

# 安徽大学

## 本科毕业论文（设计、创作）

题目：基于深度学习的心脏图像分割算法研究与系统实现

学生姓名：凌兴 学号：X02014156

院（系）：计算机科学与技术学院 专业：计算机科学与技术

入学时间：2020 年 9 月

导师姓名：杜秀全 职称/学位：教授

导师所在单位：安徽大学

完成时间：2024 年 4 月

# 基于深度学习的心脏图像分割算法研究与系统实现

## 摘要

在医生对人体心脏影像分析前，对心脏各个结构进行分割通常是至关重要的一步，其目的是图像数据按心脏结构分成不同的区域，以便后续进行医学检查。然而对心脏各个结构进行分割是耗时又费力的，且这一过程易受到医疗人员的工作状态的影像，因此实现自动化的心脏各结构分割就成为了一项有价值的研究内容。

近年来，深度学习方法在许多传统领域都创造了巨大的成功，同样地，在医学图像分析领域中，深度学习方法同样取得了许多成就。然而，深度学习方法通常需要取得大量的标注数据以用作训练，而医学图像由于伦理问题等，导致其具有相对稀缺的特点，而且由于医学场景的复杂，导致虽然都是对于心脏疾病的检查，可以采用的心脏成像方式种类却很多，这更加剧了医学影像数据稀缺的问题，因为大部分方法都是针对某一类数据所提出，无法有效利用其他种类的数据，因此这一问题难以忽视。此外，在心脏影像成像中，不同心脏结构之间往往存在某些难以辨别的分界边缘，这无疑降低了深度学习算法对心脏结构分割的性能。因此，基于上述问题，本文提出了一种基于生成对抗网络的数据增强以及边缘感知的心脏分割算法，该方法一方面利用循环一致性生成对抗网络缓解数据稀缺的问题，另一方面又通过对心脏结构边缘添加额外监督以及设计一种基于注意力的特征融合模块来减轻心脏结构边缘模糊对深度学习模型带来的负面影响。

**关键词：**深度学习；医学影像处理；心脏结构分割；对抗生成网络；特征融合

# **Research and System Implementation of Heart Image Segmentation Algorithm Based on Deep Learning**

## **Abstract**

Before doctors analyze cardiac imaging, it is usually crucial to segment various structures of the heart. The purpose is to divide the image data into different regions according to the cardiac structures for subsequent medical examinations. However, segmenting various structures of the heart is time-consuming and laborious, and this process could be easily influenced by the working status of medical personnel. As a result, achieving automated segmentation of various cardiac structures has become a valuable research topic.

In recent years, deep learning methods have achieved lots of great successes in many traditional fields. Similarly, in the field of medical image analysis, deep learning methods have also made many achievements. However, deep learning methods typically require a large amount of annotated data for training. Due to ethical issues and other factors, medical images tend to be relatively scarce. More importantly, due to the complexity of medical scenarios, although all are examinations for cardiac diseases, there are many types of cardiac imaging techniques that can be used. This exacerbates the problem of scarce medical imaging data because most methods are proposed for a certain type of data and cannot effectively utilize other types of data. Therefore, this problem is difficult to ignore. Additionally, in cardiac imaging, there are often some indistinguishable boundaries between different cardiac structures, which undoubtedly reduces the performance of deep learning algorithms in segmentation of cardiac structures. Therefore, based on the above issues, this paper proposes a cardiac segmentation algorithm based on generative adversarial networks for data augmentation and edge perception. This method utilizes a cycle-consistent generative adversarial network to alleviate the problem of data scarcity while adding extra supervision to the edges of cardiac structures and designing an attention-based feature fusion module to mitigate the negative impact of blurred edges of cardiac structures on deep learning models.

**Keywords:**word1; deep learning; medical image analysis; segmentation of cardiac structures; generative adversarial networks; feature fusion

# 目 录

1 绪论 .....	1
1.1 研究背景和意义 .....	1
1.2 国内外研究现状 .....	2
1.2.1 对于网络结构做出的改进 .....	2
1.2.2 对于数据增强方法做出的改进 .....	3
1.3 本文工作和贡献 .....	4
1.4 章节安排 .....	4
2 医学影像语义分割技术背景 .....	4
2.1 基于深度学习的心脏图像语义分割技术相关知识 .....	4
2.1.1 注意力机制 .....	4
2.1.2 生成对抗网络 .....	5
2.1.3 常用损失函数 .....	6
2.1.3.1 交叉熵损失 .....	6
2.1.3.2 Dice 损失 .....	6
2.1.3.3 交叉熵和 Dice 的混合损失 .....	6
2.1.4 常用评价指标 .....	7
2.2 数据集介绍 .....	7
3 基于生成对抗网络的数据增强以及边缘感知的心脏分割方法 .....	7
3.1 基于循环一致性生成对抗网络的 MRI 域至 CT 域的数据转换方法 .....	7
3.1.1 循环一致性生成对抗网络 .....	8
3.2 基于额外边缘任务以及特征融合模块的边缘感知方法 .....	9
3.2.1 额外边缘任务设计 .....	9
3.2.2 基于注意力的特征融合模块设计 .....	10
3.2.3 损失函数设计 .....	10
3.3 整体架构 .....	11
3.4 相关实验 .....	12
3.4.1 实验设置 .....	12
3.4.2 对比实验结果与分析 .....	12
3.5 本章小结 .....	14
4 心脏图像分割系统设计与实现 .....	15
4.1 开发环境 .....	15
4.2 架构设计 .....	15
4.3 用户使用流程 .....	16

4.4 系统展示 .....	17
4.5 本章小结 .....	17
5 总结与展望 .....	17
5.1 工作总结 .....	17
5.2 展望 .....	17
主要参考文献 .....	19
致 谢 .....	22

# 1 绪论

## 1.1 研究背景和意义

世界卫生组织的调查显示，全球超 30%的死亡是由心血管疾病引起，而根据《中国心血管病报告》一文，我国目前心血管疾病患者的人数高达 3.2 亿<sup>[1]</sup>，因此目前，心血管疾病在中国已经成为了人们健康的主要杀手之一。

心血管疾病种类很多，包括冠心病、高血压、肺源性心脏病、先天性心脏病等，且大部分心血管疾病都伴随有心血管结构的改变，例如肺源性心脏病是一种由于肺动脉高压引起右心室肥大和左心室射血减少的一种心脏疾病<sup>[2]</sup>；风湿性心脏病则会有左心房和右心室增大等<sup>[3]</sup>。

医学影像是一种通过不同的成像技术，如计算机断层扫描（CT）、核磁共振成像（MRI）、超声波、X 射线等，来获取人体内部结构、功能和病变的影像。这些影像可以帮助医生进行诊断、治疗和监测疾病的进展。实际上，在临床上对于心血管疾病的诊断，计算机断层扫描（CT）和磁共振成像（MRI）是两种最常见的检查方法。近年来，随着医学影像技术的不断发展和普及，越来越多心血管疾病患者会进行医学影像检查，而传统的人工手动处理影像样本的方法需要具有专业知识的医生参与，且易受医生主观因素影响，因此难免存在标准不一、效率不高等缺点，在国内这种医生少而病人多的情况下，大量的心血管疾病病患给医疗系统带来了巨大压力。

医生在诊断时，常常需要对心脏影像进行分析，而分析的一项内容就包括区分出不同心脏结构，这与心脏各结构分割任务目标相同。心脏各结构分割是指从心脏影像中对

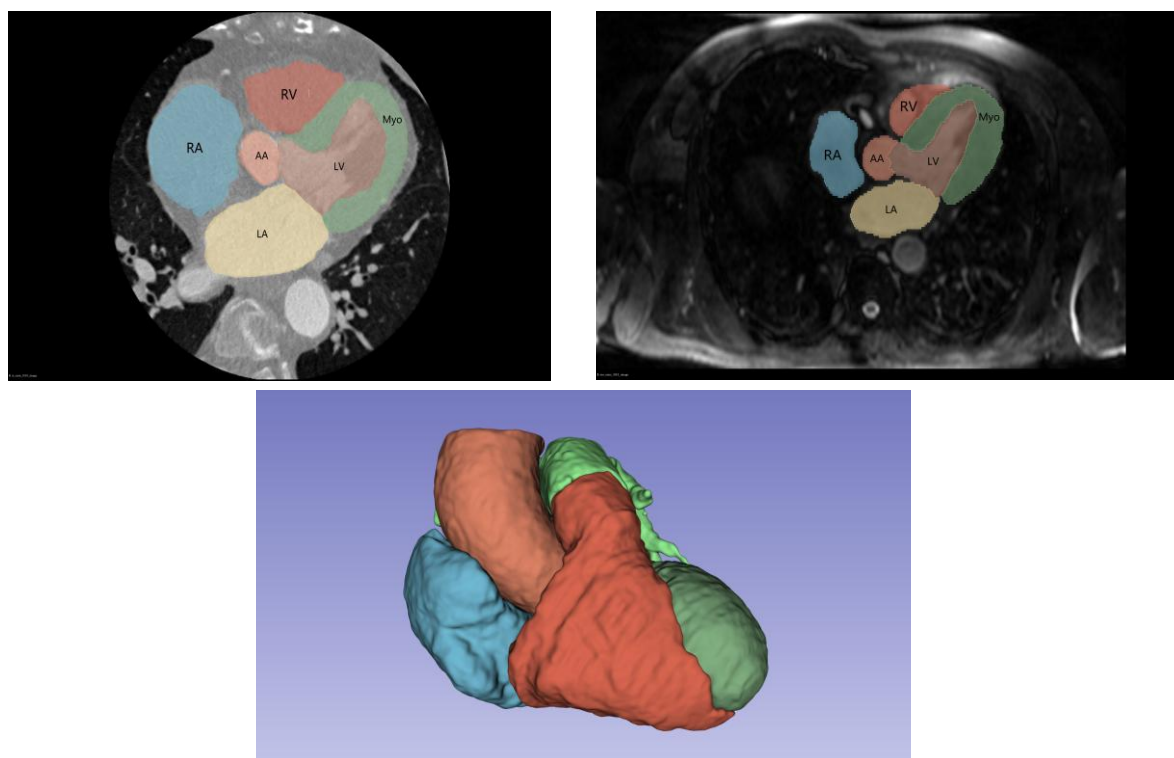


图 1 带像素级标注心脏 CT 图像（左上）、带像素级标注 MRI 图像（右上）、心脏 3D 模型（下）

所有像素按照所属心脏结构类别进行分类，以划分出不同的心脏区域，并以此为基础对心脏进行 3D 建模以便后续的病例分析，这些心脏区域包括左心室（LV）、右心室（RV）、左心房（LA）、右心房（RA）、升主动脉（AO）和左心室心肌（Myo）等部位，如图 1 所示。

