

山东农业大学

# 毕 业 论 文

题 目 基于卷积神经网络的肾小球图像分割方法研究

学 院 信息科学与工程学院

专业班级 数据科学与大数据技术 2021-2

届 次 2025 届

学生姓名 郭鹏鹏

学 号 2021213895

指导教师 芦旭 副教授

二〇二五年五月十五日

# 基于卷积神经网络的肾小球图像分割方法 研究

## Research on Glomerulus Image Segmentation Method Based on Convolutional Neural Network

专业 Speciality	数据科学与大数据技术 Data Science and BigData Technology
学生 Undergraduate	郭鹏鹏 Guo Peng Peng
指导教师 Supervisor	芦旭 Lu Xu

# 目 录

<b>1 引言 .....</b>	<b>1</b>
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	1
1.3 论文主要工作.....	3
1.4 论文组织结构.....	3
<b>2 相关技术介绍 .....</b>	<b>3</b>
2.1 卷积神经网络介绍.....	5
2.1.1 卷积层.....	5
2.1.2 激活函数.....	6
2.1.3 池化层.....	8
2.1.4 全连接层.....	9
2.2 常见的卷积神经网络.....	10
2.2.1 AlexNet.....	10
2.2.2 ResNet.....	10
2.3 常见语义分割网络.....	11
2.3.1 全卷积网络（FCN） .....	11
2.3.2 PSPNet（金字塔场景解析网络） .....	13
2.4 注意力机制与卷积注意力模块.....	13
2.5 辅助损失（Auxiliary Loss） .....	15
2.6 本章小结.....	15
<b>3 基于 CBAM 增强的多尺度金字塔分割网络 .....</b>	<b>17</b>
3.1 多尺度金字塔分割网络的构建.....	18
3.1.1 CBAM 增强的 ResNet50 网络 .....	18
3.1.2 金字塔池化模块.....	19
3.2 损失函数.....	19
3.2.1 交叉熵损失.....	19
3.2.2 Dice 损失 .....	20
3.2.3 交叉熵-Dice 联合损失.....	21
3.3 本章小结.....	21
<b>4 实验 .....</b>	<b>22</b>
4.1 数据集.....	22
4.1.1 数据集简介.....	22
4.1.2 数据增强.....	23
4.2 评价指标.....	24

4.2.1 混淆矩阵.....	24
4.2.2 核心评估指标.....	25
4.3 肾小球分割网络的设计.....	27
4.4 实验设计与结果分析.....	27
4.4.1 实验环境和网络参数设置.....	27
4.4.2 实验结果.....	29
4.4.3 实验数据可视化分析.....	29
4.4.4 模型预测结果可视化.....	35
4.5 肾小球病理分析系统.....	36
4.5.1 Web 功能实现路径.....	37
4.5.2 运行效果展示.....	37
4.6 本章小结.....	40
<b>5 总结和展望 .....</b>	<b>41</b>
5.1 全文总结.....	41
5.2 未来展望.....	41
<b>参考文献 .....</b>	<b>43</b>
<b>致谢 .....</b>	<b>45</b>

# Contents

<b>1 Introduction.....</b>	<b>1</b>
1.1 Research Background and Significance.....	1
1.2 Research Status at Home and Abroad.....	1
1.3 Main work of the paper.....	3
1.4 Organizational structure of the paper.....	3
<b>2 Introduction to relevant technologies .....</b>	<b>3</b>
2.1 Introduction to Convolutional Neural Networks .....	5
2.1.1 Convolutional Layer .....	5
2.1.2 Activation function .....	6
2.1.3 Pooling layer .....	8
2.1.4 Fully connected layer.....	9
2.2 Common Convolutional Neural Networks .....	10
2.2.1 AlexNet.....	10
2.2.2 ResNet.....	10
2.3 Common semantic segmentation networks .....	11
2.3.1 Fully Convolutional Network (FCN).....	11
2.3.2 PSPNet (Pyramid Scene Analysis Network) .....	13
2.4 Attention mechanism and convolutional attention module .....	13
2.5 Auxiliary Loss.....	15
2.6 Summary of this chapter .....	15
<b>3 Multi scale Pyramid Segmentation Network Enhanced by CBAM.....</b>	<b>17</b>
3.1 Construction of Multi scale Pyramid Segmentation Network .....	18
3.1.1 CBAM enhanced ResNet50 network.....	18
3.1.2 Pyramid pooling module.....	19
3.2 Loss function.....	19
3.2.1 Cross entropy loss .....	19
3.2.2 Dice loss.....	20
3.2.3 Cross entropy Dice joint loss .....	21
3.3 Summary of this chapter .....	21
<b>4 Experiments.....</b>	<b>22</b>
4.1 Datasets .....	22
4.1.1 Dataset Introduction.....	22
4.1.2 Data enhancement.....	23

4.2 Evaluating Index .....	24
4.2.1 Confusion matrix .....	24
4.2.2 Core evaluation indicators .....	25
4.3 Design of glomerular segmentation network.....	27
4.4 Experimental design and result analysis .....	27
4.4.1 Experimental environment and network parameter settings.....	27
4.4.2 Experimental result .....	29
4.4.3 Visualization analysis of experimental data .....	29
4.4.4 Visualization of Model Prediction Results .....	35
4.5 Glomerular Pathology Analysis System .....	36
4.5.1 Web function implementation.....	37
4.5.2 Performance display.....	37
4.6 Summary of this chapter .....	40
<b>5 Summary and Outlook .....</b>	<b>41</b>
5.1 Full text summary .....	41
5.2 Future prospects .....	41
<b>Reference .....</b>	<b>43</b>
<b>Acknowledgement.....</b>	<b>45</b>

# 基于卷积神经网络的肾小球图像分割方法研究

**摘要：**随着数字病理学领域的快速发展，肾小球图像自动分割成为热门研究领域。PSP-Net 可用于医学图像分割，但存在多尺度特征融合不够、空间细节留存效果不好以及类别不平衡处理能力较弱等问题。结合肾小球病理切片分割时目标小、边界模糊、形态多变的难点，本文提出基于 PSPNet 架构的改进网络 CBAM-PSPNet。此网络运用双重优化策略，一方面嵌入通道-空间双路注意力机制即 CBAM，提高对肾小球复杂纹理以及微小边界的特征聚焦，另一方面设计交叉熵-Dice 联合损失函数，减轻背景类主导导致的梯度偏移问题。实验结果表明，改进后的 CBAM-PSPNet 于 Hubmap 数据集上有着出色表现，其中肾小球类 Dice 系数提升到了 91.53%，mIOU 达到 91.07%，有效克服了传统方法存在的漏检率高以及边界分割粗糙等问题，借助特征表示与损失函数的协同优化，该模型对肾小球解剖结构和病理特征的识别能力得到明显提高，使得医学图像解析精度得以提升，为肾脏疾病辅助诊断与病理分析给予了可靠的技术支持。

**关键词：**PSP-Net；语义分割；医学图像；注意力机制

## Research on Glomerulus Image Segmentation Method Based on Convolutional Neural Network

**Abstract:** With the rapid development of digital pathology, automatic segmentation of glomerular image has become a hot research field. PSP net can be used for medical image segmentation, but there are some problems, such as the lack of multi-scale feature fusion, the poor effect of spatial detail preservation, and the weak ability of category imbalance processing. Considering the difficulties of small target, fuzzy boundary and variable shape in segmentation of glomerular pathological sections, this paper proposes an improved network cbam-pspnet based on pspnet architecture. This network uses a dual optimization strategy. On the one hand, it embeds the channel space dual attention mechanism, i.e. CBAM, to improve the feature focus on the complex texture and micro boundary of glomerulus. On the other hand, it designs a cross entropy dice joint loss function to reduce the gradient shift caused by the dominant background class. The experimental results show that the improved CBAM pspnet has excellent performance on hubmap data set, in which the dice like coefficient of glomerulus has increased to 91.53%, and the Miou has reached 91.07%, which effectively overcomes the problems of high missed detection rate and rough boundary segmentation existing in the traditional methods. With the collaborative optimization of feature representation and loss function, the recognition ability of the model for glomerular anatomical structure and pathological features has been significantly improved, which improves the accuracy of medical image analysis, and provides reliable technical support for the auxiliary diagnosis and pathological analysis of kidney diseases.

**Keywords:** PSP-Net; Semantic segmentation; Medical images; Attention Mechanism

# 1 引言

## 1.1 研究背景及意义

随着全切片扫描也就是 Whole Slide Imaging 技术不断革新,数字病理学领域正遭遇海量图像数据的急剧增长,这项技术运用先进数字化方式把传统玻璃病理切片转变为高分辨率数字图像,和传统切片保存与管理模式相比,它在应用便捷性、功能多样性以及突破时空限制等方面有着明显优势。美国国立卫生研究院也就是 NIH 所发起的癌症基因组图谱即 TCGA 项目,已经积累了超过 10 万张病理 WSI 图像,如此数量众多的数据为疾病研究提供了极为丰富的资源,同时也让高效的图像分析技术成为了迫切需要解决的技术难题。

肾小球是肾脏执行过滤功能的关键核心结构,其形态方面的变化和肾小球肾炎、糖尿病肾病等重大病症的发展紧密相连,有研究说明,糖尿病肾病患者在患病早期的时候,肾小球基底膜的厚度会增加 10%至 30%,而对于局灶节段性肾小球硬化症患者而言,肾小球硬化的比例跟肾功能下降的速率呈现出十分突出的正相关关系<sup>[1]</sup>。然而在病理图像里,对肾小球进行分割会遇到诸多难题:其有复杂的组织学特征,这使得肾小球的边界因为炎症细胞的浸润而变得模糊不清,在不同的病理阶段,肾小球的形态会从规则的球形逐渐演变成塌陷纤维化的状态,并且它与肾小管、间质组织之间的灰度差异非常小,正是这些因素致使传统的图像分割技术很难达成精准识别。

面对该难题,传统解决方案大多依赖人工标注或者阈值分割技术,然而人工标注需要病理专家耗费大量时间,并且不同专家的标注结果存在主观差异,传统算法在复杂场景下的处理能力也比较有限,近些年,深度学习技术的兴起为肾小球图像分割开辟了新途径,其中监督学习方法表现得格外突出。凭借引入病理专家标注的高质量数据集,基于监督学习的深度学习模型可自动提取肾小球的纹理、形状等特征,提高分割精度以及模型泛化能力。

本文把深度学习技术和医学图像分割相互结合,鉴于现有的深度学习模型在多尺度特征融合以及小样本泛化能力等方面存在欠缺,努力去构建精确的自动分割算法,研究得出的成果可形成标准化的量化分析工具以供临床医生运用,可以缩短诊断周期,还可降低误诊漏诊的概率,为临床诊断给予客观的量化依据。

## 1.2 国内外研究现状

肾小球分割技术的发展历程可追溯至传统图像处理算法主导的早期阶段。在数字病理学尚未普及的背景下,研究者主要依赖人工设计的特征与经典算法实现肾小球区域的粗略划分。例如 Otsu 算法<sup>[2]</sup>,通过全局阈值或者局部阈值分离肾小球与背景,但受限于肾小球与肾小管、间质等周围组织的灰度重叠性,分割结果



的误分割率居高不下。Adams 等<sup>[3]</sup>提出的连通域生长策略, 依赖人工种子点初始化, 通过相似性准则扩散生成候选区域, 然而病理图像中噪声与伪影的存在导致区域生长过程易受干扰, 难以保证分割结果的稳定性。Canny 算子<sup>[4]</sup>、Sobel 梯度算子<sup>[5]</sup>, 通过捕捉灰度突变定位肾小球边界, 但是肾小球边缘的模糊性及形态学异质性经常引发轮廓断裂问题。随后为了提高复杂场景下的分割精度, 多特征融合策略开始逐渐流行起来, 研究者们尝试将颜色直方图、纹理描述符等技术相结合, 例如 LBP<sup>[6]</sup>、Gabor 滤波器<sup>[7]</sup>, 以及形状特征, 像 Hu 矩<sup>[8]</sup>, 来构建高维特征空间, 并借助支持向量机 SVM<sup>[9]</sup>、随机森林<sup>[10]</sup>等分类器实现像素级语义分类。这类方法在特定数据集上获得了不错的效果, 然而由于其性能在很大程度上依赖于手工设计的特征, 并且模型难以对肾小球与周围组织的依赖关系进行建模, 致使分割结果呈现碎片化, 无法契合临床诊断对于高精度分割的要求。

随着卷积神经网络技术不断发展, 深度学习基础上的语义分割技术给肾小球分割带来全新范式, 国内外诸多研究学者都开始对语义分割任务进行研究, 2015 年, <sup>[11]</sup>Long 等人所提出的全卷积网络也就是 Fully Convolutional Network, 简称为 FCN, 首次把端到端像素级预测引入到语义分割任务之中, 全卷积化架构取代了传统 CNN 里的全连接层, 借助反卷积操作逐步恢复特征图分辨率。全卷积神经网络引入跳跃连接机制, 有效整合了浅层高分辨率特征以及深层语义信息, 一定程度上改善了肾小球边缘的定位性能, 不过在处理医学影像时, 该网络依旧面临着一些分割挑战, 这主要是因为网络的多级下采样过程, 使得微观结构特征出现衰减, 同时其全局上下文建模能力不足, 难以适应肾小球形态学的多样性。为解决这些局限性, Ronneberger 等学者同期提出了 U-Net<sup>[12]</sup>架构, 此架构采用对称式编解码结构, 借助跨层跳跃连接实现多尺度特征的深度融合, 编码器端凭借卷积与池化运算, 逐级提取层次化抽象特征, 解码器端采用上采样与特征拼接技术, 逐步恢复空间细节信息。该网络在医学图像分割任务中表现出卓越性能, 迅速成为领域基准, 但它自身也存在一些有待改进的地方。卷积操作会对所有特征通道以及空间位置进行同等处理, 无法自适应地聚焦于肾小球关键区域, 在复杂背景干扰的情况下, 误分割率会比较高, 交叉熵损失函数对类别不平衡较为敏感, 在小目标分割时, 召回率会出现下降情况, 多尺度特征融合策略依靠简单相加操作, 忽视了通道间以及空间维度上的语义关联性。国内研究人员针对这些问题展开了积极探索, 比如温爱卿等<sup>[13]</sup>针对腹部 CT 图像提出上下文拼接网络, 依靠多尺度特征融合保留空间细节并提高上下文语义, 在复杂解剖结构分割中取得提升, 温佳圆等<sup>[14]</sup>针对肾小球滤过膜提出了改进 DeepLab-v3, 把肾小球滤过膜的 3 个组成部分同时分割出来, 且可达到较好的分割效果。赵旭团队<sup>[15]</sup>优化 MaskR-CNN, 凭借调整锚框尺寸、均衡特征金字塔融合、相位一致性定位及空洞卷积提高多尺度肿瘤检测, 结合条件随机场细化边缘, 提升乳腺超声肿瘤分割精度, 王佳等<sup>[16]</sup>

提出 PSSNet 网络, 结合 ResNet-101、DDSPP 多尺度特征和 DP 模块融合, 依靠数据优化提高实现前列腺 MR 图像高精度分割, DSC 达 0.954。王晗<sup>[17]</sup>等提出了一种新医学图像分割模型 COPANet 采用空洞卷积替换普通卷积, 增大感受野范围以获取更多上下文全局语义信息。添加跳跃连接和注意力机制, 以提高图像的分割精度。与其他分割模型相比, 所提模型具有较好的分割精度和泛化性能, Precision 可达 92.76%, IOU 可达 91.08%。郭宇辰<sup>[18]</sup>等提出 MFPNet 模型, 该模型通过多分支融合模块统一不同尺度特征、全局匹配注意力模块增强多尺度表征、带注意力门控的解码器抑制冗余特征, 其在 Synapse 与 ACDC 数据集上 Dice 指标分别为 79.28%和 89.96%, Synapse 数据集 HD95 达 18.28%, 分割精度优于先前方法。

然而, 现有方法在临床应用中仍面临诸多挑战: 标注数据的稀缺性限制了监督学习模型的泛化能力, 另外, 不同染色方案与切片制备差异导致的色彩偏移问题同样会影响分割精度, 除此之外实时处理全切片图像的计算效率需求对模型轻量化设计提出了很高的要求。

### 1.3 论文主要工作

肾小球图像面临组织结构尺度差异大、和周围组织对比度低、病理特征复杂以及存在类别不平衡等挑战, 本文构建了深度学习分割框架, 该框架融合多尺度上下文提高与注意力优化机制, 研究核心工作有以下这些方面:

在肾小球分割领域存在着类别不平衡这一较为突出的难题, 为解决该问题本研究特意设计了联合损失函数, 此函数是交叉熵与 Dice 损失共同发挥作用的结果, 其中交叉熵损失是从像素级概率分布优化方面出发, 以此保障模型输出有整体语义一致性, 而 Dice 损失是基于集合相似度的计算方式, 能强化模型对于稀疏小目标的特征敏感性。凭借引入权重分配机制, 在模型训练进程中动态平衡两类损失的贡献比例, 使得模型对小目标区域的细节捕捉能力得到提升。

采用 Streamlit Web 开发框架, 设计并且实现了交互式肾小球病理分析系统, 该系统支持多终端设备进行访问, 此系统集成移动终端图像采集预测功能, 可借助移动设备拍摄肾小球显微图像, 随后即时上传至系统以完成智能分析预测。

### 1.4 论文组织结构

本文对基于卷积神经网络的肾小球分割网络进行了研究, 各部分内容如下: 第一章介绍了医学图像分割技术的研究背景及意义。

第二章详细介绍了卷积神经网络和语义分割网络以及其它使用到的技术。

第三章对肾小球分割网络的整体架构设计做了系统阐述, 展开来说, 此章节先从宏观角度展示了网络模型的整体框架, 接着详细分析了各核心模块的功能特性以及实现机制, 最后运用严谨的数学建模方法对关键组件的运行原理进行了形

式化描述并开展了理论推导。

第四章先对所运用的数据集以及其预处理办法做了较为系统的阐述,紧接着探讨了语义分割任务里常用的性能评价指标体系,基于此,依靠可视化对比的方式来分析原始 PSPNet 网络架构和改进之后网络模型的性能差异,最终完成了一套有临床应用价值的肾小球病理图像智能分析系统的设计与实现工作。

