基于双监督反向注意力的医学图像分割*

胡博程 [1] 季葛鹏 [2] 邵典 [3] 范登平 [1]
[1] 南开大学 [2] 澳大利亚国立大学 [3] 西北工业大学

摘要

精准的医学图像分割在辅助诊断和治疗中具有至关重要的作用。此前，PraNet-V1 通过引入反向注意力（RA）模块来有效利用背景信息，显著提升了息肉分割性能。然而，其在多类别分割任务中的表现仍存在一定局限性。为此，我们提出了全新的 PraNet-V2 框架。相比 PraNet-V1，PraNet-V2 在应对多类别分割等更复杂任务时展现出更强的适应性。其核心组件——双监督反向注意力（DSRA）模块——引入了显式的背景监督机制，能够独立建模背景特征，并在语义更明确的空间实现更精细的注意力融合。实验结果表明，PraNet-V2 在四个主流息肉分割数据集上均取得了优异性能。同时，将 DSRA 模块集成到三种主流语义分割模型中，可进一步提升其前景分割质量，最高带来**1.36%**的平均 Dice 分数增益。相关代码与模型已开源，欢迎访问：https://github.com/ai4colonoscopy/PraNet-V2/tree/main/binary_seg/jittor。

1. 引言

医学图像分割在现代医学诊断与治疗中发挥着至关重要的作用，其核心任务是从医学图像中识别出目标区域，如病灶、器官或组织结构。随着医学领域对计算机图像处理技术的依赖日益增强，医学图像分割任务也在不断发展，由最初的二值分割任务[10, 12, 3, 28, 21, 35]，逐步扩展至更加复杂的多类别分割任务 [5, 37, 33, 40]。

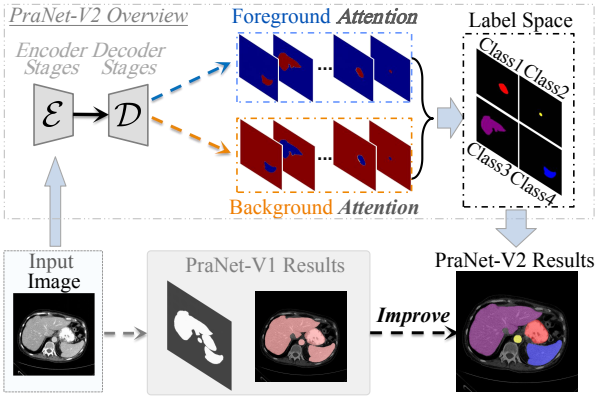

图 1. PraNet-V1 和 PraNet-V2 在背景建模和适配任务方面的关键区别图示。

例如，U-Net 系列模型 [27, 39, 14, 7]采用编码器-解码器架构，并通过多样的跳跃连接机制来捕获多尺度信息，从而为分割任务提供更精细的语义与空间特征。在此基础上，nnU-Net [16]致力于提升模型的通用性与自适应能力，不依赖于结构上的复杂改进。另一方面，DeepLab 系列模型 [6]利用空洞卷积扩大感受野，增强了上下文信息的处理能力，但其整体建模能力仍受限于卷积神经网络（CNN）的固有局限。近年来，基于 Transformer 的模型 [5, 4, 15, 32]则通过融合局部特征与全局上下文，有效建模长距离依赖关系，从而进一步提升了分割性能。

尽管现有方法在特征提取与注意力机制方面已取得显著进展，但上述模型却**忽视了背景特征建模**。这限制了模型对边界的精准刻画，尤其在前景与背景对比度较低的场景下，分割性能会明显下降。