如前文所述，若由一位拥有专业知识的医生对一位病患的心脏影像进行各结构分割，这一过程一方面耗时长——至少需要消耗数十分钟，另一方面结果易受主观因素影响——疲劳或心情差对工作产生影响，因此不利于病情的诊断。而在计算机上运行的心脏的自动分割系统则能对心脏影像进行高效且客观的处理，因此其能有效地辅助医生进行医学诊断，对于病患治疗有很大帮助。

## 1.2 国内外研究现状

本章将分别在网络结构和数据增强方法两个方面总结目前国内外科研界在医学影像语义分割领域做出的工作。

### 1.2.1 对于网络结构做出的改进

Ronneberger 等人<sup>[4]</sup>在对医学影像数据进行详细分析后，针对医学影像的特点提出了经典的 U-Net 网络，U-Net 具有编码器和解码器两部分，这两部分都由卷积层、池化层和上采样层组成。编码器部分用于提取图像特征，将图像从高维度降到低维度表示。而解码器部分则用于恢复分割图像的分辨率和精度，将低维度的特征向量转换为与原始图像大小相同的输出图像。

U-Net 在医学图像分割任务中表现出了优异的性能，主要原因在于它能够利用分割任务中的先验知识，即图像的局部空间信息在分割任务中起到重要的作用。U-Net 的结构允许网络有效地利用图像中的上下文信息，以提高分割结果的准确性。

因此，U-Net 已成为医学图像分割任务中的常用算法，并已被广泛应用于许多医学应用，如肿瘤分割、血管分割、心脏分割等。

后来又出现了许多不同的 U-Net 变体。Zhou 等人<sup>[5]</sup>提出的 U-Net++网络，旨在进一步提高医学图像分割的性能。与 U-Net 不同，U-Net++网络采用了一种新的多分辨率分割方法，可以更好地利用不同分辨率下的特征信息。在 U-Net++中，每个下采样和上采样模块都被扩展成一个完整的 U-Net 子网络，其中包括一个附加的嵌套分支，用于捕获不同分辨率下的细节信息。这种分层和多分辨率的设计使 U-Net++能够更好地捕获图像中的结构细节和上下文信息，从而实现更准确的分割。Milletari 等人<sup>[6]</sup>提出的 V-Net 网络，引入了 3D 卷积来处理 3D 医学图像，通过一个具有两个 3D 卷积层和一个下采样层的编码器来提取特征。在解码器中，V-Net 通过上采样将编码器的特征图扩大到原始大小，然后通过一个具有两个 3D 卷积层和一个上采样层的解码器来逐步恢复分割掩码。Huang 等人<sup>[7]</sup>提出的 U-Net3+网络，引入了轴对称模块（Axial Symmetric Module, ASM），用于对三维卷积中的轴向信息进行捕捉，从而提高模型的特征表达能力。其次，U-Net3+采用了多尺度金字塔池化（Multi-scale Pyramid Pooling, MPP）模块<sup>[8]</sup>，用于对不同尺度的特征进行融合，从而提高模型的语义表达能力。受到 He 等人<sup>[9]</sup>提出的 ResNet 网络中的残差连接思想的启

发, Ibte haz 等人<sup>[10]</sup>在 U-Net 的基础上引入了多尺度残差连接并形成了 MultiResUNet 网络, 该网络利用多个分辨率下的特征图进行跨尺度信息交互, 从而增强了模型对于不同尺度信息的提取和利用能力。在跨尺度残差连接模块中, 模型利用了多个分辨率下的特征图进行信息交互。具体来说, 编码器中每个池化层之后额外增加了一个残差连接模块, 以将对应分辨率的特征图进行跨尺度的融合, 最终提升模型准确率。

近年来, 基于自注意力机制的 transformer 在其原本提出的自然语言处理领域取得了巨大的成功, 因此许多研究都试图将其应用于视觉领域, 如 Alexey 等人提出的基于原始 transformer 的 Vision Transformer<sup>[11]</sup>, 以及后续 Wang 等人在此基础上提出的 Pyramid Vision Transformer<sup>[12]</sup>, 于是又有许多研究者尝试将这些视觉 transformer 方法用于改进基础 U-Net, 如 Sun 等人提出的 DA-TransUNet<sup>[13]</sup>, Lan 等人提出的 BRAU-Net++等<sup>[14]</sup>。这些基于 transformer 的方法在视觉领域同样取得了极好的成果。

然而, 尽管上述方法在一定程度上提高了模型的准确率, 然而它们很少针对心脏不同结构的边界边缘存在模糊不清的区域导致模型准确率降低的问题提出对应解决方案。

### 1.2.2 对于数据增强方法做出的改进

前文提到基于深度学习的方法通常都需要获取大量具有像素级标注的数据以用于提高模型的泛化能力, 然而医学影像的数据相当稀缺, 因为对数据的标注需要消耗具有专业知识医生的大量时间和精力, 因此想获得大规模数据具有难度。目前, 针对该问题通常有如下三种处理方法:

1) 传统数据增强: 通过对原始数据添加噪声和尺度变换等方式产生更多数据, 但这类处理方式过于简单, 在处理医学影像的复杂数据时不足以让模型获得更多的泛化性能。

2) 迁移学习: 这种方法通常包括预训练和微调两部分, 首先将模型在易于获得的大规模影像数据上训练, 如自然影像数据, 之后再将训练好的模型在特定数据集上进行训练<sup>[15]</sup>, 然而自然影像数据和医学影像数据的差别通常很大, 因此其提供的无效特征很多而有效特征有限。

3) 生成对抗网络: 这种方式能通过对抗训练的方式在现有数据的基础上生成更多的数据, 这种方式能够增加数据的多样性, 提高模型的泛化能力, 然而利用这种方式生成的数据有可能存在质量不高, 图像模糊的问题, 因此对于此方法的使用也需要保持谨慎。

可以看出, 目前存在的数据增强的方法都或多或少有缺陷, 而且很少针对医学图像中

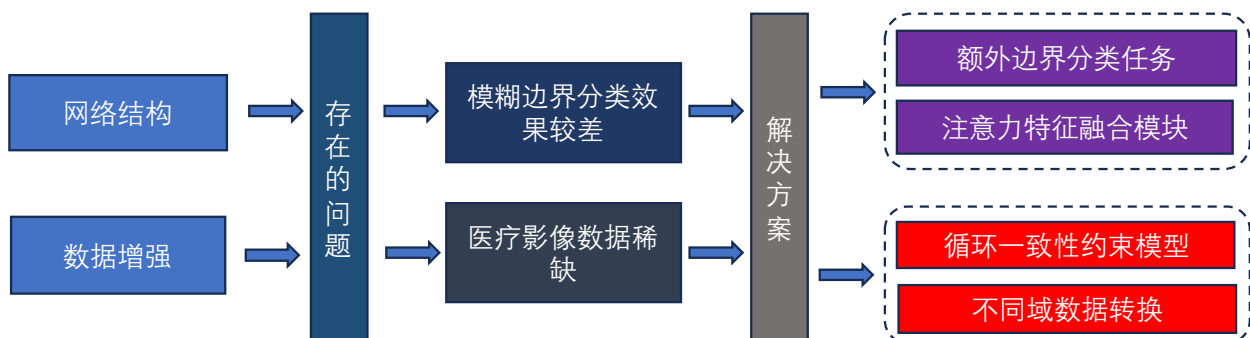


图 2 针对两方面问题给出的解决方案



数据模态多，而大部分模型却仅能有效利用一种模态数据的问题，如心脏图像就有 MRI 和 CT 两种不同模态，而这些不同模态的数据却很难用于同一深度学习模型的训练。

### 1.3 本文工作和贡献

针对目前存在的心脏不同结构边界模糊不清以及医学影像数据稀缺的问题，本文提出了一种基于生成对抗网络的数据增强以及边缘感知的心脏分割方法，一方面其通过添加额外边缘任务以对分界区域施加额外的监督以及设计一种独特的注意力特征融合模块有效的提高了模型的准确率，另一方面又对于医学图像数据稀缺且不同模态类型数据多的问题提出了使用循环一致性生成对抗网络进行不同模态间数据转换的方法，其能有效地将有标签的 MRI 数据转为有标签的 CT 数据，从而扩充 CT 数据集，增强模型泛化能力，从而提升准确率。

此外，本文为了方便对这些改进方法的使用，设计和实现了心脏医学影像语义分割系统，其配备有本文改进的深度学习模型用于影像语义分割，以及本文提出的数据增强方法用于提高深度学习模型的泛化能力，且该系统生成的数据能够直接用于各种医学影像 3D 建模软件的使用，如医学图像分析领域经常使用的开源软件 3D Slicer[16]。

