epoch，保留验证集 Dice 最高的模型。所有对比实验均未加载 ImageNet 预训练权重。

3.2. 消融实验

为验证 DASH-Net 各组件的有效性，本研究在 ISIC 2018 上进行了系统的消融实验，结果如表 1 所示。基线 U-Net 的 Dice 为 87.01%。添加深度注意力分支($\lambda_d = 0.5$)后，Dice 提升至 88.39%，表明深度辅助监督能为分割提供有效的空间层次先验。进一步添加边界监督($\lambda_b = 0.1$)后，Dice 提升至 88.59%。完整的 DASH-Net 取得最优性能(Dice = 88.99%，Jaccard = 82.02%，Sensitivity = 94.20%)。

Table 1. Segmentation performance of different modules on ISIC 2018 dataset (mean \pm std)

表 1. 不同模块在 ISIC 2018 数据集上的分割性能(均值 \pm 标准差)

模块	JI (%)	DC (%)	Acc (%)	Sen (%)	Spe (%)
Baseline	78.99 \pm 0.81	87.01 \pm 0.57	91.17 \pm 0.97	92.27 \pm 0.71	90.87 \pm 0.74
w/深度注意力($\lambda_d = 0.5$)	81.08 \pm 0.68	88.39 \pm 0.34	91.80 \pm 0.29	93.12 \pm 0.58	91.10 \pm 0.78
w/深度注意力 + 边界监督($\lambda_b = 0.1$)	81.46 \pm 0.56	88.59 \pm 0.31	92.03 \pm 0.20	93.18 \pm 0.48	90.17 \pm 0.28
DASH-Net (完整)	82.02 \pm 0.16	88.99 \pm 0.38	92.20 \pm 0.72	94.20 \pm 0.58	89.14 \pm 0.64

DC: Dice Coefficient, JI: Jaccard Index, Acc: Accuracy, Sen: Sensitivity, Spe: Specificity.

3.3. 损失函数权重敏感性分析

为确定深度损失权重 λ_d 和边界损失权重 λ_b 的最佳组合,本研究在 ISIC 2018 上进行了网格搜索实验,结果如表 2 所示。当 $\lambda_d=0.5$ 、 $\lambda_b=0.1$ 时,模型取得最优综合性能(Dice = 88.99%, Jaccard = 82.02%)。 $\lambda_d=0.3$ 时 Dice 降至 88.69%,说明深度监督强度不足会导致几何先验的辅助作用减弱; λ_d 过大(如 0.5 配合 $\lambda_b=0.3$)时 Dice 降至 88.78%,表明过度强调边界监督会干扰深度特征的协同学习。 $\lambda_b=0.1$ 是一个较为均衡的取值,能够在不显著影响区域分割精度的前提下有效提升边界质量。

Table 2. Performance of different loss weight combinations on ISIC 2018 (mean \pm std)

表 2. 不同损失权重组合在 ISIC 2018 上的性能(均值 \pm 标准差)

$\lambda_{\text{boundary}}$	λ_{depth}	JI (%)	DC (%)	Acc (%)	Sen (%)	Spe (%)
0.3	0.5	81.59 \pm 0.81	88.69 \pm 0.49	91.75 \pm 0.76	94.58 \pm 0.45	87.24 \pm 0.78
0.15	0.3	81.74 \pm 0.85	88.78 \pm 0.50	92.14 \pm 0.28	93.38 \pm 0.30	91.25 \pm 0.51
0.1	0.5	82.02 \pm 0.16	88.99 \pm 0.38	92.20 \pm 0.72	94.20 \pm 0.58	89.14 \pm 0.64

λ_d : depth loss weight, λ_b : boundary loss weight.

3.4. 不同骨干网络对性能的影响

为评估 DSCA 模块对不同编码器骨干的兼容性,本研究在 ISIC 2018 上对比了三种编码器:原生 U-Net [6] (7.85M 参数)、ResNet-50 [36] (23.5M 参数)和 PVTv2 [21] (26.3M 参数),所有骨干均从头训练(其中 ResNet-50 及 PVTv2 骨干使用官方的预训练权重),结果如表 3 所示。PVTv2 骨干取得最优性能(Dice = 90.61%, Jaccard = 84.12%, HD95 = 16.18, ASSD = 6.66),显著优于 U-Net 和 ResNet-50。PVTv2 的金字塔视觉 Transformer 架构通过分层 Patch 嵌入与空间缩减注意力机制提供了更丰富的多尺度表示,与 DSCA 模块形成良好协同。ResNet-50 作为传统 CNN 骨干,性能介于两者之间(Dice = 89.80%)。U-Net 虽然参数量最小,但仍能达到 88.99% Dice,在计算资源受限场景下具有实用价值。

Table 3. Segmentation performance and computational cost of DASH-Net with different encoder backbones on ISIC 2018 (mean \pm std)

表 3. DASH-Net 采用不同编码器骨干时的分割性能与计算成本(ISIC 2018, 均值 \pm 标准差)

Backbone	训练参数 (M)	推理参数 (M)	FLOPs (G)	速度 (ms)	JI (%)	DC (%)	Acc (%)	Sen (%)	Spe (%)	HD95	ASSD
U-Net [5]	17.49	13.50	48.22	22.57	82.02 \pm 0.16	88.99 \pm 0.38	92.20 \pm 0.72	94.20 \pm 0.58	89.14 \pm 0.64	20.93 \pm 0.17	8.11 \pm 0.64
ResNet-50 [18]	41.76	32.69	17.40	20.98	82.99 \pm 0.55	89.80 \pm 0.60	93.26 \pm 0.78	94.23 \pm 0.31	91.29 \pm 0.98	18.19 \pm 0.37	7.18 \pm 0.18
PVTv2 [15]	27.69	26.29	7.01	15.97	84.12 \pm 0.62	90.61 \pm 0.56	93.67 \pm 0.51	94.64 \pm 0.29	91.31 \pm 0.24	16.18 \pm 0.19	6.66 \pm 0.34

Params: parameters, FLOPs: floating point operations.

