第 9 卷 第 6 期
2020 年 11 月
集 成 技 术
JOURNAL OF INTEGRATION TECHNOLOGY
Vol. 9 No. 6
Nov. 2020

引文格式：
魏建华, 李佳颖, 黄成健, 等. 用于图像分割的强制召回特征注意力网络 [J]. 集成技术, 2020, 9(6): 59-70.
Wei JH, Li JY, Huang CJ, et al. Attention network with forced recall feature for image segmentation [J]. Journal of Integration Technology, 2020, 9(6): 59-70.

用于图像分割的强制召回特征注意力网络

魏建华[1,2] 李佳颖[1] 黄成健[1,2] 胡庆茂[1,3]

[1]（中国科学院深圳先进技术研究院 深圳 518055）
[2]（中国科学院大学深圳先进技术学院 深圳 518055）
[3]（中国科学院大学人工智能学院 北京 100049）

摘 要 为解决医学图像中前景背景比例严重失衡及小目标区域难以分割的问题，该文提出了一种基于高斯图像金字塔的注意力网络。具体地，首先在特征解码阶段将空间信息与抽象信息进行特征融合；其次，设计了一个特征召回器以强制编码器减少遗漏感兴趣区域的特征；最后，引入分类精度和全局区域重叠项组成的混合损失函数来处理医学图像前景背景严重不平衡问题。所提出的方法在膝关节软骨数据集和 COVOID-19 胸部 CT 数据集中进行了验证，其分割区域分别占 2.08% 和 10.73%。与 U-Net 及其主流变体相比，该方法在两个数据集上都得到了最佳的 Dice 系数，分别为 0.884±0.032 和 0.831±0.072。

关键词 图像分割；高斯图像金字塔；注意力网络；特征召回器；混合损失函数
中图分类号 TP 391 **文献标志码** A doi: 10.12146/j.issn.2095-3135.20200803001

Attention Network with Forced Recall Feature for Image Segmentation

WEI Jianhua[1,2] LI Jiaying[1] HUANG Chengjian[1,2] HU Qingmao[1,3]

[1](*Shenzhen Institutes of Advanced Technology*, *Chinese Academy of Sciences*, *Shenzhen* 518055, *China*)
[2](*Shenzhen College of Advanced Technology*, *University of Chinese Academy of Sciences*, *Shenzhen* 518055, *China*)
[3](*School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing* 100049, *China*)

Abstract To solve the problem of serious imbalance between the foreground and background in medical images and small objects segmentation, we propose an attention network based on Gaussian image pyramid to fuse spatial information and abstract information in the feature decoding stage. In addition, a feature recaller is designed to force the encoder to avoid missing features of the region of interest. Finally, a hybrid loss function composed of classification accuracy and global overlapping terms is employed to deal with

收稿日期：2020-08-03 **修回日期**：2020-09-11
基金项目：国家重点研发计划项目（2018YFB1105600）；国家自然科学基金项目（61671440）
作者简介：魏建华，硕士研究生，研究方向为医学图像处理；李佳颖，硕士，助理工程师，研究方向为图像处理；黄成健，硕士研究生，研究方向为医学图像处理；胡庆茂（通讯作者），博士，研究员，研究方向为图像处理与模式识别，E-mail：qm.hu@siat.ac.cn。

the serious imbalance between the foreground and background. The proposed method was validated on a knee articular cartilage dataset and the COVOID-19 chest CT dataset where the foreground proportions are 2.08% and 10.73%, respectively. The proposed method achieves the highest Dice coefficients on both datasets as compared with U-Net and its state-of-the-art variants, which are 0.884±0.032 and 0.831±0.072, respectively.

Keywords medical image segmentation; Gaussian image pyramid; attention network; feature recaller; hybrid loss function

1 引 言

随着成像技术和重建算法的发展[1-3]，基于图像的计算机辅助诊断在临床诊断和治疗中发挥着越来越重要的作用。尤其是近几年人工智能技术的迅速发展，使基于深度学习的算法在医学图像分析领域被广泛使用，并取得了较好的效果。但是该算法也还存在一些问题，如在分割任务中需要分割的前景区域只占据整个图像的很小一部分，甚至小于 1%。这种数据的不平衡会导致现有生成和区分框架的不稳定[4]。其次，由于医学数据涉及隐私问题，使得数据采集受到严格限制。因此，如何在图像小样本数据上训练得到一个泛化性能较好的模型成为亟需解决的问题[5]。

近年来，卷积神经网络(Convolutional Neural Networks，CNN)已成功地应用于二维和三维生物医学数据的自动分割。U-Net[6]是一个经典的 CNN 分割框架，目前依然被广泛应用于医学图像分割任务。U-Net 与全卷积网络(Fully Convolutional Network，FCN)[7]非常相似，是 FCN 的变体。与 FCN 相比，U-Net 首先是完全对称的，即左边和右边类似；其次跳跃连接也有区别，FCN 用的是加操作，U-Net 用的是叠操作。U-Net 的这种编码-解码结构不仅可以把深层提取到的抽象特征还原解码到原图的尺寸，而且可以把浅层提取到的信息融入到抽象特征中。针对这种编码-解码结构，一些研究者提出使用注意力门控制网络(Attention Gate Networks，AGNs)[8-9]来提高模型对于小感兴趣区(Region of Interest，ROI)的鉴别性。带有注意门的神经元可以使特征的提取聚焦于目标区域，以突出显著的 ROI 特征，抑制不相关区域的特征激活。在后续发展中，基于 U-Net 的改进网络克服了 U-Net 存在的缺点并在医学图像分割中取得了显著的效果，如 UNet＋＋[10]。UNet＋＋模型可以根据数据自身问题的难度自动选择下采样层数，采用了长短连接填补 U-Net 的空心部分，并利用深度监督机制为不同水平层的子网络设计损失函数。

针对前景背景不平衡问题，一些研究者聚焦于目标函数的设计。焦点损失(Focal Loss，FL)[11]在交叉熵损失的基础上加入了调节因子，用于对分类良好的样本进行指数降权，防止了大量简单的负样本支配梯度。Salehi 等[12]针对医学图像分割任务中的病变体素数量远低于非病变体素数量所导致的训练模型高精确度(Precision)、低召回率(Recall)的问题，提出了一种基于 Tversky 指数的广义损失函数，即 Tversky Loss(TL)，以解决数据不平衡的问题，在精确度和召回率之间寻求更好的平衡。Abraham 等[9]利用深度监督机制在解码阶段的每一层设置 TL 迫使中间层在每个尺度上都具有语义上的区分性。Zhang 等[13]通过将分类项损失与区域项损失组成混合损失，来解决单一损失不能处理的类不平衡及小细节平滑问题。