### 1.4 章节安排

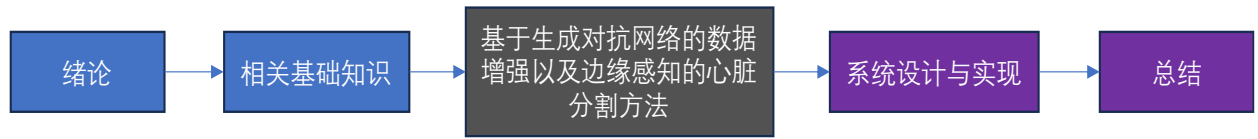


图 3：章节安排

如图 3，本文分为六章，第一章主要介绍研究背景和研究现状，第二章主要介绍与医学影像语义分割方向的有关深度学习知识以及数据集，第三章主要介绍一种本文设计的基于边缘形状感知以及注意力特征融合的心脏分割方法，第四章则介绍本文提出的一种基于对抗学习的心脏影像数据增强方法，第五章则介绍本文基于上述两种方法设计与实现的系统，最后第六章进行总结以及对未来研究方向做出展望。

## 2 医学影像语义分割技术背景

### 2.1 基于深度学习的心脏图像语义分割技术相关知识

#### 2.1.1 注意力机制

注意力机制主要可分为三类，分别是通道注意力机制、空间注意力机制以及自注意力机制，这里由于仅使用了前两类，因此这里不介绍自注意力机制。

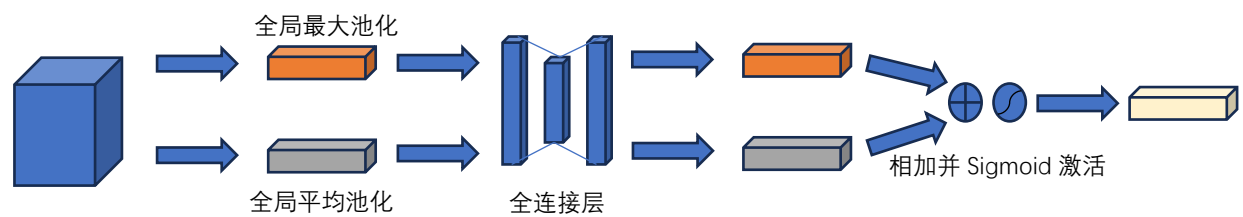


图 4：通道注意力计算过程

$$CAM(Input) = \sigma \left( MLP(Avg_c(Input)) + MLP(Max_c(Input)) \right) \quad (1)$$

1) 通道注意力机制：通道注意力通过挤压特征图，将空间维度进行压缩后送入全连接网络以得到注意力权重矩阵，最后通过权重矩阵对原特征在通道上矫正特征，如图 4。其计算过程可写为式 1，其中 $Avg_c$ 表示在每幅特征图上进行全局平均池化， $Max_c$ 表示在每幅特征图上进行全局最大池化， $MLP$ 表示全连接层， $\sigma$ 代表Sigmoid激活函数。

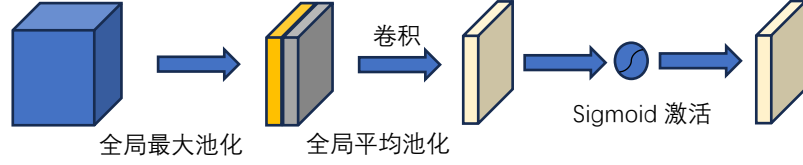


图 5：空间注意力计算过程

$$SAM(Input) = \sigma \left( COV \left( CAT(Avg_s(Input), Max_s(Input)) \right) \right) \quad (2)$$

2) 空间注意力机制：空间注意力通过在不同通道上进行挤压特征图实现，具体来说，其通过平均池化和最大池化来获取每个空间位置的最大值和平均值，然后将得到的两个矩阵拼接以及卷积来学习每个空间位置的权重，如图 5。其计算过程可写为式 2，其中 $Avg_s$ 表示在每个通道上进行全局平均池化， $Max_s$ 表示在每个通道上进行最大平均池化， $CAT$ 表示拼接操作， $COV$ 表示卷积运算，一般使用 $7 \times 7$ 的卷积核， $\sigma$ 代表Sigmoid激活函数。

### 2.1.2 生成对抗网络

2014 年，受到博弈论中零和博弈思想启发的 Goodfellow 等人<sup>[17]</sup>提出了生成对抗网络（GAN），该模型框架由两个相互对抗的网络组成，一个网络进行数据的生成，另一个则用于判断输入的数据是否为真实数据，且由于特殊的损失函数设计，它们的竞争是零和博弈，因此它们在训练中相互竞争学习，这种竞争学习使得其能够生成逼真的伪数据。实际上，由于其独特的设计思路和训练方式，GAN 一经提出便受到了广泛关注，给深度学习许多领域带去了重大的突破。

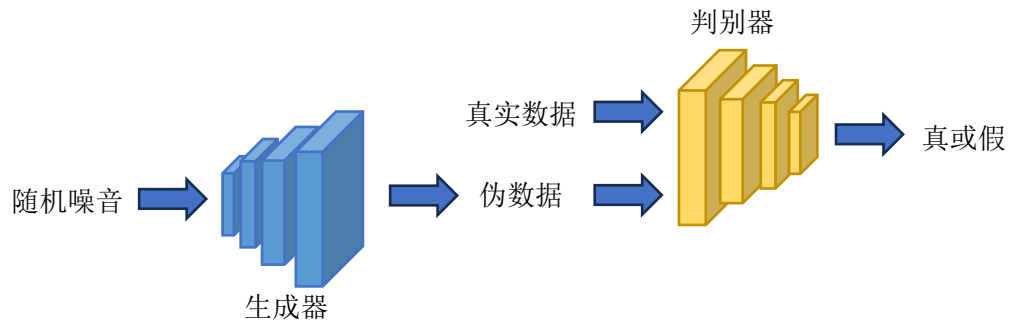


图 6：生成对抗网络

GAN 的网络结构可分为两部分，分别是生成器 G 和判别器 D，结构如图 6 所示。生成器以随机噪音作为输入，并尝试生成与真实数据样本分布尽可能接近的数据以使得判别器无法识别出其输入的数据是否为真实数据。由于精心设计的损失函数，G 和 D 的竞争为零和博弈，因此它们为严格的竞争关系，永远无法共赢，具体来说，生成器 G 的损失函数 $L_G$ 如式 3 所示：

$$L_G = \log(1 - D(G(z))) \quad (3)$$

其中 $z$ 为随机噪音。对于辨别器  $D$  来说，其损失函数 $L_D$ 如式 4 所示：

$$L_D = \log D(x) + \log(1 - D(G(z))) \quad (4)$$

其中 $x$ 为真实数据。从它们的损失函数中也可以看出，辨别器  $D$  对真实数据尽可能输出 1，对伪数据尽可能输出 0 时，其损失最大，而对于生成器  $G$  则需要尽可能生成与真实数据相近的伪数据，并使得辨别器输出为 1 时，其损失最小，可以看出，在对抗过程中，判别器  $D$  追求损失最大，生成器  $G$  追求损失最小。

在 GAN 诞生之后，许多基于对抗训练的网络纷纷被设计出来，如 StyleGAN<sup>[18]</sup>，CycleGAN<sup>[19]</sup>等。

### 2.1.3 常用损失函数

#### 2.1.3.1 交叉熵损失

交叉熵损失是一种多分类常用的损失函数。对于 $n$ 分类任务来说，第 $i$ 类分类对象的真实标签和预测结果可以表示为两个长度相同的 one-hot 编码向量 $y_i$ 和 $P(x)_i$ ，其中 $y_i$ 第 $j$ 个位置的值为 1 表示其是目标对象，0 则不是， $P(x)_i$ 第 $j$ 个位置的值同理，因此它的数学表达式可以写成式 5 所示：

$$CELoss(y, P(x)) = - \sum_i^n y_i \log P(x)_i \quad (5)$$

其中， $x$ 表示输入， $y$ 表示真实标签， $n$ 表示总类别数， $P(x)$ 表示模型预测值， $y_i$ 和 $\log P(x)_i$ 则分别表示第 $i$ 类的真实标签和模型预测值。显然，当模型输出越接近真实标签时，损失越小。

#### 2.1.3.2 Dice 损失

Dice 损失是语义分割领域使用的损失函数，其非常适合处理正负样本不平衡的问题，尤其对于某些小目标的预测。具体来说，它的数学表达式如式 6 所示：

$$DiceLoss(y, P(x)) = 1 - \sum_i^n \frac{2|y_i P(x)_i|}{|y_i| + |P(x)_i|} \quad (6)$$

其中， $x$ 表示输入， $y$ 表示真实标签， $n$ 表示总类别数， $P(x)$ 表示模型预测值， $y_i$ 和 $\log P(x)_i$ 则分别表示第 $i$ 类的真实标签和模型预测值， $||$ 表示向量中 1 的个数。值得注意的是，该损失不是一个凸函数，且其导数的波动大，不利于模型的收敛，因此其一般与正则化方法或其他损失函数一起使用。

#### 2.1.3.3 交叉熵和 Dice 的混合损失

实际训练时，Dice 损失常常与交叉熵损失一起使用组成混合损失，一方面可以解决 Dice 损失导致的模型收敛难的问题，另一方面又可以缓解正负样本不平衡的问题。具体来说，该混合损失可以写成式 7 所示：

$$UnionLoss(y, P(x)) = \beta CELoss(y, P(x)) + (1 - \beta) DiceLoss(y, P(x)) \quad (7)$$

其中 $x$ 表示输入， $y$ 表示真实标签， $\beta$ 是人工设置的超参数，用于调整两个损失的权重。

#### 2.1.4 常用评价指标

Dice 系数是目前医学图像分割中最常用的评估指标，其显示了预测结果和真实标签的差别，其计算公式如式 8:

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|} \quad (8)$$

其中 $X$ 表示预测的正样本集合， $Y$ 表示真实的正样本集合， $| \cdot |$ 表示集合中元素的个数。可以看出，Dice 系数的取值范围在 0 到 1 之间，当预测样本和真实样本完全重合时为 1，没有重叠时为 0。

