

针对皮肤病变分割任务的改进 U-Net

陈睿

统计与管理学院

上海财经大学

摘要: 医学图像分割一直以来都是被深入研究的课题，但由于医学图像样本较少、特征提取困难的特点，这一任务很有挑战。随着 Attention 机制的提出，这为常用与图像处理领域的卷积神经网络提供了更多的可能。本文针对皮肤病变图像分割问题，基于常用的卷积网络结构 Unet，综合使用注意力门控和循环卷积残差模块来改进网络结构，并进一步通过定义新的损失函数以及双阈值边界确定方式，使得模型能在一方面减少参数的同时，获得更强的表征能力。实验证明我们的改进取得了良好的效果，并提供向更多网络结构扩展的可能。

关键词: Unet 改进 循环卷积 注意力机制 图像分割

1 课题介绍

在图像处理领域，医学图像分割一直以来都是被深入研究的课题。由于对医学图像打标签要求极高的专业知识，以及对病例隐私的保护，可供训练的样本相对其他任务较少，再加上医学图像（如 CT 图像、核磁共振图像等）的独特特征，这一任务很有挑战。精确可靠的解决方案可以快速有效地提供量化的结果来支持临床诊断效率，并支持诊断决策。

皮肤癌是最常见的癌症之一，全世界每年有超过 140 万新增病例，超过其他所有癌症的总和。黑色素瘤是最为致命的皮肤癌，在过去的十年中，每年诊断出的新发黑色素瘤病例增加了 47%(Bray et al., 2018)。尽管绝大多数皮肤癌是有患者自己首先发现的 (Brady et al., 2000)，在没有仪器辅助的情况下，专业的医生通过肉眼的诊断准确率仅有 60%(Kittler et al., 2002)，对与可能导致黑色素瘤的早期皮肤病变来说，诊断的准确率将更低。因此，针对皮肤病变的图像分割一直是生物医学图像领域的热点方向之一。然而，这项任务并不简单。因为大小、形状、颜色以及皮肤类型上的差异，病变的外观表现差异很大。于此同时，病变不规则、模糊的边界以及固有的皮肤特征（血管、毛发等）使得模型学习更具挑战。

随着卷积神经网络 (CNNs) 的出现，深度学习模型在包括心血管磁共振成像 (Bai et al., 2017) 和肺部 CT 结核检测任务 (Liao et al., 2019) 中，达到了接近，甚至超过人类专家的表

现。目前来说，U-Net 是一种较为常用的网络结构(图1)，尤其在针对医学图像问题中提取小特征的能力很好。尽管这种结构具有良好的提取特征的能力，但是当样本的差异较大时，模型的效果往往不够理想。目前有一种解决这种的思路是 Cascade R-CNN(Cai et al., 2018)，这种框架的结构如图2所示。但是这种结构会造成模型参数的冗余，例如相似的低级别特征在每一个级联内部的子模型都会被重复地提取，从而大大增加计算需求。

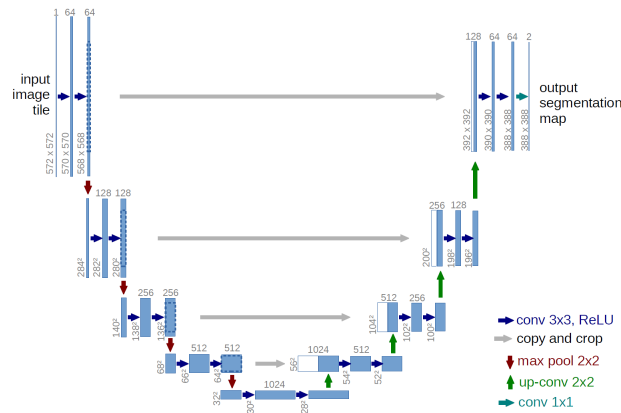


图 1: U-Net 结构

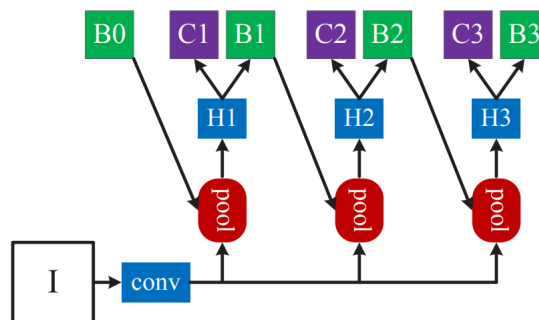


图 2: Cascade 框架

在本文中,我们将针对皮肤病变分割问题,基于 U-Net 结构,综合使用注意力机制(Jetley et al., 2018)、循环卷积残差单元 (Alom et al., 2018) 来改进网络结构,并进一步通过定义新的损失函数以及双阈值边界确定方式,使得模型能在一方面减少参数的同时,获得更强的表征能力。在实验阶段,我们将用这种模型与常规的 U-Net 以及 ISIC2018 挑战的排名第一的方案进行比较。

2 数据

本文的数据来自于 *ISIC Challenge 2018 - Skin Lesion Analysis Towards Melanoma Detection*。该数据集共包括 2594 个皮肤镜图片样本，如图3所示，每一个样本包括 jpg 格式的原始图片以及对应的 png 格式的病变皮肤掩码。图片的尺寸和大小是不统一的。数据集包括的病变类型包括痣、脂溢性角化以及恶性黑色素瘤。任务的目标是针对各种背景下的皮肤病变生成精确的 0-1 掩码。我们将数据集按照 7:1:2 的比例划分训练集、验证集和测试集，训练集有 1815 个样本，验证集有 259 个样本，测试集有 520 个样本。

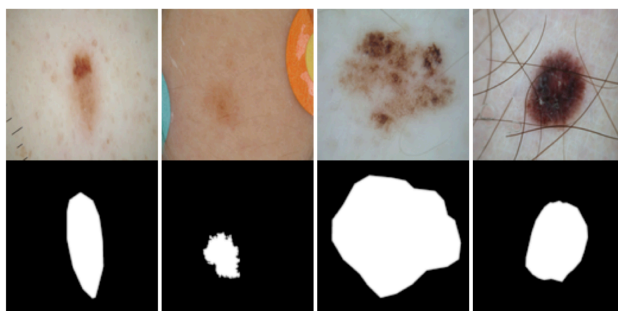


图 3: ISBI2017 数据集。第一行：原始图片，第二行：分割掩码

3 模型与方法

3.1 网络结构

我们设计了一种改进后的 U-Net 模型来输出每一个像素点的病变概率。我们的模型结构如图4所示。在 U-Net 的基础上，我们将每一个卷积模块改进成循环卷积残差模块，在上卷积部分中，我们用通过 Attention Gate 单元处理后的结果取代了经典 U-Net 模型的简单拼接。为了减小模型的复杂程度，我们将 U-Net 整体的下采样的次数减少了一次，这将大大减少模型的参数个数以及占用的内存，但是实验结果表明改进后的结构相较于层数更多的常规 U-Net 有更好的效果。接下来我们将具体介绍网络中用到的两种改进方法。

3.1.1 循环卷积残差模块

Alom et al. (2018) 最早提出了循环卷积残差模块，模块的结构如图5所示。相比于常规的卷积层，这种结构采用了循环卷积层，图5中展示了当时时间步 t 为 2 时的循环卷积操作，包括一个单卷积层，后接两个连续的循环卷积层，每做一次卷积都会加一层 Batch Normalization 和 ReLu。除此之外，这一模块还具有残差联通性，模块的输入会与循环卷积层的输出直接

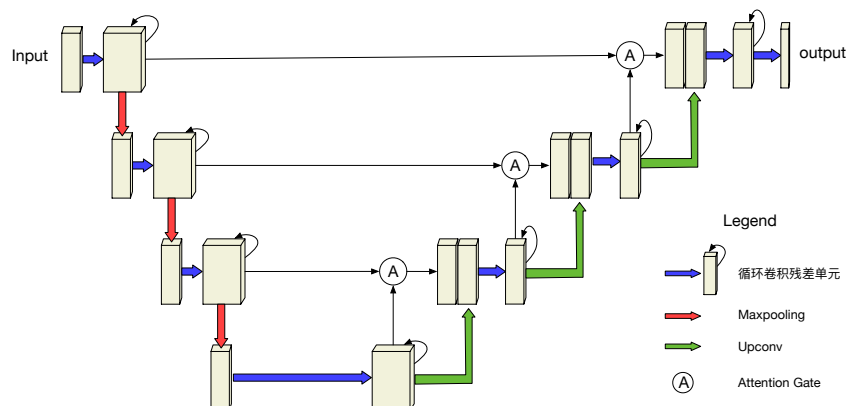


图 4: 改进后的 U-Net

连接。因此这一模块集合了卷积网络、残差网络、循环网络三者的特点和优势。用循环卷积残差模块代替编解码部分的常规前向卷积层有助于生成更有效的、更深入的模型。另外，循环结构采用不同时间步的特征积累方式，使得模型能够得到更强的特征表示。因此，它有助于提取非常低层次的特征，这些特征对于医学图像的分割任务是必不可少的。与传统的 U-Net 相比，该体系结构能有效地提高网络参数数量方面的效率，循环和残差操作并不会增加网络参数的数量，但是却能有效地增加网络的深度。