我们此前的工作 PraNet-V1 [10]引入了反向注意力（RA）机制，可以显式对背景区域建模，从而在前景与背景对比度低、类别分布严重不均的

*本文为 CVMJ 2025-Submission 论文 [13]的中文翻译版。

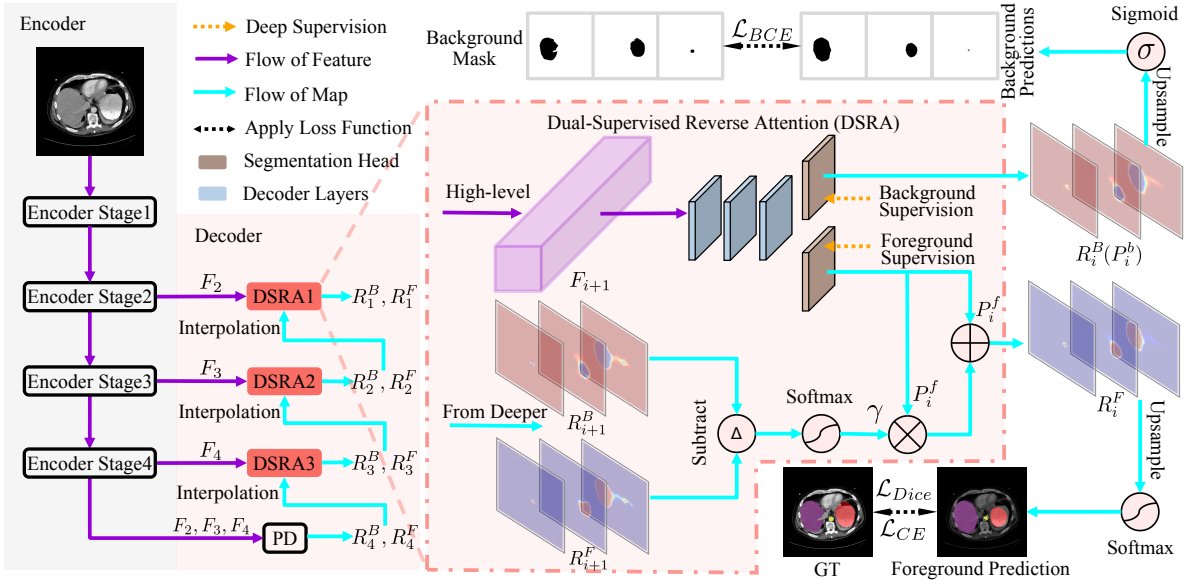

图 2. PraNet-V2 框架及 DSRA 模块的整体结构示意图。编码器产生的高级特征（F_2, F_3, F_4）依次由并行部分解码器（PD）及三个 DSRA 模块处理。每个 DSRA 模块对第 $i+1$ 层的高级特征 F_{i+1} 进行解码，生成前景与背景的分割图，同时融合来自更深层 DSRA 模块或 PD 的输出（R_{i+1}^F, R_{i+1}^B），以进一步优化前景分割结果。

场景下实现有效的息肉分割。然而，尽管 PraNet-V1 开创性地将反向注意力机制引入医学图像分割任务，后续研究和评估结果仍暴露出其部分局限性：**(1) 适用场景有限。** PraNet-V1 专为二值息肉分割任务设计，难以扩展至更复杂的多类分割场景；**(2) 基于规则的直接取反。** 在 PraNet-V1 中，反向注意力权重是通过将每个像素的前景概率从 1 中相减得到的，既未引入额外的上下文信息，也容易沿用前向注意力中的预测误差。**(3) 背景信息融合的方式缺乏明确的语义指导。** PraNet-V1 在特征空间中将反向与前向注意力进行融合，导致前景与背景的高维特征交织混杂，缺乏清晰的语义边界。

针对上述问题，我们对反向注意力（RA）模块进行了重构。如图1所示，我们提出的双监督反向注意力（DSRA）模块引入独立参数，可以对每一类别的前景与背景注意力进行显式建模。此外，DSRA 模块在语义增强的标签空间中融合前景与背景信息，相较于传统 RA 模块，在可解释性方面进一步提升。

总体而言，本文的主要贡献包括：

1. **提出全新模块：** 我们设计了 DSRA 模块，通过前景与背景的解耦处理，提升了对背景区域与边界的识别能力。
2. **构建新型框架：** 基于 DSRA 模块，我们构建了 PraNet-V2 模型，其在息肉分割任务中显著优于 PraNet-V1。
3. **达成最新 SOTA 结果：** 我们将 DSRA 模块集成至三种主流分割模型中，有效提升了其分割性能，平均 Dice 分数可以提高 0.50% 至 1.36%。

2. 方法

2.1. 双监督反向注意力模块

在息肉分割任务中，PraNet-V1 通过引入反向注意力（RA）有效地捕捉背景信息，于二分类场景表现出色。然而，在多类别分割任务中，单一的反向注意力计算无法有效区分不同类别；同时，其基于规则的反向注意力提取方式也难以与多通道的像素级置信度输出相兼容。

因此，我们在 PraNet-V1 的反向注意力（RA）模块基础上提出了双监督反向注意力（DSRA）模块。正如图2所示，我们将 DSRA 模块作为解码器的一部分嵌入 U-Net 架构中，充分利用多尺度特征与跳跃连接的优势。

对于输入图像 $X \in \mathbb{R}^{H\times W\times C}$，编码器首先提取多尺度特征表示 $\{F_i \,|\, i=1,2,3,4\}$，其中 $F_i \in \mathbb{R}^{\frac{H}{2^{i+1}}\times\frac{W}{2^{i+1}}\times C_i}$ 表示第 i 个编码阶段提取的特征。随后，解码器选取其中的三个高级特征 $\{F_i \,|\, i=2,3,4\}$ 进行下一步处理，通过四个阶段逐步生成分割结果 $\{R_i \,|\, i=1,2,3,4\}$。解码的整体流程由并行部分解码器（PD）[10] 和 DSRA3 至 DSRA1 完成。每个阶段的分割输出均包含各类别前景分割结果（R_i^F）及其对应的背景分割结果（R_i^B）。我们对 PraNet-V1 的并行部分解码器（PD）输出层进行了调整，并将其用于第一个解码阶段，以聚合高级特征 (F_2, F_3, F_4)，从而生成粗糙的分割结果 $R_4 = \{R_4^F, R_4^B\}$。