#### 2.2 数据集介绍

多模态全心脏分割数据集（MM-WHS 2017）<sup>[20-23]</sup>是一项针对心脏分割任务的数据集，其数据包括了从上腹部到主动脉弓的整个心脏，且包括 CT 和 MRI 两种模态的数据，其中包含有 60 例心脏 CT 图像，20 例有标注，40 例无标注，使用常规心脏 CT 成像技术获取，像素之间距离为 0.78 毫米左右，切片厚度则为 1.60 毫米左右，60 例心脏 MRI 图像，20 例有标注，40 例无标注，使用 3D 平衡稳态自由进动序列获取，重采样前各体素之间距离约为 2 毫米，经过重采样后的各体素距离约为 0.78 毫米，如图 7 所示。

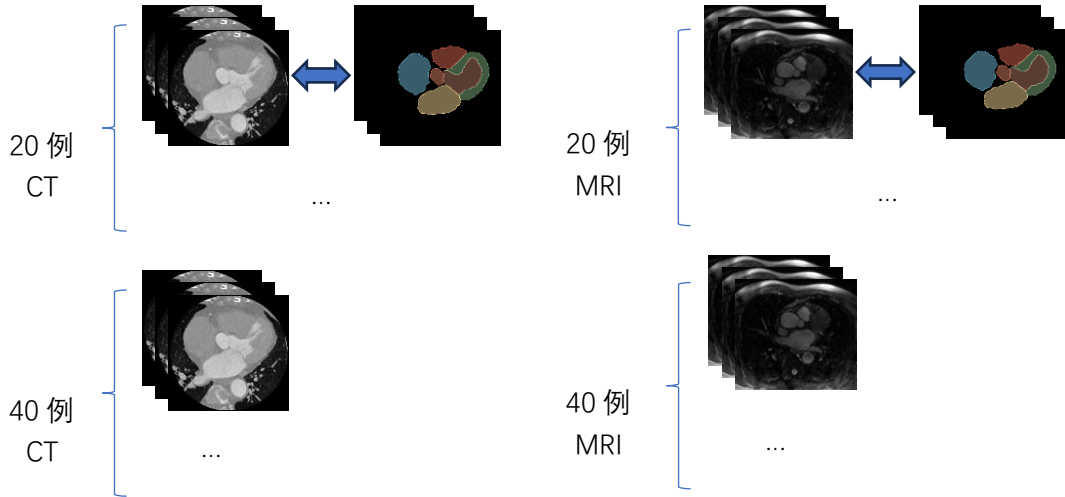


图 7: MMWHS 2017 数据集，网络地址为 <https://zmiclab.github.io/zxh/0/mmwhs/>

### 3 基于生成对抗网络的数据增强以及边缘感知的心脏分割方法

#### 3.1 基于循环一致性生成对抗网络的 MRI 域至 CT 域的数据转换方法

在复杂医学场景中，对于人体心脏部位的成像检查类型通常有多种，它们的成像效果也不尽相同，Yang 等人<sup>[25]</sup>指出在前列腺的 MRI 成像检查中，目标前列腺区域的平均像素值与非前列腺区域的平均像素值相差大，对比度较大，而 CT 成像检查中，平均像素值相差小，对比度较小。事实上，在一般情况下，大多数 MRI 的成像效果都要高于 CT 图像，且在 CT 检查时人体会吸收少量辐射，因此不适合孕妇或婴儿使用。然而 MRI 成像费用相较 CT 成像费用更高，因此也有许多患者会选择 CT 成像。因此这种复杂情况也就造成了心脏影像存在多模态数据的结果。上文提到过，由于医学伦理等原因，能获取到的心脏影像数据稀缺，而有标注的数据更加稀少，因此，如何将不同模态数据统一起来并

充分利用就成为了一项重要的研究课题。

在本节中，本文会介绍一种基于循环一致性生成对抗网络的心脏影像增强方法，该方法利用循环一致性生成对抗网络（CycleGAN）<sup>[26]</sup>将 MRI 模态心脏影像数据迁移至 CT 模态以扩充 CT 数据集。

### 3.1.1 循环一致性生成对抗网络

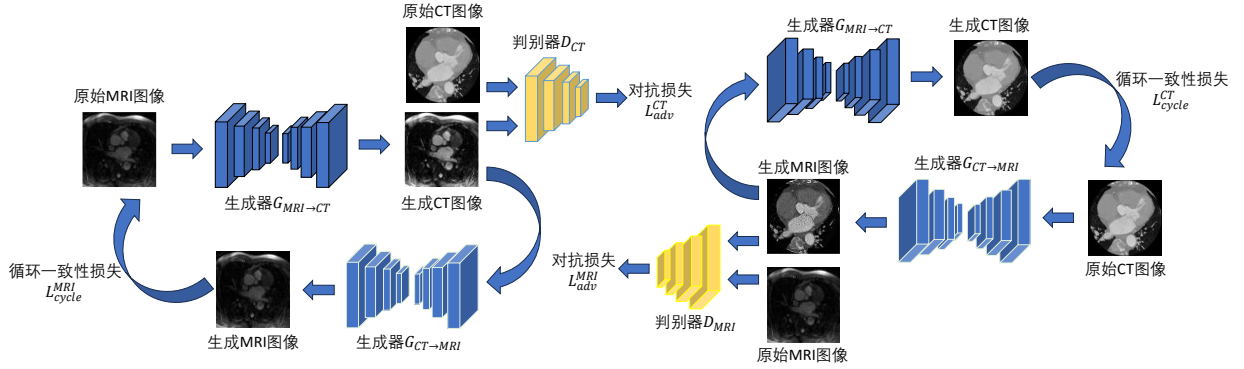


图 8: 循环一致性生成对抗网络框架图

心脏的 MRI 和 CT 模态数据之间有很大的外观差异，而循环一致性生成对抗网络最初设计用于对图像风格进行转换，且由于使用了循环一致性约束，训练过程中不需要配对的不同风格图像数据，因此这里很适合用于将 MRI 域数据转移至 CT 域，从而减轻不同域间的语义鸿沟，使得模型能更好的理解从 MRI 域转换的 CT 数据。

循环一致性生成对抗网络相比传统对抗生成网络的差异在于其利用两组传统对抗生成网络以及独特设计的循环一致性损失函数实现了非配对图像的转换，其具体框架如图 8，可以看到，循环一致性生成对抗网络包括两个生成器，分别为生成器  $G_{MRI \rightarrow CT}$  和  $G_{CT \rightarrow MRI}$ ，它们分别用于将 MRI 域数据转为 CT 域数据和将 CT 域数据转为 MRI 域数据，除此之外，还包括两个判别器网络  $D_{CT}$  和  $D_{MRI}$ ，它们分别用于判断 CT 数据的真伪以及 MRI 数据的真伪。对于将 MRI 域数据转换至 CT 域的部分，即图 8 左半部分，首先原 MRI 域数据  $X_{MRI}$  被输入进生成器  $G_{MRI \rightarrow CT}$  中，得到生成器  $G_{MRI \rightarrow CT}$  的输出数据  $G_{MRI \rightarrow CT}(X_{MRI})$ ，又由于该数据特征应当与原 CT 域数据相同，因此可以使用判别器  $D_{CT}$  进行对抗训练，并产生对抗损失  $L_{adv}^{CT}$ ，此外，生成器  $G_{MRI \rightarrow CT}$  的输出数据  $G_{MRI \rightarrow CT}(X_{MRI})$  还被送入生成器  $G_{CT \rightarrow MRI}$  中产生数据  $G_{CT \rightarrow MRI}(G_{MRI \rightarrow CT}(X_{MRI}))$ ，该生成器能将 CT 域数据转换至 MRI 域，因此数据  $G_{CT \rightarrow MRI}(G_{MRI \rightarrow CT}(X_{MRI}))$  应当与原 MRI 域数据  $X_{MRI}$  相同，所以此处还可以构造约束损失  $L_{cycle}^{MRI}$ ，而对于原 CT 域转换 MRI 域部分，即图 8 右半部分同理。根据上述过程可以得出循环一致性生成对抗网络的损失组成如式 9

$$L = L_{adv}^{CT} + L_{cycle}^{MRI} + L_{adv}^{MRI} + L_{cycle}^{CT} \quad (9)$$

其中具体每一项的表示如式 10、11、12、13。

$$L_{adv}^{CT} = \log D_{CT}(X_{CT}) + \log (1 - D_{CT}(G_{MRI \rightarrow CT}(X_{MRI}))) \quad (10)$$



$$L_{adv}^{CT} = \log D_{MRI}(X_{MRI}) + \log (1 - D_{MRI}(G_{CT \rightarrow MRI}(X_{CT}))) \quad (11)$$

$$L_{cycle}^{MRI} = \|G_{CT \rightarrow MRI}(G_{MRI \rightarrow CT}(X_{MRI})) - X_{MRI}\|_1 \quad (12)$$

$$L_{cycle}^{CT} = \|G_{MRI \rightarrow CT}(G_{CT \rightarrow MRI}(X_{CT})) - X_{CT}\|_1 \quad (13)$$

上述式中， $\|\cdot\|_1$ 为像素级 $L_1$ 约束， $X_{CT}$ 为原 CT 域数据。

该基于循环一致性对抗网络的不同域间数据转换方法能有效地将不同域之间的数据进行转换，减小语义鸿沟，从而使模型能有效地利用 MRI 域数据，从而提高模型的泛化能力和准确率。