本文针对医学图像的前景、背景比例严重

失衡及小目标区域难以分割的问题，提出了一种新的基于 U-Net 编码-解码结构强制召回特征的注意力网络以及适合于小病灶分割的混合损失函数。主要贡献包括：(1) 提出了多尺度输入图像金字塔，利用图像的局部不变性 (尺度不变性、旋转不变性) 在编码阶段的不同尺度层上提取 ROI 的轮廓与边界特征，并在解码阶段通过注意力门将这些信息与编码器提取到的类别抽象特征进行融合以提高分割的准确性；(2) 设计了 ROI 特征召回网络，该网络的输入是只保留 ROI 特征的编码器输出特征图，并使用了一个 FCN 分类器进行特征预测。目标预测的召回损失 (Recall Loss，RL) 使得网络从分类器向编码器传播梯度，并强制编码器避免遗漏与 ROI 相关的特征；(3) 设计了混合损失函数——特征召回损失、基于分类项损失与区域项损失组成的分割损失共同优化模型，灵活地平衡了精确度与召回率。

2　方　法

本文所提出分割框架的体系结构由编码器、解码器、ROI 特征召回器组成，具体如图 1 所示。编码阶段，在 U-Net 编码器的 4 个不同尺度层上分别加入了相同尺度的输入图像。这些输入图像是基于高斯图像金字塔的，用于提取与 ROI 相关的边界轮廓等空间信息。为了提取更深层的特征和得到更大的感受野，本文在下采样的最后一层特征图后接入了一个空洞空间金字塔池化模块[14]。解码阶段，在原 U-Net 上下层特征跳跃连接的基础上，引入了注意力机制使网络模型可以很好地处理少量的训练样本。首先，分别从空间和通道两个维度计算出高分辨率特征图的加权映射矩阵；然后，依次加权到高分辨率特征图上；最后，使用跳跃连接结合高分辨率的局部特性和低分辨率的全局特性，从而鼓励语义上更有意义的输出。本设计的特征召回器用于对编码器提取

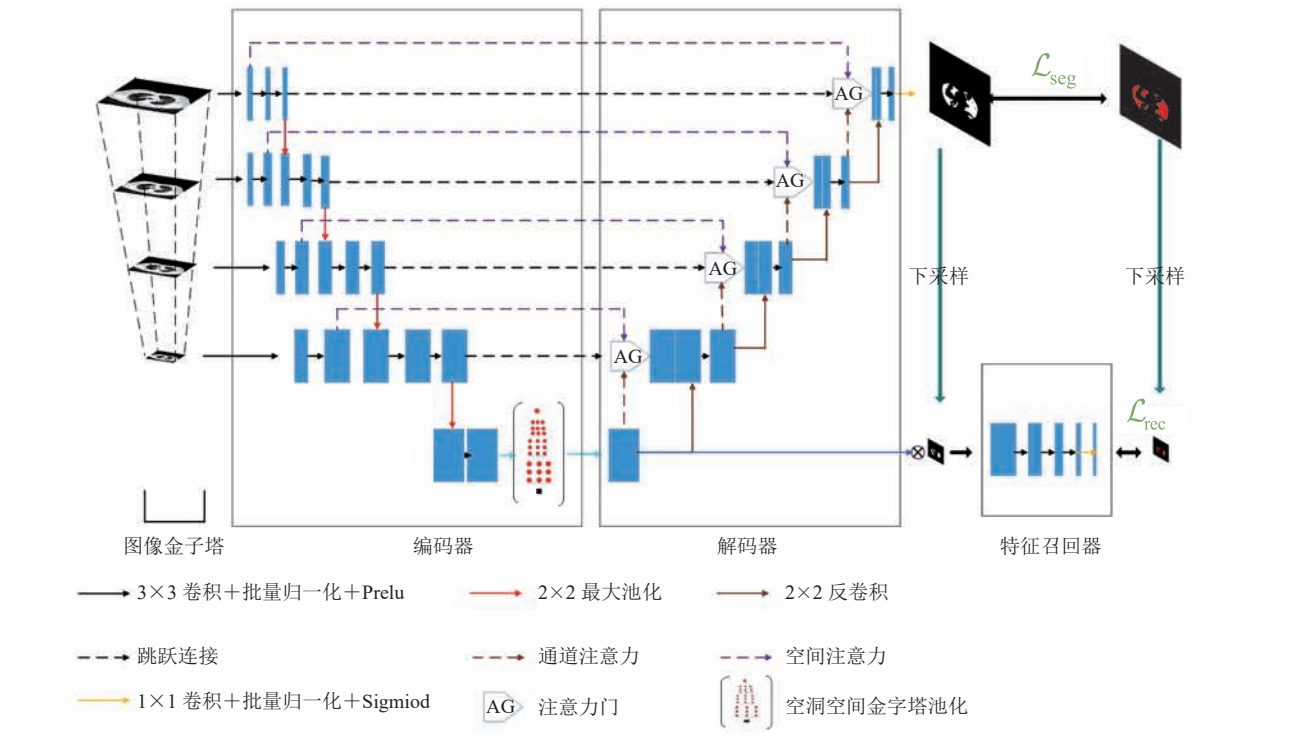


图 1　提出的基于注意力及特征召回的 U-Net 网络结构图

Fig. 1　The proposed U-Net based on the attention and feature recall

到的 ROI 特征信息进行鉴定，并通过召回损失使得网络将梯度从召回器传播到编码器，促使编码器避免遗漏 ROI 有效的特征表示。

2.1 高斯图像金字塔注意力门控制网络

随着编码器下采样次数的增加，网络提取到的特征表示也越来越丰富。但是由于池化层、归一化层和非线性激活函数的存在，导致深层特征输出图的空间细节信息丢失，这会使得小目标物体的分割变得非常困难。因为相较于大的目标物体，小目标物体形状变化较大，对空间的细节信息也更加敏感，特别是随着网络下采样的不断加深、卷积核已经大于目标区域时，网络已无法利用周围局部信息提取到有用的空间特征表示[15]。为了解决这一问题，使用 AGNs 从图像金字塔特征图中识别相关的空间信息，并将其传播到解码阶段。然后利用级联得到的重采样特征信号来获取通道(分类)信息，类似于卷积模块注意力模型[16]从空间与通道两个维度实施高分辨率跳跃连接特征图的加权以突出 ROI 的显著特征，其结构如图 2 所示。空间信息提取支路用于从输入的图像金字塔特征图 H 中确定 ROI 的位置信息，通道信息提取支路可以从提供上下文信息的粗尺度门控信号 L 处获取全局特征信息。