第五章对全文做出了总结,并对未来的工作做出了展望。

## 2 相关技术介绍

### 2.1 卷积神经网络

卷积神经网络<sup>[19]</sup>（Convolutional Neural Network, CNN）是一种专为处理网格状拓扑数据设计的深度学习模型，如图 2.1 所示为卷积神经网络结构图，其核心结构由输入层、卷积层、池化层、激活函数、全连接层及输出层构成，每一层具有特定的功能。输入层作为深度神经网络架构的起始部分，肩负着接收原始图像数据的关键要点，其输入特征维度涉及了图像宽度、高度、通道数以及 RGB 色彩空间的相关信息，卷积层属于卷积神经网络的核心构成部分，它借助可学习的卷积核，针对输入特征图开展局部区域的卷积运算，以此达成对图像多层次特征的逐步提取，那些经由卷积核滑动扫描得到的局部特征，会被传送到后续的网络层次当中。为了优化网络的计算效率并对模型复杂度加以控制，池化层大多时候被设置在卷积层之后，这一结构依靠降采样操作，使得特征图的空间维度大幅降低，在减少参数量的运算速度得以提高，过拟合现象也得到了缓解，在卷积运算之后加入非线性激活函数，可以有效提高模型的非线性表达能力，提升其复杂特征的学习能力。经过多级卷积-池化层进行特征提取后，网络把高维特征张量展开成为一维向量，经过全连接层实现从特征空间到类别空间的映射转变，最终借助 Softmax 函数输出各个类别的概率分布状况。这种层与层之间交替堆叠的结构，可以逐步提取从边缘、纹理到语义对象的层次化特征，因此被广泛应用于图像处理领域。

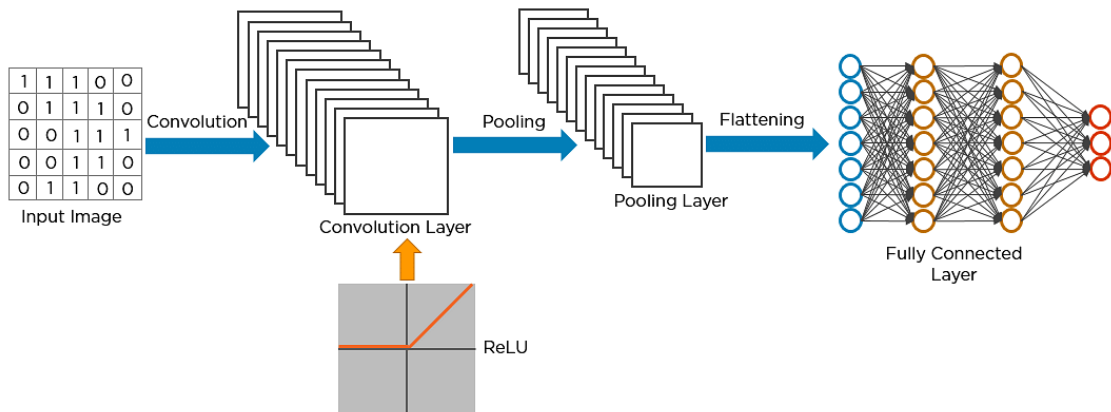


图 2.1 卷积神经网络结构图

#### 2.1.1 卷积层

卷积层身为卷积神经网络的关键构成部分，凭借其独有的特征提取方式达成了对图像数据的多层面呈现，展开来说，卷积运算借助可学习的卷积核于输入特征图上开展滑动窗口动作，针对局部感受野里的像素实施加权求和运算来获得局部特征，就特征提取的层次特性而言，浅层网络主要提取像边缘、纹理这类的低

级视觉特征，深层网络则是经由对这些基础特征进行非线性组合，一步步构建出有高级语义信息的特征表示，例如人脸或者车辆等复杂目标的识别特征。卷积神经网络借助共享卷积核参数的途径达成对输入数据不同空间位置的一致性处理，这般参数共享机制降低了模型的复杂度，还提高了计算效率，为处理海量数据提供了可行办法。

对于给定的输入特征图  $X \in R^{H \times W \times C}$ （高度  $H$ 、宽度  $W$ 、通道数  $C$ ），卷积核  $K \in R^{k \times k \times C \times C'}$ （核大小  $k \times k$ 、输入通道  $C$ 、输出通道  $C'$ ）的运算公式如下：

$$Y_{i,j,c'} = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} \sum_{c=1}^C K_{m,n,c,c'} \cdot X_{i+m,j+n,c} \quad (2-1)$$

其中  $Y \in R^{H' \times W' \times C'}$  为输出特征图， $H'$  与  $W'$  由步长（Stride）与填充（Padding）决定，步长控制滑动步幅，填充通过边缘补零维持特征图尺寸。

要解决卷积神经网络里因步长设置致使的特征图边界信息丢失问题，一般的做法是在输入图像的外围开展零值填充操作，此项技术可保证卷积核可完整地覆盖输入特征图的边缘区域，维持输出特征图的空间维度。填充卷积操作示意图如图 2.2 所示：

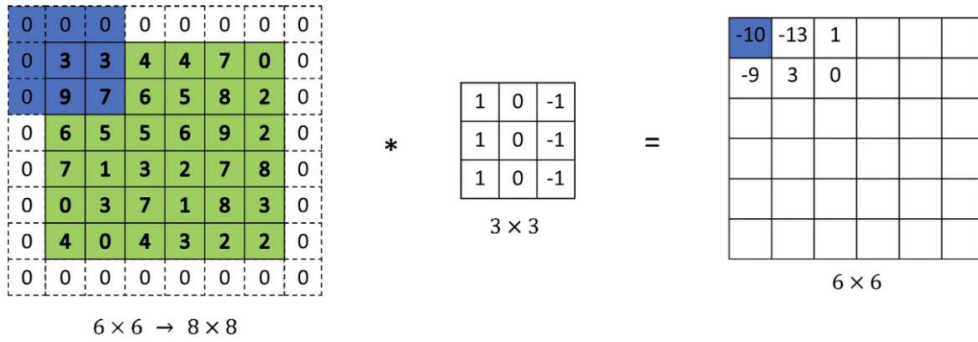


图 2.2 卷积运算示意图

### 2.1.2 激活函数

为了增加神经网络的非线性拟合能力，卷积操作后通常引入非线性激活函数，增强模型表达能力。线性整流函数<sup>[20]</sup>（Rectified Linear Unit, ReLU）因其计算高效性与缓解梯度消失的特性被广泛应用，公式如下：

$$ReLU(x) = \max(0, x) \quad (2-2)$$

ReLU 函数在  $X < 0$  的区间上是导数恒为 0 的线性函数，在  $X > 0$  的区间上是导数恒为 1 的函数，如图 2.3 所示。但是从整体来看，在定义域  $-\infty < X < +\infty$  区间上却是一个非线性函数，虽然 ReLU 激活函数在定义域为负无穷到正无穷的范围内呈现出非线性的特质，然而其分段线性的结构特点让该函数在每个区间的导数计算过程得到了简化，展开来说导数值只会出现 0 或者 1 这两种情况，这种数学特性降低了计算的复杂程度，还提升了采用梯度下降算法时的收敛速率。

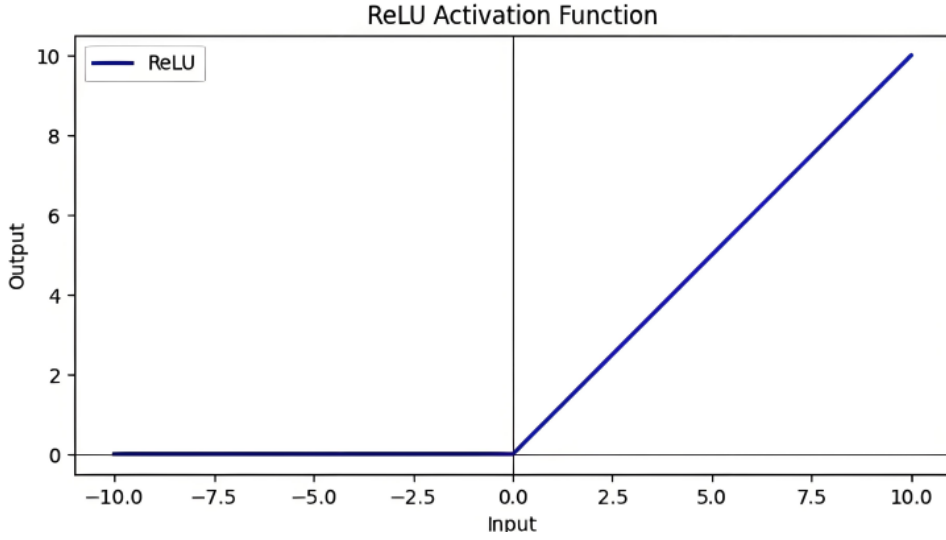


图 2.3 ReLU 函数

凭借理论分析可证实，要是神经网络模型运用线性激活函数，那么其多层架构的输出结果可被表示成输入数据的多层线性组合，展开来说，拿包含两个隐藏层的神经网络结构当作例子，当各层激活函数都是线性变换时，这个模型实际上构建起了一个输入向量的多层线性映射关系，推导过程如下：

$$g(z) = \lambda z \quad (2-3)$$

推导其输出表达式：

第一层计算：

$$Z^{[1]} = W^{[1]}X + B^{[1]} \quad (2-4)$$

$$A^{[1]} = g(Z^{[1]}) = \lambda Z^{[1]} = \lambda(W^{[1]}X + B^{[1]}) \quad (2-5)$$

第二层计算：

$$Z^{[2]} = W^{[2]}A^{[1]} + B^{[2]} \quad (2-6)$$

$$A^{[2]} = g(Z^{[2]}) = \lambda Z^{[2]} = \lambda(W^{[2]}(\lambda(W^{[1]}X + B^{[1]})) + B^{[2]}) \quad (2-7)$$

展开并简化：

$$A^{[2]} = \lambda^2 W^{[2]}W^{[1]}X + \lambda^2 W^{[2]}B^{[1]} + \lambda B^{[2]} \quad (2-8)$$

令：

$$W' = \lambda^2 W^{[2]}W^{[1]} \quad (2-9)$$

$$B' = \lambda^2 W^{[2]}B^{[1]} + \lambda B^{[2]} \quad (2-10)$$

则输出可表示为：

$$A^{[2]} = W'X + B' \quad (2-11)$$

借助上述理论分析可得出这样的结论，在神经网络结构里，就算网络深度持续增加，只要各个隐藏层都运用线性激活函数来开展运算，那么网络最终的输出

$A^{[L]}$ 和输入数据  $X$  之间就会一直维持线性映射关系。此时，隐含层的作用被完全抵消，整个网络等效于一个单层线性回归模型。在这种情况下，网络无法学习非线性关系，无法解决异或（XOR）等非线性可分问题。

### 2.1.3 池化层

卷积神经网络架构里的池化层是关键部分，它借助特征图下采样办法，有效减少了网络参数量以及计算复杂度，以及抑制模型过拟合的优势，这一操作可以进行特征维度压缩，还可以凭借局部区域特征聚合，提取更高层次的抽象表征，提升模型的泛化性能，在具体实现方面，池化层一般部署在卷积层后面，用预设尺寸的池化核，以滑动窗口的形式遍历输入特征图，再依据设定的步长参数生成输出特征。目前深度学习领域主要有两种经典池化策略：最大池化是提取局部区域极值来保留较大特征，平均池化是依靠计算区域均值实现特征平滑。

#### (1) 最大池化

借助提取局部感受野范围内的特征极值，最大池化运算可切实保留输入特征图里最具判别性的信息，生成维度缩减后的特征表示，此操作使得输出特征图的空间分辨率相较于原始输入特征图出现了一定程度的降低。如图 2.4 所示为  $4 \times 4$  大小的输入，使用  $2 \times 2$  大小的池化和且步长设置为 2 的最大池化操作。

对于输入区域  $R$ ，最大池化输出为：

$$Y_{i,j,c} = \max_{(m,n) \in R} X_{m,n,c} \quad (2-12)$$

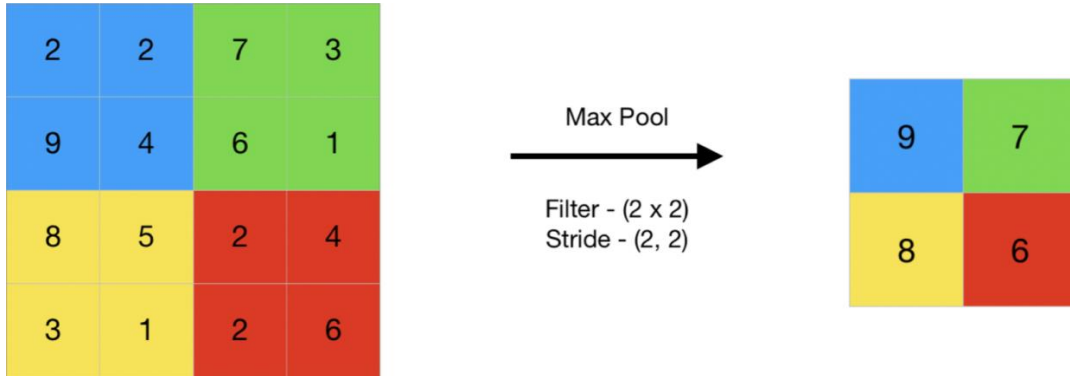


图 2.4 最大池化操作

#### (2) 平均池化

平均池化操作是凭借计算局部窗口内数值的算术平均值来实现的，它能有效对特征图进行平滑处理，降低噪声干扰给特征表达带来的影响，提取出更具鲁棒性的平滑特征表征，此方法的优势在于能完整保留局部区域里的全局统计信息，让提取的特征维持了空间连续性，又保证了语义完整性。如图 2.5 所示为  $4 \times 4$  大小的输入，使用  $2 \times 2$  大小的池化和且步长设置为 2 的平均池化操作。

对于输入区域  $R$ ，平均池化输出为：



$$Y_{i,j,c} = \frac{1}{|R|} \sum_{(m,n) \in R} X_{m,n,c} \quad (2-13)$$

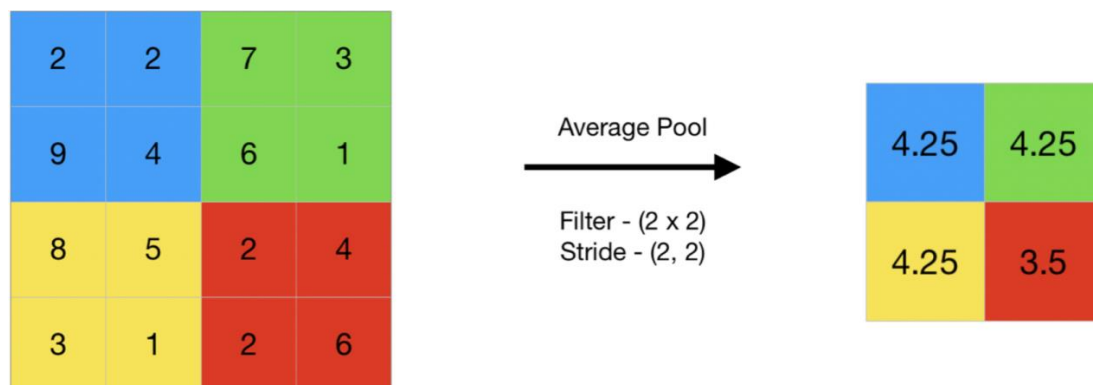


图 2.5 平均池化操作

#### 2.1.4 全连接层

全连接层属于深度神经网络架构里的关键部分，借助构建前一层所有神经元跟当前层各个神经元间的完全连接方式，达成了对输入数据全局特征的建模与整合，这种全互连网络结构特点，让其可有效提取并融合前一层的所有输出信息，在特征表示方面有强大的全局信息捕获能力，图 2.6 为全连接层示意图。

其运算可表示为：

$$y = W \cdot x + b \quad (2-14)$$

其中  $x \in R^D$  为输入向量， $W \in R^{C \times D}$  为权重矩阵， $b \in R^C$  为偏置向量， $C$  为类别数。

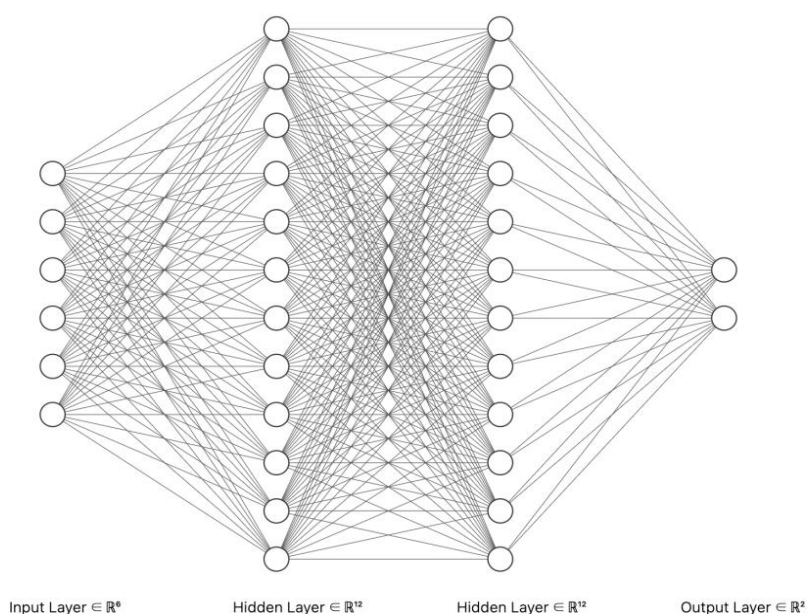


图 2.6 全连接层



## 2.2 常见的卷积神经网络

### 2.2.1 AlexNet

AlexNet<sup>[21]</sup>网络由五层卷积层、三层池化层和三层全连接层构成，如图 2.7 所示，其中，每个卷积层都包含卷积核、偏置项、ReLU 激活函数和局部响应归一化（LRN）模块。在模型架构这一方面，该网络于前三个卷积层里的第 1 层、第 2 层以及第 5 层之后都设置了最大池化层，之后配置三个全连接层，最终借助 Softmax 输出层把网络输出转变为概率分布，以此来达成对图像类别的预测，和同期的神经网络相比较，AlexNet 在深度上有了突破，它的深层结构设计让特征提取能力得到了提升。此模型创造性地使用修正线性单元当作激活函数，和传统的 sigmoid 以及 tanh 函数相比，ReLU 计算效率更高，还可有效缓解深层网络里的梯度消失问题，让训练效率有了大幅提升，为了应对过拟合问题，该模型引入了数据提高以及 Dropout 这两项关键技术：数据提高依靠对训练图像进行旋转、翻转以及裁剪等几何变换，有效扩充了数据的多样性，而 Dropout 技术则是在训练阶段随机屏蔽部分神经元，迫使网络学习多个独立的子网络结构，提高了模型的泛化性能。

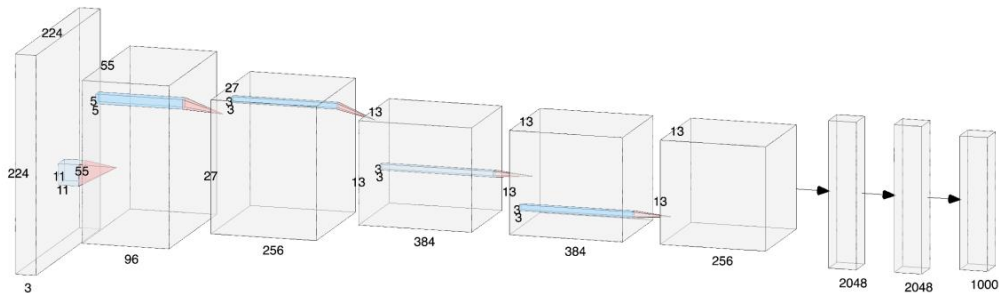


图 2.7 AlexNet 网络示意图

### 2.2.2 ResNet

ResNet<sup>[22]</sup>的全称为残差神经网络（Residual Neural Network），该网络由 He 等人提出。实验研究显示，神经网络深度增加时模型性能呈现出先上升而后下降的非单调变化态势，当网络层数超过某一临界数值后，准确率出现下降情况，这一现象被定义为网络退化问题，经过分析，研究者指出退化问题的本质是深层网络难以有效学习恒等映射，其根源在于过度的非线性变换对网络实现基本线性变换功能产生了妨碍。为解决此关键问题，ResNet 创新性地提出了残差学习机制，具体实现方式如同公式(2-15)所展示，凭借在网络结构里引入跳跃连接来重构特征映射路径。

$$H(x) = F(x) + x \quad (2-15)$$

假设某个层的输入为 $x$ ，其中 $H(x)$ 是理想输出，传统的网络会直接学习 $H(x)$ ，而 ResNet 则学习  $F(x) = H(x) - x$ ，即残差。残差结构如图 2.8 所示：

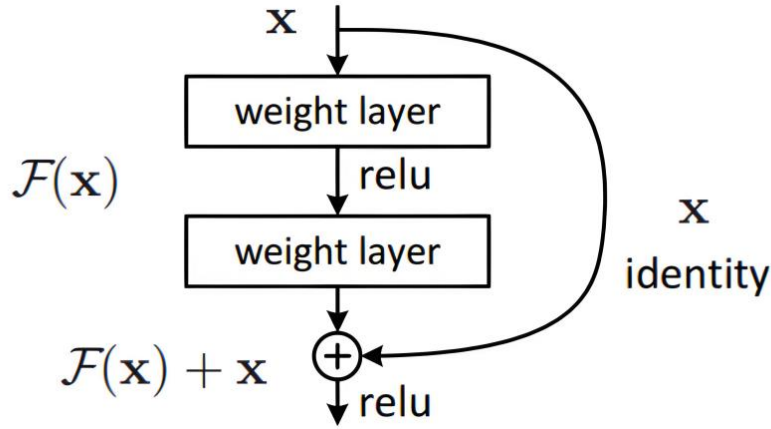


图 2.8 残差块结构示意图

### 2.3 常见语义分割网络

语义分割任务的关键之处在于针对图像里的每一个像素点都要做出精准类别标注，其面临的技术难题主要聚焦在达成像素层面的精细预测，传统卷积神经网络的全连接层结构会致使空间特征信息出现损失，这种网络架构不太适合直接用于像素级分类任务，全卷积网络被提出之后，该领域的研究正式迈入了以深度学习技术作为主导的全新阶段。

#### 2.3.1 全卷积网络（FCN）

全卷积网络<sup>[23]</sup>（Fully Convolutional Network, FCN）通过全卷积化架构替代传统卷积神经网络中的全连接层，保留输入图像的空间维度信息，实现端到端的像素级密集预测。其核心创新在于引入跳跃连接，将编码器的浅层高分辨率特征与解码器的深层高语义特征融合，从而解决因多次下采样导致的细节信息丢失问题。FCN 结构如图 2.9 所示，FCN 的主要改进如下：

##### （1）转置卷积

假设编码器输出的深层特征图为  $F_{deep} \in R^{H/32 \times W/32 \times C}$ ，输入图像尺寸为  $H \times W$ ，经 5 次步长为 2 的下采样后尺寸缩小至  $H/32 \times W/32$ ，通过转置卷积（Transposed Convolution）逐步上采样恢复空间分辨率。转置卷积的运算可定义为公式（2-16）：

$$F_{up}^{(l)} = \sigma(W_{transpose}^{(l)} * F_{deep}^{(l-1)} + b^{(l)}) \quad (2-16)$$

其中： $l$  表示网络层索引， $*$  表示转置卷积操作， $W_{transpose}^{(l)} \in \mathbb{R}^{k \times k \times C_{in} \times C_{out}}$  为第  $l$  层转置卷积核， $k$  为卷积核大小， $b^{(l)}$  为第  $l$  层的偏置项， $\sigma$  为激活函数，通

常情况下为 ReLU。输出特征图尺寸  $H_{up} \times W_{up}$  由转置卷积的步长 (Stride) 与填充 (Padding) 决定, 计算公式为 (2-17):

$$H_{up} = (H_{in} - 1) \times \text{stride} - 2 \times \text{padding} + k \quad (2-17)$$

### (2) 跳跃连接与特征融合

为补偿上采样过程中的细节损失, FCN 将编码器中对应层的高分辨率特征  $F_{\text{shallow}} \in R^{H/2 \times W/2 \times C'}$  通过跳跃连接与上采样结果融合。具体步骤如下: 对浅层特征  $F_{\text{shallow}}$  进行  $1 \times 1$  卷积调整通道数, 使其与上采样特征  $F_{up}$  的通道数一致:

$$F'_{\text{shallow}} = \text{Conv}_{1 \times 1}(F_{\text{shallow}}) \quad (2-18)$$

$$\text{Conv}_{1 \times 1}: R^{H/2 \times W/2 \times C'} \rightarrow R^{H/2 \times W/2 \times C} \quad (2-19)$$

然后将调整后的浅层特征与上采样特征融合:

$$F_{\text{fused}} = F_{up} + F'_{\text{shallow}} \quad (2-20)$$

最后通过 ReLU 增强特征表达能力:

$$F_{\text{fused}} = \text{ReLU}(F_{\text{fused}}) \quad (2-21)$$

### (3) 像素级分类与 Softmax

最终, 融合后的特征图  $F_{\text{fused}} \in R^{H \times W \times C}$  经过  $1 \times 1$  卷积映射至类别空间, 并通过 Softmax 函数生成像素级概率分布, 概率如公式 (2-22):

$$P(y_{i,j} = c) = \frac{\exp(W_c^T F_{\text{fused}}(i,j) + b_c)}{\sum_{c'=1}^C \exp(W_{c'}^T F_{\text{fused}}(i,j) + b_{c'})} \quad (2-22)$$

其中  $W_c \in R^C$  为类别  $c$  的权重向量,  $b_c$  为偏置项。

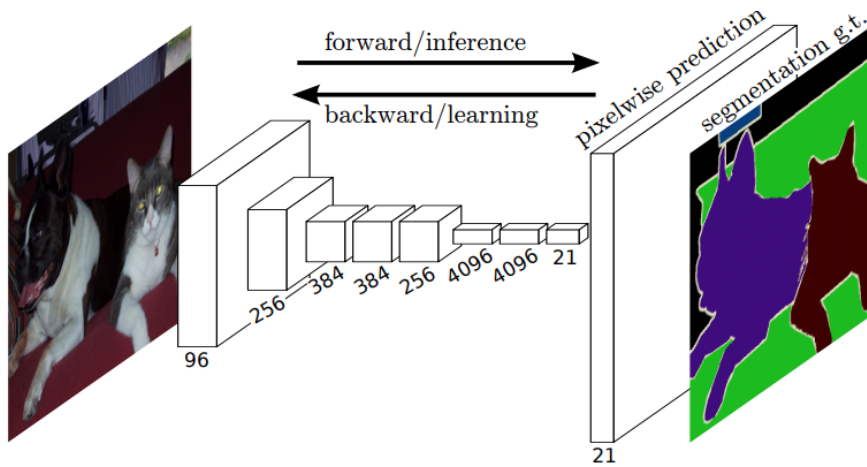


图 2.9 FCN 结构图

### 2.3.2 PSPNet（金字塔场景解析网络）

PSPNet<sup>[24]</sup>网络分为两大部分，即编码器和解码器，编码器通常采用 ResNet 或 VGG 等骨干网络提取特征，解码器通过金字塔池化模块与上采样操作实现多尺度特征融合。参照图 2.10 可以看到，PSPNet 运用了金字塔池化模块结构，借由多级自适应池化操作达成了对不同尺度上下文信息的有效提取，该网络架构的特征提取流程可具体细分为以下几个步骤。

首先对编码器输出特征  $F_{enc}$  进行  $N$  种不同尺寸的全局平均池化，生成多尺度特征金字塔，公式如下：

$$F_{pool}^{(n)} = AdaptiveAvgPool2d(s_n)(F_{enc}), \quad n = 1, 2, \dots, N \quad (2-23)$$

其中  $s_n$  表示第  $n$  级池化的输出尺寸，例如  $1 \times 1, 2 \times 2, 3 \times 3, 6 \times 6$ 。

对于每级池化结果通过  $1 \times 1$  卷积降维，再上采样至原特征图尺寸，公式如下：

$$F_{ppm}^{(n)} = Conv_{1 \times 1}(F_{pool}^{(n)}) \Rightarrow \tilde{F}_{ppm}^{(n)} = Upsample\left(F_{ppm}^{(n)}, size = \left(\frac{8}{H}, \frac{8}{W}\right)\right) \quad (2-24)$$

其中  $Conv_{1 \times 1}: R^{s_n \times s_n \times C} \rightarrow R^{s_n \times s_n \times C'}$ 。

最后将原始特征与多尺度池化特征沿通道维度拼接：

$$F_{ppm} = Concat(F_{enc}, \tilde{F}_{ppm}^{(1)}, \tilde{F}_{ppm}^{(2)}, \dots, \tilde{F}_{ppm}^{(N)}) \quad (2-25)$$

最终  $F_{ppm} \in R^{\frac{8}{H} \times \frac{8}{W} \times (C+NC')}$ ，实现多尺度特征融合。

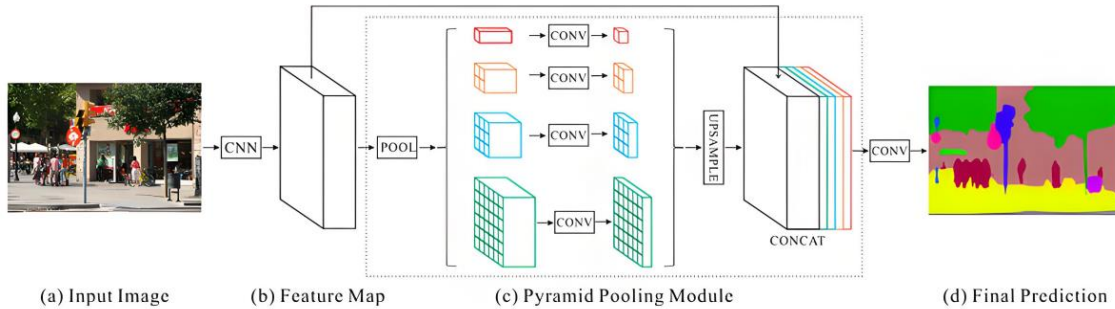


图 2.10 PSP-Net 结构图

### 2.4 注意力机制与卷积注意力模块

(1) 注意力机制（Attention Mechanism）是一种模拟人类认知系统的仿生计算范式，最早由 Bahdanau 等人<sup>[25]</sup>在神经机器翻译任务中系统提出。通过动态权重分配策略，建立输入序列中不同位置间的关联性权重，使得模型能够自适应地聚焦于与当前解码状态最相关的语义信息。对于给定的源语言序列  $X = \{x_1, x_2, \dots, x_n\}$  和目标序列  $Y = \{y_1, y_2, \dots, y_m\}$ ，注意力权重  $\alpha_{ij}$  通过可学习的对齐模型计算：

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^n \exp(e_{ik})} \quad (2-26)$$

$$e_{ij} = f_{align}(h_i^{enc}, h_j^{dec}) \quad (2-27)$$

在相关内容中,  $h_i^{enc}$ 与 $h_j^{dec}$ 分别代表编码器以及解码器的隐藏状态, 而 $f_{align}$ 则是对齐函数。这种借助软选择的注意力机制, 成功突破了传统序列模型在处理长程依赖建模时所面临的瓶颈, 成为 Transformer<sup>[26]</sup>架构里的关键组件, 于医学图像分割领域而言, 注意力机制可借助抑制那些无关的背景噪声, 同时提高病灶区域的响应, 以此来提升模型对于低对比度解剖结构的敏感性。

(2) CBAM<sup>[27]</sup> (Convolutional Block Attention Module) 是由韩国科学技术院研究人员在论文“CBAM: Convolutional Block Attention Module”中提出的, 其核心思想在于借助注意力机制, 使模型自主学习特征图中通道与空间位置维度下重要特征, 进而增强重要特征、抑制非重要特征。如图 2.11 所呈现的那样, 此模块运用了双分支并行处理架构, 其中含有通道注意力机制(CA)以及空间注意力机制(SA)这两个核心子模块, 借助级联的方式针对输入特征图展开逐层的优化处理。

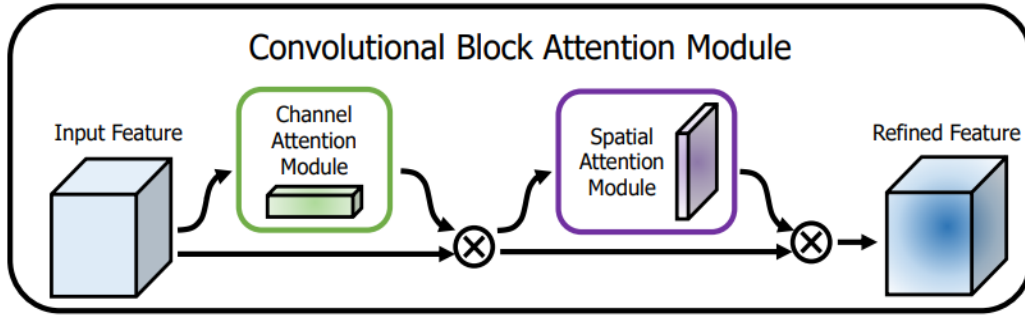


图 2.11 CBAM 模块示意图

从图 2.12 中可清晰看出, 通道注意力机制主要围绕特征图的通道维度展开处理工作, 其具体的实现过程可细分为下面这些步骤: 一开始会针对输入特征图开展全局平均池化以及全局最大池化操作, 以此来提取出可表征通道维度平均特征和最大特征的两个全局特征向量, 紧接着把这两个特征向量输入到带有共享参数的多层感知机当中, 这个网络含有一个隐藏层, 其神经元数量一般会被设置为输入通道数的  $1/ratio$ , 经过 MLP 处理之后所得到的权重表示, 借助相加运算并结合 Sigmoid 激活函数来生成各个通道的最终权重系数, 最后借助将权重系数与输入特征图在通道维度上进行逐元素相乘, 达成特征图的加权调整, 其数学表达式如下所示:

$$Mc(F) = \sigma \left( MLP(AvgPool(F)) + MLP(MaxPool(F)) \right) \quad (2-28)$$

$$Mc(F) = \sigma(W1(W0(F_{avg}^c)) + W1(W0(F_{max}^c))) \quad (2-29)$$



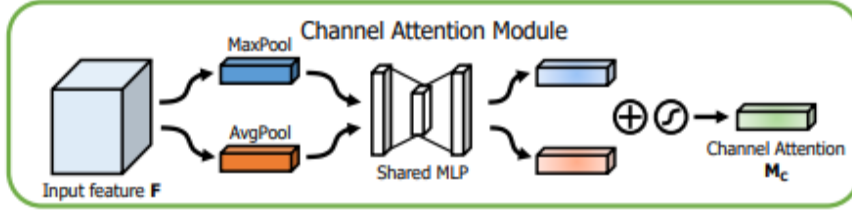


图 2.12 通道注意力模块示意图

如图 2.13 呈现的空间注意力模块着重处理特征图的空间维度特征，它先对经过通道注意力模块处理后的特征图开展通道维度的平均池化以及最大池化操作，以此分别提取出代表空间维度平均特征与最大特征的两类空间特征图，接着这两个特征图于通道维度进行拼接，借助一个输出通道数是 1 的卷积层来达成特征融合，随后运用 Sigmoid 激活函数生成各个空间位置的最终权重系数。该权重系数和经通道注意力模块处理的特征图在空间维度上进行逐元素相乘，完成特征加权过程，此 CBAM 模块可依据实际情况优化特征图在通道和空间维度的特征响应强度，提升模型的特征表征能力，其数学表达式如下所示：

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \quad (2-30)$$

$$M_s(F) = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (2-31)$$

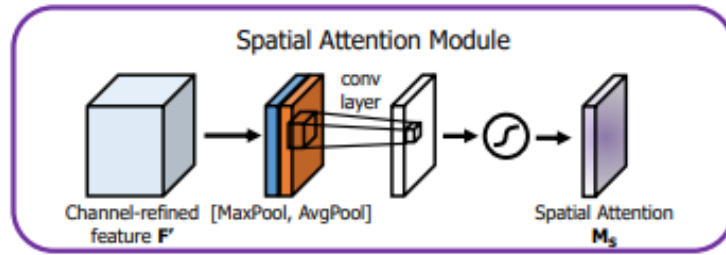


图 2.13 空间注意力模块示意图

## 2.5 辅助损失 (Auxiliary Loss)

辅助损失 (Auxiliary Loss) 策略，除了主分割头的损失，还引入了辅助分割头的损失，用于在网络的中间层提供额外的监督信息，帮助网络在训练早期更好地学习到特征，加速收敛并提高模型的性能。整体损失函数是主分割头损失和辅助分割头损失的加权和，其数学表达式如下：

$$L = L_{main} + \alpha L_{aux} \quad (2-32)$$

其中， $L$  是总的损失值， $L_{main}$  是主分割头的损失， $L_{aux}$  是辅助分割头的损失， $\alpha$  是辅助损失的权重系数。

## 2.6 本章小结

在这一章节之中，对语义分割领域的技术发展脉络展开了系统且全面的梳理，一开始是从数学方面入手，阐释了卷积神经网络所有的核心运算机制，针对卷积

层、池化层以及全连接层的数学表达式进行了详细的推导，紧接着将重点放在语义分割任务的特性上面，剖析了全卷积网络以及 **PSP-Net** 的模型架构创新之处，还对其核心算法公式做了严谨的数学推导。最后对注意力机制以及辅助损失技术给予介绍，此两项技术为后续肾小球分割模型的优化奠定了理论根基，还为医学图像分析的算法设计给予了指引。

### 3 基于 CBAM 增强的多尺度金字塔分割网络

传统的注意力方法通过全局平均池化压缩空间信息，再计算通道权重以增强判别性特征，但是这类方法在医学场景中面临一些局限：首先，全局池化操作削弱了病灶区域的局部空间关联性，会导致空间细节丢失，降低分割精度；另外，单一尺度的注意力权重难以适应多尺度结构的特征表达需求，网络的多尺度建模能力不足。

为切实解决上述提及的各类问题，本文独具创新地搭建起一种经过改进的医学图像分割网络 CBAM - PSPNet，该网络是在 PSPNet 架构的基础之上展开优化设计工作的，其核心创新之处主要呈现于以下这些方面：在 ResNet50 的 Stage3 与 Stage4 嵌入通道-空间双路注意力模块（CBAM），强化多尺度特征的判别性表示；采用空洞卷积金字塔捕获多尺度上下文信息；构建交叉熵-Dice 联合损失函数，优化小目标的分割性能。

图 3.1 为肾小球图像分割的典型流程示意图。

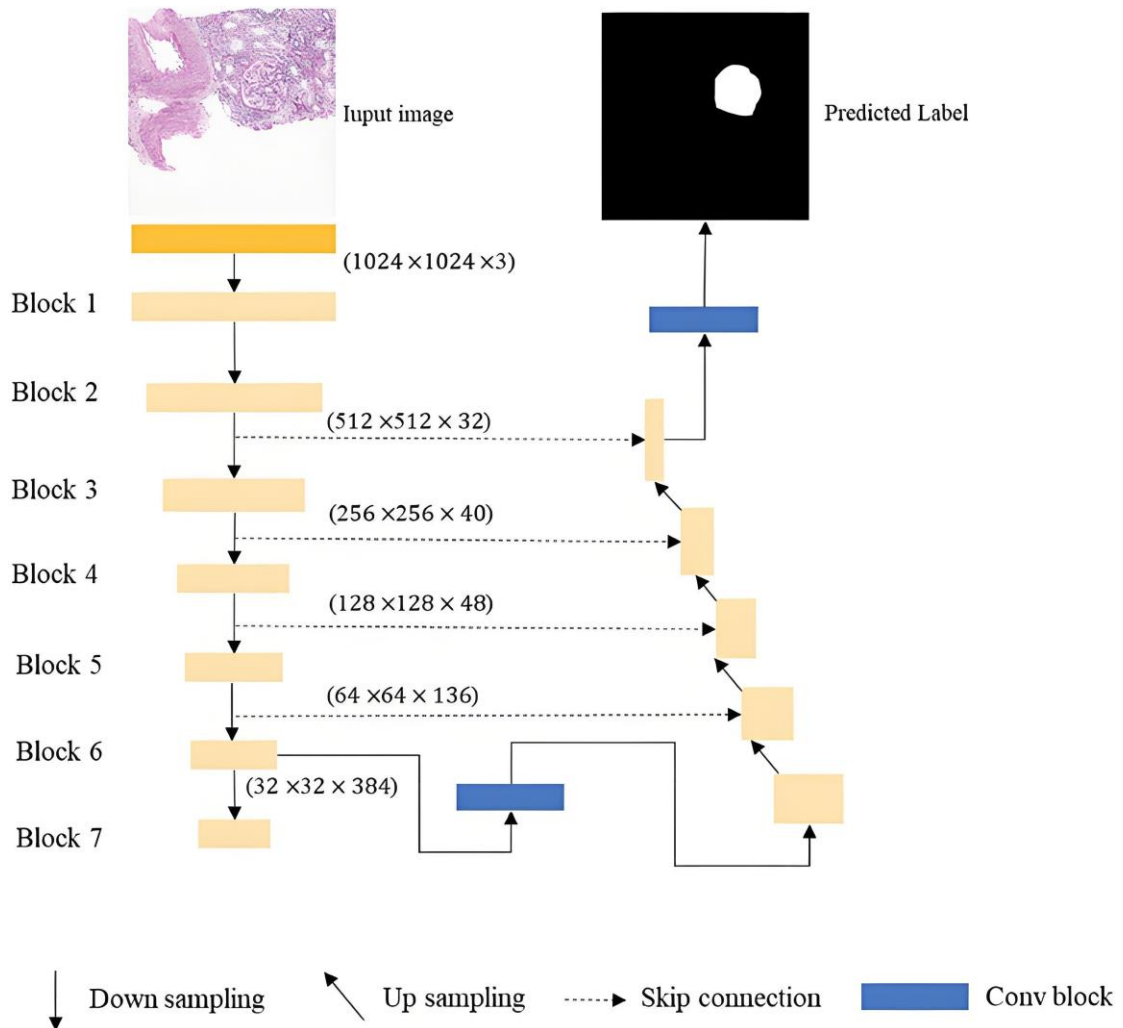


图 3.1 肾小球图像分割流程示意图



### 3.1 多尺度金字塔分割网络的构建

#### 3.1.1 CBAM 增强的 ResNet50 网络

CBAM-PSPNet 网络以 ResNet50 为骨干网络，如图 3.2。输入图像经预处理固定为  $I \in R^{256 \times 256 \times 3}$ ，首先通过  $7 \times 7$  卷积（stride=2, padding=3）将空间维度下采样至  $128 \times 128$ ，输出特征图  $F'_0 = \text{Conv}_{7 \times 7}(I) \in R^{128 \times 128 \times 64}$ ，随后通过  $3 \times 3$  最大池化（stride=2）进一步压缩  $64 \times 64$ ，得到  $F_0 = \text{MaxPool}_{3 \times 3}(F'_0) \in R^{64 \times 64 \times 64}$ ，完成初始  $\times 4$  下采样。

Stage1 采用 stride=1 的残差块保持特征图尺寸不变，输出  $F_1 \in R^{64 \times 64 \times 256}$ ；Stage2 通过 stride=2 的卷积操作将分辨率降至  $32 \times 32$ ，输出  $F_2 \in R^{32 \times 32 \times 512}$ ；Stage3 与 Stage4 采用空洞卷积策略（dilation=2 和 4）并维持 stride=1，避免空间信息丢失，其空洞卷积定义为：

$$\text{Conv}_{3 \times 3}^{\text{dilate}=d}(X)_{i,j} = \sum_{m=-1}^1 \sum_{n=-1}^1 W_{m,n} \cdot X_{i+dm,j+dn} \quad (3-1)$$

分别输出  $F_3 \in R^{32 \times 32 \times 1024}$  和  $F_4 \in R^{32 \times 32 \times 2048}$ 。

在本研究里，于第三阶段以及第四阶段的残差网络结构之中融入了卷积块注意力模块也就是 CBAM，该模块的数学建模进程可被分解成通道注意力机制和空间注意力机制这两个子模块，其中通道注意力子模块借助全局平均池化操作并结合多层感知器来达成特征通道权重的动态分配。

$$M_c(F) = \sigma \left( W_1 \cdot \delta \left( W_0 \cdot (GAP(F) + GMP(F)) \right) \right) \quad (3-2)$$

其中  $GAP(F) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F(i,j,:)$ ， $GMP(F) = \max_{i,j} F(i,j,:)$ ， $W_0 \in R^{\frac{C}{r} \times \frac{C}{r}}$ ， $W_1 \in R^{C \times \frac{C}{r}}$ （压缩比  $r=16$ ），空间注意力通过拼接平均池化与最大池化特征并卷积生成权重  $M_s(F) = \sigma(\text{Conv}_{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)]))$ ，最终特征校准为  $F'' = M_c(F) \otimes M_s(F') \otimes F$ 。Stage3 的中层特征（ $32 \times 32 \times 1024$ ）通过 CBAM 筛选重要通道并聚焦局部区域，避免金字塔池化引入噪声；Stage4 的深层特征（ $32 \times 32 \times 2048$ ）通过 CBAM 强化硬化区域相关通道并补偿空洞卷积导致的边界模糊。

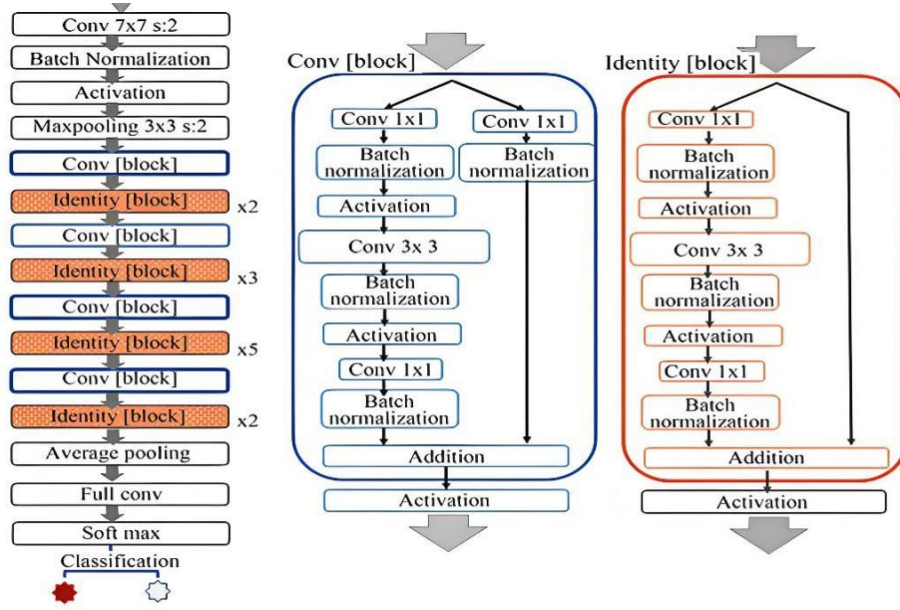


图 3.2 ResNet50 结构图

### 3.1.2 金字塔池化模块

金字塔池化模块（PPM）接收 $F_4$ ，通过多尺度自适应池化：

$$F_{pool}(k) = AdaptiveAvgPool2d(s_k)(F_4), \quad s_k \in \{1, 2, 3, 6\} \quad (3-3)$$

生成 $1 \times 1, 2 \times 2, 3 \times 3, 6 \times 6$ 的特征图，经 $1 \times 1$ 卷积降维后上采样至 $32 \times 32$ ，拼接并融合为：

$$F_{ppm} = Conv_{3 \times 3} \left( Concat \left( F_4, \tilde{F}_{ppm}^{(1)}, \tilde{F}_{ppm}^{(2)}, \tilde{F}_{ppm}^{(3)}, \tilde{F}_{ppm}^{(4)} \right) \right) \in R^{32 \times 32 \times 4096} \quad (3-4)$$

解码器通过两次上采样（ $\times 4$ 与 $\times 2$ ）逐步恢复分辨率。

首次上采样融合 Stage2 特征：

$$F_{up1} = Conv_{3 \times 3} \left( Concat \left( Upsample(F_{ppm}, 4), F_2 \right) \right) \in R^{128 \times 128 \times 512} \quad (3-5)$$

二次上采样融合 Stage1 特征：

$$F_{up2} = Conv_{3 \times 3} \left( Concat \left( Upsample(F_{up1}, 2), F_1 \right) \right) \in R^{256 \times 256 \times 256} \quad (3-6)$$

最终通过 $1 \times 1$ 卷积与 Softmax 生成分割概率图：

$$P = Softmax \left( Conv_{1 \times 1}(F_{up2}) \right) \in R^{256 \times 256 \times 2} \quad (3-7)$$

Stage3-4 的 CBAM 其插入位置位于残差块瓶颈结构( $1 \times 1 \rightarrow 3 \times 3 \rightarrow 1 \times 1$ )的末端，对通道压缩后的特征进行学习，避免过早引入注意力而导致不可逆的信息损失。

## 3.2 损失函数

在深度学习范畴内，损失函数作为用以评估模型预测输出和真实标签之间差异程度的量化指标，被界定为非负实值函数，它的数学表达式一般呈现为

$L(Y, f(x))$ ), 此函数值的减小大多时候和模型鲁棒性的提升呈现出正相关的关系。对于语义分割任务而言, 研究者们研发出了多种有代表性的损失函数类型, 主要有:

### 3.2.1 交叉熵损失

交叉熵是信息论里的关键概念, 它主要作用是衡量模型预测概率分布跟真实标签分布之间的差异程度, 在优化模型时, 不断把交叉熵损失函数降到最小, 系统就能动态调整内部参数设置, 让模型输出的概率分布慢慢接近真实标签的统计特征, 这种优化方式可精准学习分类决策边界, 最后提升分类性能。对于多类别分类任务, 它的数学表达式是:

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K [y_{i,k} \log(p_{i,k}) + (1 - y_{i,k}) \log(1 - p_{i,k})] \quad (3-8)$$

其中 $N$ 表示图像中像素总数,  $K = 2$ 表示类别只有肾小球和背景。 $y_{i,k} \in \{0,1\}$ 表示像素 $i$ 的真实标签 (one-hot 编码),  $p_{i,k} \in [0,1]$ 表示模型预测像素 $i$ 属于类别 $k$ 的概率。

梯度计算公式如下:

$$\frac{\partial L_{CE}}{\partial p_{i,k}} = -\frac{y_{i,k}}{p_{i,k}} + \frac{1 - y_{i,k}}{1 - p_{i,k}} \quad (3-9)$$

### 3.2.2 Dice 损失

Dice 系数作为一种评估指标, 是基于集合相似度度量的, 在语义分割任务里, 它把预测结果和真实标签当作像素级集合来比对, 其核心计算原理重点考察两者的交集区域, 可反映模型预测结果的精确度以及召回性能。Dice 损失对于前景和背景不均衡的数据集有较好的适应性, 能够避免模型在训练过程中对多数类的过度拟合, 从而更有效地捕捉到少数类的特征。Dice 损失的计算公式为:

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N \sum_{j=1}^C y_{ij} p_{ij}}{\sum_{i=1}^N \sum_{j=1}^C y_{ij} + \sum_{i=1}^N \sum_{j=1}^C p_{ij}} \quad (3-10)$$

梯度计算公式如下:

对单个预测概率 $p_{i,k}$ 求偏导, 可得:

$$\frac{\partial L_{Dice}}{\partial p_{i,k}} = -\frac{2[y_{i,k}(\sum y + \sum p) - \sum y p]}{(\sum y + \sum p)^2} \quad (3-11)$$

其中:  $\sum y = \sum_{i,k} y_{i,k}$  为真实标签总和;  $\sum p = \sum_{i,k} p_{i,k}$  为预测概率总和;  $\sum y p =$

$\sum_{i,k} y_{i,k} p_{i,k}$  为交集总和。

### 3.2.3 交叉熵-Dice 联合损失

针对医学图像分割中肾小球区域与背景的严重类别不平衡问题，本文提出一种动态加权的联合损失函数，其数学形式定义为：

$$L = L_{CE} + \lambda L_{Dice} \quad (3-12)$$

其中  $\lambda \in [0,1]$  为动态平衡系数， $L_{CE}$  为交叉熵损失， $L_{Dice}$  为 Dice 损失。

引入交叉熵损失函数可加快模型训练的收敛速度，结合 Dice 损失函数可提升模型提取小尺度目标特征的能力，缓解因训练数据分布不均衡造成的模型性能下降问题。可根据不同数据集的特点和任务需求，通过对  $\lambda$  的优化调节，灵活分配两种损失的权重，使得模型在保持整体分类准确性的同时，提升对复杂场景和目标结构的分割精度与鲁棒性。

## 3.3 本章小结

本文精心搭建了 CBAM-PSPNet 网络的完整框架，深入分析了其关键模块的技术实现方式，在特征提取阶段，网络以改进型 ResNet50 作为基础架构，于 Stage3 和 Stage4 层级引入注意力机制，结合空洞卷积技术达成感受野扩展以及特征图分辨率维持，在全局特征提取方面，金字塔池化模块借助多尺度自适应池化操作，有效整合不同层次的上下文信息，提升了模型对肾小球多尺度特征的判别能力。在特征重建环节，解码器采用逐级上采样与跨层特征融合策略，逐步恢复空间细节信息，最终经 Softmax 函数输出像素级语义分割结果。

在肾小球显微图像分割任务里，类别分布不均衡这一问题较为突出，针对此，本研究构建了一种联合损失函数，此函数创新性地融合了交叉熵损失与 Dice 系数。通过动态权重系数  $\lambda$  分配权重，平衡逐像素分类精度与区域重叠度优化目标，在训练初期以交叉熵稳定收敛，后期以 Dice 损失强化小目标的敏感性。

## 4 实验

### 4.1 数据集

#### 4.1.1 数据集简介

在本文中，选择的原始数据集合来自 HuBMAP (Human BioMolecular Atlas Program)-Hacking the Kidney。数据集包含 11 张新鲜冷冻和 9 张福尔马林固定石蜡包埋 (Formalin Fixed Paraffin Embedded, FFPE) PAS 染色的肾脏图像。所有的 20 个组织样本均标注了肾小球 FTU。数据集被划分为 15 个训练集与 5 个公开测试集，原始图像以及标注图像如图 4.1：

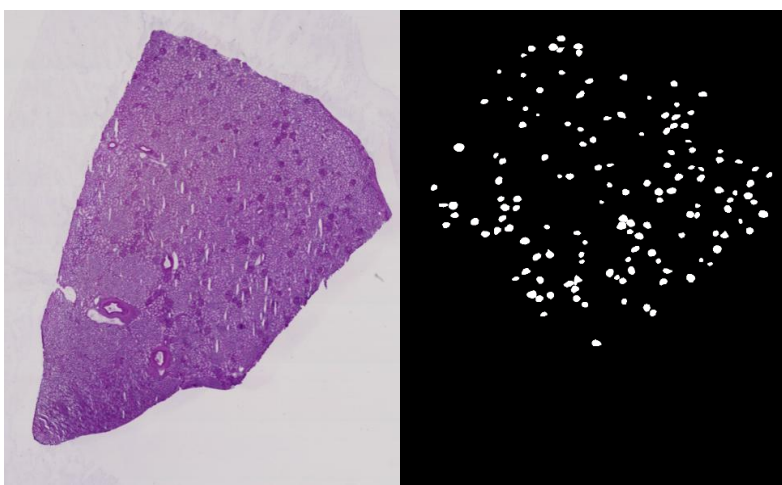


图 4.1 原始图像和肾小球标注图像

由于 HuBMAP 数据集的原始长度、宽度范围从 50000 到 10000 不等，难以直接处理，且原始数据存在大量无细胞组织的空白区域，为了方便实验的进行，将其分割成  $1024 \times 1024$  的大小，并删除完全空白区域的图像，得到共计 2575 组肾小球切片图像。每组图像包含 1 张 3 通道肾小球切片图像和 1 张单通道掩码图像，样例如图 4.2：

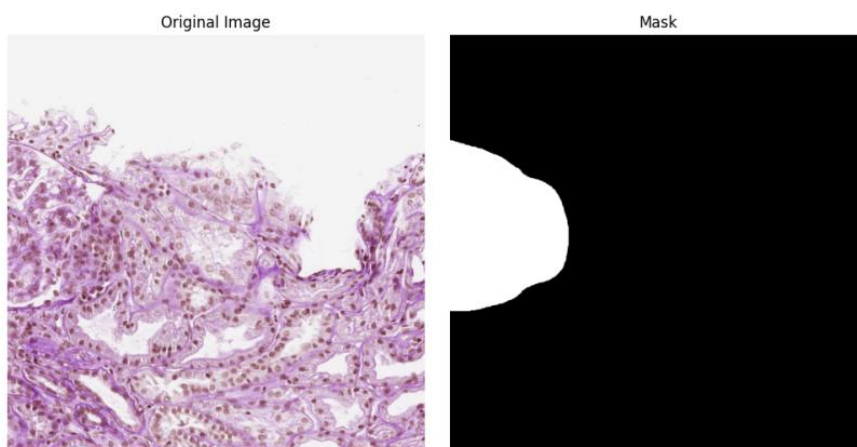


图 4.2 数据集样例图



### 4.1.2 数据增强

为了提高模型训练的泛化性能，有效避免出现过拟合的情况，本研究以由 2575 组肾小球切片图像以及与之对应的掩码图像所组成的原始数据集为对象，系统地采用了多种数据提高策略，运用随机裁剪、水平翻转以及多尺度随机缩放等数据扩增技术来增加样本数量，之后按照 8:2 的比例把提高后的数据集分成训练集和测试集，以此保障模型评估的可靠性。

#### (1) 随机裁剪

随机裁剪裁剪原图中的子区域作为新的数据，可以帮助模型更好的学习图像中的局部特征，避免了模型对整体图像特征的依赖，如图 4.3:

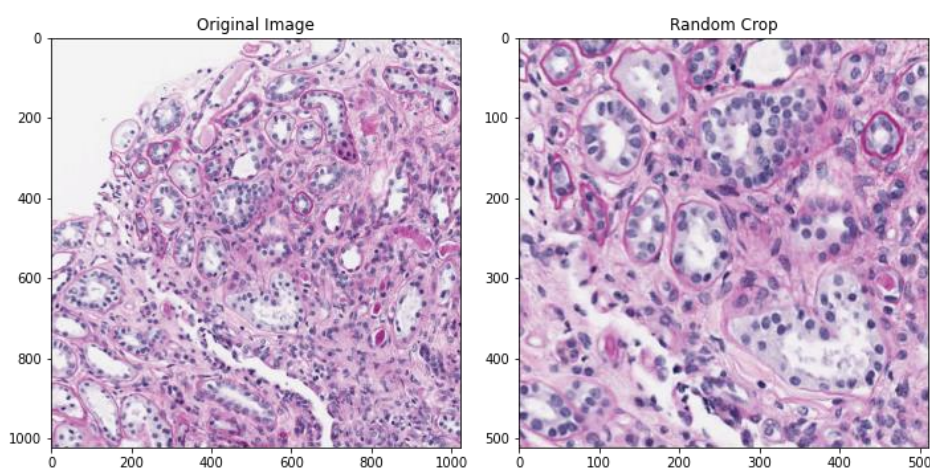


图 4.3 随机裁剪前后的图像

#### (2) 随机水平翻转

在现实世界中，很多物体的方向是随机的，通过随机水平翻转可以模拟这种真实场景中的变化。例如，在拍摄人物照片时，人物可能会向左或向右转身，随机水平翻转图像可以模拟这种不同的拍摄角度，使模型更好地适应真实世界中的各种情况，以增强模型的泛化性能，如图 4.4:

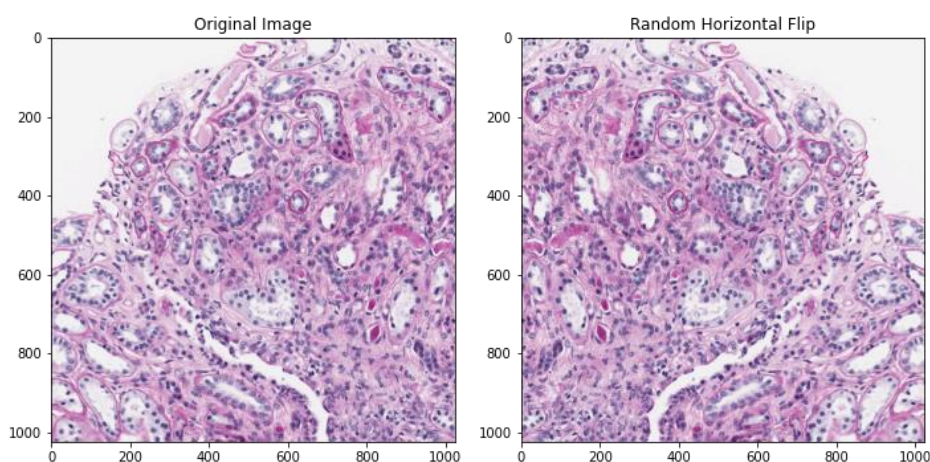


图 4.4 随机水平翻转前后的图像

### (3) 多尺度随机缩放

实验得出的结果显示,运用多尺度随机缩放策略可有效地扩充训练样本的尺寸多样性,如同图 4.5 呈现的那样,此方法可以生成有不同比例特征的图像样本,而且能让模型学习到目标物体在多种尺度下的表征特性,提升检测系统对于不同尺寸目标的识别性能:

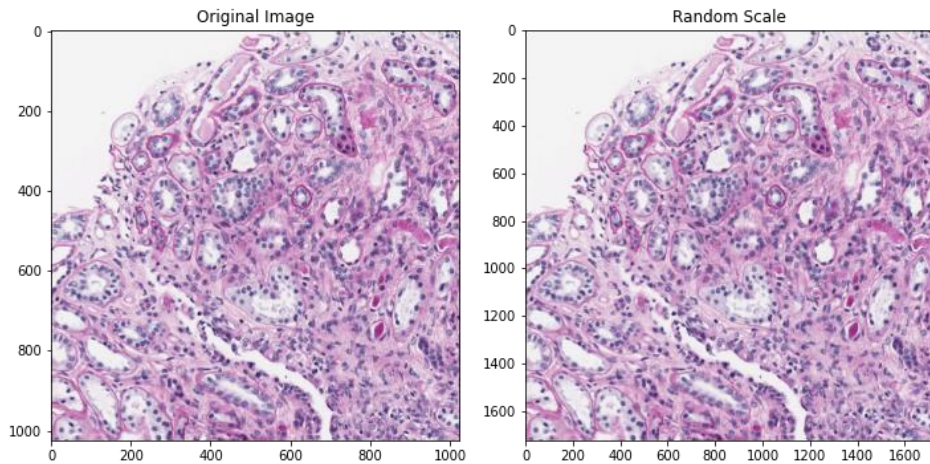


图 4.5 多尺度随机缩放前后的图像

## 4.2 评价指标

### 4.2.1 混淆矩阵

混淆矩阵是评估分类模型性能的一项关键工具,它凭借直观呈现预测结果和真实标签之间的对应关系来呈现其核心价值,该矩阵当中,行向量代表着样本的真实类别归属,列向量反映的是模型的预测分类结果,主对角线元素所表征的是被正确识别的样本数量,而非对角线元素指示的是各类别间的误判情况。

在处理二分类任务的过程中,研究人员一般会选用如表 4.1 所呈现的那种  $2 \times 2$  维度的混淆矩阵,将其当作评估模型性能的基础工具来使用。

表 4.1 混淆矩阵

混淆矩阵		真实值	
		Positive	Negative
预测值	Positive	TP	FN
	Negative	FP	TN

混淆矩阵基础指标如下:

在针对分类模型所展开的性能评估工作里,真正例指的是该模型对于实际属于正类的那些样本做出正确预测的具体数量,假反例意味着模型把实际属于正类

的样本错误地判定为负类的数量情况,假正例体现出模型将实际属于负类的样本错误地归类到正类的数量状况,而真反例对应的则是模型对实际为负类的样本做出正确判别的数量情形。

混淆矩阵二级指标:

准确率是分类模型性能评估里相当关键的一项指标,它所指的是模型预测结果当中正确分类的样本数量和总样本量的比值,该指标具体的计算公式可参考公式(4-1):

$$Accuracy = \frac{TP + TN}{P + N} \quad (4-1)$$

精确率指的是真阳性占预测输出中阳性样本的比值,计算公式如(4-2):

$$Precision = \frac{TP}{TP + FP} \quad (4-2)$$

召回率是评估模型性能时的一项关键指标,它指的是被准确识别出来的阳性样本数量在实际阳性样本总数里所占的比例,该指标的计算方式可借助公式(4-3):

$$Recall = \frac{TP}{TP + FN} \quad (4-3)$$

特异度是评估模型性能的一项关键指标,它被定义为在真实的负样本当中,那些被准确判断为负例的样本在总体中所占的比例,借助这个指标,可对分类器针对负类样本的判别能力给予有效衡量,关于它的数学表达式可以在公式(4-4)中详细看到:

$$Specificity = \frac{TN}{TN + FP} \quad (4-4)$$

#### 4.2.2 核心评估指标

本研究选用 **aAcc** (平均准确率)、**mIoU** (平均交并比)、**mRecall** (平均召回率)、**mPrecision** (平均精确率)、**mFscore** (平均 F1 值) 及 **mDice** (平均 Dice 系数) 作为核心评估指标,各指标具体介绍如下:

(1) **aAcc** (平均准确率): 平均准确率是用来评估模型预测性能的一项关键指标,它的计算原理是依据模型正确分类的像素数量与总像素数量之间的比值关系,这个指标借助对预测结果和真实标签一致性程度的量化,可切实反映出分类模型的整体精度状况。公式为:

$$aAcc = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C \delta(p_{ij}, y_{ij}) \quad (4-5)$$

其中  $N$  为像素总数,  $C$  为类别总数,  $y_{ij}$  表示像素  $i$  的真实类别是否为  $j$ ,  $\delta$  为克罗内克函数,当预测类别  $p_{ij}$  与真实类别  $y_{ij}$  一致时为 1, 否则为 0。  $aAcc$  综合考虑所有类别,数值越高表明模型在整体像素分类上的准确性越好,但在处理类别不



均衡问题时，可能因多数类主导而掩盖少数类的分类效果。

(2) **mIoU** (平均交并比): 在语义分割领域里，平均交并比属于关键的性能评价指标，它是借助计算预测结果跟真实标注之间的区域重合程度，以此来对模型性能给予评估，展开来说，对于该指标中每个类别  $k$  的交并比，可借助下面的数学表达式加以量化：

$$IoU_k = \frac{|P_k \cap G_k|}{|P_k \cup G_k|} \quad (4-6)$$

表示预测为类别  $k$  的像素集合与真实属于类别  $k$  的像素集合， $|\cdot|$  表示集合中元素的数量。 $mIoU$  则是所有类别  $IoU$  的平均值。 $mIoU$  数值越高，意味着模型对各类目标的分割精度越高，能够有效反映模型在不同类别上的综合分割性能，对目标边界的划分准确性较为敏感。

(3) **mRecall** (平均召回率): 平均召回率作为一项用于评估的指标，它的核心功能主要是对模型识别真实目标像素的正确能力加以量化，具体所呈现出来的是，被准确检测到的目标像素和全部真实目标像素之间形成的平均比率，借助以下数学表达式可对这个指标进行精确计算：

$$mRecall = \frac{1}{C} \sum_{k=1}^C \frac{TP_k}{TP_k + FN_k} \quad (4-7)$$

其中， $TP_k$  为类别  $k$  的真正例（模型正确预测为类别  $k$  的像素数）， $FN_k$  为类别  $k$  的假反例（模型错误预测为其他类别的真实类别  $k$  像素数）。 $mRecall$  关注模型对目标的捕捉能力，较高的  $mRecall$  表明模型能够尽可能多地检测到真实目标，减少漏检情况，在处理小目标或稀疏目标时，该指标尤为重要。

(4) **mPrecision** (平均精确率):  $mPrecision$  衡量模型预测为正类的像素中，真正属于该类别的平均比例，其公式为：

$$mPrecision = \frac{1}{C} \sum_{k=1}^C \frac{TP_k}{TP_k + FP_k} \quad (4-8)$$

其中， $FP_k$  为类别  $k$  的假正例（模型错误预测为类别  $k$  的像素数）。 $mPrecision$  反映了模型预测结果的可靠性，较高的  $mPrecision$  表示模型预测的目标像素中，真实目标的占比较大，有效降低了误检率。

(5) **mFscore** (平均 F1 值):  $mFscore$  是综合考虑精确率和召回率的调和平均数，其公式为：

$$mFscore = \frac{2 \cdot mPrecision \cdot mRecall}{mPrecision + mRecall} \quad (4-9)$$

$mFscore$  能够平衡模型在精确率和召回率之间的表现，避免单一指标的片面性。当  $mFscore$  数值较高时，表明模型在准确识别目标的同时，也能有效捕捉到大部分真实目标，是评估模型综合性能的关键指标。

(6) **mDice** (平均 Dice 系数): 平均 Dice 系数作为一项用于衡量相似性的指标，它是基于集合重叠度来计算的，其数学表达式为公式 (4-10)，该指标主要

是依靠对模型预测结果和真实标注之间的重合程度进行量化, 评估二者的相似性水平。

$$mDice = \frac{2}{C} \sum_{k=1}^C \frac{2|G_k \cap P_k|}{|G_k| + |P_k|} \quad (4-10)$$

$mDice$  的取值范围在 0 到 1 之间, 1 表示预测结果与真实标签完全一致。 $mDice$  对前景和背景类别不均衡的数据具有良好的适应性。

### 4.3 肾小球分割网络的设计

本文对 PSPNet 的改进主要从两个方面进行: 在网络设计层面引入通道-空间双路注意力机制; 在损失函数层面, 设计交叉熵-Dice 联合损失函数。

将网络改进方案和损失函数损失方案进行两两组合, 得到四种模型如表格 4.2 所示, 然后将这四种模型在 HuBMAP 数据集上进行实验, 对实验的 Loss 和评价指标进行分析。最后将原 PSPNet 模型与优化方案进行比较, 分析实验结果, 确定改进方案是否在 HuBMAP 数据集上各项评估指标的变化情况。

表 4.2 网络改进方案和损失函数损失方案

PSPNet	/	CBAM
CE	PSP+CE	PSP+CE+CBAM
CE+Dice	PSP+CE+Dice	PSP+CE+Dice+CBAM

### 4.4 实验设计与结果分析

#### 4.4.1 实验环境和网络参数设置

##### (1) 实验环境

本研究将 PyTorch 用作核心深度学习框架, 此框架由 Facebook 人工智能研究院牵头开发, 如今已成为深度学习领域的主流工具之一, 实验还整合了 OpenMMLab 项目旗下的 MMSegmentation 工具箱, 这个基于 PyTorch 的开源语义分割平台有模块化架构设计、标准化评估体系以及多算法支持等特点, 其应用场景包括自动驾驶系统开发、医学影像分析等多个领域, 能为用户自定义分割模型的快速构建给予便利的技术支持。

本次研究以 Featurize 云计算平台为依托来展开实验, 运用 Python 编程语言达成系统开发工作, 表 4.3 当中详细地罗列了在实验进程里所涉及到的关键软硬件配置参数。

表 4.3 软硬件配置

软硬件配置	版本
GPU	RTX 3060(12GB)
内存	32GB
Python	3.7
PyTorch	1.10.1
MMCV	2.0.0rc4
GCC	9.3
ftfy	6.1.1
regex	2024.4.16
mmsegmentation	1.2.2
torchvision	0.11.2

(2) 网络参数设置

为了实验的可重复性，如表 4.4 所示为部分网络参数设置。本实验使用 SGD 优化器，初始学习率设置为 0.01，为了抑制震荡设置动量为 0.9，为防止过拟合，设置 L2 正则化权重衰减为 0.0005。学习率调度器为多项式衰减策略 (PolyLR)，衰减率为 0.9，最小学习率下限为 0.0001。此外，添加 500 次迭代的线性预热，避免初期梯度不稳定。卷积核初始化方案为 Kaiming 初始化。

表 4.4 网络参数设置

参数	数值
优化器	SGD
学习率	0.01
卷积核初始化方案	Kaiming
batch_size	8
iter (迭代次数)	24000
动量	0.9
权重衰减	0.0005
学习率调整器	PolyLR
img_size (图像尺寸)	256
in_chans (输入通道数)	3
num_works (数据加载器的数量)	2

#### 4.4.2 实验结果

四组模型在 Hubmap 数据集上的评估关键指标如表 4.5 所示：

**表 4.5 四个模型在数据集上的结果**

模型设置	测试集 mIOU	测试集 mDice	肾小球类 IOU	肾小球类 Dice
CE	90.72	94.95	84.26	91.46
CE+Dice	91.07	95.11	84.38	91.53
CE+CBAM	86.22	92.04	84.4	91.54
CE+Dice+CBAM	88.38	93.45	84.21	91.43

实验得出的结果显示，运用交叉熵与 Dice 结合而成的复合损失函数，使得模型的整体性能有了提升，尽管引入注意力机制让模型的学习复杂度有所增加，致使训练集上部分评估指标出现了下降，不过凭借对比分析肾小球类别的 IOU 以及 Dice 系数等关键指标可发现，集成注意力模块的模型在整体性能方面并不比基准模型差，并且在特定指标上还呈现出更优的表现。

从数据稳定性方面来考察，运用注意力机制的模型呈现出更加稳健的性能表现，实验结果显示，该模型架构提升了泛化性能，在面对未知数据时能保持较高的预测准确性以及可靠性。

在本文中，将网络改进方案和损失函数方案进行两两组合，得到四种模型，并在 HuBMAP 数据集上进行实验。通过对实验结果的分析，交叉熵-Dice 联合损失函数对模型性能的提升作用得到了验证，引入注意力机制的模型在数据稳定性和泛化能力方面表现更优，证明了所提方法的有效性。

#### 4.4.3 实验数据可视化分析

##### （1）各组训练损失和评估指标分析

##### 第一组：CE 模型

通过实验验证 CE 模型，图 4.6 以直观的方式呈现出了该模型在训练过程里损失函数的收敛趋势，图 4.7 则系统地展示了模型于测试数据集上针对肾小球类别以及背景类别的各项性能指标均值情况，图 4.8 是详细地报告了该模型在肾小球这一特定类别上的各项评估指标的具体表现。

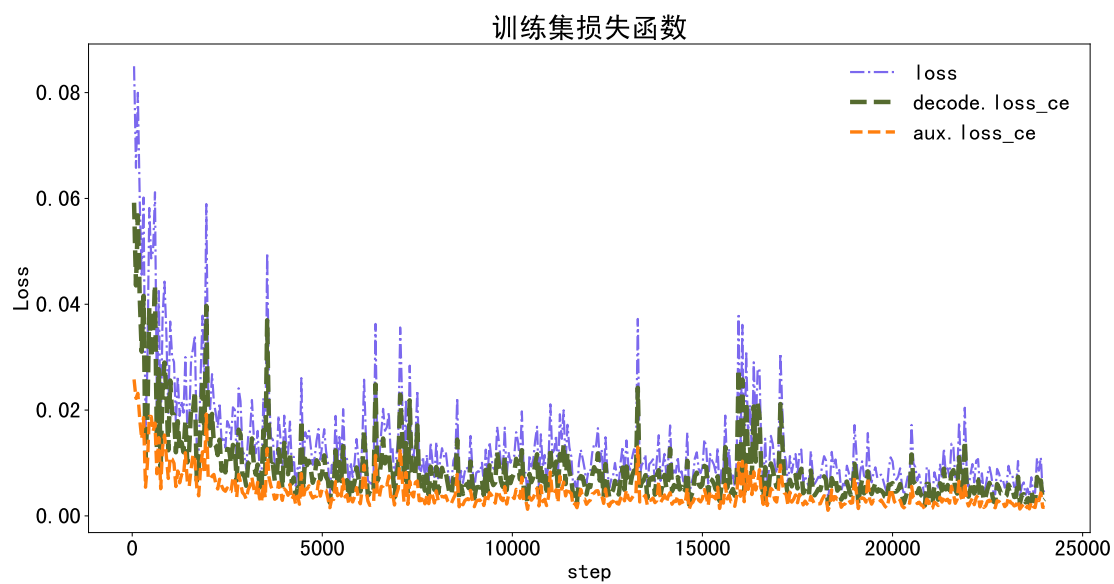


图 4.6 CE 模型的损失

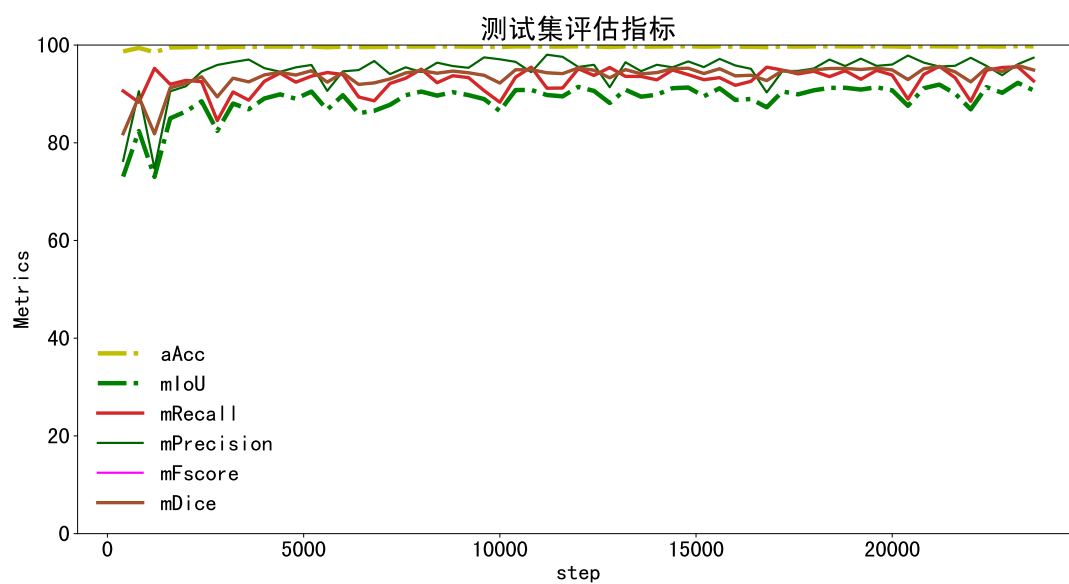


图 4.7 CE 模型在测试集各项评估指标

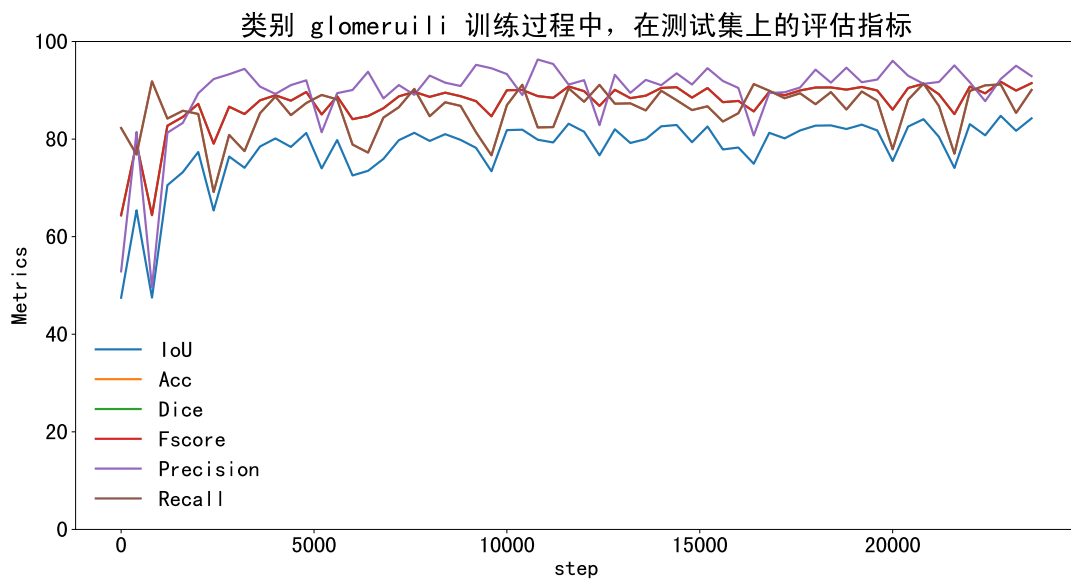


图 4.8 CE 模型在肾小球类别上的各项评估指标

使用交叉熵损失函数时，即原始的 PSPNet，经过 24000 次迭代，训练集上的损失从 0.084 降低到了 0.004，测试集上的 mIOU 从 73.08 提升到了 90.72，mDice 从 81.86 提升到了 94.95，细分到类别时，肾小球类的 IOU 从 47.48 提升到了 84.26，Dice 从 64.38 提升到了 91.46。

#### 第二组：CE+Dice 混合损失函数模型

通过实验验证 CE+Dice 混合损失函数模型，图 4.9 以直观的方式呈现出了该模型在训练过程里损失函数的收敛趋势，图 4.10 则系统地展示了模型于测试数据集上针对肾小球类别以及背景类别的各项性能指标均值情况，图 4.11 是详细地报告了该模型在肾小球这一特定类别上的各项评估指标的具体表现。

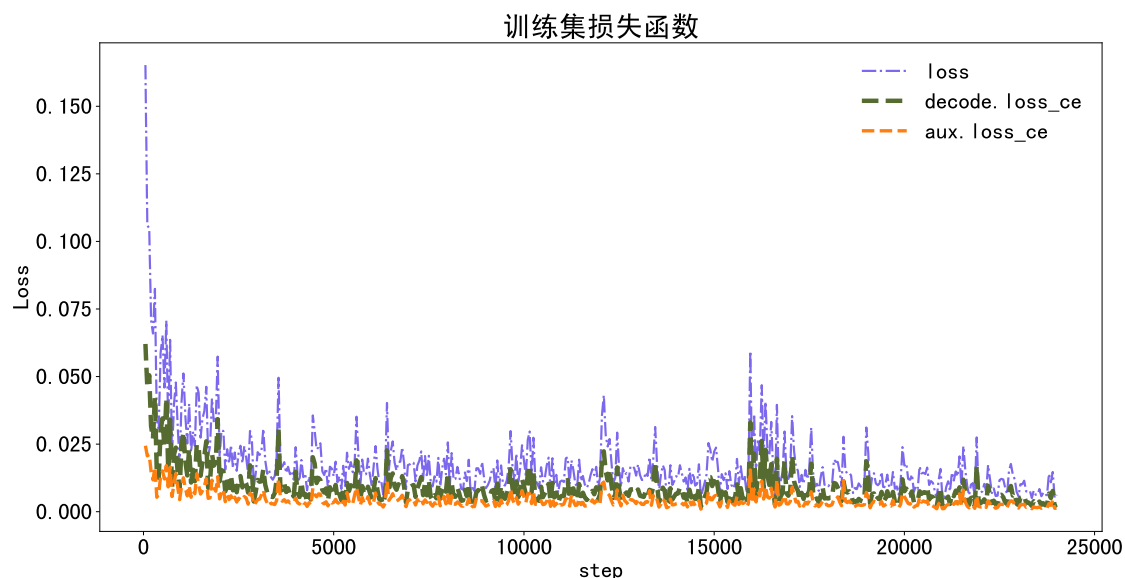


图 4.9 CE+Dice 模型的损失

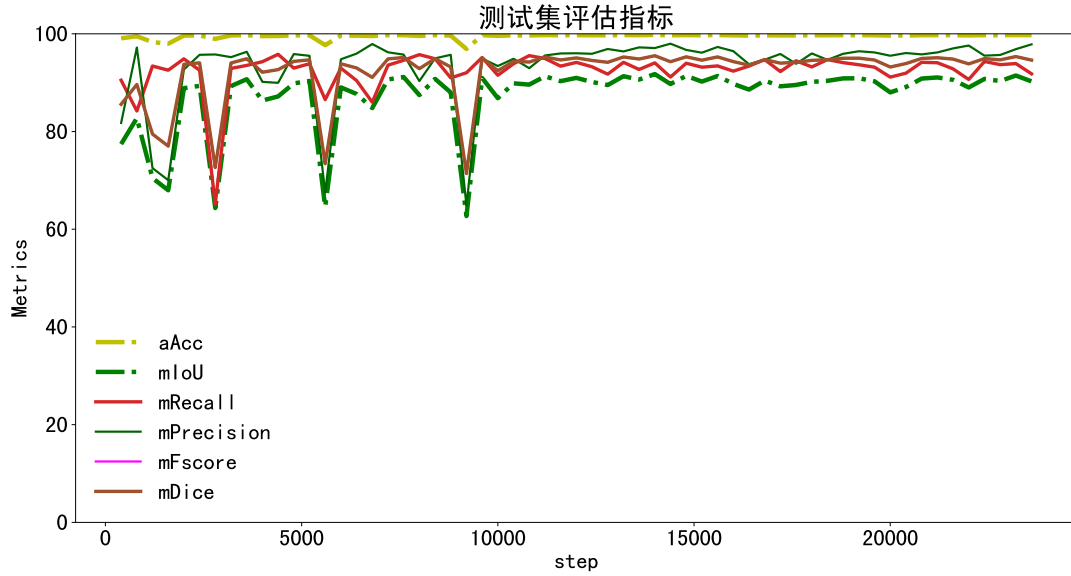


图 4.10 CE+Dice 模型在测试集各项评估指标

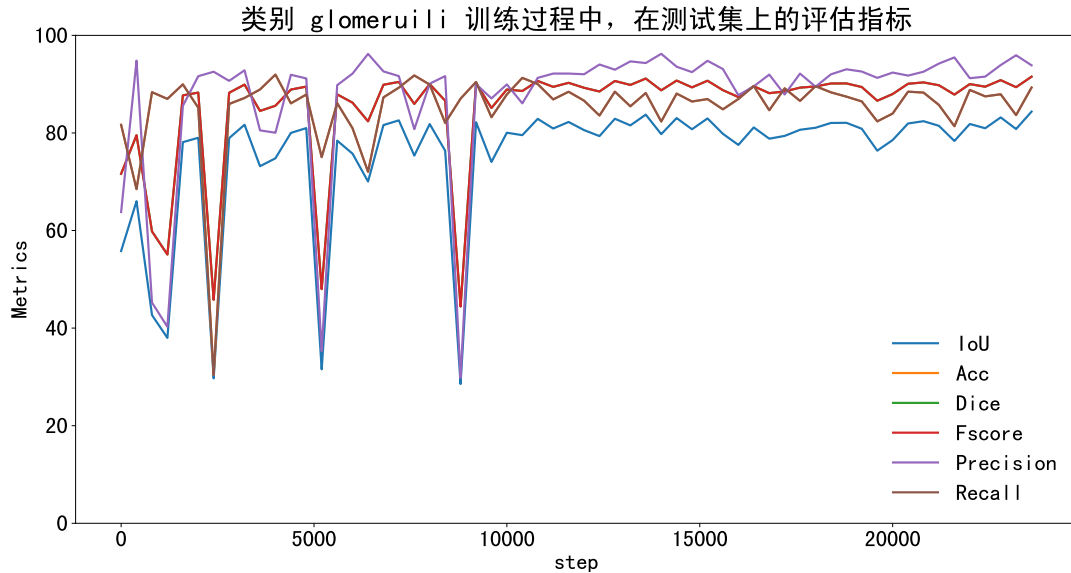


图 4.11 CE+Dice 模型在肾小球类别上的各项评估指标

使用交叉熵-Dice 联合损失函数时，经过 24000 次迭代，训练集上的损失从 0.165 降低到了 0.007，测试集上的 mIoU 从 77.43 提升到了 91.07，mDice 从 85.58 提升到了 95.11，细分到类别时，肾小球类的 IOU 从 55.8 提升到了 84.38，Dice 从 71.63 提升到了 91.53。

### 第三组：CE+CBAM 模型

通过实验验证 CE+CBAM 模型，图 4.12 以直观的方式呈现出了该模型在训练过程里损失函数的收敛趋势，图 4.13 则系统地展示了模型于测试数据集上针对肾小球类别以及背景类别的各项性能指标均值情况，图 4.14 是详细地报告了该模型在肾小球这一特定类别上的各项评估指标的具体表现。

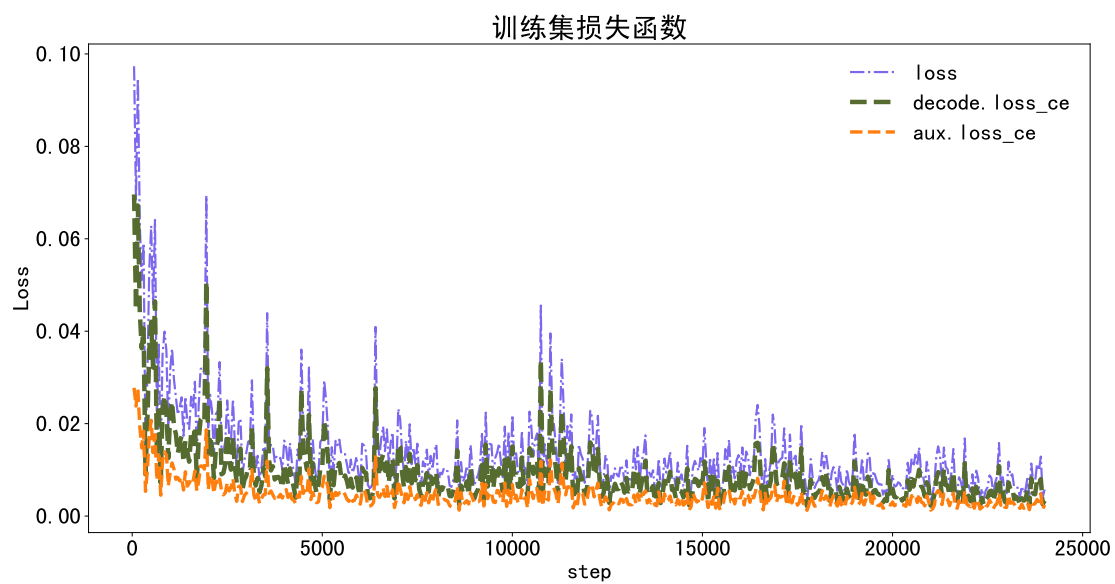


图 4.12 CE+CBAM 模型的损失

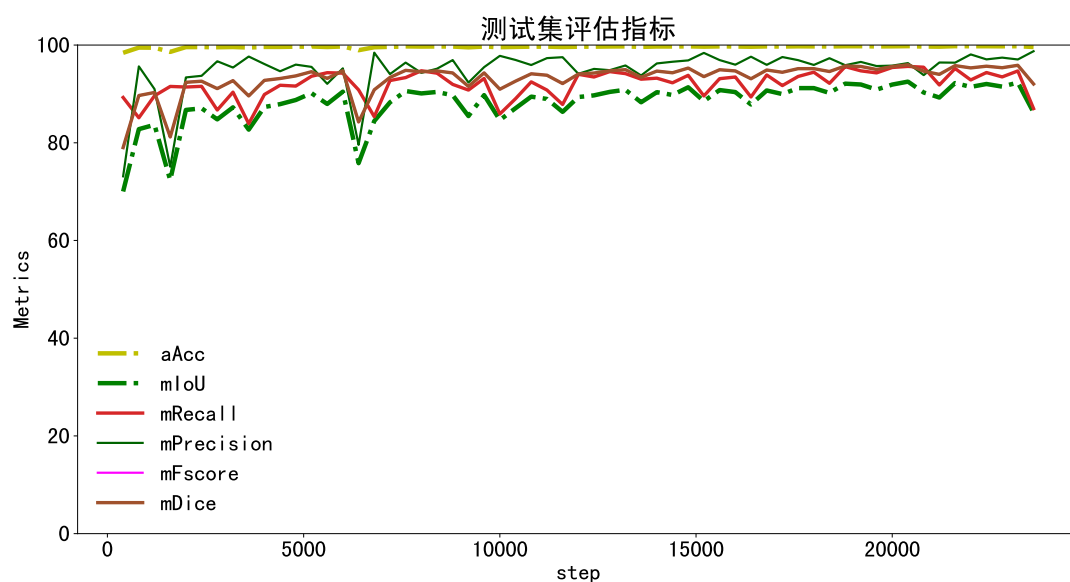


图 4.13 CE+CBAM 模型在测试集各项评估指标



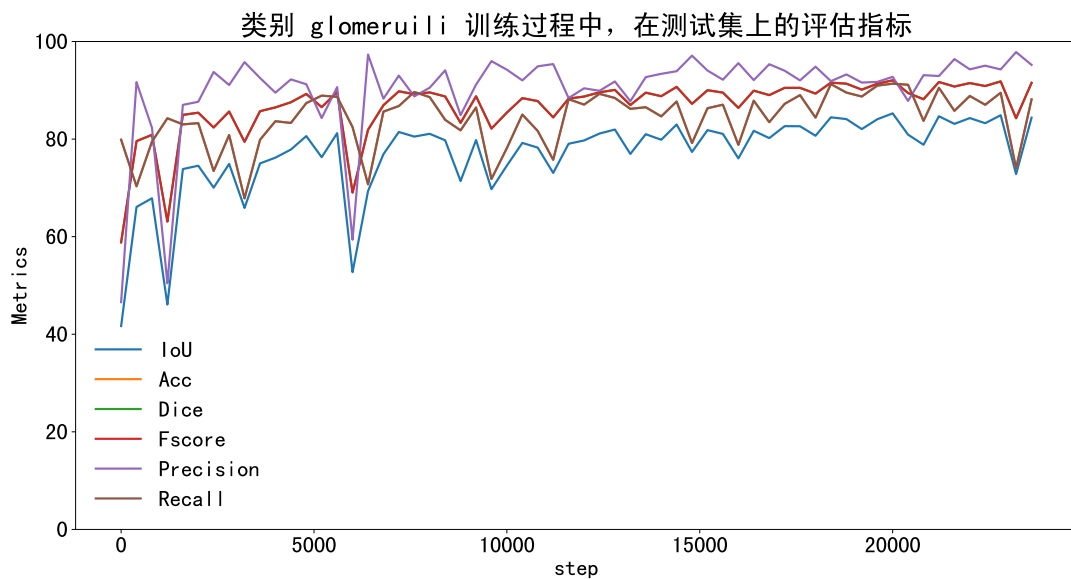


图 4.14 CE+CBAM 模型在肾小球类别上的各项评估指标

使用交叉熵+注意力机制时，经过 24000 次迭代，训练集上的损失从 0.097 降低到了 0.005，测试集上的 mIOU 从 70.03 提升到了 86.22，mDice 从 79.01 提升到了 92.04，细分到类别时，肾小球类的 IOU 从 41.68 提升到了 84.4，Dice 从 58.83 提升到了 91.54。

#### 第四组：CE+Dice+CBAM 模型

通过实验验证 CE+Dice+CBAM 模型，图 4.15 以直观的方式呈现出了该模型在训练过程里损失函数的收敛趋势，图 4.16 则系统地展示了模型于测试数据集上针对肾小球类别以及背景类别的各项性能指标均值情况，图 4.17 是详细地报告了该模型在肾小球这一特定类别上的各项评估指标的具体表现。

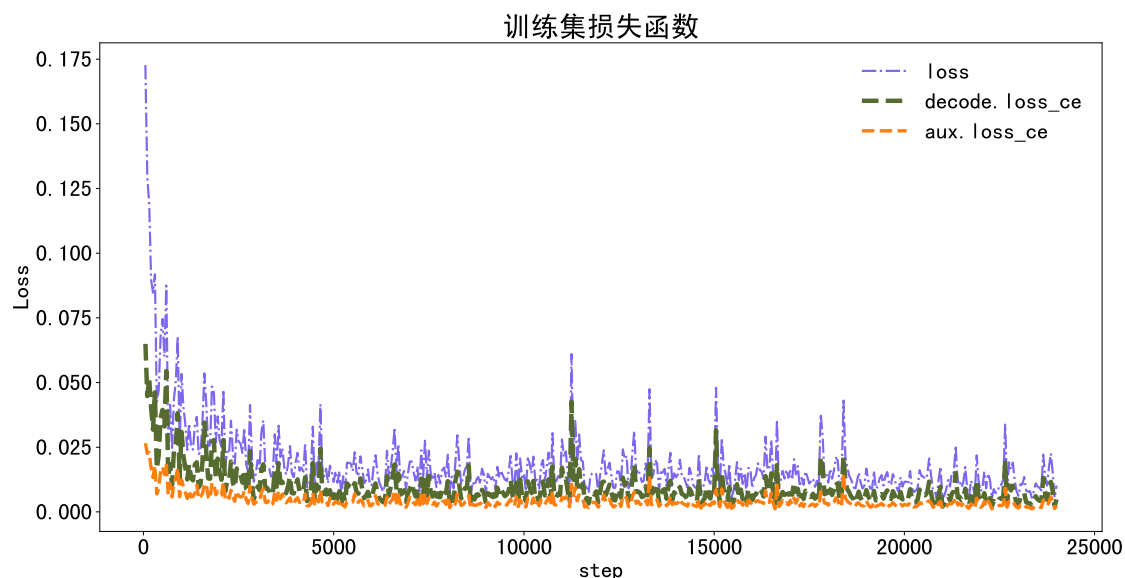


图 4.15 CE+Dice+CBAM 模型的损失

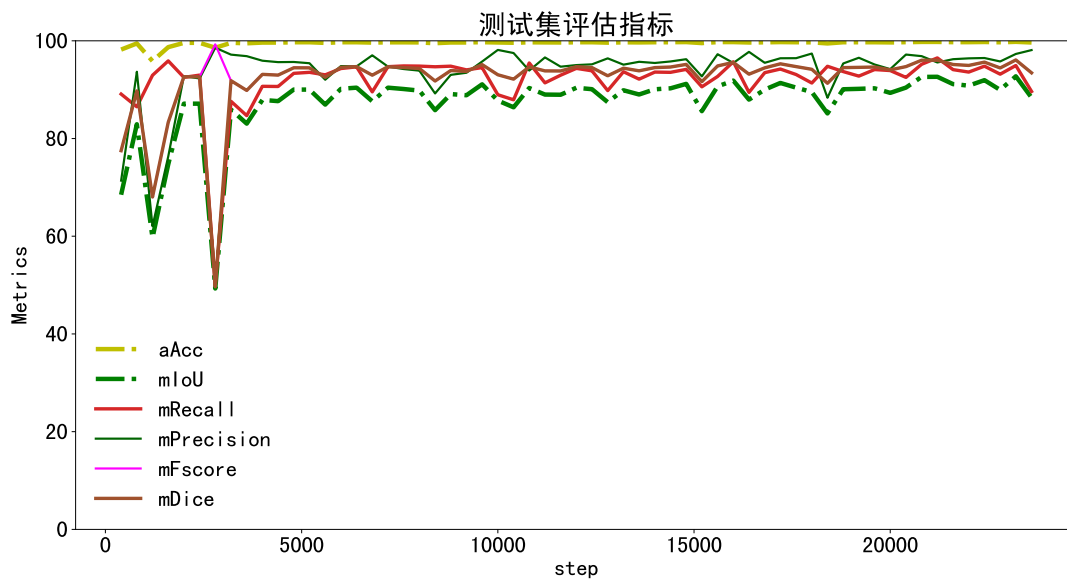


图 4.16 CE+Dice+CBAM 模型在测试集各项评估指标

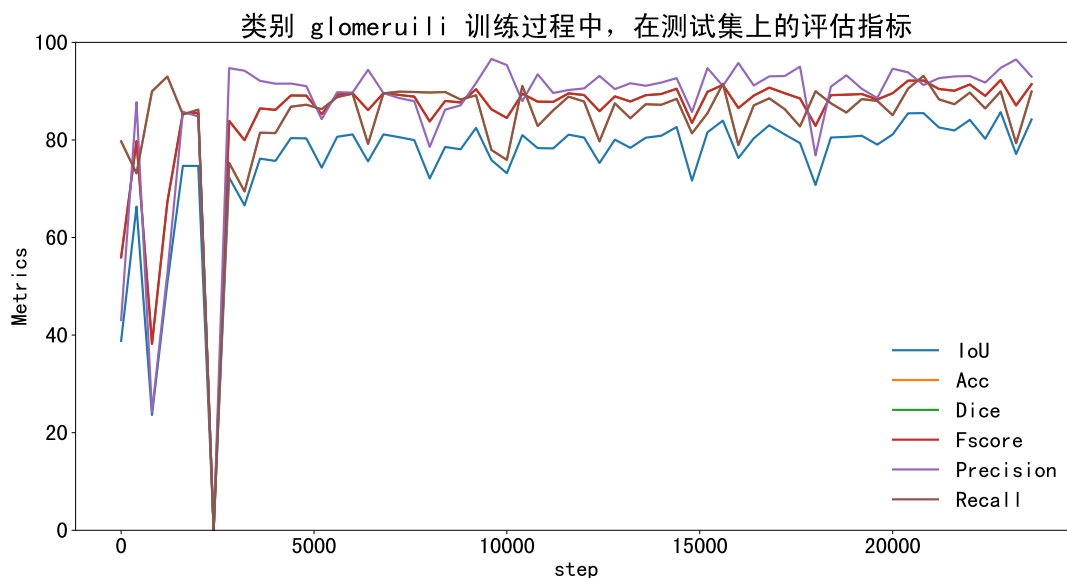


图 4.17 CE+Dice+CBAM 模型在肾小球类别上的各项评估指标

使用交叉熵-Dice 联合损失+注意力机制时，经过 24000 次迭代，训练集上的损失从 0.172 降低到了 0.010，测试集上的 mIOU 从 68.50 提升到了 88.38，mDice 从 77.51 提升到了 93.45，细分到类别时，肾小球类的 IOU 从 38.82 提升到了 84.21，Dice 从 55.93 提升到了 91.43。

#### 4.4.4 模型预测结果可视化

模型训练完成后，构建语义分割模型以实现肾小球图像的分割预测，并将语义分割结果可视化如图 4.18 所示：

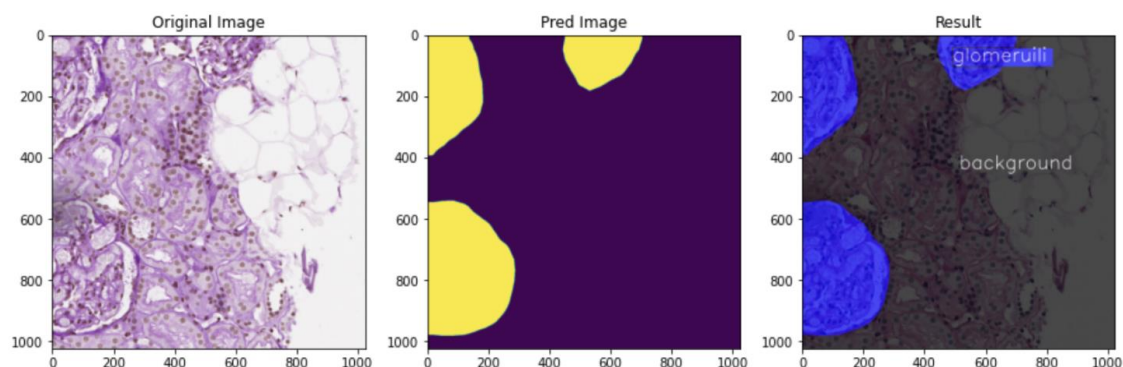


图 4.18 肾小球分割结果可视化

对应的混淆矩阵如图 4.19:

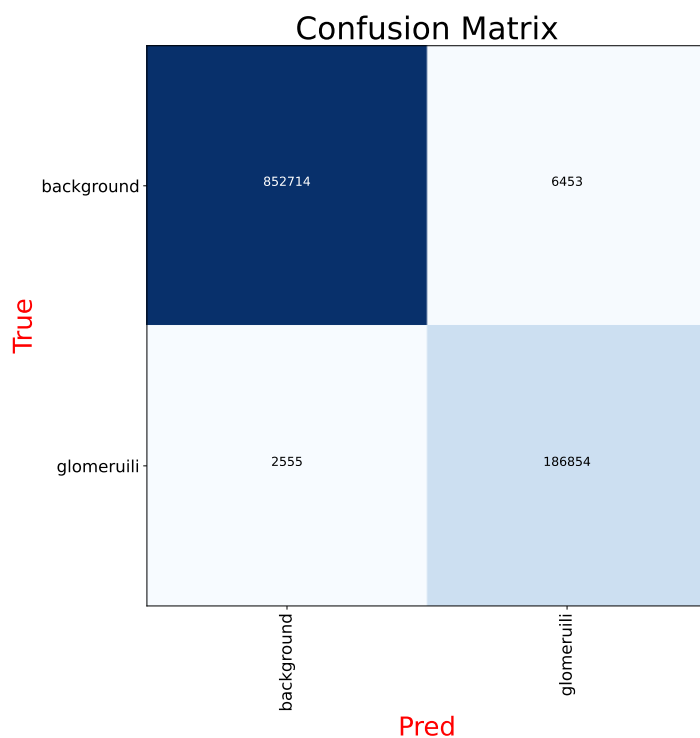


图 4.19 混淆矩阵

#### 4.5 肾小球病理分析系统

为了满足临床和科研对肾小球病理快速分析的需求，本文基于优化的 PSPNet 网络，借助 Streamlit 开发框架设计并实现了一套交互式肾小球病理分析系统。本研究一开始运用经过专业标注的肾小球病理图像数据集来训练深度学习模型，构建出有病理分析功能的神经网络架构，之后借助 Web 平台部署训练好的智能诊断模型，以此达成肾小球病理图像的自动化分析以及高效诊断功能。

本文以 Python 3.7 作为开发环境，将 Streamlit 框架和 PyTorch 深度学习库相结合，打造出专业的病理图像处理平台，之采用 Streamlit 框架，是因为它有高效

的交互式 Web 界面开发能力，可用简洁代码完成快速原型开发，极大提升了用户上传病理图像以及可视化分析结果的便利程度。PyTorch 框架凭借其动态计算图的独特优势，为多尺度病理图像特征的深度学习模型训练与部署提供了灵活的解决办法，特别适合处理病理图像分析里复杂的特征提取任务。

开发环境如表 4.4 所示：

表 4.4 开发环境

操作系统	Windows11
深度学习框架	Pytorch1.10.2
Streamlit	1.30.0
Python 版本	3.7

#### 4.5.1 Web 功能实现路径

本文中的软件功能实现如图 4.20 所示：

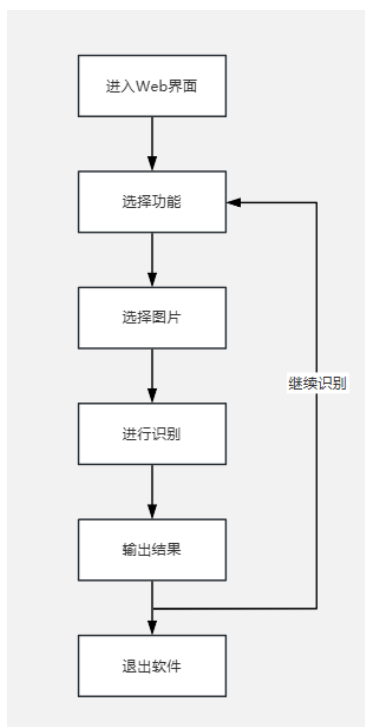


图 4.20 软件功能实现

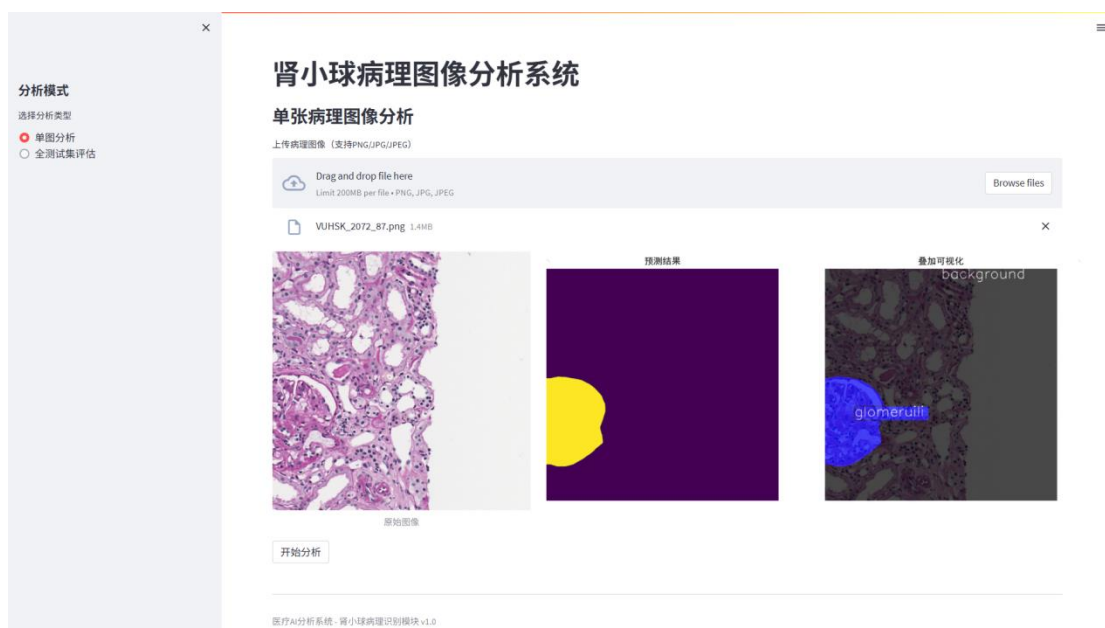
#### 4.5.2 运行效果展示

进入肾小球病例分析系统 Web 界面，左侧是功能选择，包括单图分析和测试集评估两大功能。如图 4.21 所示：



图 4.21 肾小球病理图像分析系统展示

单张图像分析功能下，首先需要上传肾小球切片图像，再点击开始分析即可得到预测结果，并实现了预测图像与原图像的叠加可视化功能。如图 4.22 所示：



4.22 单张图像分割结果展示

在全测试集性能评估模式下，需要上传肾小球切片图像和对应的掩码图像，同时支持多文件上传评估。点击开始全面评估即可得到评估结果，包括混淆矩阵和各类别的分类指标。单文件、多文件评估如图 4.23、图 4.24 所示：

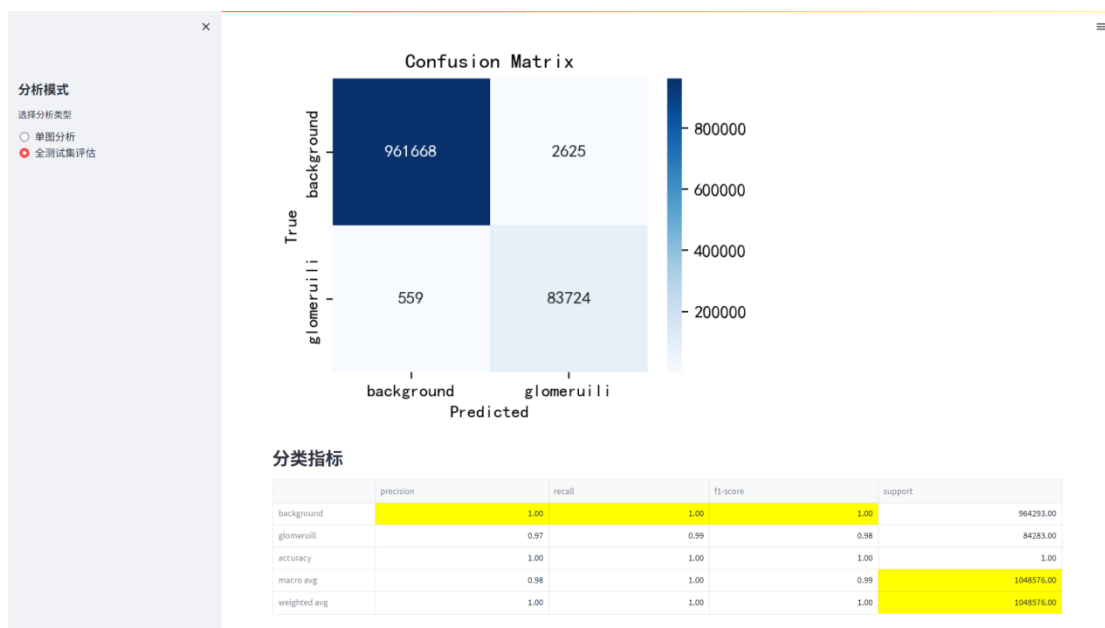


图 4.23 单张图像分割评估指标

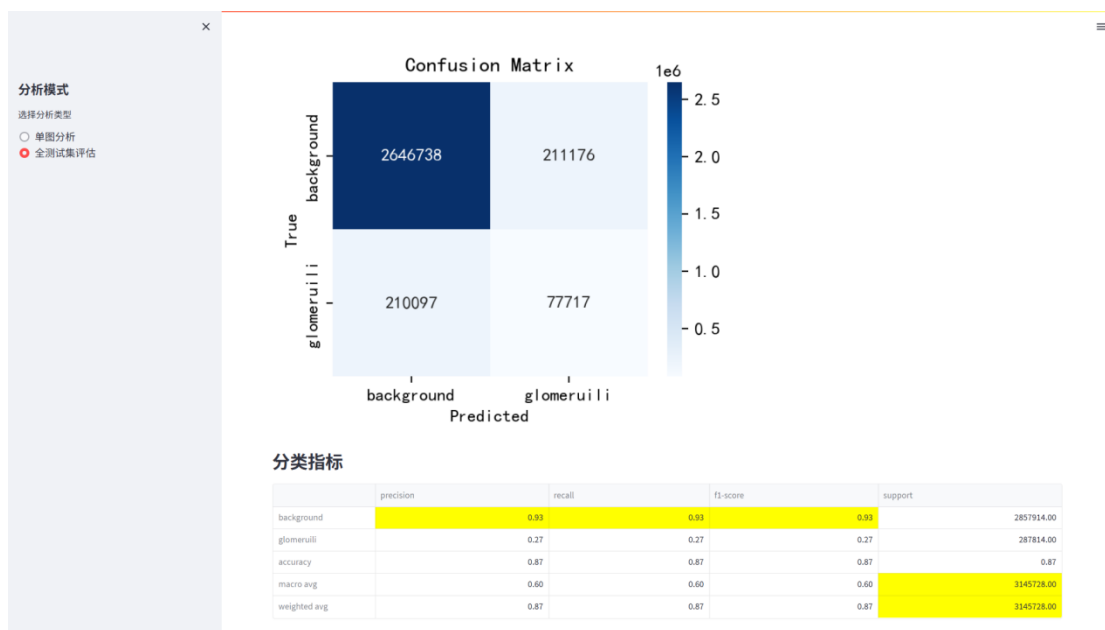


图 4.24 多张图像分割评估指标

同时，借助 Streamlit 强大的跨设备协同能力，该系统支持在同一局域网环境下实现多终端无缝连接。医生可以通过手机、平板等移动设备访问系统。另外本系统集成移动终端图像采集预测功能，医生可直接使用手机摄像头拍摄肾小球显微图像，并即时上传至系统进行智能分析预测，如图 4.25。该系统将原本需要在实验室特定设备上完成的图像分析流程，转化为可以随时随地操作，提升了医疗人员诊断效率与临床工作的灵活性。

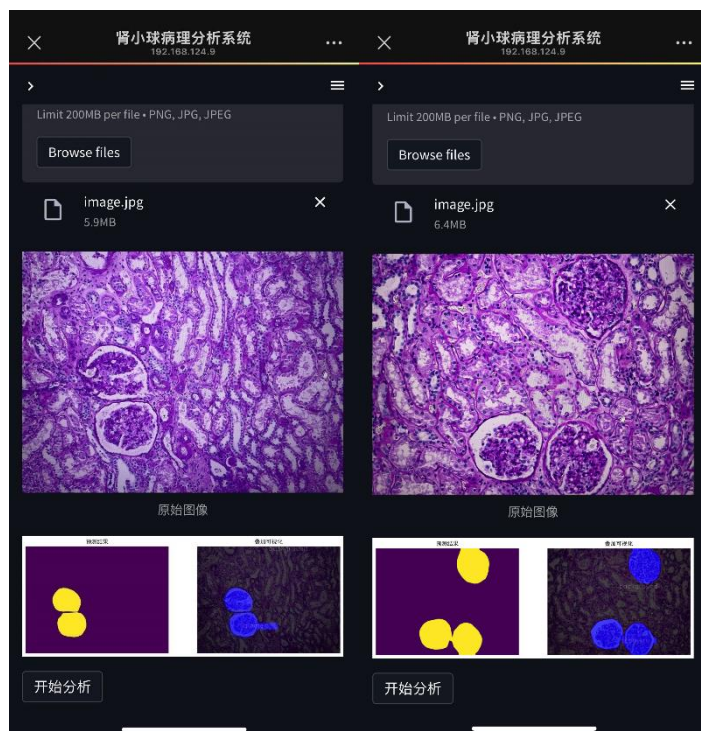


图 4.25 移动端肾小球分割展示

#### 4.6 本章小结

在本章中首先介绍了本实验所使用的数据集，图像初步处理以及所使用的数据增强技术，然后对语义分割领域使用的评估指标进行了详细介绍，接着对网络的改进策略进行了两两组合，得到四个模型。后面介绍了网络的参数设置、实验环境与硬件配置。接着进行对比实验，对多项指标进行了分析并将其可视化。最后设计了肾小球病理分析系统，介绍了开发环境与功能实现，并对肾小球病理分析系统的功能进行了介绍，将预测结果以及各项指标、混淆矩阵进行可视化。



## 5 总结和展望

### 5.1 全文总结

本文旨在使用卷积神经网络实现医学肾小球图像的自动分割,本文对于原始的 PSP-Net 做出了改进,以用于医学肾小球图像分割任务中。针对 PSPNet 网络模型在图像分割上存在的问题,本文从网络架构和损失函数两个方面对原 PSPNet 模型进行了改进。相较于原模型,虽然加入注意力机制后的模型的学习难度有所增加,但是新模型的整体表现仍然持平或高于原始模型。本文的主要工作如下:

(1) 本研究一开始就对实验数据的获取途径给予清晰说明,接着又对数据预处理的具体操作流程进行了细致的阐述,鉴于样本容量存在不足的这种情况,运用数据提高技术来对原始数据集开展扩充工作。

(2) 分别从特征融合网络以及损失函数对原 PSPNet 模型进行改进,提出 CBAM 增强的 ResNet50 网络和交叉熵-Dice 联合损失函数,分别对这两种改进策略进行对比试验,通过对其性能对比和分割效果图的综合分析,改进方案相较原模型在肾小球分割上的不足之处有所改善。

(3) 基于改进的 PSPNet 训练语义分割模型并搭建肾小球病理分析系统,介绍了整个训练环境,包括深度学习框架、实验所用到的软件以及实验过程中的参数设置。此肾小球病理分析系统主要由功能模块以及应用模块共同组成,它的关键功能是针对肾小球组织图像展开高精度的分割工作,并且还会将分割结果进行可视化呈现,同时展示多维度的定量评估指标。

### 5.2 未来展望

尽管上述模型提升了部分实验指标,在实际应用过程中仍暴露出一些问题:首先是在训练数据集中未能充分考虑到真实场景下数据的异质性和多模态性;其次是对网络结构的设计仍有待于完善和提高。本研究未来的改进方向有以下几点:

(1) 注意力机制具有良好的鲁棒性和稳定性,在很多情况下都取得了很好的结果。但是在实际应用过程中发现,如果数据集中样本特征比较均衡或者样本中的关键特征因为一些干扰因素不能有效的被提取出来的话,则容易造成注意力机制所赋予的权重不准确的问题,从而降低了整个网络的效果。所以接下来可以继续尝试着去改进网络使得它能够在各种环境下都能够更好的工作。

(2) 引入注意力机制之后,网络的模型参数规模有所提升,计算复杂度也跟着提高,这在很大程度上增加了网络出现过拟合的概率。因此未来研究可以探索新型轻量化网络架构设计,或使用神经架构搜索技术自动优化网络,以降低模型复杂度。

(3) 注意力机制还存在着一个重要的缺点就是它的超参选择会影响最终的

结果。而目前大部分研究人员还是采用手动的方式来进行调优，这样不仅耗时而且效率很低下。所以接下来也可以结合贝叶斯优化,强化学习等先进的智能算法来建立超参的自适应调节器，将它们融入到整体的网络中，实现超参的自动化调节。

## 参考文献

- [1] Xu Ricong , Cao Tao , Xu Yi , Liao Ying , Li Zhijian , Wan Qijun. Analysis of association between segmental glomerulosclerosis and renal function decline in IgA nephropathy[J]. Chinese Journal of Nephrology, 2020, 36(11): 851-857.
- [2] Otsu N. A threshold selection method from gray-level histograms. IEEE TSMC, 1979.
- [3] Adams R, et al. Seeded region growing. IEEE TPAMI, 1994.
- [4] Canny J. A computational approach to edge detection. IEEE TPAMI, 1986.
- [5] Sobel I, et al. A  $3\times 3$  isotropic gradient operator for image processing. Stanford AI Project, 1968.
- [6] Ojala T, et al. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE TPAMI, 2002.
- [7] Jain A K, et al. Unsupervised texture segmentation using Gabor filters. PR, 1991.
- [8] Hu M K. Visual pattern recognition by moment invariants. IRE TIT, 1962.
- [9] Cortes C, et al. Support-vector networks. ML, 1995.
- [10] Breiman L. Random forests. ML, 2001.
- [11] Long J, et al. Fully convolutional networks for semantic segmentation. CVPR, 2015.
- [12] Ronneberger O, et al. U-Net: Convolutional networks for biomedical image segmentation. MICCAI, 2015.
- [13] 温爱卿. 基于改进 FCN 的腹部图像分割方法研究 [D]. 广东: 华南理工大学, 2020.
- [14] 温佳圆, 林国钰, 张逸文, 等. 应用深度学习网络实现肾小球滤过膜超微病理图像的语义分割[J]. 中国医学物理学杂志, 2020, 37(02): 195-204.
- [15] 赵旭. 基于医学先验的多尺度乳腺超声肿瘤实例分割方法[D]. 哈尔滨工业大学, 2019.
- [16] 王佳. 基于卷积神经网络的前列腺 MR 图像语义分割方法研究[D]. 天津工业大学, 2019.
- [17] 王晗, 杨晶东, 吕江涛, 等. 一种新颖的医学图像分割 CNN 模型 [J/OL]. 电子科, 1-9[2025-06-04].
- [18] 郭宇辰, 罗会兰. 基于全局多分支融合匹配注意力的医学图像分割模型 [J]. 软件导刊, 2025, 24 (04): 147-153.
- [19] Lecun Y, Bottou L. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [20] Nair V, Hinton G E. Rectified Linear Units Improve Restricted Boltzmann

- Machines Vinod Nair[J].Omnipress, 2010.
- [21] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60 (6): 84-90.
- [22] He K , Zhang X , Ren S ,et al. Deep Residual Learning for Image Recognition[J].IEEE, 2016.
- [23] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [24] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.
- [25] Bahdanau D, et al. Neural machine translation by jointly learning to align and translate. ICLR 2015.
- [26] Vaswani A, et al. Attention is all you need. NeurIPS 2017.
- [27] Woo S, Park J, Lee J Y,et al. CBAM:Convolutional Block Attention Module[C]//European Conference on Computer Vision. Springer, Cham, 2018: 3-19.

## 致谢

在这篇论文完成之际，内心感慨万千，这段写作论文的时光，是我最难忘的日子，许多人在这个过程中给予了我无私的帮助与支持，在此我要向所有人表示我的感谢。

非常感谢我的导师芦旭副教授。从论文的选题、开题到资料的收集、框架的搭建，再到初稿的修改以及最终的定稿，导师都始终给予我悉心的指导。每一次与导师的交流，都能让我从迷茫中找到方向，她严谨的治学态度和耐心的教导，让我顺利完成了论文。

感谢大数据专业的各位老师，在大学期间，是你们的授课让我打下了坚实的专业基础，你们为我上课时毫无保留地传授知识，分享经验，增加了我的专业知识，是你们教会了我在课堂上学到的知识，使我能够把学到的知识充分运用在我的这篇论文上，并以此来完成它。

同时也要感谢一起学习生活的同学们，在我论文完成之际经常与我探讨，你们的分享、帮助给予了我很大的灵感。你们在我不懂的时候给我鼓劲加油，让我感觉到集体的力量。还有我的家人，一直都在背后默默地支持着我，在我辛苦撰写论文的这一段日子里，他们一直在默默无闻地为我提供支持，没有他们的帮忙，我不敢想象我能独自一人完成这篇论文。

最后感谢审核论文的各位老师，感谢你们在百忙之中抽出时间对我的论文进行评审和指导，你们的意见和建议将成为我今后学习和研究的宝贵财富。

行文已毕，论文终已完成，它并非是一个终点，而是代表了另一个起点，我会继续前进，我会尽全力让自己的生活更加精彩！