### 3.2 基于额外边缘任务以及特征融合模块的边缘感知方法

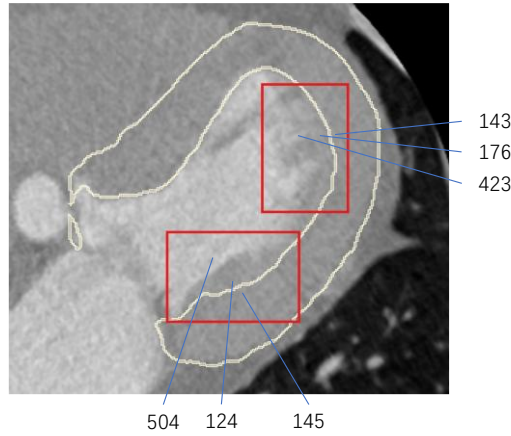


图 9: Myo 结构边缘灰度图像

实际对心脏医学图像进行分割时，采集数据的机器种类多，且不同患者的心脏结构差异大，因此得到数据的质量也不尽相同，对于某些质量较低的心脏图像，常常存在心脏目标结构与背景的边缘模糊不清的情况，如图 8 所示，标出的红框中，Myo 结构边缘内侧的灰度值约为 145 左右，而边缘外侧的部分区域灰度值变化不大，因此易导致模型对于这些区域的分割不准确，从而降低模型性能，给心脏结构分割带来负面影响，而且在使用上文提出的数据增强方法将 MRI 域数据转换至 CT 域数据后，该问题更加明显。因此，本节提出了通过添加额外的边界分类任务以及特征循环的方式增强模型对于边缘的感知能力，此外，由于额外的边缘分类任务目标极小，分类难度大，考虑到 UNet 网络较为基础，可能无法有效的完成难度较高的边缘任务，因此这里设计了基于注意力的特征融合模块用于增强模型的性能，提高模型对边缘区域的感知能力。

#### 3.2.1 额外边缘任务设计

为了增强模型对心脏结构边缘的感知能力，可以通过额外添加结构边缘分类任务来实现，这一步首先我们需要通过对原标签进行边缘提取，由于不同的心脏结构可能彼此相连，因此原标签中不同心脏结构的分割结果可能相接，如图 9 中原始标签图所示，因此我们需要提取结构的内边缘，以免边缘相互重合。具体来说，先将不同类别的标签分离，对于一张有 6 个分类类别的标签图来说，先按照标签类别的不同生成 6 张标签图，每张标

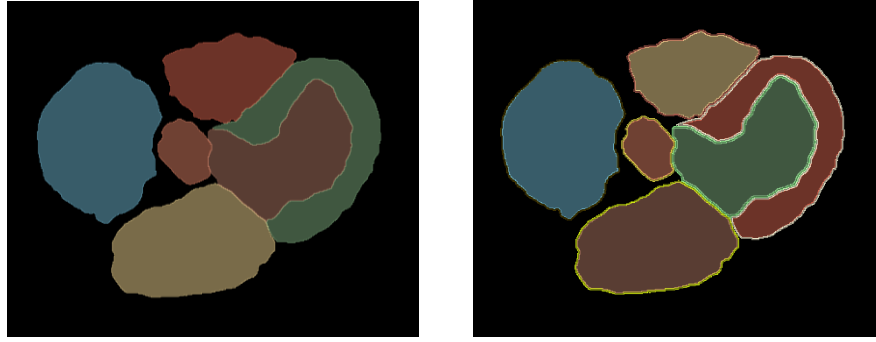


图 10: 原始标签图 (左), 带边缘的标签图 (右)

签图仅有一个分类类别的标签且为二值图，之后对于每一张标签图 $x_i$ ，对其做形态学腐蚀 Erosion 操作得到 $\hat{x}_i$ ，即将其自外向内收缩，再用标签图 $x_i$ 减去 $\hat{x}_i$ 即得到边缘标签图 $y_i$ ，在此之后我们将得到的 6 张边缘标签图与 6 张原标签图合并为一张带有 12 个分类类别的标签图，最终结果如图 9 中的带边缘的标签图所示。

值得注意的是，对原标签图做形态学腐蚀操作的次数也有可能会影响结果，由于腐蚀次数越多，得到的边缘就会越粗，正样本变多，模型就能对更好的学习边缘特征，然而这样也会导致原标签正样本变少，模型对于原任务的学习就会减弱，因此，腐蚀次数 $m$ 也是一项需要实验的重要参数。

### 3.2.2 基于注意力的特征融合模块设计

为了帮助模型完成额外边缘任务，提高边缘感知能力，这里设计了基于通道注意力和空间注意力的特征融合模块，该模块结构如图 10 所示，该模块主要思想为分别融合输入的通道注意力特征和空间注意力特征，并将它们分别与融合的输入特征图进行拼接卷积运算，从而将输入的特征图不仅在通道上进行拼接卷积，而且在它们的通道注意力和空间注意力尺度上进行特征融合。具体来说，当将该模块用于 U-Net 网络解码器部分的低维特征与高维特征融合时，解码器能够更有效的利用通过跳跃连接传递来的低维特征。

### 3.2.3 损失函数设计

对于当前心脏结构分割任务，一般较多使用交叉熵损失函数。本文考虑到该心脏结

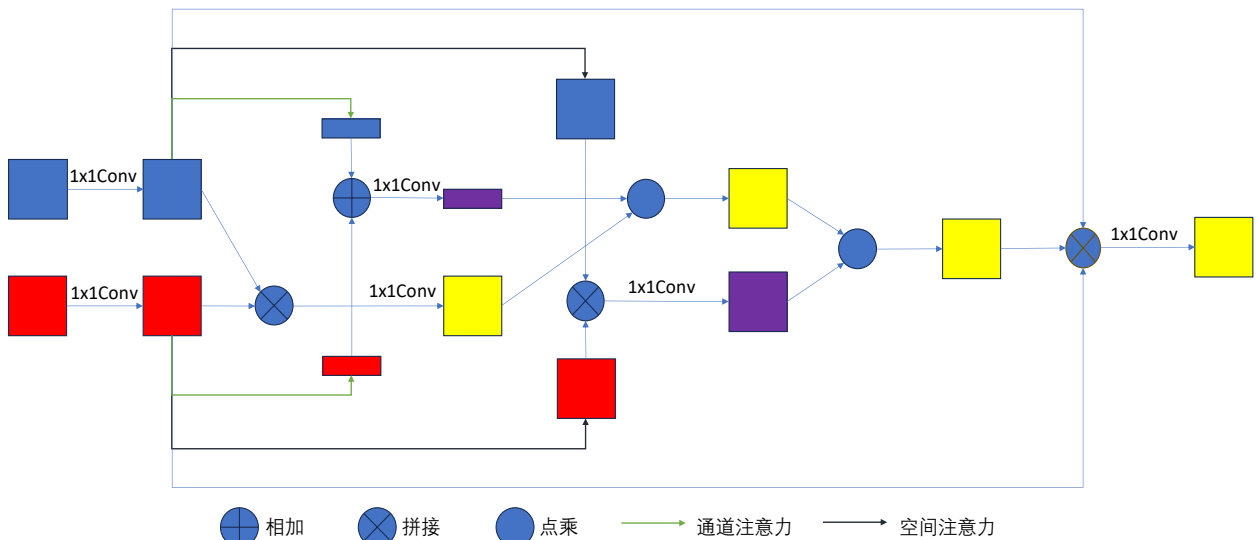


图 11: 特征融合模块

构分割任务中目标尺寸变化大，且常存在细小目标的问题，因此这里对于心脏结构分割任务采用的损失函数为交叉熵损失和 Dice 损失的混合损失函数。此外，这里实际使用的 Dice 损失相比前文提到的有所不同，这是为了避免在实际计算损失时分母为 0 的情况出现，具体 Dice 损失如式 9：

$$SoftDiceLoss(y, P(x)) = 1 - \sum_i^n \frac{2|y_i P(x)_i| + \delta}{|y_i| + |P(x)_i| + \delta} \quad (9)$$

其中 $\delta$ 为平滑系数，其余符号意义与式 6 中相同。

此外，对于边缘分类任务来说，本文考虑到边缘分类的目标相比原本的分类目标小得多，正样本极少，导致实际训练中这些目标易被模型忽视，因此这里同样需要使用上文提到的交叉熵损失和 Dice 损失的混合损失函数，以提高模型对这些边缘分类任务的重视程度，否则其极易被模型忽视，从而降低模型对心脏结构边缘的感知效果。因此总损失函数 $Loss$ 可以写成如式 10：

$$Loss = Loss_{Seg}(x, P(z)_x) + \alpha Loss_{Edge}(y, P(z)_y) \quad (10)$$

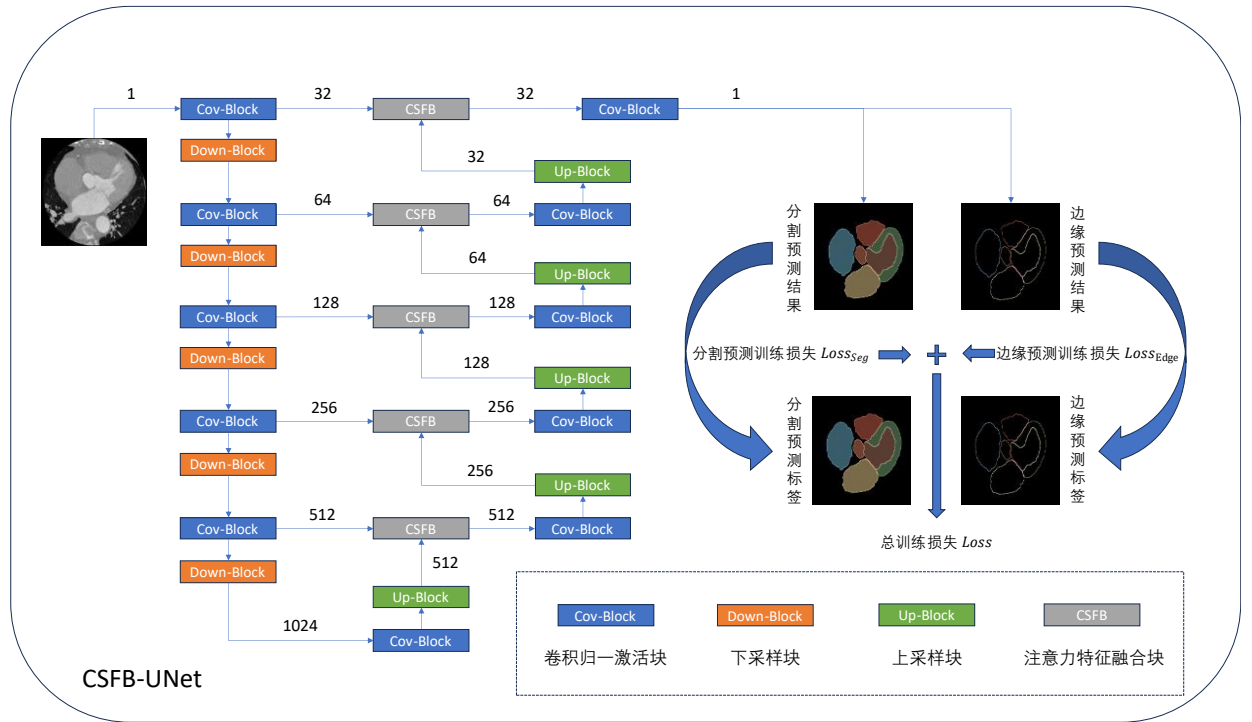


图 12：整体架构

其中， $\alpha$ 为用于平衡两项任务损失的超参数， $x$ 为心脏结构分割任务标签， $P(z)_x$ 表示模型输出的分割任务预测结果， $y$ 为边缘分类任务标签， $P(z)_y$ 表示模型输出的边缘任务预测结果， $Loss_{Seg}$ 为心脏结构分割任务损失函数， $Loss_{Edge}$ 为边缘分割任务损失函数。

