

PraNet: 基于并行反向注意力的息肉分割网络

范登平¹, 季葛鹏², 周涛¹, 陈耿¹, 付华柱¹ ✉, 沈建冰¹ ✉, and 邵岭^{3,1}

¹ 阿联酋起源人工智能研究院, 阿联酋阿布扎比

² 武汉大学计算机学院, 中国湖北

³ Mohamed bin Zayed 人工智能大学, 阿联酋阿布扎比

{huazhu.fu, jianbing.shen}@inceptioniai.org

<https://github.com/DengPingFan/PraNet>

Abstract. 结肠镜检查是一种有效的结直肠息肉检测技术, 而结直肠息肉与结直肠癌密切相关。在临床实践中, 从结肠镜图像中分割息肉具有重要意义, 因为它为诊断和手术提供了非常有价值的信息。然而, 息肉的精确分割是一项具有挑战性的任务, 主要有两个原因: (i) 同一类型的息肉具有不同的大小、颜色和纹理; (ii) 息肉与其周围粘膜之间的边界不清晰。为了解决这些问题, 我们提出了一种基于并行反向注意力的网络 (名为 *PraNet*) 来精确分割结肠镜图像中的息肉。具体来说, 我们首先使用并行的部分解码器 (PPD) 在高级层中聚合特征。基于这个组合特征, 我们生成一个全局映射图作为以下组件的初始引导区域。此外, 我们使用反向注意模块 (RA) 挖掘边界线索, 该模块能够建立区域和边界线索之间的关系。由于区域和边界之间的迭代交互机制, 我们的 *PraNet*¹ 能够校正预测结果中相互矛盾的区域, 从而提高分割精度。在五个具有挑战性的数据集, 六个指标上定量和定性的实验结果表明, 本文的 *PraNet* 显著地提高了分割精度, 并且在泛化性能和分割实时性 (~50fps) 方面具有明显优势。

Keywords: 结肠镜检查 · 息肉分割 · 结直肠癌

1 介绍

结直肠癌 (CRC) 是全球第三大最常见的癌症 [24]。因此, 通过筛查和切除癌前病变 (结直肠腺瘤) 来预防结直肠癌至关重要, 并已成为世界范围内公共卫生的重点。结肠镜检查可以提供结直肠息肉的位置和外观信息, 使医生能够在结直肠息肉发展为结直肠癌之前将其切除, 是一种有效的结直肠癌筛查和预防技术。多项研究表明, 早期结肠镜检查使结直肠癌的发病率下降了 30% [15]。因此, 在临床上, 精确的息肉分割是非常重要的。然而, 这是一项具有挑战性的任务, 有两个主要原因。首先, 息肉的外观通常不同, 例如大小、颜色和质地, 即使它们是同一类型的。其次, 在结肠镜图像中, 息肉与其周围黏膜之间的边界通常是模糊的, 并且缺乏分割方法所需的强烈对比度。这些问题导致了息肉区

¹ 本文为 MICCAI20 [8] 的中文翻译版。

域的不精确分割,有时甚至会导致息肉的漏检。因此,一种能够在早期发现所有潜在息肉的自动而准确的息肉分割方法对预防结肠直肠癌具有重要意义 [18]。

在息肉分割方法中,早期基于学习的方法依赖于人工提取的特征 [19,25],如:颜色、纹理、形状、外观或者这些特征的组合。这些方法通常通过训练分类器来区分息肉和周围组织。然而,这些模型往往存在较高的漏检率。其主要原因是,在处理息肉区域与高度相似区域之间的强类内差异性和弱类间差异性时,人工提取特征的特征能力相当有限 [31]。近年来,许多基于深度学习的息肉分割方法被开发出来 [31,33]。虽然这些方法已经取得了一定的进展,但它们只使用边界框来标注检测的息肉,所以无法准确定位息肉的边界。为了解决这个问题,Brandao 等 [3] 使用了基于预训练模型的 FCN 来识别和分割息肉。Akbari 等 [1] 使用了一种改进的 FCN 来提高息肉分割的准确性。受 U-Net [23] 在生物医学图像分割中成功应用的启发,U-Net++ [40] 和 ResUNet++ [17] 被应用在息肉分割任务中并取得了良好的效果。这些方法侧重于对息肉的整个区域进行分割,但忽略了区域边界约束,这对提高分割性能至关重要。为此,Psi-Net [20] 在息肉分割中同时利用了区域和边界信息,但并没有完全捕捉到区域和边界之间的关系。此外,Fang 等 [11] 还提出了一种同时具有区域和边界约束的三步选择性特征聚集网络,用来分割息肉。该方法显示地考虑了区域和边界之间的依赖性,并通过附加的边缘监督获得了良好的结果;但是,它非常耗时 (>20 小时),并且有过拟合的风险。

在这篇论文中,我们针对息肉分割任务提出一种全新的深度神经网络,称为并行反向注意力网络 (*PraNet*)。我们的动机源于这样一个事实:在息肉标注过程中,临床医生首先粗略地定位息肉,然后根据局部特征准确地勾勒出轮廓。因此,我们认为区域和边界是区别正常组织和息肉的两个关键特征。与 [11] 不同的是,我们先预测粗糙区域,然后用反向注意力的方法对边界进行隐式建模。该策略有三个优点:学习能力强,泛化能力强,训练效率高。更多细节见本文的实验部分 (§ 3)。简而言之,本文的贡献体现在三方面。(1) 我们提出了一种新的深度神经网络用于实时精确的息肉分割。通过使用并行部分解码器 (PPD) 在高层聚合特征,组合特征获取上下文信息并生成全局映射图作为后续步骤的初始**引导区域**。为了进一步挖掘**边界线索**,我们反复利用反向注意力 (RA) 模块来建立区域和边界线索之间的关系。由于区域和边界之间的这种迭代的交互机制,我们的模型能够校正预测结果中一些不一致的区域。(2) 我们介绍了几种新的息肉分割评估指标,并基于现有的公开可用的先进模型上提出了一个全面的基准测试。(3) 大量实验表明,在五个具有挑战性的数据集上,本文的 *PraNet* 优于大多数先进模型并大幅度提高性能,可以实时预测结果且训练时间较短。

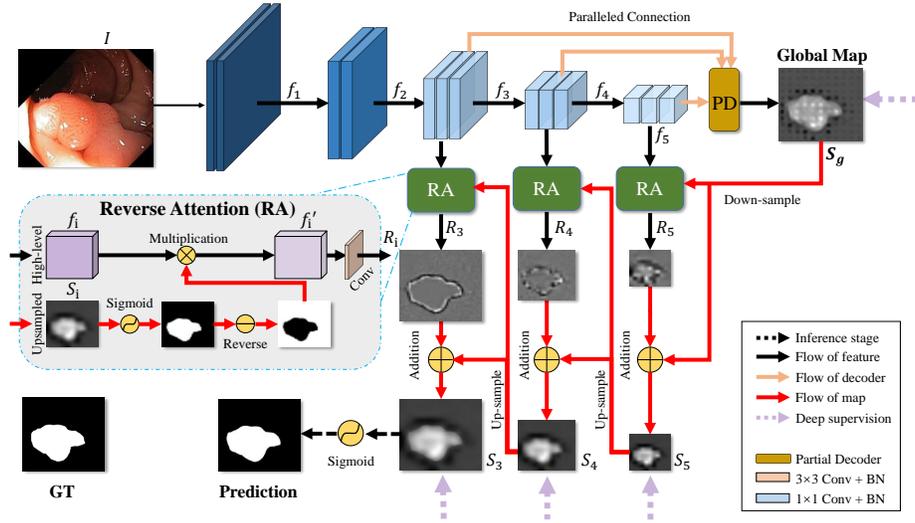


Fig. 1: *PraNet* 模型的概要。它由三个反向注意力模块和一个并行的部分解码器连接组成。详见 § 2。

2 方法

图. 1展示了我们的模型，它使用一个并行的部分解码器来生成高级语义全局映射图和一组反向注意力模块来从结肠镜图像中精确分割息肉。每个组成部分将详述如下。

2.1 基于并行的部分解码器的特征聚合

目前流行的医学图像分割网络通常依赖于 U-Net [23] 或类似 U-Net 的网络 (例如: U-Net++ [40], ResUNet [36] 等)。这些模型本质上是编解码框架, 通常将从卷积神经网络中提取的所有多级特征集合起来。如 Wu 等 [30] 指出, 与高级特征相比, 低级特征由于其更大的空间分辨率而需要更多的计算资源, 但对性能的贡献较小。受此启发, 我们提出用一个并行的部分解码器组件来聚合高级特征。更具体地说, 将一个尺寸为 $h \times w$ 的息肉图像 I 作为输入, 可以从基于 Res2Net[13] 的主干网中提取分辨率为 $[h/2^{k-1}, w/2^{k-1}]$ 的五个层次的特征 $\{f_i, i = 1, \dots, 5\}$ 。然后, 我们将 f_i 特征分为低级特征 $\{f_i, i = 1, 2\}$ 和高级特征 $\{f_i, i = 3, 4, 5\}$ 。我们引入了一个全新的先进译码器组件——部分解码器 $pd(\cdot)$ [30], 通过并行连接来聚合高级特征。部分解码器特征由 $\mathbf{PD} = pd(f_3, f_4, f_5)$ 来计算, 从而得到全局映射图 S_g 。

2.2 反向注意力模块

在临床上，医生首先大致定位息肉区域，然后仔细检查局部组织，以准确标记息肉。如 § 2.1 所述，我们的全局映射图 S_g 来自最深的卷积神经网络层，它只能捕捉息肉组织的相对粗略位置，而没有结构细节（见图. 1）。为了解决这个问题，我们提出了一种通过擦除前景对象的方式逐步挖掘有区别的息肉区域的策略 [28,4]。与文献 [4,14,37,34] 不同，我们提出可以在三个平行的高级特征中自适应地学习的**反向注意力**机制。换言之，我们的体系结构可以通过从高级侧输出特征中擦除现有的已估计的息肉区域，从而顺序地挖掘互补区域和细节，其中现有的估计是从更深层向上采样的。

具体地说，我们通过将高级侧输出的特征 $\{f_i, i = 3, 4, 5\}$ 乘以反向注意力权重 A_i 来获得输出的反向注意力特征 R_i ，如下所示：

$$R_i = f_i \odot A_i, \quad (1)$$

反向注意力权重 A_i 已广泛应用在计算机视觉领域中的显著目标检测任务 [4,35]，它被描述为：

$$A_i = \ominus(\sigma(\mathcal{P}(S_{i+1}))), \quad (2)$$

其中 $\mathcal{P}(\cdot)$ 表示上采样操作， $\sigma(\cdot)$ 是 Sigmoid 函数， $\ominus(\cdot)$ 是从全 1 矩阵 \mathbf{E} 中减去输入的反向运算符。图. 1 (RA) 展示出了该过程的细节。值得注意的是，由反向注意力驱动擦除机制最终可以将不精确和粗略的估计细化为准确而完整的边缘预测图。

2.3 过程学习和实现细节

损失函数。我们的损失函数定义为 $\mathcal{L} = \mathcal{L}_{IoU}^w + \mathcal{L}_{BCE}^w$ ，其中， \mathcal{L}_{IoU}^w 和 \mathcal{L}_{BCE}^w 分别表示基于全局约束和局部（像素级）约束的加权 IoU 损失和二进制交叉熵（BCE）损失。与目前广泛应用于分割任务中的标准 IoU 损失不同，加权 IoU 损失通过增加难样本像素的权重来突出其重要性。另外，与标准的 BCE 损失函数相比， \mathcal{L}_{BCE}^w 更注重难样本像素，而不是对所有像素进行等权重分配。这些损失函数的定义与 [22,27] 中相同，其有效性已在显著性目标检测领域得到验证。在这里，我们对三个输出（即： S_3 , S_4 , 和 S_5 ）和全局映射图 S_g 采用深监督策略。映射图都被逐一上采样（例如 S_3^{up} ）到与真值图 G 相同的大小。从而提出的 *PraNet* 的整体损失函数可表示为： $\mathcal{L}_{total} = \mathcal{L}(G, S_g^{up}) + \sum_{i=3}^{i=5} \mathcal{L}(G, S_i^{up})$ 。

实施细节。我们在 PyTorch 框架中实现了该模型，它由 NVIDIA TITAN RTX GPU 加速。所有输入统一调整为 352×352 ，采用多尺度训练策略 $\{0.75, 1, 1.25\}$ 替代数据增强。使用 Adam 优化算法来优化整体参数，学习率设定为 $1e-4$ 。整个网络以端到端的方式进行训练，批大小设定为 16 个，训练 20 个周期仅用 32 分钟即可达到收敛状态。最终预测图 S_p 则由 S_3 经 sigmoid 运算后生成。

3 实验

3.1 息肉分割实验

在本节中，我们将本文提出的模型与现有模型在学习能力、泛化能力、复杂性和定性结果方面进行比较。

数据集和对比模型。实验在五个息肉分割数据集上进行：ETIS [24]，CVC-ClinicDB/CVC-612 [2]、CVC-ColonDB [25]、EndoScene [26] 和 Kvasir [16]。前四个是标准的评测数据集，最后一个为近期公开的最大规模的具有挑战性数据集。我们将本文模型与四种先进的医学图像分割方法进行比较：U-Net [23]，U-Net++ [40]，ResUNet-mod [36] 和 ResUNet++ [17]。此外，我们也对比了最新的息肉分割模型，即：SFA [11]。SFA 模型的分割结果由公开的代码，采用原始默认设置得到。

训练设置和评价指标。除非另有说明，我们遵循与 [17] 中相同的训练设置，即来自 Kvasir 和 CVC-ClinicDB 数据集的图像被随机地分成三部分，其中 80% 用于训练，10% 用于验证，10% 用于测试。类似于 [17,16]，我们使用两个指标（即平均 Dice 和平均 IoU）进行定量评估。为了更深入地了解模型性能，我们进一步引入了在目标检测领域 [7,32,9,12,38,39] 中广泛使用的其他四种度量方法。加权 Dice 度量 F_β^w 用于修正 Dice 中的“同等重要的缺陷”问题。MAE 度量用来评估像素级的精度。为了评估像素级和全局级的相似性，我们采用了最近发布的增强对齐度量 E_ϕ^{max} [6]。由于 F_β^w 和 MAE 是基于像素级别的评价度量体系，忽略了目标结构的相似性，所以我们采用 S_α [5] 来评估预测与真值图之间的相似性。本文使用的评估工具箱位于 <https://github.com/DengPingFan/PraNet>。

学习能力。在这一部分中，我们进行了两个实验来验证我们的模型在两个见过的数据集上的学习能力，即 Kvasir 和 CVC-612。*Kvasir* 是最近发布的一个具有挑战性的数据集，它包含从 Kvasir 数据集 [21] 的子类（息肉类）中选择的 1000

Table 1: Kvasir [16] 和 CVC-612 [2] 数据集的定量结果。‘n/a’ 表示结果不可获取。‘†’ 表示来摘自 [17] 的评估分数。

Methods		mean Dice	mean IoU	F_{β}^w	S_{α}	E_{ϕ}^{max}	MAE
Kvasir	U-Net (MICCAI'15) [23]	0.818	0.746	0.794	0.858	0.893	0.055
	U-Net++ (TMI'19) [40]	0.821	0.743	0.808	0.862	0.910	0.048
	ResUNet-mod [†] [36]	0.791	n/a	n/a	n/a	n/a	n/a
	ResUNet++ [†] [17]	0.813	0.793	n/a	n/a	n/a	n/a
	SFA (MICCAI'19) [11]	0.723	0.611	0.670	0.782	0.849	0.075
	<i>PraNet (Ours)</i>	0.898	0.840	0.885	0.915	0.948	0.030
CVC-612	U-Net (MICCAI'15) [23]	0.823	0.755	0.811	0.889	0.954	0.019
	U-Net++ (TMI'19) [40]	0.794	0.729	0.785	0.873	0.931	0.022
	ResUNet-mod [†] [36]	0.779	n/a	n/a	n/a	n/a	n/a
	ResUNet++ [†] [17]	0.796	0.796	n/a	n/a	n/a	n/a
	SFA (MICCAI'19) [11]	0.700	0.607	0.647	0.793	0.885	0.042
	<i>PraNet (Ours)</i>	0.899	0.849	0.896	0.936	0.979	0.009

个图像。CVC-ClinicDB, 也称为 CVC-612, 包括从 31 张结肠镜图像中剪裁出的 612 张公开图像。如表. 1 所示, 我们的 *PraNet* 在两个数据集中的所有指标上都远远优于所有的先进模型 (平均 Dice: 大约 > 7%)。这表明我们的模型具有很强的学习能力, 能够有效地分割息肉。

泛化能力。我们进行了三个实验来检验模型的泛化能力。这三个没见过的数据集有它们自己具有挑战性的情况和性质。CVC-ColonDB 是一个小规模数据集, 包含来自 15 个短结肠镜视频序列的 380 张图像。所有图像都作为我们的测试集数据。ETIS 是一个早期构建的数据集, 包含 196 个息肉图像, 用于结直肠癌的早期诊断。Endosece 是 CVC-612 和 CVC300 的组合。我们根据 Fang 等人 [11] 的方式将其分为训练、验证和测试子集。本实验只使用 Endosece-CVC300 测试集, 因为 CVC-612 数据集的一部分可能被用于训练。*PraNet* 再次优于现有的经典医学分割基线 (即: U-Net、U-Net++ 以及 SFA)。*PraNet* 在三个没见过的数据集中, 都有显著的改进 (见表. 2)。一个值得注意的发现是, SFA 在这些没见过的数据集上评估指标数值急剧下降, 侧面证明了模型的泛化能力较差。

定性结果。在图. 2 中, 我们给出了 *PraNet* 在 Kvasir 测试集上的息肉分割结果。我们的模型可以在许多挑战性场景下精确定位和分割息肉组织, 例如: 不同大小、均匀区域、不同质地等。

训练和推理分析。在表. 3 中, 我们给出了 *PraNet* 和现有先进方法的训练时间和推理时间。所有用于比较的模型的运行时间是在 Intel i9-9820X CPU 和单张 24 GB 内存的 TITAN RTX 显卡上测试的。如图所示, 我们的模型只需要 20 个周

Table 2: CVC-ColonDB [25]、ETIS [24] 和 EndoScene [26] 的测试集 (CVC-T) 的定量结果。SFA 结果是使用发布的代码生成的。

Methods		mean Dice	mean IoU	F_{β}^w	S_{α}	E_{ϕ}^{max}	MAE
ColonDB	U-Net(MICCAI'15) [23]	0.512	0.444	0.498	0.712	0.776	0.061
	U-Net++(TMI'19) [40]	0.483	0.410	0.467	0.691	0.760	0.064
	SFA (MICCAI'19) [11]	0.469	0.347	0.379	0.634	0.765	0.094
	PraNet (Ours)	0.709	0.640	0.696	0.819	0.869	0.045
ETIS	U-Net (MICCAI'15) [23]	0.398	0.335	0.366	0.684	0.740	0.036
	U-Net++ (TMI'19) [40]	0.401	0.344	0.390	0.683	0.776	0.035
	SFA (MICCAI'19) [11]	0.297	0.217	0.231	0.557	0.633	0.109
	PraNet (Ours)	0.628	0.567	0.600	0.794	0.841	0.031
CVC-T	U-Net (MICCAI'15) [23]	0.710	0.627	0.684	0.843	0.876	0.022
	U-Net++ (TMI'19) [40]	0.707	0.624	0.687	0.839	0.898	0.018
	SFA (MICCAI'19) [11]	0.467	0.329	0.341	0.640	0.817	0.065
	PraNet (Ours)	0.871	0.797	0.843	0.925	0.972	0.010

Table 3: 基于 CVC-ClinicDB [2] 数据集的训练和推理分析 (同一平台)。我们记录了模型收敛的迭代数 #epochs。Lr= 学习率。

Methods		Epoch	Lr	Training	Inference	mean Dice
CVC-612	U-Net (MICCAI'15) [23]	30	3e-4	~40 minutes	~8fps	0.823
	U-Net++ (TMI'19) [40]	30	3e-4	~45 minutes	~7fps	0.794
	SFA (MICCAI'19) [11]	500	1e-2	>20 hours	~40fps	0.700
	PraNet (Ours)	20	1e-4	~30 minutes	~50fps	0.899

期 (~0.5 小时) 的训练就可以实现收敛。一个原因是我们的 *PraNet* 的并行结构提供了一种短连接方式将损耗反向传播到解码器路径中的早期层 (图. 1 中的红色数据流)。另外, 侧输出还解决了梯度消失问题, 指导了早期的分层训练。注意, 对于 352×352 输入, 我们的模型以 ~50fps 的实时速度运行, 这保证了我们的方法可以应对结肠镜视频中的息肉分割任务。

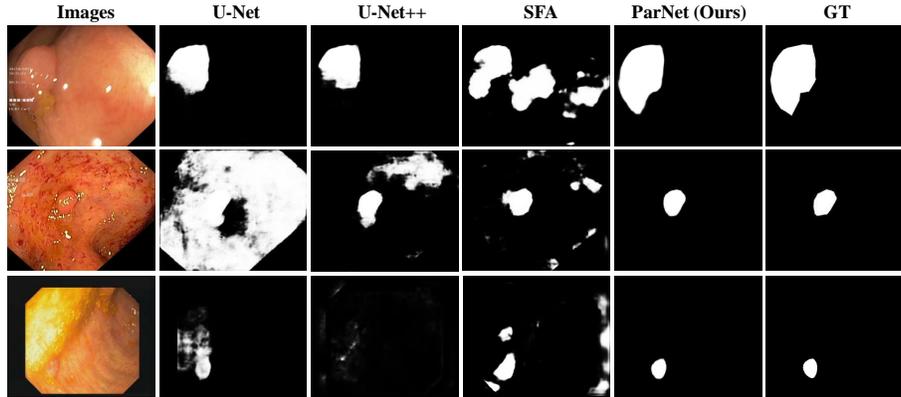
3.2 消融实验

在本节中, 我们将在见过和没见过的数据集上测试模型的每个组件, 以深入了解我们的模型。

PPD 的有效性。我们研究了级联机制 (并行部分解码器, PPD) 的重要性。我们从表. 4 观察到 No.2 (主干 + PPD) 优于 No.1 (主干), 清楚地表明级联机制

Table 4: *PraNet* 在 CVC-612 和 CVC300 数据集上的消融实验

Settings	CVC-612 (<i>seen</i>)			CVC300 (<i>unseen</i>)		
	mean Dice	mean IoU	S_α	mean Dice	mean IoU	S_α
Backbone (No.1)	0.747	0.668	0.735	0.726	0.631	0.670
PPD + Backbone (No.2)	0.865	0.798	0.902	0.824	0.734	0.893
RA + Backbone (No.3)	0.888	0.845	0.912	0.871	0.800	0.888
PPD + RA + Backbone (No.4)	0.899	0.849	0.936	0.871	0.797	0.925

**Fig. 2:** 不同方法的定性结果。

对于提高性能是必要的。注意，我们的 PPD 只部署在高级特征上，这大大减少了模型的训练时间 (见表. 3, $Inference = \sim 50\text{fps}$)。

RA 的有效性。我们进一步探讨反向注意力的贡献。结果列在表. 4 的第一列和第三列。我们观察到，No.3 改善了 CVC-612 的主干 (No.1) 性能，将平均 Dice 从 0.747 增加到 0.888，并且结构测量值 S_α 从 0.735 增加到 0.912。这些改进表明，引入反向注意力机制可以使我们的模型能够准确区分真正的息肉组织。

PPD & RA 的有效性。为了评估 PPD 和 RA 模块的组合，我们测试了 No.4 (PPD+RA+ 主干) 的性能。如表. 4 所示，本文的模型 (No.4) 通常优于其他设置 (No.1~No.3)。此外，*PraNet* 在所有测试的数据集上以显著的改进 (>5%) 优于四个先进模型，成为了一个健壮、统一的体系结构，有助于促进息肉分割的未来研究。

4 总结

我们提出了一种新的结构，命名为 *PraNet*，用于从结肠镜图像中自动分割息肉。大量实验表明，在五个具有挑战性的数据集中，我们的模型始终比所有最先进的方法表现出更大的优势 (>5%)。此外，我们的模型实现了非常高的精度 (Kvasir 数据集上的平均 Dice=0.898)，无需任何预/后处理。另一个优点是 *PraNet* 具有通用性和灵活性，这意味着可以添加更有效的模块来进一步提高精度。与目前排名靠前的 SFA 模型相比，*PraNet* 具有较强的学习能力、泛化能力和实时分割效率。我们希望这项研究能为社区提供一个机会，在相关课题上探索更强大的模型，如肺部感染的分割 [10]/分类 [29]，甚至上游任务等。

References

1. Akbari, M., Mohrekeesh, M., Nasr-Esfahani, E., Soroushmehr, S.R., Karimi, N., Samavi, S., Najarian, K.: Polyp segmentation in colonoscopy images using fully convolutional network. In: IEEE EMBC. pp. 69–72 (2018)
2. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. CMIG 43, 99–111 (2015)
3. Brandao, P., Mazomenos, E., Ciuti, G., Calì, R., Bianchi, F., Menciassi, A., Dario, P., Koulaouzidis, A., Arezzo, A., Stoyanov, D.: Fully convolutional neural networks for polyp segmentation in colonoscopy. In: Medical Imaging 2017: Computer-Aided Diagnosis. vol. 10134, p. 101340F (2017)
4. Chen, S., Tan, X., Wang, B., Hu, X.: Reverse attention for salient object detection. In: ECCV. pp. 234–250 (2018)
5. Fan, D.P., Cheng, M.M., Liu, Y., Li, T., Borji, A.: Structure-measure: A new way to evaluate foreground maps. In: IEEE ICCV. pp. 4548–4557 (2017)
6. Fan, D.P., Gong, C., Cao, Y., Ren, B., Cheng, M.M., Borji, A.: Enhanced-alignment Measure for Binary Foreground Map Evaluation. In: IJCAI (2018)
7. Fan, D.P., Ji, G.P., Sun, G., Cheng, M.M., Shen, J., Shao, L.: Camouflaged object detection. In: IEEE CVPR (2020)
8. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranet: Parallel reverse attention network for polyp segmentation. In: MICCAI (2020)
9. Fan, D.P., Liu, J.J., Gao, S.H., Hou, Q., Borji, A., Cheng, M.M.: Salient objects in clutter: Bringing salient object detection to the foreground. In: ECCV. pp. 1597–1604. Springer (2018)
10. Fan, D.P., Zhou, T., Ji, G.P., Zhou, Y., Chen, G., Fu, H., Shen, J., Shao, L.: InfNet: Automatic COVID-19 Lung Infection Segmentation from CT Images. IEEE TMI (2020)
11. Fang, Y., Chen, C., Yuan, Y., Tong, K.y.: Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: MICCAI. pp. 302–310. Springer (2019)
12. Fu, K., Fan, D.P., Ji, G.P., Zhao, Q.: JI-dcf: Joint learning and densely-cooperative fusion framework for rgb-d salient object detection. In: IEEE CVPR. pp. 3052–3062 (2020)

13. Gao, S.H., Cheng, M.M., Zhao, K., Zhang, X.Y., Yang, M.H., Torr, P.: Res2net: A new multi-scale backbone architecture. *IEEE TPAMI* pp. 1–1 (2020)
14. Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J.: CE-Net: Context encoder network for 2d medical image segmentation. *IEEE TMI* 38(10), 2281–2292 (2019)
15. Haggard, F.A., Boushey, R.P.: Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clinics in colon and rectal surgery* 22(04), 191–197 (2009)
16. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., de Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: *MMM*. pp. 451–462. Springer (2020)
17. Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., De Lange, T., Halvorsen, P., Johansen, H.D.: Resunet++: An advanced architecture for medical image segmentation. In: *IEEE ISM*. pp. 225–2255 (2019)
18. Jia, X., Xing, X., Yuan, Y., Xing, L., Meng, M.Q.H.: Wireless capsule endoscopy: A new tool for cancer screening in the colon with deep-learning-based polyp recognition. *Proceedings of the IEEE* 108(1), 178–197 (2019)
19. Mamonov, A.V., Figueiredo, I.N., Figueiredo, P.N., Tsai, Y.H.R.: Automated polyp detection in colon capsule endoscopy. *IEEE TMI* 33(7), 1488–1502 (2014)
20. Murugesan, B., Sarveswaran, K., Shankaranarayana, S.M., Ram, K., Joseph, J., Sivaprakasam, M.: Psi-Net: Shape and boundary aware joint multi-task deep network for medical image segmentation. In: *IEEE EMBC*. pp. 7223–7226 (2019)
21. Pogorelov, K., Randel, K.R., Griwodz, C., Eskeland, S.L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.T., Lux, M., Schmidt, P.T., et al.: Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In: *ACM MSC*. pp. 164–169 (2017)
22. Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M.: Basnet: Boundary-aware salient object detection. In: *IEEE CVPR*. pp. 7479–7489 (2019)
23. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: *MICCAI*. pp. 234–241. Springer (2015)
24. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery* 9(2), 283–293 (2014)
25. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. *IEEE TMI* 35(2), 630–644 (2015)
26. Vázquez, D., Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., López, A.M., Romero, A., Drozdal, M., Courville, A.: A benchmark for endoluminal scene segmentation of colonoscopy images. *Journal of Healthcare Engineering* 2017 (2017)
27. Wei, J., Wang, S., Huang, Q.: F3Net: Fusion, Feedback and Focus for Salient Object Detection. In: *AAAI* (2020)
28. Wei, Y., Feng, J., Liang, X., Cheng, M.M., Zhao, Y., Yan, S.: Object region mining with adversarial erasing: A simple classification to semantic segmentation approach. In: *IEEE CVPR*. pp. 1568–1576 (2017)
29. Wu, Y.H., Gao, S.H., Mei, J., Xu, J., Fan, D.P., Zhao, C.W., Cheng, M.M.: JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation. *arXiv preprint arXiv:2004.07054* (2020)
30. Wu, Z., Su, L., Huang, Q.: Cascaded partial decoder for fast and accurate salient object detection. In: *IEEE CVPR*. pp. 3907–3916 (2019)
31. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A.: Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos. *IEEE JBHI* 21(1), 65–75 (2016)

32. Zhang, J., Fan, D.P., Dai, Y., Anwar, S., Sadat Saleh, F., Zhang, T., Barnes, N.: UC-Net: Uncertainty Inspired RGB-D Saliency Detection via Conditional Variational Autoencoders. In: IEEE CVPR (2020)
33. Zhang, R., Zheng, Y., Poon, C.C., Shen, D., Lau, J.Y.: Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker. *Pattern Recognition* 83, 209–219 (2018)
34. Zhang, S., Fu, H., Yan, Y., Zhang, Y., Wu, Q., Yang, M., Tan, M., Xu, Y.: Attention Guided Network for Retinal Image Segmentation. In: MICCAI, pp. 797–805 (2019)
35. Zhang, Z., Lin, Z., Xu, J., Jin, W., Lu, S.P., Fan, D.P.: Bilateral attention network for rgb-d salient object detection. *arXiv preprint arXiv:2004.14582* (2020)
36. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters* 15(5), 749–753 (2018)
37. Zhang, Z., Fu, H., Dai, H., Shen, J., Pang, Y., Shao, L.: ET-Net: A generic edge-attention guidance network for medical image segmentation. In: MICCAI. pp. 442–450. Springer (2019)
38. Zhao, J.X., Cao, Y., Fan, D.P., Cheng, M.M., Li, X.Y., Zhang, L.: Contrast prior and fluid pyramid integration for rgb-d salient object detection. In: IEEE CVPR. pp. 3927–3936 (2019)
39. Zhao, J.X., Liu, J.J., Fan, D.P., Cao, Y., Yang, J., Cheng, M.M.: EGNNet: Edge guidance network for salient object detection. In: IEEE ICCV. pp. 8779–8788 (2019)
40. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. *IEEE TMI* pp. 3–11 (2019)