在此基础上，后三个解码阶段进一步引入 DSRA 模块，将粗糙的分割结果逐步细化为更精确的结果。如图2右侧所示，第 i 个 DSRA 模块 $(i=1,2,3)$ 的输出为 $R_i = \{R_i^F, R_i^B\}$，其计算方式如下：

$$R_i^F = P_i^f + P_i^f \circ \gamma, \tag{1}$$

$$R_i^B = P_i^b. \tag{2}$$

其中，P_i^f 和 P_i^b 分别为在前景监督与背景监督分支的输出；符号 $\circ$ 表示逐元素乘法操作。γ 表示反向增益，用于引入来自更深层 DSRA (或 PD) 模块的细化信息，从而更有效地利用级联结构与背景信息。

与 PraNet-V1 中的 RA 模块不同，DSRA 模块使用了专门的监督信号与独立的结构**分别**生成前景与背景的分割结果，缓解了共享结构与参数引发的语义混杂问题。P_i^f、P_i^b 与 γ 的具体计算方式如下：

$$\{P_i^f, P_i^b\} = \phi(F_{i+1}), \tag{3}$$

$$\gamma = \mathrm{Softmax}(\mathcal{I}(R_{i+1}^F; F_{i+1}) - \mathcal{I}(R_{i+1}^B; F_{i+1})), \tag{4}$$

其中，$\phi(\cdot)$ 表示卷积层，它们组成了解码器层与分割头；$\mathcal{I}(x; y)$ 表示通过双线性插值将张量 x 的宽度与高度调整为与 y 相同。R_{i+1}^F 与 R_{i+1}^B 分别表示第 $i{+}1$ 层 DSRA 模块生成的前景与背景分割图。在上述公式所描述的层级级联结构中，来自更深层 DSRA 模块的输出被逐级整合，用于不断细化粗糙的分割结果。这一设计使得 DSRA 模块能够以参数学习的方式提取反向注意力，并在分割标签空间中与前向注意力进行融合，从而有效缓解 PraNet-V1 直接融合压缩特征所带来的空间与语义对齐误差问题。

总的来说，相较于传统的 RA 模块，DSRA 在以下三个关键方面表现更为出色：

(1) **结构独立性：** DSRA 为前景与背景分别设计了**独立**的分割分支，便于捕捉更细致的特征信息。

(2) **背景建模能力：** DSRA 引入额外的监督信号，通过参数拟合的方式学习每一类别的背景特征，从而实现对多类别分割任务的有效支持。

(3) **语义信息精细化：** DSRA 在标签空间中融合前景与背景信息，充分利用像素级置信度，进一步提升了边界定位与背景区域的分割精度。

2.2. 背景监督与损失函数

背景掩码。 为了对每个类别的背景区域进行监督，我们引入了一个多通道背景掩码，其中的每个通道对应一个目标类别。在该掩码中，像素值为 1 表示背景区域，值为 0 表示对应的目标区域。如图2顶部的 “Background Mask” 所示，背景掩码基于真实标签（Ground Truth）构建，用于提取每个类别对应的背景区域，其中白色区域表示 1，黑色区域代表 0。

损失函数。 基于背景掩码，我们将总损失函数定义为 $L_{\text{total}} = w_1 \times L_{\text{Dice}} + w_2 \times L_{\text{CE}} + w_3 \times L_{\text{BCE}}$，其中 w_1、w_2 和 w_3 为加权系数。对于前景监督，Dice 损失（L_{Dice}）用于缓解类别不平衡问题并提升区域级分割效果，交叉熵损失（L_{CE}）则在局部细粒度地优化像素级分割性能。对于背景监督，二值交叉熵损失（L_{BCE}）用于将模型输出的背景预测与背景掩码对齐，从而实现对各类别背景的独立建模。

实现细节。 在 DSRA 模块的基础上，我们提出了 PraNet-V2 框架，其在息肉分割任务中的性能详见第3.1节。此外，DSRA 展现出很强的通用性，多数主流分割网络均可借鉴 DSRA 的双分支结构，通过两个分割头分别生成前景与背景的分割结果，并按照式1与式4中所述的方式，对前景分割进行迭代优化。鉴于 DSRA 良好的灵活性，我们在第3.2节中进一步评估了其在三种主流多类别医学图像分割模型中的集成效果。