### 3.3 整体架构



本文提出的基于边缘形状感知以及注意力特征融合的心脏分割方法整体框架图如图 11 所示。其中箭头上的数字表示特征图数量，卷积归一激活块由卷积层、归一化层和激活层组成，下采样块与卷积归一激活层类似，但是卷积的步长为 2，上采样块则由卷积层、归一化层、激活层和转置卷积实现，上述各块具体结构如图 12，注意力特征融合块结构如图 10。

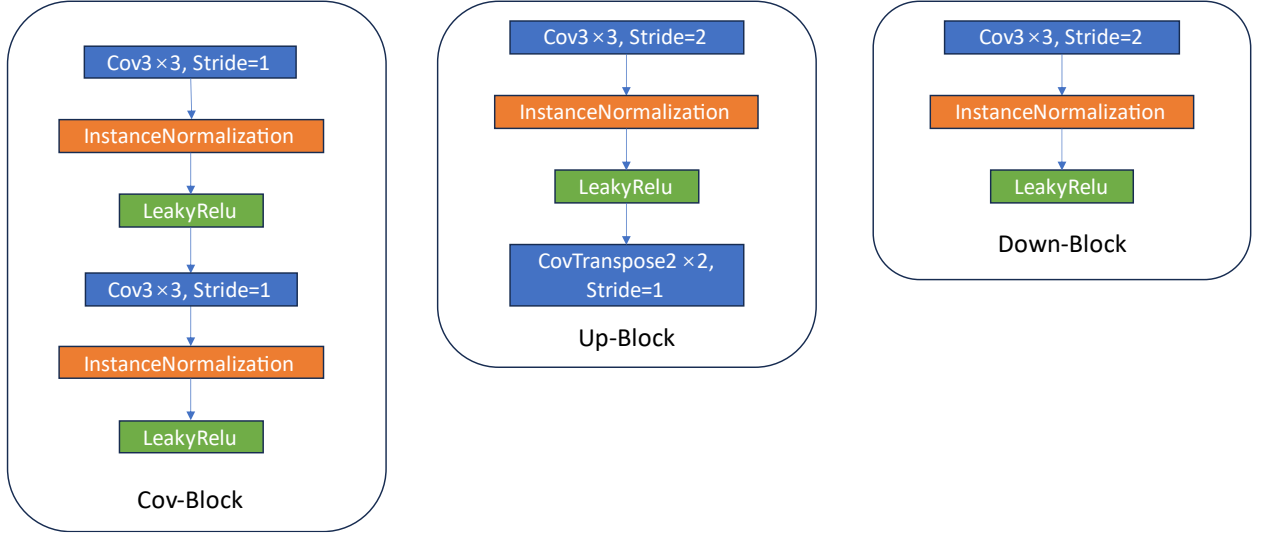


图 13: 卷积归一激活块、上采样块和下采样块的具体组成

### 3.4 相关实验

为了验证本文所提出的方法性能，本文在公开数据集 MM-WHS 2017<sup>[20-23]</sup>上进行了一系列对比实验，且为了验证提出的方法有效性，进行了消融实验。

#### 3.4.1 实验设置

本文实验第一步为训练循环一致性对抗生成模型，取 MM-WHS 2017 的 40 例无标签 CT 图像和 40 例无标签 MRI 图像作为训练数据，共训练 50 轮，初始学习率为 0.01, 采用线性衰减方式，训练 50 轮学习率降为 0，批大小为 2，输入块大小为  $384 \times 384$ 。第一步训练完成后利用训练好的循环一致性对抗生成模型将 20 例有标注的 MRI 数据转换为 CT 数据。之后训练 $\alpha$ 设置为 1， $\beta$ 设置为 0.5，采用随机梯度下降优化器作为本文中的模型优化器，训练 200 轮，初始学习率为 0.01，采用线性衰减方式，第 200 轮后学习率降为 0，批大小为 14，这里不包括数据集的划分的说明，因为不同对比实验中所使用的训练和测试数据可能不同，这是为了与对比的其他论文保持一致，后续不同的对比实验中会详细说明。

实验所用 CPU 为 Intel i7-12700KF，GPU 为 NVIDIA GeForce RTX 3090 Ti 24G，内存为 64GB 4800MHz，使用操作系统为 Windows10 企业版 LTSC，代码为 Python 编写，使用 PyTorch 软件包搭建模型、训练与测试，使用 nnU-Net 软件包<sup>[27]</sup>进行数据预处理。

#### 3.4.2 对比实验结果与分析

首先，本文对于前文提到的边缘任务中使用形态学腐蚀的最优次数进行实验，实验数据为 MM-WHS 2017 的 CT 训练集，其中 16 例做为训练集，4 例做为测试集，五折交叉验证，取最好结果，训练模型为 nnUNet 中提供的基础 UNet，结果如表 1，可以看到当仅使

用一次形态学腐蚀时得到最高的 Dice 系数，且相比不添加边缘任务提升了约 3 个百分点。由于次数为 1 时效果最好，所以后续所有实验的边缘任务采用的腐蚀次数均为 1。

表 1: 进行不同次数的形态学腐蚀得到的结果

次数	LV	RV	LA	RA	Myo	AO	Mean
<b>0</b>	<b>93.1</b>	83.8	91.5	85.4	<b>90.9</b>	83.1	87.9
<b>1</b>	92.8	<b>86.8</b>	<b>93.1</b>	<b>86.9</b>	89.3	<b>95.1</b>	<b>90.7</b>
<b>2</b>	92.8	86.7	92.8	86.1	88.3	94.5	90.2
<b>3</b>	92.5	86.2	92.7	86.0	86.2	87.2	89.8

本文将提出的边缘感知方法与其他改进方法进行比较，以验证本文方法的有效性。表 2 中列出了心脏各个部分的百分比 Dice 系数以及它们的平均值和与其他方法的比较，其中基准方法为 nnUNet<sup>[27]</sup>，此外，这里为了对比公平，表 2 中的 CSFBU-Net 训练数据未使用 3.1 小节中提出的数据增强方法，即仅使用了 MM-WHS 2017 的 CT 训练集中的 16 个，余下 4 个作为测试集，采用五折交叉验证方法，取最好的结果。可以看到，本文的方法在除了右心室 RV 部分落后于 Habijan 等人<sup>[28]</sup>的方法外，其他各个部分均有较大提升，且平均 Dice 系数相较其他方法略有提升。

表 2: 本文边缘感知方法与其他改进方法的比较

方法	LV	RV	LA	RA	Myo	AO	Mean
基准方法	93.1	83.8	92.0	86.4	90.9	83.1	88.4
Chen et al. <sup>[29]</sup>	91.9	——	91.1	——	87.7	92.7	90.8
Habijan et al. <sup>[28]</sup>	91.7	<b>88.2</b>	<b>93.1</b>	86.1	87.9	94.5	90.2
CSFBU-Net	<b>93.1</b>	87.2	92.8	<b>87.0</b>	<b>90.0</b>	<b>95.4</b>	<b>90.9</b>

此外，为了验证本文提出的数据增强方法，这里使用 MM-WHS 2017 中 20 例带标注的 MRI 转换所得的 CT 数据以及带标注的 20 例原 CT 数据组成数据集训练，取 10 例带标注的原 CT 数据作为测试集，其余作为训练集，取最好的结果，结果如表 3，其中 nnUNet 作为基准方法，其未使用数据增强方法，而是直接在 20 例原始 MRI 和 10 例 CT 图像上进行训练。可以看到，对于基准方法使用数据增强的效果与 Zhang 等人提出的方法相近，相较于 Cheng 等人提出的方法仅在 LV, RA 部位有略微超越，且平均 Dice 系数低约 1 个百分点。但在使用了本文提出的全部方法后，几乎所有部位的 Dice 系数都有巨大提升，且平均值也有所上升。

表 3: 本文数据增强方法与其他改进方法的比较

方法	LV	RV	LA	RA	Myo	AO	Mean
基准方法	<b>93.1</b>	83.8	91.5	85.4	90.9	83.1	87.9
Jiang et al. <sup>[30]</sup>	90.7	85.2	89.7	83.3	87.2	94.8	88.5
Zhang et al. <sup>[31]</sup>	91.3	86.3	91.1	85.1	87.8	94.3	89.3
Li et al. <sup>[32]</sup>	92.6	88.1	<b>91.9</b>	87.4	89.3	95.9	90.8
基准方法+数据增强	92.7	86.2	91.3	<b>87.6</b>	88.2	90.9	89.5
CSFBU-Net	92.9	<b>89.4</b>	91.2	87.4	<b>91.1</b>	<b>96.2</b>	<b>91.3</b>