尽管 DASH-Net 在训练阶段因多任务监督需要额外的深度解码器(参数量增加 28.2%~30.6%,见表 3),

但在推理阶段该分支被完全丢弃，有效参数量较原始骨干仅增加 0.1%~0.8% (U-Net [6]: +0.11 M, ResNet-50 [36]: +0.11 M, PVTv2 [21]: +0.03 M)，这全部来源于 DSCA 模块的轻量级 1×1 卷积与 MLP 操作。从计算效率看，PVTv2 [21] 骨干表现最优：FLOPs 仅 7.01 G，推理速度 15.97 ms (62.6 FPS)，较 U-Net [6] 骨干 (48.22 G, 22.57 ms) 分别降低 85.5% 和 29.2%。值得注意的是，PVTv2 的 FLOPs 优势主要源于其解码器设计：编码器 FLOPs (5.10 G) 与 ResNet-50 (5.38 G) 处于同一量级，但其解码器仅含 3 层上采样且最终通过 $4 \times$ 双线性插值恢复分辨率，避免了 U-Net 中 4 层逐卷积上采样的密集计算 (编码器 FLOPs 即达 14.68 G)。综合来看，DASH-Net 以极小的推理开销 (参数量增加 $< 1\%$ ，DSCA < 0.11 M) 换取了显著的分割精度提升，且通过选择轻量骨干 (如 PVTv2) 可进一步大幅降低计算成本，具有良好的临床部署灵活性。

3.5. 伪深度图精度对性能的影响

为探究伪深度图存储精度对最终分割性能的影响，本研究对比了 float16 和 float32 两种精度格式，结果如表 4 所示。使用 float32 精度存储伪深度图时，模型取得更优的分割性能 (Dice = 88.58% vs. 87.80%)。float16 量化虽然可减少 50% 的存储开销，但深度值的精度损失会传播至 DSCA 模块，导致通道注意力权重计算出现偏差，最终影响分割精度。这反过来证明了伪深度图确实携带了有意义的几何信号，而非随机噪声。

Table 4. Performance of different pseudo-depth map storage precisions on ISIC 2018 (mean \pm std)

表 4. 不同伪深度图存储精度在 ISIC 2018 上的性能 (均值 \pm 标准差)

存储精度	Jl (%)	DC (%)	Acc (%)	Sen (%)	Spe (%)
float16	80.37 \pm 0.51	87.80 \pm 0.35	91.82 \pm 0.23	91.37 \pm 0.68	91.66 \pm 0.96
float32	81.71 \pm 0.79	88.58 \pm 0.44	92.30 \pm 0.21	90.96 \pm 0.78	93.36 \pm 0.53

3.6. ISIC 2018 数据集上的对比实验

为全面评估 DASH-Net 的性能，本研究将其与六种主流分割网络在 ISIC 2018 上进行对比，包括 U-Net [6]、UNet++ [7]、Ms RED [11]、PFD-Net [12]、WinGraphUNet [22] 和 DFAFormer [23]，结果如表 5 所示。DASH-Net (U-Net 骨干) 在 Dice (88.99%)、Jaccard (82.02%) 和 Sensitivity (94.20%) 三项指标上均优于所有对比方法，且参数量仅 7.85 M。使用 PVTv2 骨干的 DASH-Net 进一步将 Dice 提升至 90.61%，Jaccard 提升至 83.58%。可视化对比如图 4 所示。DASH-Net 的病灶边界更为平滑且连续，尤其在边界模糊区域 (如第 1、3 行) 和毛发遮挡区域 (如第 2 行) 表现突出。

Table 5. Comparison of segmentation performance among different networks on ISIC 2018 (mean \pm std). Bold indicates the best result, underline indicates the second-best result

表 5. 不同网络在 ISIC 2018 上的分割性能对比 (均值 \pm 标准差)。粗体为最优结果，下划线为次优结果

网络	参数量 (M)	Jl (%)	DC (%)	Acc (%)	Sen (%)	Spe (%)	HD95	ASSD
U-Net [5]	9.42	78.99 \pm 0.81	87.01 \pm 0.57	91.17 \pm 0.97	92.27 \pm 0.71	90.87 \pm 0.74	25.28 \pm 0.87	9.70 \pm 0.43
UNet++ [6]	9.16	79.45 \pm 0.61	87.45 \pm 0.38	91.59 \pm 1.07	92.72 \pm 0.52	91.49 \pm 0.67	25.66 \pm 0.68	9.40 \pm 0.20
Ms RED [10]	4.02	80.60 \pm 0.83	88.00 \pm 0.79	91.44 \pm 0.32	90.75 \pm 0.86	91.96 \pm 0.36	23.73 \pm 0.45	9.04 \pm 0.46

续表

PFD-Net [11]	28.89	81.83 ± 0.57	88.94 ± 0.50	92.26 ± 0.64	93.64 ± 0.28	91.34 ± 0.49	21.80 ± 0.45	8.17 ± 0.93
WinGraphUNet [25]	16.90	79.08 ± 0.49	87.09 ± 0.61	91.21 ± 0.87	93.80 ± 0.81	89.48 ± 0.48	20.61 ± 0.31	8.40 ± 0.14
DFAFormer [26]	53.30	80.57 ± 0.33	87.69 ± 0.44	91.57 ± 0.90	90.81 ± 0.92	93.03 ± 0.34	18.37 ± 0.54	7.59 ± 0.74
DASH-Net (ours)	17.49	82.02 ± 0.16	88.99 ± 0.38	92.20 ± 0.72	94.20 ± 0.58	89.14 ± 0.64	20.93 ± 0.17	8.11 ± 0.64

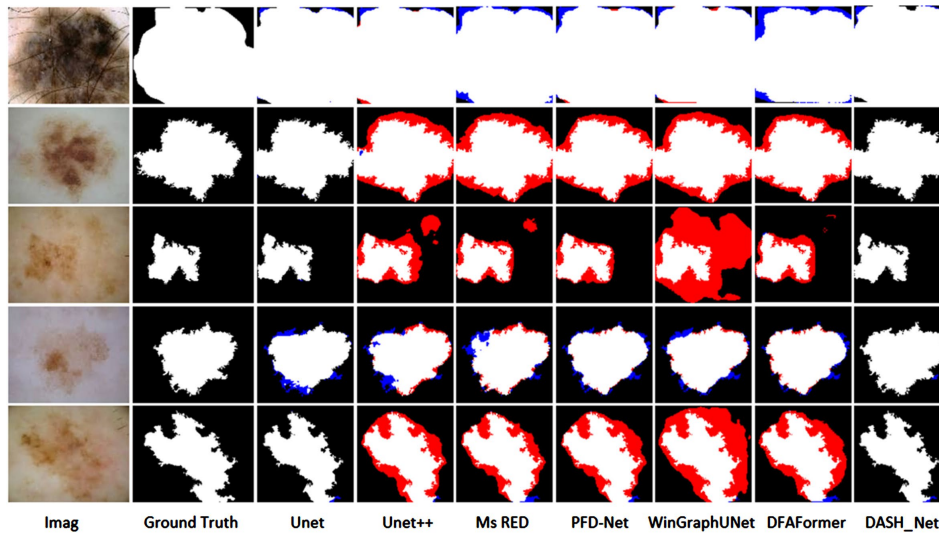


Figure 4. Visual comparison of segmentation results on ISIC 2018 dataset (red: false positives, blue: false negatives)
图 4. ISIC 2018 数据集上不同方法的分割可视化对比(红色: 假阳性, 蓝色: 假阴性)

3.7. 静脉曲张数据集上的实验结果

为验证 DASH-Net 及 DASH-Light 在静脉曲张分割中的有效性, 本研究在私有静脉曲张数据集上进行了对比实验, 结果如表 6 所示。DASH-Light 在该数据集上取得最优 Dice (88.60%)和 Jaccard (79.89%), 较当前最优方法(MALUNet)分别提升 0.62%和 0.91%。DASH-Net (使用亮度代理)同样表现优异(Dice = 87.98%), 验证了亮度代理策略的有效性。

Table 6. Performance comparison of different networks on the venous malformation dataset. Bold indicates the best result
表 6. 不同网络在静脉曲张数据集上的性能对比。粗体为最优结果

网络	JI (%)	DC (%)	Acc (%)	Sen (%)	Spe (%)	HD95	ASSD
U-Net [5]	45.79	57.73	97.70	71.98	98.37	58.60	20.12
UNet++ [6]	41.96	53.05	97.61	66.03	98.38	61.89	22.95
Ms RED [10]	42.75	55.07	97.68	67.80	98.55	70.89	24.70
PFD-Net [11]	53.76	65.70	98.37	76.12	98.92	43.89	14.16
WinGraphUNet [25]	51.72	64.17	98.26	76.65	98.80	45.23	14.46
DFAFormer [26]	49.91	61.98	98.13	71.56	98.98	42.61	14.89
DASH-Net (ours)	50.15	66.80	98.35	71.69	98.91	30.21	10.78

可视化对比如图 5 所示。DASH-Net 在边界精度和小目标检测方面具有明显优势。

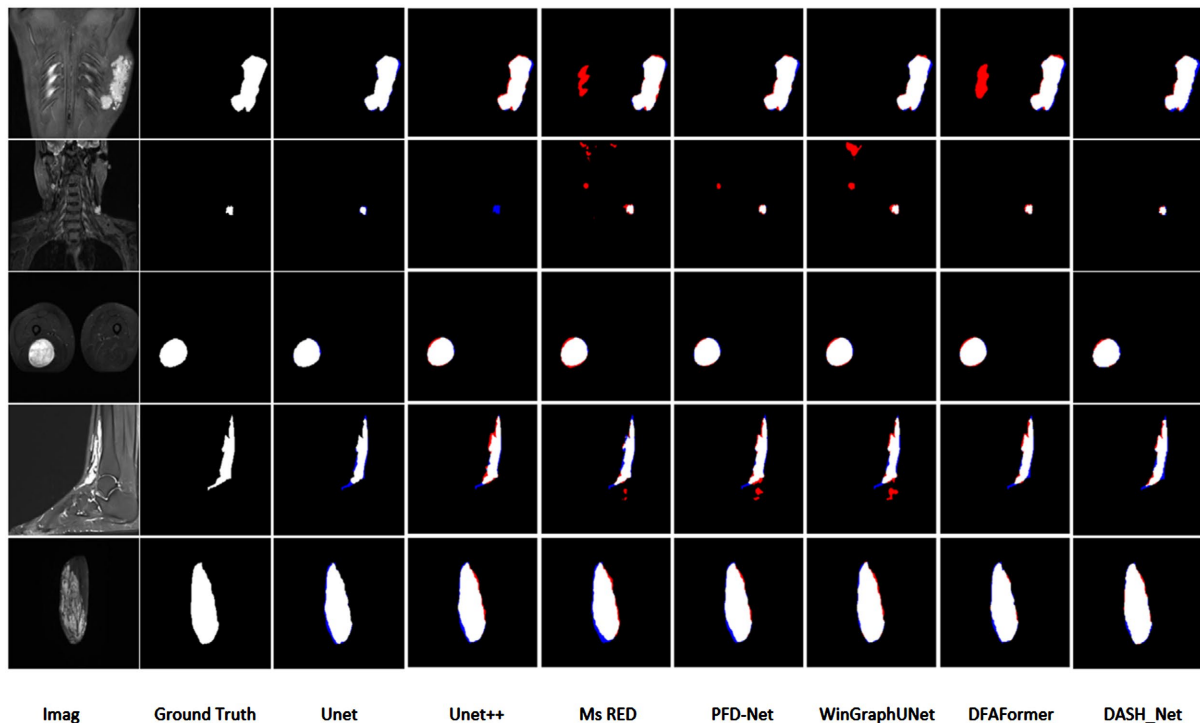


Figure 5. Visual comparison of segmentation results on the venous malformation dataset (red: false positives, blue: false negatives)

图 5. 静脉畸形数据集上不同方法的分割可视化对比(红色：假阳性，蓝色：假阴性)

3.8. 优化器及学习策略的比较

本研究对比了两组优化配置在 ISIC 2018 上的表现，结果如表 7 所示。SGD + 阶梯式衰减策略取得最优 Dice (88.99%) 和 Sensitivity (94.20%)，而 AdamW + 余弦退火策略在 Specificity (91.24%) 上略优。两种策略的差异主要体现在收敛速度上：SGD 配置在约 80 epoch 即达到稳定，而 AdamW 配置需要约 100 epoch。

Table 7. Performance of different optimizers and learning strategies on ISIC 2018 (mean \pm std)

表 7. 不同优化器及学习策略在 ISIC 2018 上的性能(均值 \pm 标准差)

优化器 + 学习策略	JI (%)	DC (%)	Acc (%)	Sen (%)	Spe (%)
AdamW + 余弦退火	81.81 \pm 0.90	88.81 \pm 0.66	92.18 \pm 0.51	92.51 \pm 0.69	91.24 \pm 0.57
SGD + 阶梯式衰减	82.02 \pm 0.16	88.99 \pm 0.38	92.20 \pm 0.72	94.20 \pm 0.58	89.14 \pm 0.64

3.9. 通道 - 深度关联的可视化验证

为验证 DSCA 模块是否真正建立了二维特征通道与视觉远近层次之间的显式关联，本研究设计了一项后处理分析实验。从 ISIC 2018 测试集中随机选取 50 张图像，依据 DPT 生成的伪深度图将像素划分为三个互斥区域：深层区域(深度值排名前 20%)、浅层区域(深度值排名后 20%)和背景区域(掩膜外部)。对

末层 DSCA (attn_out, $C_s = 64$) 的每个通道, 分别计算其在三个区域上的平均激活强度, 并进行 Min-Max 归一化。

依据激活模式将 64 个通道分为三类: 1) 深层偏好通道(Deep-Prefer, $n = 23$): 深层区域平均激活 > 0.5 且背景区域平均激活 < 0.1 ; 2) 背景偏好通道(BG-Prefer, $n = 20$): 背景区域平均激活 > 0.5 且深层区域平均激活 < 0.1 ; 3) 非专门化通道(Non-specialized, $n = 21$): 不满足上述任一条件。统计结果如表 8 所示。

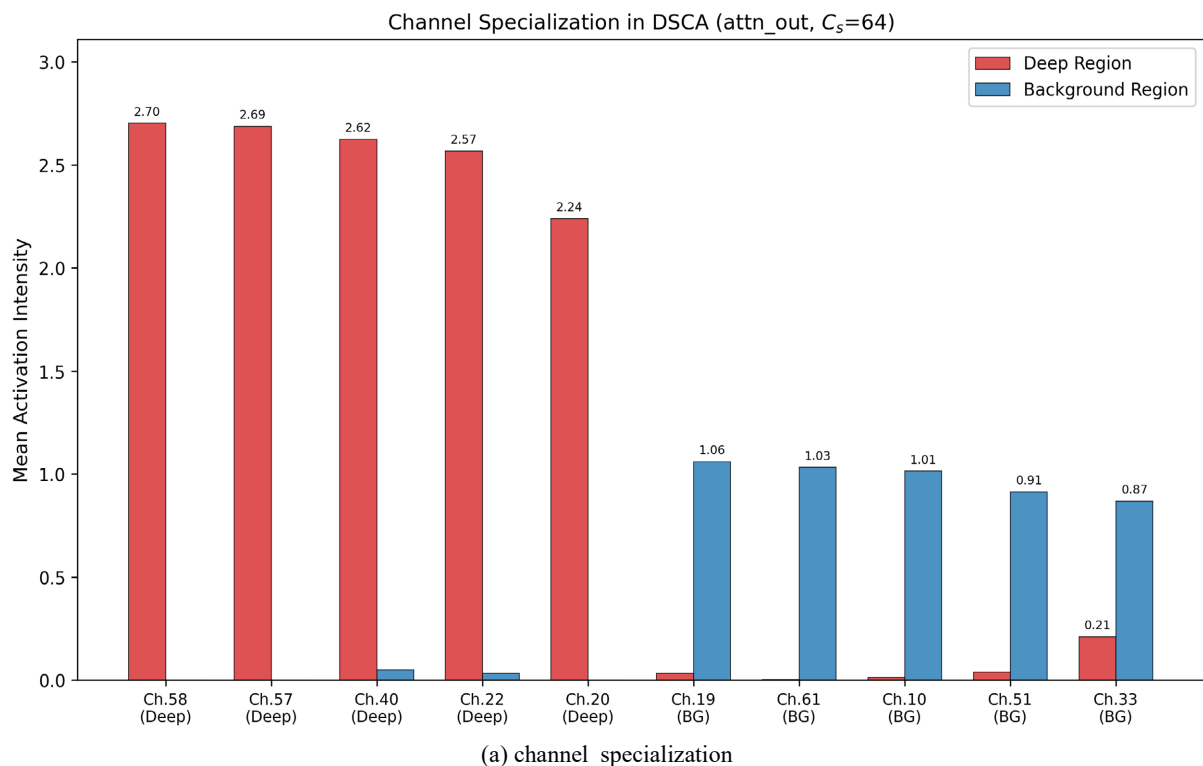
Table 8. Depth semantic specialization statistics of the last-layer DSCA channels (attn_out, $C_s = 64$)

表 8. 末层 DSCA 通道的深度语义专门化统计(attn_out, $C_s = 64$)

通道类别	代表通道	深层平均激活	浅层平均激活	背景平均激活	样本占比
Deep-Prefer	58, 57, 40, 22, 20	2.24~2.70	< 0.001	$< 0.001 \sim 0.050$	35.9% (23/64)
BG-Prefer	19, 61, 10, 51, 33	0.003~0.21	< 0.001	0.87~1.06	31.3% (20/64)
Non-specific	其余	混合	-	混合	32.8% (21/64)

Top-5 深层偏好通道(Ch58, 57, 40, 22, 20)在深层区域的平均激活强度达 2.24~2.70, 而在背景区域几乎为零(< 0.001), 说明这些通道专门编码“深层/病灶核心”相关的视觉特征。相反, Top-5 背景偏好通道(Ch19, 61, 10, 51, 33)在背景区域的平均激活达 0.87~1.06, 而在深层区域几乎为零, 说明这些通道负责编码“表层/背景皮肤”特征, 如图 6 所示。非专门化通道(35.9%, 23/64)在三个区域上的激活较为均匀, 可能编码通用的纹理或结构特征。

这一发现具有重要的方法学意义: DSCA 模块并非简单地对所有通道施加相同的深度加权, 而是自发地学习到一种“功能分工”——部分通道专门负责深层区域的分割, 部分通道专门负责背景区域的识别, 这与人类视觉系统中“what”通路和“where”通路的功能分离具有概念上的相似性。



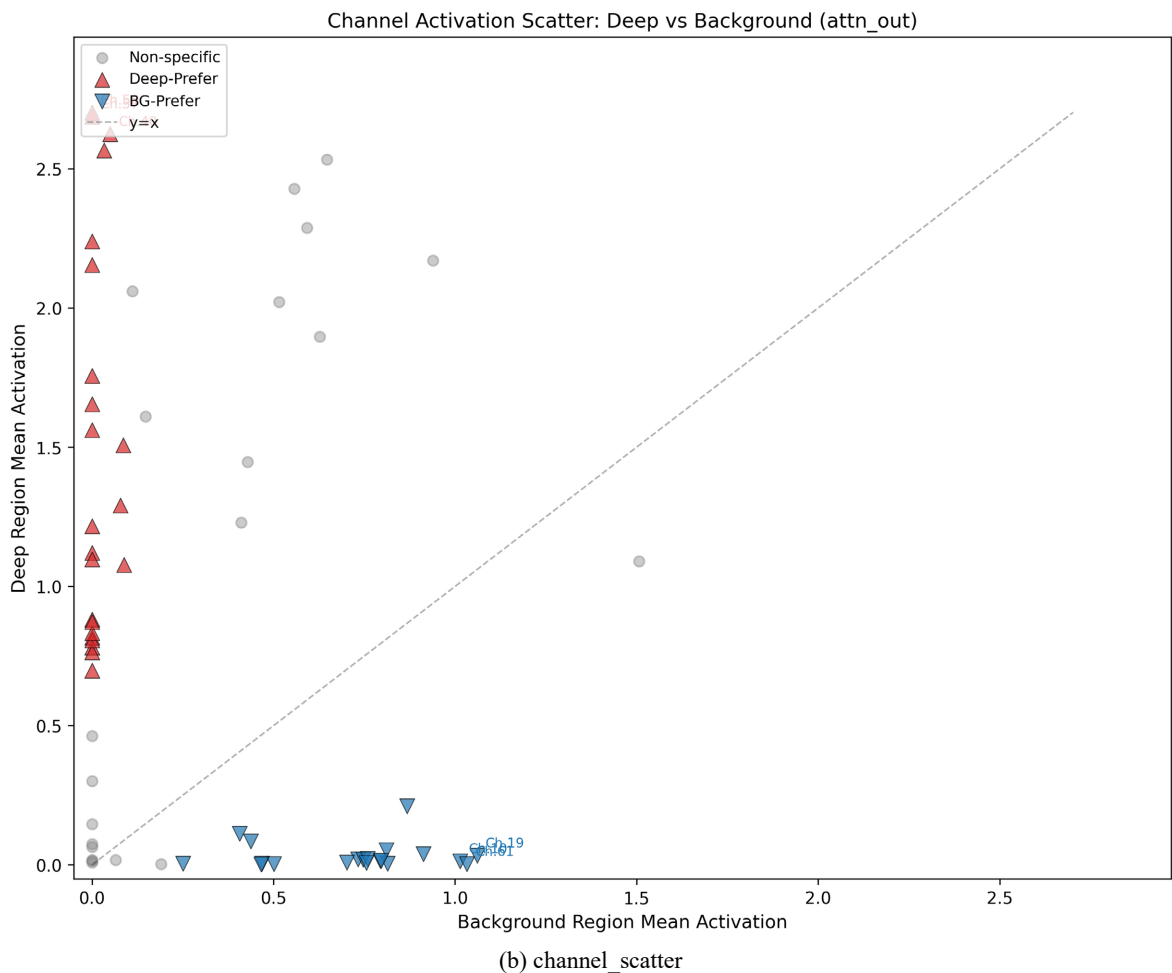


Figure 6. Visualization of channel-depth association. (a) Average activation intensity comparison of Top-5 deep-preferring channels (red) and Top-5 background-preferring channels (blue); (b) Deep-BG activation scatter plot of all 64 channels, with red triangles for deep-preferring channels, blue inverted triangles for background-preferring channels, gray dots for non-specialized channels, and dashed line $y = x$ as reference

图 6. 通道 - 深度关联可视化。(a) Top-5 深层偏好通道(红色)与 Top-5 背景偏好通道(蓝色)的平均激活强度对比；(b) 全部 64 个通道的 Deep-BG 激活散点图，红色三角为深层偏好通道，蓝色倒三角为背景偏好通道，灰色圆点为非专门化通道，虚线为 $y = x$ 参考线

4. 讨论

本研究提出并验证了一种基于深度 - 通道关联的医学图像分割新范式。传统分割方法将图像的所有特征通道视为等同，忽略了不同通道可能编码不同“视觉层次”的信息这一事实。DASH-Net 的核心思想在于：通过 DSCA 模块，让网络自动学习“哪些通道应该关注深层区域，哪些通道应该关注表层区域”，从而实现一种数据驱动的“深度语义分工”。

4.1. 伪深度图的质量分析与有效性来源

一个关键且常被质疑的问题是：DPT 在皮肤镜图像上生成的伪深度图质量究竟如何？它并非真实的物理深度，能否为分割提供可靠先验？本研究从三个角度分析其有效性。从深度估计模型的先验知识角度，DPT-SwinV2-Large-384 在大量自然场景图像上进行了预训练，学习到了丰富的场景几何先验，包括表面朝向、遮挡边界和相对距离关系[40]。当应用于皮肤镜图像时，这些先验能够帮助模型识别病灶区域

相对于周围皮肤表面的“隆起”或“凹陷”特征。从数值精度角度，表 4 的实验表明 float32 精度存储的伪深度图(Dice = 88.58%)显著优于 float16 (Dice = 87.80%)，差距达 0.78%，说明伪深度图中的精细数值变化对 DSCA 模块的特征选择具有实质影响。从辅助监督的作用机制角度，深度解码器作为辅助任务，其梯度通过共享编码器反向传播，迫使编码器在早期卷积层就学习与深度相关的空间层次特征，类似于多任务学习中的正则化效应。

4.2. 深度与通道关系的方法学分析

DSCA 模块的本质是一种“跨模态通道选择”机制。通道分支将深度描述符 $z \in R^{C_d}$ 通过两层 MLP 映射为通道权重 $\alpha_c \in R^{C_s}$ ，实现从“深度空间”到“特征空间”的线性变换。 W_1 提取最具代表性的深度语义， W_2 将其分配到各分割特征通道，使网络自动学习不同通道与不同深度模式的关联。

“图层即远近”的物理解释是本文的核心创新思想。分割特征 F_{seg} 的 C_s 个通道可视为 C_s 个“虚拟图层”，通道注意力 α_c 对其进行加权求和，权重由深度特征决定。 $\alpha_c(k)$ 较大的通道被保留(与当前深度层次相关)，较小的被抑制。当 α_c 在深层通道激活时，模型“选择”远景信息；在浅层通道激活时，聚焦近景信息。所有通道的加权和构成完整的深度层次表示，实现了从“通道维度”到“远近维度”的映射。

空间注意力 α_s 对深度不连续区域(即病灶边界)赋予更高权重，与通道注意力形成“空间-通道”联合选择：首先在空间上定位边界区域，然后在通道上选择适合的深度层次特征。残差缩放因子 γ 初始化为 0，训练初期 DSCA 不改变原始特征，随着深度估计可靠性提升， γ 逐渐增大，深度注意力逐步被主分割任务采纳。这种渐进式融合策略避免了训练初期深度误差对分割任务的干扰。

4.3. 边界监督机制的临床意义

在皮肤癌诊断中，病灶边界的规则性是 ABCD 法则的核心指标之一[2]。本研究的边界监督分支通过动态 Canny 边缘生成与 Hausdorff 距离损失，显式地将临床关注的边界质量纳入模型优化目标。动态 Canny 边界的自适应阈值策略($T_{low} = \mu_G - 0.5 \cdot \sigma_G$, $T_{high} = \mu_G + 0.5 \cdot \sigma_G$)相比固定阈值的优势在于：不同病灶具有不同的最佳边缘检测参数，大面积、边界模糊的病灶需要较低的阈值以捕获完整边界，而小面积、边界清晰的病灶则可使用较高阈值以避免噪声干扰。从实验结果看，添加边界监督后 HD95 从 25.28 降至 20.93 (↓17.2%)，ASSD 从 9.70 降至 8.11 (↓16.4%)，边界质量改善显著。

4.4. 局限性

本研究存在以下局限性：伪深度图质量受限于 DPT 模型的领域泛化能力，针对皮肤镜图像微调 DPT 以生成更高质量的领域特定深度图是未来方向。当前实验仅在 2D 图像上进行，DSCA 模块沿通道维度的注意力机制天然可扩展至 3D 分割(如 CT、MRI)。此外，静脉畸形数据集规模相对较小(87 张图像)，需在更大规模临床数据集上进一步验证 DASH-Light 的泛化能力。

5. 结论

本文提出 DASH-Net 及 DASH-Light 两种深度感知分割框架，核心创新在于通过深度引导的空间-通道双注意力模块(DSCA)建立二维图像特征通道与视觉远近层次之间的自适应关联。具体而言，DSCA 通过空间注意力图定位深度不连续区域(病灶边界)，通过通道注意力向量选择性地增强与当前深度层次相关的特征通道，并通过可学习的残差缩放参数实现渐进式融合。从伪深度到亮度代理的成功迁移进一步验证了“通道-视觉层次关联”作为一种通用计算范式的广阔应用前景。

参考文献

- [1] Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., *et al.* (2017) Dermatologist-Level Classification

- of Skin Cancer with Deep Neural Networks. *Nature*, **542**, 115-118. <https://doi.org/10.1038/nature21056>
- [2] Nachbar, F., Stolz, W., Merkle, T., *et al.* (1994) The ABCD Rule of Dermatoscopy: High Prospective Value in the Diagnosis of Doubtful Melanocytic Skin Lesions. *Journal of the American Academy of Dermatology*, **30**, 551-559.
- [3] 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). 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>
- [4] Sultana, F., Sufian, A. and Dutta, P. (2018) Advancements in Image Classification Using Convolutional Neural Network. 2018 *Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICR-CICN)*, Kolkata, 22-23 November 2018, 122-129. <https://doi.org/10.1109/icrcicn.2018.8718718>
- [5] Shelhamer, E., Long, J. and Darrell, T. (2017) Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 640-651. <https://doi.org/10.1109/tpami.2016.2572683>
- [6] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Lecture Notes in Computer Science*, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [7] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N. and Liang, J. (2020) UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging*, **39**, 1856-1867. <https://doi.org/10.1109/tmi.2019.2959609>
- [8] Oktay, O., Schlemper, J., Folgoc, L.L., *et al.* (2018) Attention U-Net: Learning Where to Look for the Pancreas.
- [9] 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, 18-23 June 2018, 7132-7141. <https://doi.org/10.1109/cvpr.2018.00745>
- [10] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: *Lecture Notes in Computer Science*, Springer, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1
- [11] Dai, D., Dong, C., Xu, S., Yan, Q., Li, Z., Zhang, C., *et al.* (2022) MS RED: A Novel Multi-Scale Residual Encoding and Decoding Network for Skin Lesion Segmentation. *Medical Image Analysis*, **75**, Article 102293. <https://doi.org/10.1016/j.media.2021.102293>
- [12] Yang, C. and Zhang, Z. (2024) PFD-Net: Pyramid Fourier Deformable Network for Medical Image Segmentation. *Computers in Biology and Medicine*, **172**, Article 108302. <https://doi.org/10.1016/j.compbiomed.2024.108302>
- [13] Huang, H., Xie, Q., Hu, Y., *et al.* (2023) CPCA: Channel Prior Convolutional Attention for Medical Image Segmentation.
- [14] Ruan, J., Xiang, S., Xie, M., Liu, T. and Fu, Y. (2022) MALUNet: A Multi-Attention and Light-Weight Unet for Skin Lesion Segmentation. 2022 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Las Vegas, 6-8 December 2022, 1280-1285. <https://doi.org/10.1109/bibm55620.2022.9995040>
- [15] Al-Masni, M.A., Kim, D.H. and Kim, T.S. (2020) Multiple Skin Lesions Diagnostics via Integrated Deep Convolutional Networks for Segmentation and Classification. *Computer Methods and Programs in Biomedicine*, **190**, Article 105351. <https://doi.org/10.1016/j.cmpb.2020.105351>
- [16] Sarker, M.M.K., Rashwan, H.A., Akram, F., Banu, S.F., Saleh, A., Singh, V.K., *et al.* (2018) SLSDeep: Skin Lesion Segmentation Based on Dilated Residual and Pyramid Pooling Networks. In: *Lecture Notes in Computer Science*, Springer, 21-29. https://doi.org/10.1007/978-3-030-00934-2_3
- [17] Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., *et al.* (2019) CE-Net: Context Encoder Network for 2D Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, **38**, 2281-2292. <https://doi.org/10.1109/tmi.2019.2903562>
- [18] Kaul, M.K. and Zhang, S. (2021) FocusNet: An Attentive-Based Fully Convolutional Network for Medical Image Segmentation. *Medical Image Analysis*, **67**, Article 101845.
- [19] Feng, S., Zhao, H., Shi, F., Cheng, X., Wang, M., Ma, Y., *et al.* (2020) CPFNet: Context Pyramid Fusion Network for Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, **39**, 3008-3018. <https://doi.org/10.1109/tmi.2020.2983721>
- [20] Chen, J., Lu, Y., Yu, Q., *et al.* (2021) TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation.
- [21] Wang, W., Xie, E., Li, X., Fan, D., Song, K., Liang, D., *et al.* (2022) PVT V2: Improved Baselines with Pyramid Vision Transformer. *Computational Visual Media*, **8**, 415-424. <https://doi.org/10.1007/s41095-022-0274-8>
- [22] Cao, H., Wang, Y., Chen, J., *et al.* (2021) Swin-UNet: UNet-Like Pure Transformer for Medical Image Segmentation.
- [23] Valanarasu, J.M.J., Oza, P., Hachialloglu, I., *et al.* (2021) MedT: Medical Transformer for Medical Image Segmentation.
- [24] Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J. and Maier-Hein, K.H. (2020) nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. *Nature Methods*, **18**, 203-211. <https://doi.org/10.1038/s41592-020-01008-z>
- [25] Kui, X.Y., Yan, H.N., Li, Q.S., *et al.* (2025) WinGraphUNet: Advanced Windowed Graph Modeling for Skin Lesion

- Segmentation. *Knowledge-Based Systems*, **329**, Article 114417.
- [26] Ji, Z.X., Ye, Y.X. and Ma, X. (2025) BDFormer: Boundary-Aware Dual-Decoder Transformer for Skin Lesion Segmentation. *Artificial Intelligence in Medicine*, **162**, Article 103079.
- [27] Söylemez, Ö.F. (2025) SkinAttn-Net: A Multi-Level Attention-Based Network for Skin Lesion Segmentation. *Scientific Reports*, **16**, Article No. 3700.
- [28] Toptaş, B. (2025) Enhanced Skin Lesion Segmentation via Attentive Reverse-Attention U-Net. *Symmetry*, **17**, Article 2002.
- [29] Naveed, A., Naqvi, S.S., Khan, T.M., et al. (2024) AD-Net: Attention-Based Dilated Convolutional Residual Network for Skin Lesion Segmentation.
- [30] Xu, R.T., Wang, C.W., Zhang, J.G., et al. (2024) SkinFormer: Learning Statistical Texture Representation for Efficient Skin Lesion Segmentation.
- [31] Akram, A., Rashid, J., Jaffar, M.A., Faheem, M. and Amin, R.U. (2023) Segmentation and Classification of Skin Lesions Using Hybrid Deep Learning Method in the Internet of Medical Things. *Skin Research and Technology*, **29**, e13524. <https://doi.org/10.1111/srt.13524>
- [32] Ma, C., Tian, S. and Yu, L. (2024) Crfnets: A Medical Image Segmentation Method Using the Cross Attention Mechanism and Refined Feature Fusion Strategy. In: Lin, Z., et al., eds., *Pattern Recognition and Computer Vision*, Springer Nature Singapore, 247-260. https://doi.org/10.1007/978-981-97-8490-5_18
- [33] Wang, S., Xu, L., Zhang, L., Zhang, Y., Li, C., Grzegorzec, M., et al. (2025) Hhbsnet: A Global Channel-spatial Attention and Multi-Scale Dilated Convolution Network for Automatic Melasma Segmentation. *Frontiers in Physiology*, **16**, Article 1665138. <https://doi.org/10.3389/fphys.2025.1665138>
- [34] Wang, Y., Zhang, H., Fu, J. and Tian, H. (2025) MSCB-unet: Elevating Skin Lesion Segmentation Performance with Multi-Scale Spatial-Channel Bridging Network. *Biomedical Signal Processing and Control*, **110**, 107986. <https://doi.org/10.1016/j.bspc.2025.107986>
- [35] Wang, L., Li, C., Zhang, Z., et al. (2021) Information Capacity-Based Depth Adjustment for Multiple Baselines Stereo. *IEEE Transactions on Circuits and Systems for Video Technology*, **31**, 4086-4100.
- [36] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/cvpr.2016.90>
- [37] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2017) Densely Connected Convolutional Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 2261-2269. <https://doi.org/10.1109/cvpr.2017.243>
- [38] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale.
- [39] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 9992-10002. <https://doi.org/10.1109/iccv48922.2021.00986>
- [40] Ranftl, R., Bochkovskiy, A. and Koltun, V. (2021) Vision Transformers for Dense Prediction. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 12159-12168. <https://doi.org/10.1109/iccv48922.2021.01196>
- [41] Miangoleh, S.M.H., Dille, S., Mai, L., et al. (2023) Boosting Monocular Depth Estimation Models to High-Resolution via Content-Adaptive Multi-Resolution Merging. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, 18-22 June 2023, 9685-9695.
- [42] Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J. and Zhao, H. (2024) Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 10371-10381. <https://doi.org/10.1109/cvpr52733.2024.00987>
- [43] Canny, J. (1986) A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**, 679-698. <https://doi.org/10.1109/tpami.1986.4767851>
- [44] Karimi, D. and Salcudean, S.E. (2020) Reducing the Hausdorff Distance in Medical Image Segmentation with Convolutional Neural Networks. *IEEE Transactions on Medical Imaging*, **39**, 499-513. <https://doi.org/10.1109/tmi.2019.2930068>