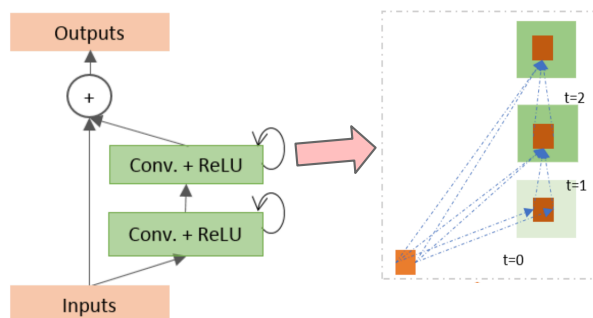


图 5: 循环卷积残差模块

3.1.2 注意力门控

Attention 机制自从最早在自然语言处理领域被提出以来，展现了非常良好的学习能力。在医疗图像中，注意力机制能提取对特定任务有用的显著特征，抑制输入图像中的不相关信息。相比于级联网络结构 (cascade network)，使用注意力机制不需要训练多个模型和大量额外参数；另外 Attention 在不相关区域抑制特征响应的机制也让这种结构不需要在网络直接裁剪出 Region of Interest。如图4中所示，解码侧的特征图上卷积后与编码侧的特征图一同送入注意力门控。注意力门控的模块结构如图6所示，解码侧上卷积后的特征图作为 g ,

编码侧的特征图作为 x^l 。具体流程如下。

- Step 1:** 令 g 通过 1×1 卷积 W_g
- Step 2:** 令 x^l 通过 1×1 卷积 W_x
- Step 3:** 两者输出逐点相加，再做 $ReLU(\sigma_1)$
- Step 4:** 再经过 1×1 卷积 ψ ，再做 $sigmoid(\sigma_2)$
- Step 5:** 计算出注意力系数 α 再与 x^l 相乘，输出 \tilde{x}^l

通过注意力门控输出的 \tilde{x}^l 再与解码侧上卷积之后的特征图拼接到一起作为下一个单元的输入。Attention 系数 α 倾向于在目标区域取得大的值，在背景区域取得较小的值，有助于提高图像分割的精度。

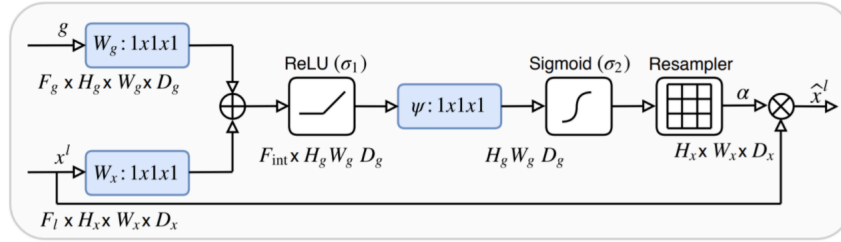


图 6: Attention Gates

3.2 预处理

我们主要使用了两种方法对数据进行预处理来提高模型的性能。首先，除了原始图片数据的 RGB 三通道，我们加入了 HSV 颜色模型（色调、饱和度、明度）中的 SV 通道和 CIELAB 表色系的 lab 通道（图7），共有 8 个通道。每一个通道的取值范围标准化为 $[0, 1]$ ，这能让模型输入更多的信息。除此之外，我们使用了数据增强，包括常规的方法如随机翻转、旋转、缩放比例、裁剪、加入高斯噪声等；另一类数据增强方法是通过随机标准化每一个通道的对比度。我们在实验中也尝试了其他的方法，但模型的结果证明上述的几种方法对提升模型效果最为有效。

3.3 损失函数

针对图片分割问题，我们基于 Jaccard 系数设计了损失函数，并加入了 L2 正则项来防止过拟合。Jaccard 系数即交并比，公式如下所示：

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

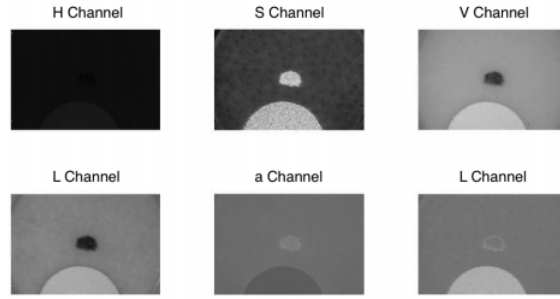


图 7: HSV 与 CIELAB 颜色模型

。我们设计的损失函数形式为

$$L_{dJ} = 1 - \frac{\sum_{i,j} (t_{i,j} p_{i,j})}{\sum_{i,j} t_{i,j}^2 + \sum_{i,j} p_{i,j}^2 - \sum_{i,j} (t_{i,j} p_{i,j})} + \lambda r$$

，其中 $t_{i,j}, p_{i,j}$ 分别是像素点 $p_{i,j}$ 的真值和模型的输出。 λ 是正则惩罚项的系数， r 为 L2 距离的惩罚项。相较于常规的交叉熵损失函数，这种损失函数于图像分割任务直接相关，因为 Jaccard 系数是评估医学图像分割准确率的常用度量。于此同时，这种损失函数十分适用于存在样本不均衡的问题（在医学图像分割中目标区域常常只占整张图片的一小部分），因为不需要一些样本平衡操作。在实验中我们发现模型很容易过拟合，因此我们加入 L2 惩罚项，来改善模型的泛化能力。

3.4 双阈值边界判定

在从网络输出的概率到二元病变掩码这一步，我们使用了双重阈值方法。我们首先使用较高的阈值（比如 0.8）进行筛选，计算所有连续区域的面积并选择最大的区域作为病变质心。接下来再使用一个相对较低的阈值（比如 0.5），并选择病变质心所在的区域作为最后的掩码输出，这一过程如图8所示。通过这种方法，我们可以很好的排除一些来源于皮肤特性的噪声和干扰，比如毛发和色斑，进一步提高图像分割的性能。

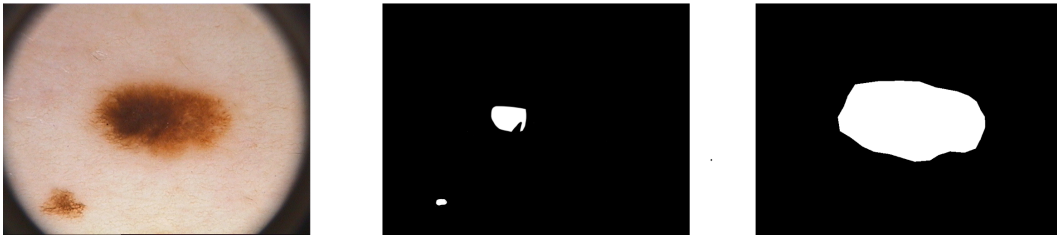


图 8: 双阈值边界判定：从左到右分别为原始图片，高阈值筛选结果以及最后输出结果

4 实验与结果

这里我们将实现我们的模型并与传统的 U-Net 和 ISIC2018 挑战的第一名的方案 (Yuan, 2017) 进行比较。该方案的主体结构是级联结构的卷积反卷积网络，同时在模型的预测阶段使用了 Bagging 集成。由于计算能力的限制，以及比较的公平，我们将比较单个模型的表现。

我们的改进后 U-Net 模型使用 Adam optimiser 进行优化。学习率设置为 $1e-4$ ，并且将在每一个 epoch 后衰减 90%，batchsize 设置为 1，正则项惩罚系数设为 0.02。

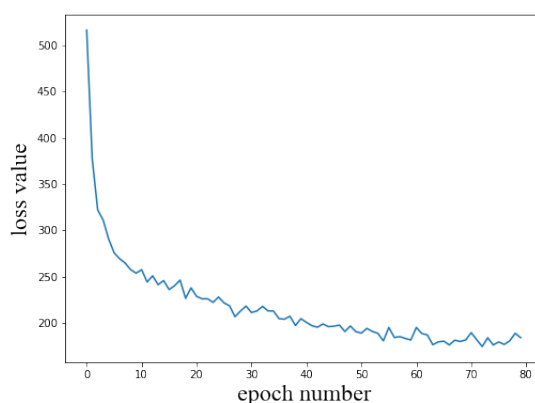


图 9: Train Loss

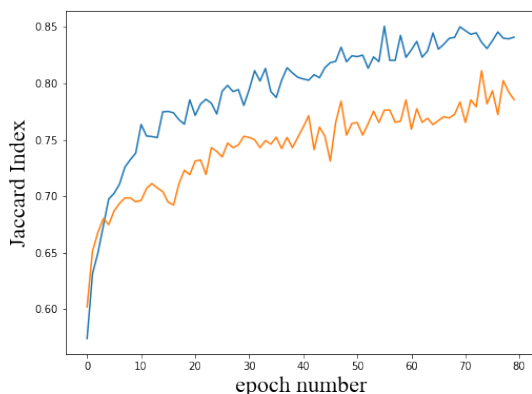


图 10: Train and Valid Dataset Jaccard Index

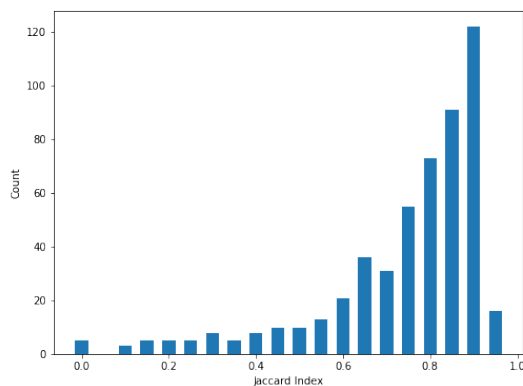


图 11: Jaccard Index 分布频数图

我们可以看到训练的 Loss 随着 epoch 次数逐渐下降并在 60 个 epoch 之后趋于平稳，收敛于 200 左右。训练集的 Jaccard 指数逐步上升后稳定在 0.84 左右。模型存在一些过拟合的现象，但是验证集的 Jaccard 指数也在逐渐上升并最后达到了 0.79 左右的水平。

三个模型在测试集上的表现如表1所示，我们可以看到改进 Unet 全面优于 Unet，在准确率和 Jaccard 指数上都超过了卷积反卷积网络。

改进 Unet 在测试集上的预测效果直方图如图 11 所示。主观上说，当 Jaccard 指数超过 0.8 时，人们会认为图像分割看起来是正确的。我们可以看到 69.4% 的图片的交并比是高于 0.8 的。

表 1: 三种模型表现

Model	Acc.	Sen.	Jaccad
Unet	0.89	0.708	0.674
CDNN	0.905	0.774	0.741
改进 Unet	0.913	0.746	0.753

5 结论

可以看到，我们的改进 U-Net 模型在测试集上的表现综合来说最佳，虽然我们模型的灵敏度略低于 ISIC2018 的最佳方案，但是考虑到 Jaccard 指数在图像分割任务中的重要性，因此我们的改进模型表现良好。我们通过实验证明了循环卷积残差模块和注意力机制能在减少模型参数的同时，有效地提高模型的提取特征能力，尤其是针对医学图像中的小特征。这两种模块也能推广到其他网络结构中，探索处理更多任务类型的可能。

参考文献

- ALOM M Z, HASAN M, YAKOPCIC C, et al., 2018. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation[J]. arXiv preprint arXiv:1802.06955.
- BAI W, SINCLAIR M, TARRONI G, et al., 2017. Human-level cmr image analysis with deep fully convolutional networks[J].
- BRADY M S, OLIVERIA S A, CHRISTOS P J, et al., 2000. Patterns of detection in patients with cutaneous melanoma: implications for secondary prevention[J]. Cancer: Interdisciplinary International Journal of the American Cancer Society, 89(2):342-347.
- BRAY F, FERLAY J, SOERJOMATARAM I, et al., 2018. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: a cancer journal for clinicians, 68(6):394-424.
- CAIZ, VASCONCELOS N, 2018. Cascade r-cnn: Delving into high quality object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.]: 6154-6162.
- JETLEY S, LORD N A, LEE N, et al., 2018. Learn to pay attention[Z]. [S.l.: s.n.].
- KITTLER H, PEHAMBERGER H, WOLFF K, et al., 2002. Diagnostic accuracy of dermoscopy[J]. The lancet oncology, 3(3): 159-165.
- LIAO F, LIANG M, LI Z, et al., 2019. Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network [J]. IEEE transactions on neural networks and learning systems, 30(11):3484-3495.

YUAN Y, 2017. Automatic skin lesion segmentation with fully convolutional-deconvolutional networks[J]. arXiv preprint arXiv:1703.05165.