除此之外，本文为了验证本文所做出不同改进方法的有效性，还进行了一系列消融实验，这里同样使用 nnUNet<sup>[27]</sup>自带的普通基础 U-Net 作为基准方法，其中不包括数据增强方法的数据集取 MM-WHS 2017 中 16 例 CT 作为训练，4 例 CT 作为测试，而对于包括数据

增强方法的数据集取 MM-WHS 2017 中 20 例 MRI 以及 16 例 CT 作为训练，4 例 CT 作为测试，取最好结果，可以看到，添加了额外边缘任务后，模型的性能得到了显著提升，尤其是对于升主动脉 AO 以及右心房 RV 部位，相比基准方法分别提升了 12 个百分点和 4 个百分点，且平均 Dice 系数也有所提升；同样，单独添加特征融合模块后模型进步依然显著；然而，在将边缘任务和特征融合模块同时使用后，相比分别使用来说提升效果较不明显，但仍有少许提升，约 0.2 个百分点；最后，在将所有方法全部使用后，平均 Dice 系数由基准的 87.9% 提升至 91.7%，提升显著。

表 4：消融实验对比

方法	LV	RV	LA	RA	Myo	AO	Mean
基准方法	93.1	83.8	91.5	85.4	90.9	83.1	87.9
基准方法+边缘任务	92.8	86.8	<b>93.1</b>	86.9	89.3	95.1	90.7
基准方法+特征融合模块	92.9	85.2	92.8	87.3	91.5	95.0	90.7
基准方法+数据增强	92.6	86.5	91.7	<b>87.9</b>	89.1	93.9	90.2
基准方法+边缘任务+特征融合模块	93.1	87.2	92.8	87.0	90.0	95.4	90.9
CSFBU-Net	<b>93.2</b>	<b>89.6</b>	91.8	87.6	<b>91.8</b>	<b>96.4</b>	<b>91.7</b>

最后，为了能更直观的展示本文提出的改进方法的有效性，本文通过使用梯度加权类激活映射（Grad-CAM）<sup>[33]</sup>方法对模型进行可视化展示，这里对输出前的最后一层卷积进行可视化，针对提升较为明显的 AO 分割任务，以展示出模型在改进前后对于图像的注意力变化，如图 13。可以看到，添加边缘任务后，模型注意力热图更加集中在要分割的部位，而基准方法的注意力明显较为分散。再添加特征融合模块后，可以明显地在注意力热图中看到不同结构的边缘部分更加突出，可以看出模型对于不同结构的边缘分辨的更加清除。

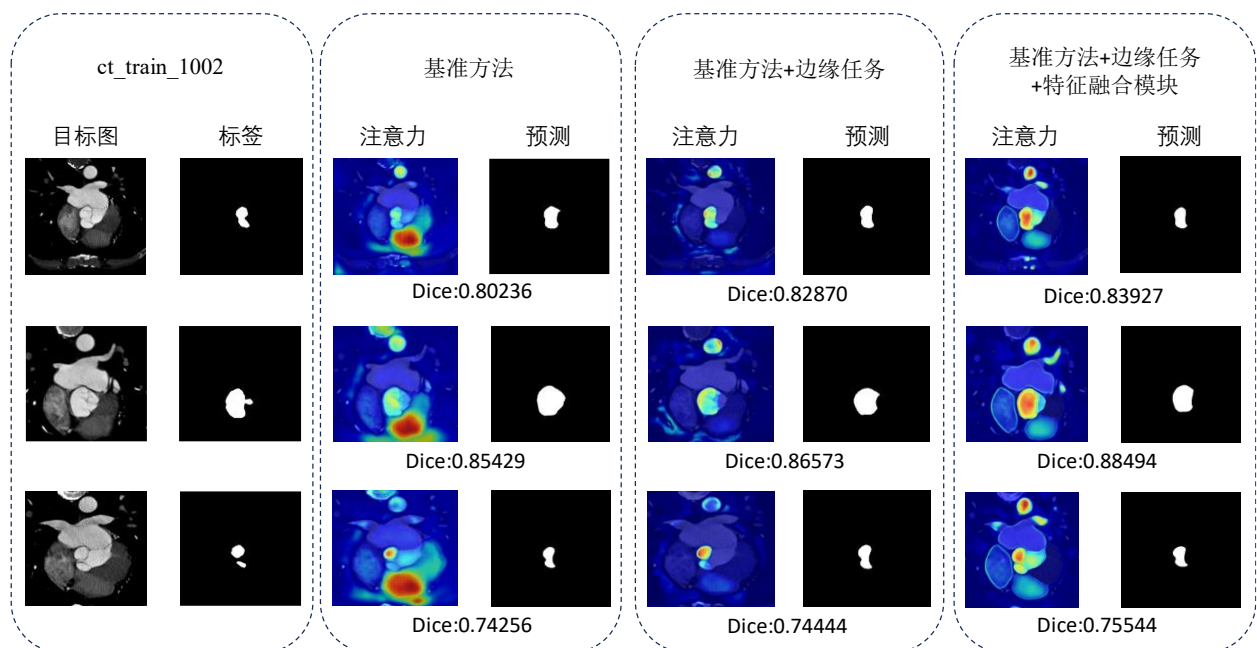


图 14：模型注意力热力图以及预测结果

### 3.5 本章小结

在本章中，首先针对目前医学影像数据稀缺且不同种类的模态数据多的问题，提出了一种基于循环一致性对抗生成网络的数据增强方法，具体来说，该方法以循环一致性对抗生成网络为核心，将 MRI 域数据转换至 CT 域数据，从而实现了 CT 域数据的扩充；其次展示了心脏影像中的不同结构分界边缘存在一些不明显的模糊区域，针对该问题，本文提出了一种基于边缘形状感知以及注意力特征融合的心脏分割方法，具体来说，该方法针对边缘区域进行额外的监督，且利用 Dice 损失针对小目标优化的性质使模型注意这些区域，此外还利用通道注意力和空间注意力机制对 U-Net 中的跳跃连接进行优化，最后，本文在公开心脏数据集上进行实验验证该方法的有效性，从而提升模型的性能。然而，需要指出的是，循环一致性对抗生成网络已经是较为落后的图像风格迁移方法，目前较为有效的新方法大部分都基于扩散模型，因此该方法改进空间很大；其次，心脏影像中不是所有结构的分界边缘都会存在一些不明显的模糊区域，因此对于是否要对所有的分割目标边缘均添加额外的监督仍然有待实验。

4 心脏图像分割系统设计与实现

第三章中，本文介绍了基于生成对抗网络的数据增强以及边缘感知的心脏分割方法，尽管本文通过实验证明了该方法的有效性，然而，在真实的医疗场景下，医生对于上述提出的方法很难直接上手，使用难度高。基于这种情况，本文构建了心脏图像检测系统，在这个系统中，医生能够直接添加要处理的目标心脏影像，该系统会快速自动的对心脏影像图像进行分割并输出结果，在此之后，医生能够方便地使用某些专业软件直接打开并查看结果，如前文提到的 3D Slicer<sup>[16]</sup>。

4.1 开发环境

本章实现的心脏影像分割系统中所有算法均依赖于 Pytorch，因此该分割系统基于 Python 编写，GUI 框架采用 PyQt5，该 GUI 框架使用 Python 语言，而 Python 的简洁性和强大的生态系统使得该 GUI 非常适合开发小型应用，此外，由于 nnUNet 提供的强大的数据处理功能，本系统还使用了 nnUNet 进行数据处理。表 5 展示了系统开发所需全部环境。

表 5：系统开发所需环境

软件名	版本号	备注
Python	3.9	编写语言
nnUNet	2.2	提供数据预处理方法
PyQt5	5.15.9	系统所用的 GUI 框架
Pytorch	2.2.1	算法使用的深度学习框架

4.2 架构设计

系统架构设计如图 15 所示。整个心脏影像分割系统可以从架构上分为三大部分，它

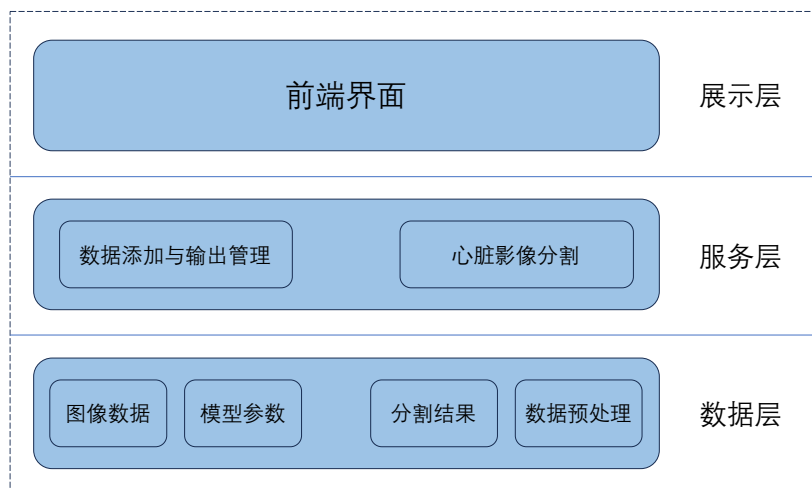


图 15: 系统架构设计图

们分别为展示层，服务层和数据层，它们自下而上进行封装，一层内相互协作实现整个系统的功能。

展示层是系统直接与用户交互的部分。该层具有图像数据添加和导出的管理功能，以及向用户展示分割结果图。用户可以通过该层添加需要分割的心脏影像、提交图像分割任务、以及观察分割后的图像效果。

服务层是系统的核心部分，该层实现数据管理的具体逻辑以及对提交的分割任务的处理逻辑，具体来说，该层从数据层获取数据以及加载模型参数以对图像进行分割，之后将得到分割结果返回值展示层。