对于不同尺度的高分辨特征图 $F^l \in \mathbb{R}^{H\times W\times C}$ 的每一个像素 F_i^l 分别在通道与空间计算通道注意力系数 $\alpha_c^i \in [0,1]$ 与空间注意力系数 $\alpha_s^i \in [0,1]$，并依次加权到高分辨率特征图上。整个过程如公式(1)～(2)所示：

$$F_1^l = M_c\left(L^l\right) \otimes F \tag{1}$$

$$F_2^l = M_s\left(H^l\right) \otimes F_1^l \tag{2}$$

其中，$M_c\left(L^l\right) \in \mathbb{R}^{1\times 1\times C}$ 为所有像素的通道注意力系数组成的一维注意力特征图；$M_s\left(H^l\right) \in \mathbb{R}^{H\times W\times 1}$ 为在空间维度的注意力特征图；$\otimes$ 为逐像素相乘。首先将 L 计算出的一维通道注意力特征图与输入的浅层特征图 F 相乘得到 F_1^l，之后计算空间信息提取支路的空间注意力特征图，并将两者相乘得到最终的输出 F_2^l。

粗尺度门控信号 L^l 的每一个通道都被视为一个特征检测器，使用全局最大池化和全局平均池化对 L^l 在空间维度上进行压缩，得到两个不同的空间背景描述 $L_{\max}^c$ 与 L_{avg}^c，并使用 ReLU 对其激活。然后将这两个不同的空间背景描述相加并进行 Sigmoid 非线性变换。计算过程如下：

$$M_c(L) = \sigma_2\left\{\sigma_1\left[\text{MaxPool}\left(L^l\right)\right] + \sigma_1\left[\text{AvgPool}\left(L^l\right)\right]\right\} \tag{3}$$

其中，σ_1、σ_2 分别表示 ReLU 变换与 Sigmoid

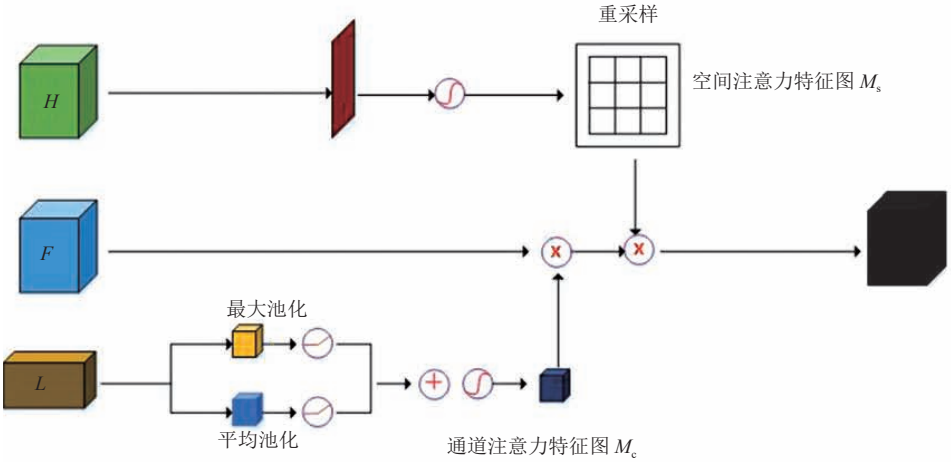


图 2 图像金字塔注意力结构图

Fig. 2 Image pyramid attention structure

变换。

由于在下采样过程中小 ROI 的特征会在卷积中丢失，因此，需要同时考虑图像在多尺度下的描述，获取感兴趣物体的空间特征。本研究通过在编码器的每个最大池化层之前添加图像金字塔的层，来检测不同尺度中的感兴趣特征。但与文献[8]中直接对原尺度的输入图像进行最大池化下采样不同，本文的图像金字塔是基于高斯的。其中，高斯金字塔是通过高斯平滑和下采样获得一系列不同尺度图像集。也就是说，第 K 层高斯金字塔通过平滑、下采样就可以获得 $K+1$层高斯图像。高斯金字塔包含了一系列低通滤波器，其频率从上一层到下一层是以因子 2 逐渐增加，所以高斯金字塔可以跨越很大的频率范围。

本研究输入图像金字塔的层数与编码器池化层相对应，并通过两次卷积得到与编码器相对应层相同通道数的特征图 H^l(如图 1 所示)。为了获得给定任务位置的空间信息，首先对 H^l 使用 1×1×1的卷积，得到一个大小为 $H\times W$ 的二维特征图，并使用 Sigmoid 变换将其映射到[0,1]。然后使用三阶线性插值对其重采样得到最终的空间注意力特征图，计算过程如公式(4)所示：

$$M_s\left(H^l\right)=\zeta\left[\sigma\left(WH^l+b\right)\right] \tag{4}$$

其中，$W\in\mathbb{R}^{1\times1/C}$；$\sigma$ 与 ζ 分别为激活函数与重采样。

2.2　感兴趣区特征召回器

在编码器的最深层，尽管网络具有非常丰富的特征表示，但依然存在着遗漏小目标物体信息的可能。为了使网络尽可能少地遗漏目标物体特征，本研究设计了一个特征召回器用于对编码器提取的特征进行查漏。特征召回器是一个 FCN，由 5 个卷积块组成，每个卷积块包括 1 个卷积层、1 个批归一化层和 1 个 Leaky-ReLU 激活层。由于编码器网络提取的特征信息并不是全部有效，因此通过一种类似于注意机制的方法去除了无效的特征：首先，在解码器得到的预测结果概率图上过滤掉可信度较低的像素(即移除背景像素)；然后，进行最大值池化并加权到编码器的最后一层，加权后的特征图只保留 ROI 的特征表示；最后，送入召回器对其进行分类。

根据网络的实际训练情况，本研究采用了一种更可靠的预测概率渐进策略。在训练开始时，预测结果的可信度较低，不足以保证训练的可靠性。随着迭代步数的增加，置信度逐渐提高，预测结果可以使用更可靠的像素点。本文设置阈值 κ 来确定哪些像素需要保存及哪些像素需要移除。其中，κ 是一个变量，在训练开始时值很小，随着训练迭代步数的增加逐渐升高。本文还设置了 κ 的下界和上界，其表示如公式(5)所示：

$$\kappa=\begin{cases}\kappa_{\text{under}} & \dfrac{\mu}{\nu}\leqslant\kappa_{\text{under}}\\ \dfrac{\mu}{\nu} & \dfrac{\mu}{\nu}>\kappa_{\text{under}}\ \text{和}\ \dfrac{\mu}{\nu}<\kappa_{\text{upper}}\\ \kappa_{\text{upper}} & \text{其他}\end{cases} \tag{5}$$

其中，μ 为当前迭代步数；ν 为总迭代步数；κ_{upper} 为阈值的上界；κ_{under} 为阈值下界。本文结果只保留预测概率图中大于阈值 κ 的像素。在迭代过程中，首先需要确定 κ 值；然后，根据 κ 值生成与金标准同尺寸大小的掩模；最后，将该掩模值与预测概率值进行比较，得到一个 one-hot 输出，即将预测结果中大于或等于 κ 的像素设置为 1，其余设置为 0。在实验中，设置 κ_{upper} 为 0.85，κ_{under} 为 0.15。

2.3　损失函数

在医学图像分割任务中，Dice 损失(Dice Loss，DL)是最广泛使用的损失函数，被用于衡量预测结果与金标准之间的重叠差异：

$$\mathcal{L}_{\text{dice}}=1-\frac{\sum_{i=1}^{N}p\left(x_{if}\right)y_{if}+\epsilon}{\sum_{i=1}^{N}p\left(x_{if}\right)+\sum_{i=1}^{N}y_{if}+\epsilon} \tag{6}$$

其中，x_{if} 表示输入图像的像素 i；$p\left(x_{if}\right)\in[0,1]$为

预测结果是前景区域 f 的概率；$y_{if}\in[0,1]$ 为 f 的金标准；ϵ 用来防止分母等于 0。DL 是一种基于全局区域相似度的度量，它平等地衡量假阳性和假阴性检测。在实际应用中，这种方法可以得到较高精确度但召回率较低的分割结果。对于医学图像来说，数据往往是高度不平衡的，需要分割的目标区域较小，因此需要提高召回率促使网络能准确地分割出较小的 ROI。

特征召回器是为了避免编码器遗漏目标物体的有效特征，即尽量地减少假阴性的存在，因此选择 DL 作为损失函数无法达到目的。召回率表示的是样本中的正样本有多少被预测正确了。以召回率作为损失可以迫使模型最大限度地提取前景区域的特征，从而提高分割准确度。其定义如公式(7)所示：

$$\mathcal{L}_{\text{recall}}=1-\frac{\sum_{i=1}^{N}p\left(x_{if}\right)y_{if}+\epsilon}{\sum_{i=1}^{N}p\left(x_{if}\right)y_{if}+\sum_{i=1}^{N}p\left(x_{i\bar{f}}\right)y_{if}+\epsilon} \tag{7}$$

其中，$p\left(x_{if}\right)$ 为像素 i 的预测结果是非前景区域 $\bar{f}$ 的概率，通过最小化 $\mathcal{L}_{\text{recall}}$ 使得梯度由召回器向编码器传播，迫使编码器能有效地避免遗漏目标物体特征。但是，这也会产生高假阳性问题，必须找到一个方法来消除高假阳性对实验结果的影响。Tversky 系数是 Dice 系数和 Jaccard 系数的一种广义系数，能灵活地平衡假阳性和假阴性，其计算如公式(8)所示：

$$TC=\frac{\sum_{i=1}^{N}p\left(x_{if}\right)y_{if}+\epsilon}{\sum_{i=1}^{N}p\left(x_{if}\right)y_{if}+\alpha\sum_{i=1}^{N}p\left(x_{if}\right)y_{i\bar{f}}+\beta\sum_{i=1}^{N}p\left(x_{i\bar{f}}\right)y_{if}+\epsilon} \tag{8}$$

其中，α、β 为其平衡因子，当 $\alpha=\beta=0.5$ 时，Tversky 系数就成为了 Dice 系数；当 $\alpha=\beta=1$ 时，Tversky 系数又变成了 Jaccard 系数。为了消除高假阳性对实验结果的干扰，使用 Tversky 系数作为分割损失，并调节平衡因子使其向假阳性倾斜，即使得假阳性对于梯度的影响更大。实验发现，当 α 设置为 0.6、β 设置为 0.4 时，可以取得最佳效果。

此外，还需要考虑像素分类的损失。最常用的分类损失是交叉熵损失：

$$\mathcal{L}_{\text{ce}}=-\left[y\log\hat{y}+\left(1-y\right)\log\left(1-\hat{y}\right)\right] \tag{9}$$

其中，$\hat{y}$ 为经过激活函数的输出，其值在 0～1。对于前景像素而言，输出概率越大损失越小；对于非前景像素而言，输出概率越小则损失越小。这会使得损失函数在大量容易区分像素的迭代中，梯度下降比较缓慢且可能无法最优化。因此，引入了焦点损失(FL)[11]作为分类损失，使得模型更加关注困难的、错分的像素。由全局区域重叠损失与分类损失组成的混合损失表示如公式(10)～(11)：

$$\mathcal{L}_{\text{FL}}=\begin{cases}-\omega\left(1-\hat{y}\right)^{\gamma}\log\hat{y} & y=1\\ -\left(1-\omega\right)\hat{y}^{\gamma}\log\left(1-\hat{y}\right) & y=0\end{cases} \tag{10}$$

$$\mathcal{L}_{\text{TCF}}=\left(1-TC\right)+\eta\mathcal{L}_{\text{FL}} \tag{11}$$

其中，ω 用于平衡前景背景像素本身的比例不均；γ 用于减少易分类像素的损失；η 是两种损失函数的平衡因子。与文献[11]一样，本研究设置 ω 为 0.25、γ 为 2。模型由特征召回损失 $\mathcal{L}_{\text{recall}}$ 与分割损失 $\mathcal{L}_{\text{TCF}}$ 共同作用，灵活地平衡了迭代过程中 ROI 的假阳性与假阴性问题，有助于整体分割精度的提高。

3 实验与结果

3.1 数据集

本研究使用了两个前景区域明显小于背景的不同数据集来验证所提出的模型，包括膝关节软骨磁共振影像数据集和 COVID-19 胸部 CT 数据集。其中，膝关节软骨数据集共有 15 例磁共振影像膝关节扫描，每个三维图像包含 46 张 512×512 的切片，在实验中随机选择了 10 例数据作为训练集并进行了相应的数据增广，其余的三维图像作为测试集。膝关节软骨数据集由广东省中山市中医院提供并由放射科拥有十年以上经验的医生勾画金标准。本研究方案已得到医院伦

理审查委员会的批准。同时所有个体都给予书面同意，并为科学和教育目的提供许可。

本文所采用的 COVID-19 胸部 CT 数据集由挪威的两名放射科医生 Tomas Sakinis 博士和 Håvard Bjørke Jenssen 博士提供(http://medicalsegmentation.com/covid19/)。该实验旨在通过人工智能算法实现对 COVID-19 的早期快速辅助筛查和预后评估。

COVID-19 早期在 CT 上最明显的表现就是双肺呈单发或多发的斑片状毛玻璃混浊，在进展期会出现毛玻璃影与实变影或条索影共存[17-18]。对毛玻璃混浊进行快速精准的分割可为 COVID-19 的诊断提高提供重要参考。本文从 COVID-19 胸部 CT 数据集提取了 12 例带金标准的有效数据，随机选择了 9 例用于训练，其余病例用于测试。实验选择横截面切片作为网络输入，所有输入都重采样到 512×512 的像素。

3.2　实施细节

本文所提出算法通过 Keras 框架与 Tensorflow 后端的 Python 语言实现，并使用 4 个 24 G 的 TITAN RTX GPU 进行训练。膝关节软骨数据训练了 50 个 epoch，batch size 设置为 16；COVID-19 数据集训练了 80 个 epoch，batch size 设置为 24。两种模型均使用自适应矩估计(Adaptive Moment Estimation，Adam)优化器[19]训练分割网络，其中 beta_1 设置为 0.9、beta_2 设置为 0.999、epsilon 设置为 10^{-8}、初始学习率设为 10^{-4}，并在每一轮进行 0.9 次幂的多项式衰减。

3.3　膝关节软骨数据集的实验结果

本研究通过在膝关节软骨数据集上进行实验来验证所提出的小 ROI 分割方法。首先，使用 U-Net 验证了不同损失函数的实验效果，接着对本文所提出的 U-Net 变体对于提高分割结果的有效性进行研究，结果如表 1 所示。本文使用 Dice 相似系数(Dice Similarity Coefficient，DSC)、精确度和召回率三个指标对分割结果进行评估，具体计算如公式(12)～(14)所示。

在与现有优秀的 U-Net 变体模型(Focal-Tversky-UNet[9]、UNet＋＋[10])对比时，为公平起见实验中未增加数据集或任何转移学习，并使用了 3 个评估指标：DSC、Jaccard 相似系数(Jaccard Similarity Coefficient，JSC)(公式(15))、平均表面对称距离(Average Symmetric Surface Distance，ASSD)(公式(16))，结果如表 2 所示。测试数据在不同模型的精确度与召回率如图 3 所示。图 4 为不同模型在膝关节软骨测试数据集上的分割结果。

$$DSC=\frac{2\times TP}{2\times TP+FP+FN} \tag{12}$$

表 1　在膝关节软骨数据集上的分割性能

Table 1　Segmentation performance on the knee cartilage dataset

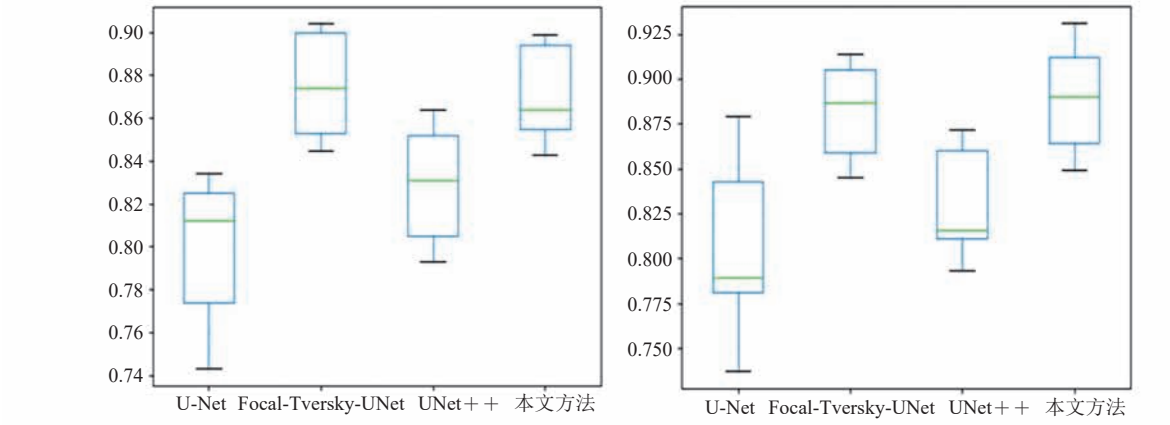
方法	Tversky 系数的平衡因子	DSC	精确度	召回率
U-Net＋ DL	$\alpha=\beta=0.5$	0.762±0.071	0.773±0.092	0.761±0.084
U-Net＋TL	$\alpha=0.3,\beta=0.7$	0.783±0.064	0.783±0.085	0.821±0.061
U-Net＋TL	$\alpha=0.3,\beta=0.7$	0.736±0.068	0.803±0.087	0.714±0.064
U-Net＋TCF	$\alpha=0.3,\beta=0.7$	0.795±0.063	0.794±0.079	0.828±0.049
U-Net＋Gau_ATTn＋DL	$\alpha=\beta=0.5$	0.806±0.065	0.839±0.072	0.827±0.061
U-Net＋Gau_ATTn＋Re＋DL	$\alpha=\beta=0.5$	0.815±0.059	0.828±0.067	**0.857±0.059**
U-Net＋Gau_ATTn＋Re＋TL	$\alpha=0.6,\beta=0.4$	0.822±0.062	0.839±0.063	0.839±0.057
U-Net＋Gau_ATTn＋Re＋TFL	$\alpha=0.6,\beta=0.4$	**0.831±0.072**	**0.841±0.069**	0.851±0.054

注：DL 为 Dice 损失；TL 为 Tvershy 损失；TCF 为 Dice 损失与 Tvershy 损失的混合损失；DSC 为 Dice 相似系数。下同表 3

表 2 U-Net 及其不同变体在膝关节软骨数据集的性能对比

Table 2 Performance comparison of U-Net and its different variants on the knee cartilage dataset

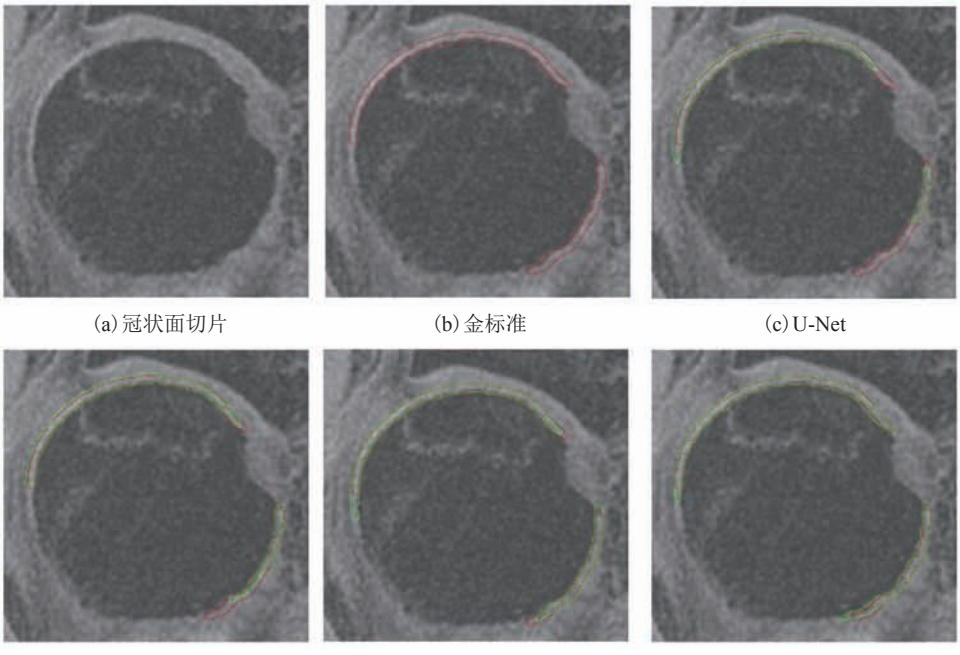
方法	DSC	Jaccard 相似系数	平均表面对称距离(mm)
U-Net[6]	0.789±0.045	0.698±0.056	1.39±0.44
UNet++[10]	0.833±0.038	0.741±0.047	0.49±0.37
Focal-Tversky-UNet[9]	0.861±0.037	0.778±0.041	0.25±0.29
本文方法	**0.884±0.032**	**0.801±0.043**	**0.17±0.31**



(a) 精确度 (b) 召回率

图 3 软骨测试数据在不同方法下的精确度与召回率表现

Fig. 3 The precision and recall performance on cartilage test data among different methods



(a) 冠状面切片 (b) 金标准 (c) U-Net

(d) UNet++ (e) Focal-Tversky-UNet (f) 本文提出方法

图 4 不同方法的软骨分割结果对比

Fig. 4 Comparison of cartilage segmentation among different methods

$$Precision=\frac{TP}{TP+FP} \tag{13}$$

$$Recall=\frac{TP}{TP+FN} \tag{14}$$

$$JSC=\frac{TP}{TP+FP+FN} \tag{15}$$

$$ASSD(A,B)=\frac{\sum_{a\in A}\left\{\min_{b\in B}\left[\mathrm{dist}(a,b)\right]\right\}+\sum_{b\in B}\left\{\min_{a\in A}\left[\mathrm{dist}(a,b)\right]\right\}}{N_A+N_B} \tag{16}$$

其中，TP、FP、FN 分别是真阳性(标记为前景预测为前景)、假阳性(标记为背景预测为前景)和假阴性(标记为前景预测为背景)体素的数目；dist 为来自预测结果 A 的表面体素 a 和来自金标准 B 的表面体素 b 之间的欧氏距离；N_A 和 N_B 分别为预测结果和金标准的表面体素的数目。

3.4　COVID-19 胸部 CT 数据集的实验结果

本文也在最新的 COVID-19 胸部 CT 数据集验证了所提出的强制召回特征的高斯金字塔注意力网络的分割效果，同时也与当前最优秀的一些算法进行了比较，使用了 DSC、Jaccard 相似系数、体积重叠误差(Volumetric Overlap Error，VOE)和相对体积差(Relative Volume Difference，RVD)作为评估指标。

$$VOE=1-\frac{A\cap B}{A\cup B} \tag{17}$$

$$RVD=1-\frac{|A|-|B|}{|B|} \tag{18}$$

其中，A、B 分别是预测结果与金标准。不同模型的评估结果如表 3 所示。图 5 为不同方法的分割结果。

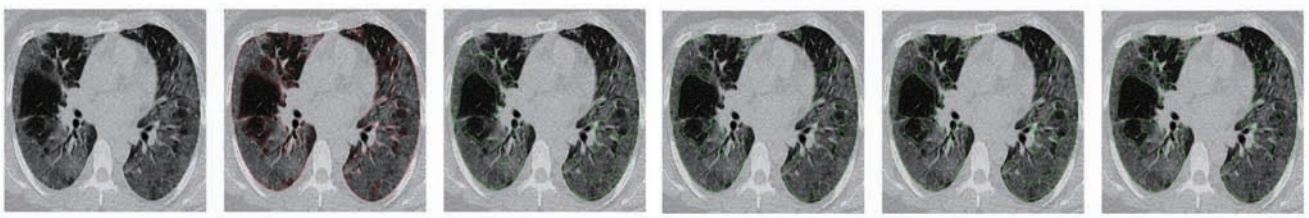
4　讨论与分析

从表 1 可以看出，对于软骨分割，使用 DL 作为损失函数的 U-Net 分割性能较差，DSC 只达到了 0.789，同时召回率与精确度的波动较大，

表 3　在 COVID-19 胸部 CT 数据集上的分割性能

Table 3　Segmentation performance on the COVID-19 chest dataset

方法	Tversky 系数的平衡因子	DSC	精确度	召回率
U-Net＋DL	$\alpha=\beta=0.5$	0.762±0.071	0.773±0.092	0.761±0.084
U-Net＋TL	$\alpha=0.3,\beta=0.7$	0.783±0.064	0.783±0.085	0.821±0.061
U-Net＋TL	$\alpha=0.3,\beta=0.7$	0.736±0.068	0.803±0.087	0.714±0.064
U-Net＋TCF	$\alpha=0.3,\beta=0.7$	0.795±0.063	0.794±0.079	0.828±0.049
U-Net＋Gau_ATTn＋DL	$\alpha=\beta=0.5$	0.806±0.065	0.839±0.072	0.827±0.061
U-Net＋Gau_ATTn＋Re＋DL	$\alpha=\beta=0.5$	0.815±0.059	0.828±0.067	**0.857±0.059**
U-Net＋Gau_ATTn＋Re＋TL	$\alpha=0.6,\beta=0.4$	0.822±0.062	0.839±0.063	0.839±0.057
U-Net＋Gau_ATTn＋Re＋TCF	$\alpha=0.6,\beta=0.4$	**0.831±0.072**	**0.841±0.069**	0.851±0.054



(a) 横截面切片　(b) 金标准　(c) U-Net　(d) UNet＋＋　(e) Focal-Tversky-UNet　(f) 本文提出方法

图 5　不同方法下的 COVID-19 胸部 CT 中磨玻璃混浊的分割结果

Fig. 5　Comparison of ground glass opacity segmentation on COVID-19 chest CT with different methods

学习模型不稳定。这是由于 DL 在追求较好的精确度时会导致较低召回率，即为了追求低假阳性而忽略了高假阴性。其次，软骨在图像中并不存在明显的特征也会导致出现高的假阴性，这些都会降低整体的分割效果。在使用 TL 替换 DL 并设置 α 为 0.3、β 为 0.7 时，使得网络更关注于假阴性后，分割性能有了一定的提高，这表明通过合理调节假阳性与假阴性的比例可以使得模型趋于更优化。作为对比，当调换 α 与 β 的值后，DSC 与召回率达到最低，分别为 0.776 与 0.754，但是精确度提高了 2%，这也从侧面说明了降低假阴性来提高召回率对分割性能的提高具有重要意义。当使用 TL 与 FL 的混合损失 TCF 去训练 U-Net 时，性能有了一个明显的提升，对比于仅仅使用 TL，DSC 从 0.803 上升到 0.815，并且精确度与召回率分别有 1.7% 与 0.8% 的提升。这可以解释为 TL 与 DL 一样是一个全局形状相似性度量，故会导致细小的 ROI 丢失，而 FL 通过考虑像素分类弥补了这一缺陷；同时也顾及到了困难的、错分的像素对于损失的影响。试验结果也证明了高斯图像金字塔注意力机制可以显著地提高分割性能，这是因为本文的注意力模块能有效地恢复下采样过程中丢失的边界轮廓等空间信息。在注意力网络的基础上加入以 DL 作为损失函数的特征召回器后，召回率达到了所有实验中最高的 0.927，而精确度反而下降至 0.828。这是由于召回网络在减少假阴性的同时也会提高假阳性的存在。对于软骨分割来说，由于 ROI 特征不明显会导致出现大量的遗漏(如图 3(c)所示)，而加入召回网络后能在最大程度上减少了假阴性，即使会造成高的假阳性，但由于减少的假阴性大于增加的假阳性，也使得 DSC 有 1.7% 的提高。当调整 α 与 β 的值使损失更倾向假阳性时，分割性能又有了一定的提升，DSC 从 0.864 上升到 0.875。同时从表 1 还可看出，召回率有了稍微的下降，而精确度有了 2.1% 的提高，这再次说明了 TL 能有效地调节假阴性与假阳性的平衡。最后使用 TCF 优化模型时，发现 DSC 是所有实验中最高的，达到了 0.884，召回率与精确度也处于较高水平。

从表 2 不同 U-Net 模型在膝关节软骨数据集的分割效果可以看出，在 3 种不同 U-Net 变体中，UNet＋＋的分割结果效果最差，DSC 只有 0.833。其原因可以归结如下：(1)UNet＋＋虽然考虑了不同数据集对于下采样次数要求不同，并使用长短连接来抓取不同层次的特征，然后和通过特征叠加的方式加以整合，但是依然没有涉及空间信息的加权处理。(2)UNet＋＋同基线 U-Net 一样没有合理考虑假阳性与假阴性的侧重问题，当目标区域特征不明显时，就会导致大量的遗漏分割目标(如图 4(d)所示)。从图 3 也可以看出，UNet＋＋虽然克服了 U-Net 存在的不同测试病例的召回率、精确度有较大的波动，但是其平均值也维持在一个较低的位置。(3)为各个子网络添加损失函数虽然可以使得梯度能更好地回传，但是也增加了额外的强约束项，也可能导致较优的参数形式被破坏，使得结果变差。

相较于 Focal-Tversky-UNet，本文所提出的 U-Net 变体模型的实验效果稍好，DSC 提高了 2.3%(如表 2 所示)。其原因可能有三个方面：(1)本文的多输入图像是基于高斯金字塔的，对比于直接最大池化，可以保留更显著的空间信息。其次，本文的输入图像金字塔不参与下采样，提取到的空间特征信息在编码阶段直接加权至浅层特征并与深层特征进行特征叠加，这更有利于避免空间信息丢失。(2)特征召回器可以迫使编码器减少遗漏 ROI 的特征信息。从图 4(f)可以看出，本文模型基本上可以分割出目标区域，但是也会存在一定过分割的问题。图 3 也证明了这一点，本文所提出的方法可以取得最高的召回率，但是由于过分割的存在，精确度稍低于 Focal-Tversky-UNet。(3)使用了混合损失，同时

顾及到了像素分类精度与全局区域重叠，克服了仅仅使用全局区域重叠损失带来的小目标平滑问题。并且 FL 通过对困难的、错分的像素的特别关注，可以有效地减少假阴性与假阳性，从而提高分割精度。

从表 3 可以看出，使用不同方法在 COVID-19 数据集上的分割结果趋势与表 1 大体一致，这进一步说明了本文提出的注意力机制、特征召回网络和混合损失的有效性，同时也证明了所提出方法的高鲁棒性，在不同的数据集上都有分割性能的提升。本文提出的模型性能表现也优于其他 U-Net[6]模型，DSC 比 UNet＋＋[10]高 2.6%，比 Focal-Tvesky-UNet[9]高 2.3%(如表 4 所示)。与 Focal-Tversky-UNet 和 UNet＋＋相比，首先，本网络引入了空洞空间金字塔池化模块，可以捕获不同尺度的感受野，能够提取区分特征进行分类，避免了由于感受野较小而导致的误报；其次，特征召回损失可以迫使网络不会因为需要分割的磨玻璃区域灰度跨越大而导致遗漏相似性差距大的特征信息(如图 5(f)所示)。在实验中发现相较于软骨分割，不管是本文提出的高斯金字塔空间通道注意力还是 Focal-Tversky-UNet 的注意力方式，对于肺部磨玻璃混浊的分割性能提升并不明显。推测是以下原因导致了这一现象，首先，磨玻璃混浊表现出高强度的变异性和不均匀性，这使得专家很难描绘出一致的金标准，这种不一致性将被传递到训练过程导致无法有效抓取特征信息。其次，在一些影像上磨玻璃混浊的边界轮廓也非常模糊，有些部分可能具有与其他组织相似的灰度，这为空间信息的提取带来了极大的干扰。最后，数据集中不同 CT 型号影像以及不同患者的个体差异也为分割带来了一定的困难。这些因素综合导致了最终分割结果的低 DSC、JSC，高体积重叠误差、相对体积差以及较大的标准差。

5 结　论

本文提出了一个基于高斯图像金字塔的通道空间注意力 U-Net 变体网络，用于弥补 U-Net 下采样导致的空间信息丢失，并在编码阶段通过特征叠加恢复了丢失的上下文信息。此外，还设计了一个 ROI 特征召回器用于迫使编码器减少遗漏目标特征。最后使用特征召回损失与基于分类项与区域项组成的分割损失共同优化模型并合理调节精确度和召回率的平衡。实验结果表明，本文提出的方法比 U-Net 及其两种变体在 Dice 得分方面更加优秀，在软骨分割中可以达到 0.884±0.032，在肺部磨玻璃浑浊分割中达到 0.831±0.072，同时能够维持精确度-召回率的平衡并保持在一个较低的标准差内。本文提出的方法可以成为医学图像分割的通用模型。

参 考 文 献

[1] Sluimer I, Schiham A, Prokop M. Computer analysis of computed tomography scans of the lung: a survey [J]. IEEE Transactions on Medical Imaging, 2006, 25(4): 385-405.

表 4　U-Net 及其不同变体在 COVID-19 胸部 CT 测试数据集的性能对比

Table 4　Performance comparison of U-Net and its different variants on the COVID-19 chest CT dataset

方法	DSC	体积重叠误差	相对体积差	Jaccard 相似系数
U-Net [6]	0.762±0.071	**37.22±13.18**	−2.10±2.37	0.631±0.132
UNet＋＋[10]	0.805±0.079	30.70±9.15	−4.30±3.35	0.692±0.083
Focal-Tversky-UNet[9]	0.808±0.066	30.23±8.86	2.70±1.81	0.699±0.088
本文方法	**0.831＋0.072**	25.43±11.17	2.42±2.11	**0.745±0.111**

[2] Niu SZ, Huang J, Bian ZY, et al. Iterative reconstruction for sparse-view X-ray CT using alpha-divergence constrained total generalized variation minimization [J]. X-ray Science and Technology, 2017, 25(4): 673-688.

[3] Niu SZ, Yu GH, Ma JH, et al. Nonlocal low-rank and sparse matrix decomposition for spectral CT reconstruction [J]. Inverse Problems, 2018, 34(2): 024003.

[4] Sudre CH, Li WQ, Vercauteren T, et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations [C] // Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2017: 240-248.

[5] 陈晨, 王亚立, 乔宇. 任务相关的图像小样本深度学习分类方法研究 [J]. 集成技术, 2020, 9(3): 15-25.

[6] Ronneberger O, FischerP, BroxT. U-Net: convolutional networks for biomedical image segmentation [C] // International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.

[7] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.

[8] Oktay O, Schlemper J, Folgoc L, et al. Attention U-Net: learning where to look for the pancreas [Z/OL]. arXiv Preprint, arXiv: 1804.03999, 2018.

[9] Abraham N, Khan NM. A novel focal Tversky loss function with improved attention U-Net for lesion segmentation [C] // International Symposium on Biomedical Imaging, 2019: 683-687.

[10] Zhou ZW, Siddquee MMR, Tajbakhsh N, et al. UNet＋＋: redesigning skip connections to exploit multiscale features in image segmentation [J]. IEEE Transactions on Medical Imaging, 2020, 39(6): 1856-1867.

[11] Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection [C] // IEEE International Conference on Computer Vision, 2017: 2980-2988.

[12] Salehi SSM, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks [Z/OL]. arXiv Preprint, arXiv:1706.05721, 2017.

[13] Zhang XD, Zhang YQ, Hu QM. Deep learning based vein segmentation from susceptibility-weighted images [J]. Computing, 2019, 101: 637-652.

[14] Chen LC, Papandreou G, Kokkinos I. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

[15] Zhou ZW, Siddquee MMR, Tajbakhsh N, et al. UNet＋＋: anested U-Net architecture for medical image segmentation [C] // International Workshop on Deep Learning in Medical Image Analysis, 2019: 3-11.

[16] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module [C] // European Conference on Computer Vision, 2018: 3-19.

[17] Fang Y, Zhang H, Xie J, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR [J]. Radiology, 2020: 200432.

[18] Bernheim A, Mei X, Huang M, et al. Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection [J]. Radiology, 2020: 200463.

[19] Diederik K, Jimmy B. Adam: a method for stochastic optimization [C] // International Conference on Learning Representations, 2015.