表 1. PraNet-V1 和 PraNet-V2 在四个息肉分割数据集上的性能比较，其中最佳性能值以**粗体**表示。

Dataset	Backbone	mDice (%)		mIoU (%)		wFm (%)		S-m (%)		mEm (%)		MAE ($\times 10^{-2}$)	
		V1	V2	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2
CVC-300	Res2Net50 [11]	87.06	**89.83**	79.61	**82.66**	84.32	**87.79**	92.55	**93.70**	94.97	**97.47**	0.99	**0.59**
CVC-ClinicDB		89.84	**92.28**	84.83	**87.22**	89.63	**91.97**	93.67	**94.87**	96.22	**97.38**	0.94	**0.91**
Kvasir		89.39	**90.70**	83.55	**85.29**	88.00	**89.59**	91.25	**91.70**	94.00	**95.07**	3.04	**2.35**
ETIS		62.75	**64.05**	**56.57**	56.54	60.07	**60.43**	79.33	**79.41**	**80.77**	79.74	3.07	**2.08**
CVC-300	PVTv2-B2 [36]	86.59	**89.89**	78.92	**83.11**	83.15	**88.48**	91.84	**93.96**	94.45	**97.04**	1.03	**0.73**
CVC-ClinicDB		90.96	**93.09**	85.42	**88.06**	89.90	**92.80**	94.34	**94.45**	96.49	**98.23**	1.02	**0.84**
Kvasir		87.09	**91.52**	81.31	**86.12**	84.52	**90.39**	89.33	**92.50**	92.58	**95.64**	4.19	**2.33**
ETIS		68.32	**76.35**	60.02	**68.72**	61.65	**72.96**	81.38	**86.50**	80.92	**88.26**	4.14	**1.45**

3. 实验

3.1. 二值分割

数据集。 我们在四个息肉分割数据集上开展实验，比较 PraNet-V1 与 PraNet-V2 的性能，并遵循 PraNet-V1 原始的数据划分方法 [10]。具体而言，我们使用的数据集包括 CVC-ClinicDB [1]、CVC-300 [31]、Kvasir [17] 以及 ETIS [29]。CVC-ClinicDB 数据集共包含 612 张图像，其中 62 张用于测试，其余用于训练。ETIS 数据集由 196 张图像组成，所含息肉通常较小且不易被检测。Kvasir 数据集包含 1,000 张图像，涵盖 700 个大息肉、48 个小息肉和 323 个中等大小的息肉 [19]，并按 8:1:1 的比例划分为训练集、验证集与测试集。为评估模型在未知数据上的泛化能力，我们未将 ETIS 和 CVC-300 的图像纳入训练，将其作为独立测试集用于性能验证。

训练细节。 所有实验均在配备有一张 NVIDIA GeForce RTX 3090 GPU 的计算节点上进行，并基于 PyTorch 2.0.1 与 CUDA 12.2 实现。我们沿用了 PraNet-V1 的训练配置，包括将输入图像分辨率统一调整为 352×352，并采用 $\{0.75\times, 1\times, 1.25\times\}$ 的多尺度训练策略。优化器使用了 Adam 并设定固定学习率 1×10^{-4}。在损失函数设计上，我们沿用了 PraNet-V1 的加权二值交叉熵损失 $\mathcal{L}^{w}_{\mathrm{BCE}}$，用于监督 DSRA 中的一个分割分支。该损失通过像素级加权机制对边界区域的误差进行更强惩罚，从而提升模型对目标边缘的辨识能力和分割精度。

评估指标。 我们采用了 PraNet-V1 [10]中使用的指标对分割性能进行量化分析，具体包括：平均 Dice 系数（mDice, %）、平均交并比（mIoU, %）、加权 F-measure（wFm, %）、结构相似性（S-m, %）[8]、平均增强度量（mEm, %）[9] 和平均绝对误差（MAE）。其中，mDice、mIoU 与 MAE 是经典的二值分割评估指标；wFm 则对息肉内部区域与边缘区域赋予更高惩罚权重。除像素级度量外，S-m 综合考虑了区域层次与目标层次的结构相似性；而 mEm 则同时评估了像素级对齐程度和整体结构准确性，提供更全面的分割质量评估。

定量分析。 公平起见，我们在相同的主干网络上实现了 PraNet-V1 与 PraNet-V2 并对比性能。正如表1所示，当使用 Res2Net50 [11] 作为主干网络时，PraNet-V2 在几乎所有数据集和评估指标上均显著优于 PraNet-V1。具体而言，在 CVC-300 与 CVC-ClinicDB 数据集上，PraNet-V2 在 mDice 指标上分别提升了 2.77%，2.44%，在 mIoU 指标上提升了 3.05%，2.39% ，可以更准确分割息肉。当主干网络更换为 PVTv2-B2 [36] 时，PraNet-V2 的性能提升变得更加显著，可以在所有基准数据集和评估指标上全面超越 PraNet-V1，进一步证明了 DSRA 模块在不同编码器架构下的强鲁棒性。特别是在未见过的 ETIS 数据集上，PraNet-V2 的 mDice 提升达 8.03%，mIoU 提升达 8.70%，充分展现了模型优秀的泛化能力。此外，PraNet-V2 在 S-m 指标上提升了 5.12%，也表明模型在保持分割区域整体形状、边界一致性与几何结构完整性方面同样具有优势。

表 2. Synapse 数据集上的性能比较。基线结果来自 [25]，我们将 DSRA 改进的数值以**粗体**显示。† 符号表示重现性能。我们还统计了每个器官的 Dice 分数 (%)。

Architectures	mDice (%)	HD95 (mm)	mIoU (%)	Aorta	GB	KL	KR	Liver	PC	SP	SM
UNet [27]	70.11	44.69	59.39	84.00	56.70	72.41	62.64	86.98	48.73	81.48	67.96
AttnUNet [22]	71.70	34.47	61.38	82.61	61.94	76.07	70.42	87.54	46.70	80.67	67.66
R50+UNet [5]	74.68	36.87	–	84.18	62.84	79.19	71.29	93.35	48.23	84.41	73.92
R50+AttnUNet [5]	75.57	36.97	–	55.92	63.91	79.20	72.71	93.56	49.37	87.19	74.95
SSFormer [34]	78.01	25.72	67.23	82.78	63.74	80.72	78.11	93.53	61.53	87.07	76.61
PolypPVT [3]	78.08	25.61	67.43	82.34	66.14	81.21	73.78	94.37	59.34	88.05	79.40
TransUNet [5]	77.61	26.90	67.32	86.56	60.43	80.54	78.53	94.33	58.47	87.06	75.00
SwinUNet [4]	77.58	27.32	66.88	81.76	65.95	82.32	79.22	93.73	53.81	88.04	75.79
MT-UNet [32]	78.59	26.59	–	87.92	64.99	81.47	77.29	93.06	59.46	87.75	76.81
MISSFormer [15]	81.96	18.20	–	86.99	68.65	85.21	82.00	94.41	65.67	91.92	80.81
PVT-CASCADE [23]	81.06	20.23	70.88	83.01	70.59	82.23	80.37	94.08	64.43	90.10	83.69
TransCASCADE [23]	82.68	17.34	73.48	86.63	68.48	87.66	84.56	94.43	65.33	90.79	83.52
MIST† [26]	81.91	14.93	–	86.15	71.43	83.09	76.43	96.02	68.2	89.39	84.59
MIST (w/ DSRA)	**83.27**	**14.11**	–	**87.54**	**75.36**	82.23	**76.53**	95.93	**71.51**	**91.66**	**85.41**
EMCAD-B2† [25]	82.71	21.74	74.65	87.24	69.56	85.23	80.88	95.59	65.88	92.62	84.64
EMCAD-B2 (w/ DSRA)	**83.75**	**17.77**	**74.81**	**88.69**	**72.79**	**85.41**	**82.91**	**95.82**	**68.47**	**93.09**	**85.85**

3.2. 多类别分割

模型与数据集。 我们在两个多类别医学图像分割数据集上开展实验，评估在集成 DSRA 模块后，目前最优的分割模型会展现出怎样的性能。所选模型包括 Cascaded MERIT [24]、MIST [26] 和 EMCAD [25]。我们使用了自动化心脏诊断挑战数据集（ACDC）[2] 与 Synapse 多器官分割数据集(Synapse)[20]。ACDC 数据集包含来自 100 名患者的心脏 MRI 图像，并标注了三类结构：右心室（RV）、左心室（LV）以及心肌（Myo）。Synapse 数据集则由 30 例增强 CT 扫描图像组成，共计 3,779 个切片，涵盖 8 类腹部器官的分割标注，包括主动脉（Aorta）、胆囊（GB）、左肾（KL）、右肾（KR）、肝脏（Liver）、胰腺（PC）、脾脏（SP）和胃（SM）。本节实验遵循上述三种模型论文中提到的数据划分策略，以确保评估结果具有可比性和一致性。

训练细节。 我们采用上述三个模型原始的训练策略，并对 batch size 进行了轻微调整，以更好适配新插入的 DSRA 模块。

评估指标。 参考文献 [24, 26, 25]的设置，我们在 ACDC 和 Synapse 数据集上采用平均 Dice 系数（mDice，%）评估模型的分割性能。此外，对于 Synapse 数据集，我们进一步借鉴 [25]的评估方案，引入平均交并比（mIoU, %）和 95 百分位 Hausdorff 距离（HD95, mm）作为补充指标。其中，HD95 用于衡量模型预测边界与真实边界之间的最大偏差（取 95 百分位），其值越小，表明预测边界与真实边界的对齐程度越高。

定量分析。 在包含八类器官的 Synapse 数据集上，集成 DSRA 模块后，模型整体性能显著提升（见表2）。以 MIST (w/ DSRA) 模型为例，平均 Dice 分数提升到了 83.27%，而 HD95 降低至 14.11，较原始版本分别提高了 1.36% 和 0.82，显示出更优的分割效果。同样地，器官级别指标提升也十分明显，例如胆囊（GB）类别的 Dice 分数由 71.43% 提升至 75.36%（+3.93%）。类似地，EMCAD-B2（w/ DSRA）在集成 DSRA 后也展现出稳定的性能增益：平均 Dice 分数由 82.71% 提高至 83.75%（+1.04%），HD95 从 21.74 降至 17.77，进一步验证了 DSRA 模块在提升分割精度和降低边界误差方面的有效性。

在 ACDC 数据集上（见表3），DSRA 同样增强了 MIST 与 Cascaded MERIT 的分割能力。引入 DSRA 后，两者的平均 Dice 分数分别达到 92.31% 和 92.28%，较原始版本均有明显提升。值得注意的是，在结构复杂、分割难度较高的右心室（RV）类别

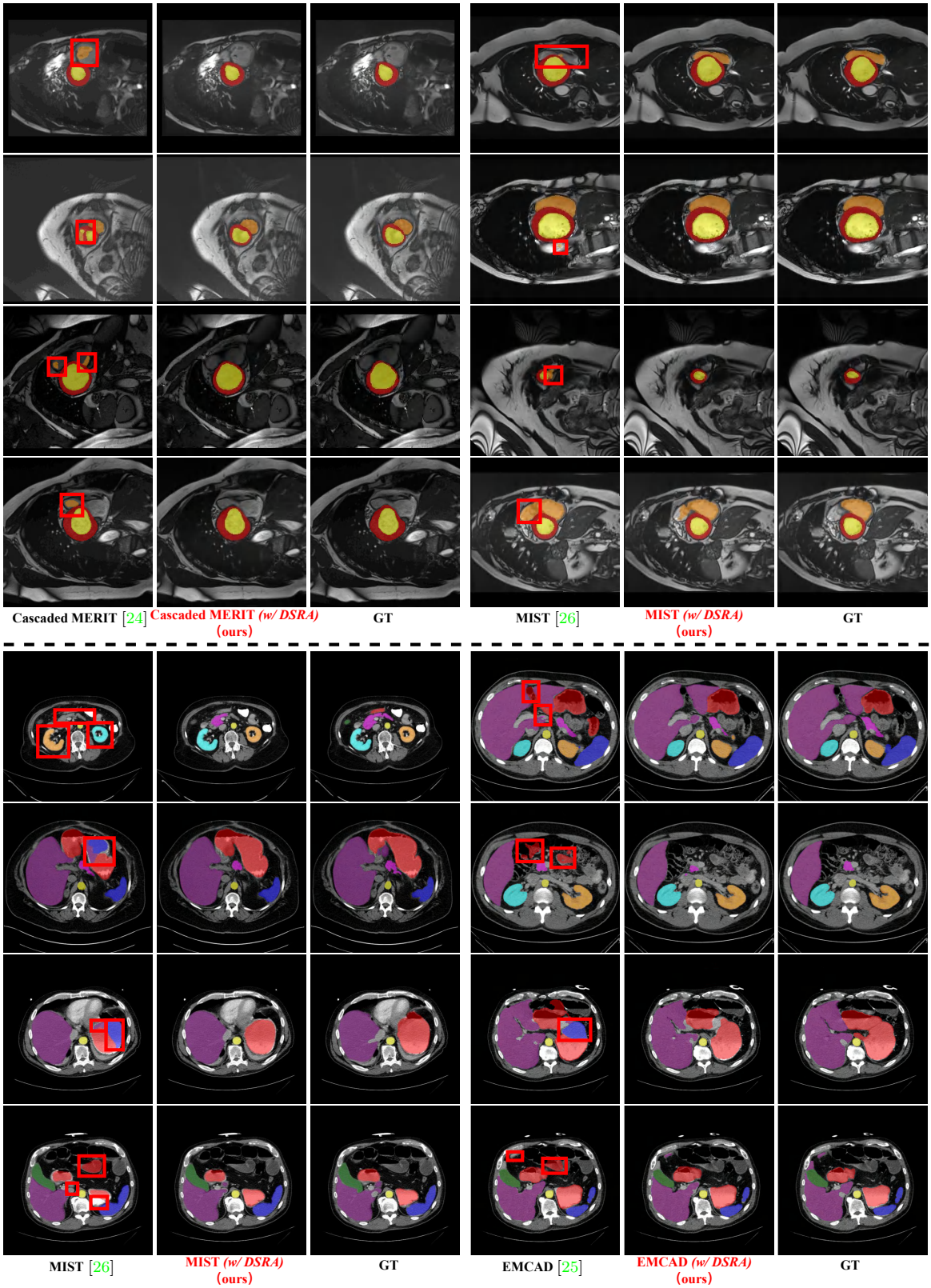

图 3. 在 ACDC 数据集（上）和 Synapse 数据集（下）的分割结果可视化，其中分割错误以红色框突显。

表 3. ACDC 数据集上的性能比较。基线结果来自 [24, 26]，我们将 DSRA 改进的数值以粗体显示。† 表示重现性能，我们还统计了每个类别的 Dice 分数 (%)

Architectures	mDice (%)	RV	Myo	LV
R50+UNet [5]	87.55	87.10	80.63	94.92
R50+AttnUNet [5]	86.75	87.58	79.20	93.47
ViT+CUP [5]	81.45	81.46	70.71	92.18
R50+ViT+CUP [5]	87.57	86.07	81.88	94.75
TransUNet [5]	89.71	88.86	84.53	95.73
SwinUNet [4]	90.00	88.55	85.62	95.83
MT-UNet [32]	90.43	86.64	89.04	95.62
MISSFormer [15]	90.86	89.55	88.04	94.99
PVT-CASCADE [23]	91.46	88.90	89.97	95.50
nnUNet [16]	91.61	90.24	89.24	95.36
nnFormer [38]	91.78	90.22	89.53	95.59
MIST† [26]	91.73	89.98	89.39	95.84
MIST (w/ DSRA)	**92.31**	**90.82**	**90.07**	**96.04**
Cascaded MERIT† [24]	91.78	90.36	89.21	95.79
Cascaded MERIT (w/ DSRA)	**92.28**	**91.27**	**89.38**	**96.19**

表 4. 针对损失函数的消融实验

Loss function			mDice (%)	mIoU (%)
$\mathcal{L}_{\text{BCE}}$	$\mathcal{L}_{\text{CE}}$	$\mathcal{L}_{\text{Dice}}$		
✓	✓		91.91	85.43
✓		✓	92.14	85.72
	✓	✓	92.16	85.84
✓	✓	✓	**92.31**	**86.02**

上，两种模型的 Dice 分数均提升了近 1%，进一步验证了 DSRA 模块在刻画不规则结构方面的优势。

定性实验。 图3展示了 ACDC 与 Synapse 数据集上的分割结果，对比了三种模型在集成 DSRA 模块前后的分割结果差异。从可视化结果可以看出，在引入 DSRA 后，三种模型均表现出更高的分割精度，预测结果中的错误与冗余显著减少。

3.3. 消融实验

我们在 MIST (w/ DSRA) 模型上进行了消融实验，以评估不同损失函数组合对分割性能的影响。如表4所示，当联合使用 $\mathcal{L}_{\text{BCE}}$、$\mathcal{L}_{\text{CE}}$ 和 $\mathcal{L}_{\text{Dice}}$ 时，网络取得了最佳性能，平均 Dice 系数（mDice）达到 92.31%，平均交并比（mIoU）为 86.02%。移除任意一项损失函数均会导致性能下降。具体而言，去除 $\mathcal{L}_{\text{Dice}}$ 会削弱模型处理类别不平衡图像的能力，而去除 $\mathcal{L}_{\text{BCE}}$ 或 $\mathcal{L}_{\text{CE}}$ 则会影响 DSRA 模块中两个分割分支将特征映射至标签空间的能力。

4. 讨论与结论

4.1. 局限性与未来工作

尽管将 DSRA 模块集成至三种主流分割模型中显著提升了它们的性能，这些模型仍受限于训练数据中预定义的类别，难以应对开放世界场景下出现的未知类别。在医学实践中，医生在面对未被归类的疾病时，往往需要依赖专家经验和共识进行判断。因此，未来研究可进一步探索异常检测方法，并尝试引入双解码器结构、特征空间操控等机制，以提升模型对未知类别的识别与适应能力 [30]。

4.2. 结论

在本文中，我们针对 PraNet-V1 的局限性，提出了双监督反向注意力（DSRA）模块，并据此构建了新版 PraNet-V2 模型。DSRA 通过解耦前景与背景特征的建模过程，有效提升了特征捕捉能力，从而显著增强了息肉分割的准确性。进一步地，我们将 DSRA 模块集成至三种表现最优的语义分割模型中，在两个基准数据集上实现了 0.50% 至 1.36% 的平均 Dice 分数提升，验证了其良好的通用性与有效性。未来，我们将探索将反向注意力机制扩展至文本引导的语义分割任务 [18]，以进一步提升模型的语义理解能力和跨场景泛化能力。

致谢

本研究得到国家自然科学基金（No. 62476143 和 No. 62306239）的支持。

参考文献

[1] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. CMIG, 43:99–111, 2015. 4

[2] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. Gonzalez Ballester, G. Sanroma, S. Napel, S. Petersen, G. Tziritas, E. Grinias, M. Khened, V. A. Kollerathu, G. Krishnamurthi, M.-M. Rohé, X. Pennec, M. Sermesant, F. Isensee, P. Jäger, K. H. Maier-

Hein, P. M. Full, I. Wolf, S. Engelhardt, C. F. Baumgartner, L. M. Koch, J. M. Wolterink, I. Išgum, Y. Jang, Y. Hong, J. Patravali, S. Jain, O. Humbert, and P.-M. Jodoin. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE TMI, 37(11):2514–2525, 2018. 5

[3] D. Bo, W. Wenhai, F. Deng-Ping, L. Jinpeng, F. Huazhu, and S. Ling. Polyp-pvt: Polyp segmentation with pyramid vision transformers. CAAI AIR, 2:9150015, 2023. 1, 5

[4] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In ECCV-W, 2022. 1, 5, 7

[5] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou. Transunet: Transformers make strong encoders for medical image segmentation. ICML-W, 2021. 1, 5, 7

[6] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE TPAMI, 40(4):834–848, 2017. 1

[7] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In MICCAI, 2016. 1

[8] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji. Structure-measure: A new way to evaluate foreground maps. In IEEE ICCV, 2017. 4

[9] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, and A. Borji. Enhanced-alignment measure for binary foreground map evaluation. In IJCAI, 2018. 4

[10] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao. Pranet: Parallel reverse attention network for polyp segmentation. In MICCAI, 2020. 1, 3, 4

[11] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr. Res2net: A new multi-scale backbone architecture. IEEE TPAMI, 43(2):652–662, 2021. 4

[12] X. Guo, C. Yang, Y. Liu, and Y. Yuan. Learn to threshold: Thresholdnet with confidence-guided manifold mixup for polyp segmentation. IEEE TMI, 40(4):1134–1146, 2020. 1

[13] B.-C. Hu, G.-P. Ji, D. Shao, and D.-P. Fan. Pranet-v2: Dual-supervised reverse attention for medical image segmentation. arXiv preprint arXiv:XXXX.XXXXX, 2025. 1

[14] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In ICASSP, 2020. 1

[15] X. Huang, Z. Deng, D. Li, X. Yuan, and Y. Fu. Missformer: An effective transformer for 2d medical image segmentation. IEEE TMI, 42(5):1484–1494, 2023. 1, 5, 7

[16] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. nmeth, 18(2):203–211, 2021. 1, 7

[17] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen. Kvasir-seg: A segmented polyp dataset. In MMM, 2020. 4

[18] G.-P. Ji, J. Liu, P. Xu, N. Barnes, F. S. Khan, S. Khan, and D.-P. Fan. Frontiers in intelligent colonoscopy. arXiv preprint arXiv:2410.17241, 2024. 7

[19] T. Kim, H. Lee, and D. Kim. Uacanet: Uncertainty augmented context attention for polyp segmentation. In ACM MM, 2021. 4

[20] B. Landman, Z. Xu, J. E. Iglesias, M. Styner, T. R. Langerak, and A. Klein. Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In MICCAI-W, 2015. 5

[21] H. Li, D. Zhang, J. Yao, L. Han, Z. Li, and J. Han. Asps: Augmented segment anything model for polyp segmentation. In MICCAI, 2024. 1

[22] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert. Attention u-net: Learning where to look for the pancreas. In MIDL, 2018. 5

[23] M. M. Rahman and R. Marculescu. Medical image segmentation via cascaded attention decoding. In IEEE WACV, 2023. 5, 7

[24] M. M. Rahman and R. Marculescu. Multi-scale hierarchical vision transformer with cascaded attention

decoding for medical image segmentation. In MIDL, 2023. 5, 6, 7

[25] M. M. Rahman, M. Munir, and R. Marculescu. Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation. In IEEE CVPR, 2024. 5, 6

[26] M. M. Rahman, S. Shokouhmand, S. Bhatt, and M. Faezipour. Mist: Medical image segmentation transformer with convolutional attention mixing (cam) decoder. In IEEE WACV, 2024. 5, 6, 7

[27] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, 2015. 1, 5

[28] H. Shao, Y. Zhang, and Q. Hou. Polyper: Boundary sensitive polyp segmentation. In AAAI, 2024. 1

[29] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. IJCARS, 9(2):283–293, 2013. 4

[30] M. Sodano, F. Magistri, L. Nunes, J. Behley, and C. Stachniss. Open-world semantic segmentation including class similarity. In IEEE CVPR, 2024. 7

[31] D. Vázquez, J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, A. M. López, A. Romero, M. Drozdzal, and A. Courville. A benchmark for endoluminal scene segmentation of colonoscopy images. JHE, 2017(1):4037190, 2017. 4

[32] H. Wang, S. Xie, L. Lin, Y. Iwamoto, X.-H. Han, Y.-W. Chen, and R. Tong. Mixed transformer u-net for medical image segmentation. In ICASSP, 2022. 1, 5, 7

[33] J. Wang, J. Chen, D. Z. Chen, and J. Wu. Lkm-unet: Large kernel vision mamba unet for medical image segmentation. In MICCAI, 2024. 1

[34] J. Wang, Q. Huang, F. Tang, J. Meng, J. Su, and S. Song. Stepwise feature fusion: Local guides global. In MICCAI, 2022. 5

[35] W. Wang, H. Sun, and X. Wang. Lssnet: A method for colon polyp segmentation based on local feature supplementation and shallow feature supplementation. In MICCAI, 2024. 1

[36] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao. Pvt v2: Improved baselines with pyramid vision transformer. CVMJ, 8(3):415–424, 2022. 4

[37] J. Wu and M. Xu. One-prompt to segment all medical images. In IEEE CVPR, 2024. 1

[38] H. Zhou, J. Guo, Y. Zhang, X. Han, L. Yu, L. Wang, and Y. Yu. nnformer: Volumetric medical image segmentation via a 3d transformer. IEEE TIP, 32:4036–4045, 2023. 7

[39] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE TMI, 39(6):1856–1867, 2019. 1

[40] W. Zhu, X. Chen, P. Qiu, M. Farazi, A. Sotiras, A. Razi, and Y. Wang. Selfreg-unet: Self-regularized unet for medical image segmentation. In MICCAI, 2024. 1