数据层是系统的底层部分，它提供许多 API 供服务层调用，包括数据预处理，模型参数加载，图像数据获取和保存等。

#### 4.3 用户使用流程

本节将介绍用户使用该软件的一般流程，从启动系统到导出分割数据需要经过以下步骤：

- 1) 启动系统
- 2) 选择数据进行添加，以及查看原始数据是否正确
- 3) 进行数据预处理，以保证后续模型对于提交的图像数据能正确且高准确率的进行分割任务
- 4) 提交分割任务，等待系统的处理
- 5) 分割任务完成后可对分割结果进行检查，检查确定分割结果无误后将分割数据导出

上述步骤如图 15 所示。

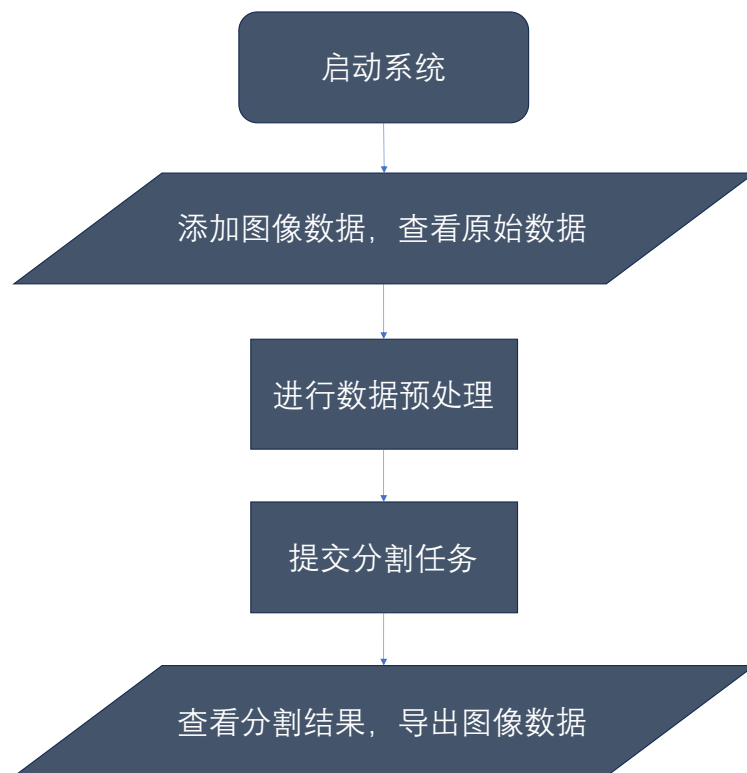


图 16: 用户使用流程图

#### 4.4 系统展示

#### 4.5 本章小结

为了探索本文提出的心脏影像分割方法的实用性价值, 本章设计和实现了一个简单的心脏影像分割系统, 该系统包括了心脏图像的添加与展示、心脏图像数据分割、心脏图像分割结果检查与导出的功能, 且在该章中介绍了该系统的开发环境和的使用流程, 其高效的心脏图像分割功能可以为后续医生对于心脏影像检查提供有效的影像数据支撑。

### 5 总结与展望

#### 5.1 工作总结

本文通过分析现有技术和方法的局限性, 提出了一种较为先进的基于生成对抗网络的数据增强以及边缘感知的心脏分割方法, 经过实验验证, 该方法一方面能有效缓解目前医学影像数据稀少的问题, 另一方面又能在心脏影像分割任务中表现出良好的准确性和鲁棒性。

在实验中, 本文将 MM-WHS 2017 数据集作为实验数据, 进行了较为详细且清楚的实验。从实验结果中可以看出, 本文提出的方法相较于目前类似的方法来说, 能够使模型提升至更高的准确率和增强模型鲁棒性, 对于心脏分割任务来说, 本文的方法能显著提升 Dice 系数, 且本文提出的方法适用性强, 无论是数据增强方法、额外边缘任务或是注意力特征融合模块都能简单的直接适用于其他模型, 且性能提升效果显著。

#### 5.2 展望

尽管本文的提出的方法具有一定的参考价值, 但是其仍然存在问题待改进, 在未来

的研究中可以针对这些问题进行进一步的深入研究和探索。

首先，本文中使用的数据增强方法是基于循环一致性神经网络[]能处理非配对数据所实现，然而目前已经出现了许多基于扩散模型的新方法，所以未来可以通过引入这些基于扩散模型的方法实现效果更好的数据增强。

其次，本文提出的额外边缘分割任务以及注意力特征融合模块其不仅可以用于医学影像领域，其同样可以用于其他的深度学习视觉任务中。

最后，为了使本文的方法能实际应用与医疗场景中，仅仅开发一项文中介绍的心脏分割系统远远不够，一方面医疗机构使用的软件中应当包括对各种不同人体部位影像的处理算法，另一方面文中开发的系统的可视化界面也存在不足，此外，由于本文的系统与算法全部基于 Python 编写，性能方面也有所欠缺，这些问题需要在未来与专业机构和医疗人员进行共同探索来完成。

总之，本文提出的深度学习算法为医学影像分割领域提供了新的性能提升方法，并由于其适用性和较好的性能效果能够轻易地用于其他先进的模型中。在未来，我们将进一步完善和扩展本文的工作，为医学影像分割领域做出贡献。

## 主要参考文献

- [1] 中国心血管健康与疾病报告编写组. 《中国心血管健康与疾病报告 2022》概要[J]. 中国介入心脏病学杂志, 2023, 31(7): 485-508.
- [2] 刘淑君. 肺源性心脏病引起的左心功能损害[J]. 江苏医药, 1991(07): 386-386.
- [3] 端知加. 风湿性心脏病症状及预防的体会[J]. 《临床心身疾病杂志》, 2015: 295-296.
- [4] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]. Medical image computing and computer-assisted intervention - MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015: 234-241.
- [5] Zhou Z, Rahman Siddiquee M M, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation[C]. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer International Publishing, 2018: 3-11.
- [6] Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]. 2016 fourth international conference on 3D vision (3DV). Ieee, 2016: 565-571.
- [7] Huang H, Lin L, Tong R, et al. Unet 3+: A full-scale connected unet for medical image segmentation[C]. ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2020: 1055-1059.
- [8] Yoo D, Park S, Lee J Y, et al. Multi-scale pyramid pooling for deep convolutional representation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2015: 71-80.
- [9] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [10] Ibtehaz N, Rahman M S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation[J]. Neural networks, 2020, 121: 74-87.
- [11] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [12] Wang W, Xie E, Li X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 568-578.
- [13] Sun G, Pan Y, Kong W, et al. DA-TransUNet: Integrating Spatial and Channel Dual Attention with Transformer U-Net for Medical Image Segmentation[J]. arXiv preprint arXiv:2310.12570, 2023.
- [14] Lan L, Cai P, Jiang L, et al. BRAU-Net++: U-Shaped Hybrid CNN-Transformer Network for Medical Image Segmentation[J]. arXiv preprint arXiv:2401.00722, 2024.
- [15] Litjens G, Kooi T, Bejnordi B E, et al. A survey on deep learning in medical image analysis[J]. Medical image analysis, 2017, 42: 60-88.
- [16] Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network[J]. Magnetic resonance imaging, 2012,



- 30(9): 1323–1341.
- [17]Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.
  - [18]Wang X, Gupta A. Generative image modeling using style and structure adversarial networks[C]. European conference on computer vision. Cham: Springer International Publishing, 2016: 318–335.
  - [19]Lu Y, Tai Y W, Tang C K. Attribute-guided face generation using conditional cyclegan[C]. Proceedings of the European conference on computer vision (ECCV). 2018: 282–297.
  - [20]Gao S, Zhou H, Gao Y, et al. BayeSeg: Bayesian modeling for medical image segmentation with interpretable generalizability[J]. Medical Image Analysis, 2023, 89: 102889.
  - [21]Zhuang X. Multivariate mixture model for myocardial segmentation combining multi-source images[J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(12): 2933–2946.
  - [22]Luo X, Zhuang X. X-Metric: An N-Dimensional Information-Theoretic Framework for Groupwise Registration and Deep Combined Computing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
  - [23]Wu F, Zhuang X. Minimizing estimated risks on unlabeled data: a new formulation for semi-supervised medical image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(5): 6021–6036.
  - [24]Liu Z, Feng Y, Yang X. Right Ventricle Segmentation of Cine MRI Using Residual U-net Convolutinal Networks[C]. 2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT). IEEE, 2019: 364–368.
  - [25]Yang W, Shi Y, Park S H, et al. An effective MR-guided CT network training for segmenting prostate in CT images[J]. IEEE journal of biomedical and health informatics, 2019, 24(8): 2278–2291.
  - [26]Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. Proceedings of the IEEE international conference on computer vision. 2017: 2223–2232.
  - [27]Isensee F, Jaeger P F, Kohl S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nature methods, 2021, 18(2): 203–211.
  - [28]Habijan M, Galić I, Leventić H, et al. Whole heart segmentation using 3d fm-pre-resnet encoder-decoder based architecture with variational autoencoder regularization[J]. Applied Sciences, 2021, 11(9): 3912.
  - [29]Chen C, Dou Q, Chen H, et al. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation[J]. IEEE transactions on medical imaging, 2020, 39(7): 2494–2505.
  - [30]Jiang J, Hu Y C, Tyagi N, et al. Tumor-aware, adversarial domain adaptation from CT to MRI for lung cancer segmentation[C]. Medical Image Computing and Computer Assisted Intervention-MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II 11. Springer International Publishing, 2018: 777–785.
  - [31]Zhang Z, Yang L, Zheng Y. Translating and segmenting multimodal medical volumes with

- cycle-and shape-consistency generative adversarial network[C]. Proceedings of the IEEE conference on computer vision and pattern Recognition. 2018: 9242–9251.
- [32]Li K, Yu L, Wang S, et al. Towards cross-modality medical image segmentation with online mutual knowledge distillation[C]. Proceedings of the AAAI conference on artificial intelligence. 2020, 34(01): 775–783.
- [33]Selvaraju R R, Cogswell M, Das A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]. Proceedings of the IEEE international conference on computer vision. 2017: 618–626.

## 致 谢

XX  
 XX  
 XX  
 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX