# 测试报告

对方小组编号：04

对方小组成员：周林宽、郑贤瑶、陈叙融、杜熙

## 1.测试方法

### 1.1 测试环境

硬件环境：NVIDIA 3090 GPU x 2，具备约50 GB显存

软件框架：Pytorch

### 1.2 数据集

**MM-WHS 2017数据集**是用于多模态全心分割的数据集。提供 20个标记和 40个未标记的 CT 图像，20个标记和 40个未标记的 MRI图像。共120张在真实临床环境中采集的多模态心脏图像。

表 1 MM-WHS 2017数据集中所包含的样本信息

| **数据集名称** | **模态** | **任务类型** | **储存格式** | **标签数量** | **已标记** | **未标记** | **组织/结构** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| MM-WHS 2017 | CT | 分割 | nii.gz | 20 | 40 | 40 | 左心室心肌/左心房血腔/左心室血腔/右心房血腔/右心室血腔/升主动脉/肺动脉 |
| MRI | 20 | 40 | 40 |

**Abdominal数据集**利用了ISBI 2019 CHAOS Challenge的 T2-SPIR MRI 训练数据（20 卷）和 Multi-atlas labeling beyond the cranial vault论文的 30 卷公共 CT 数据。两个数据集中都提供了四个腹部器官的地面实况掩模，包括肝脏、右肾、左肾和脾脏，我们用其进行多器官分割的任务。

### 1.3 测试流程

* **环境配置**：首先选择合适的硬件环境并根据说明文档对所需的软件环境进行配置，接着对MM-WHS 2017 数据集和Abdominal数据集进行预处理和划分（train/val/test）。
* **网络训练**：首先通过命令行参数和配置文件设置了UNet网络的超参数，包括学习率、批量大小、迭代周期数、类别数、随机种子等。训练过程遵循标准的监督学习范式，不涉及对抗训练或半监督训练。损失函数由MONAI库提供的DiceCELoss实现，它结合了Dice损失和交叉熵损失，权重分别由lambda\_dice和lambda\_ce参数控制，此处两者均设为0.5，表示两者权重相同。训练循环初始化了Adam优化器和指数衰减学习率调度器。每个epoch开始时，数据通过DataLoader加载，模型在每个batch上执行前向传播、计算损失、执行反向传播并更新权重。训练过程中，使用TensorBoard记录了训练损失，并且在每个epoch结束时在验证集上评估模型性能。模型性能通过一个自定义的evaluate函数评估，计算验证集上的一些指标，如Dice系数。训练过程中，如果模型在验证集上的性能在指定的patience个epoch内没有提升，则会触发早停（early stopping）机制以避免过拟合。此外，训练脚本还包括了检查点保存功能，允许在训练过程中保存模型的状态，以便后续恢复或进一步分析。训练结束后，脚本记录了总的训练时间，并关闭了TensorBoard的SummaryWriter。
* **评估指标：**训练完成后，对模型性能进行评估，使用Dice系数作为主要的评估指标。Dice系数是图像分割领域中常用的一种评价指标，用于衡量两个样本（通常是一个预测的分割图和一个真实的分割图）之间的相似度。它最初用于衡量生物学中细胞的相似性，但后来被广泛应用于图像分割任务中，尤其是在医学图像分析中。
* **结果分析：**对结果进行定量和对比分析，以评估模型性能并确定未来的改进方向。

## 2.参数配置

对于UNet分割网络，优化器采用了Adam算法，该算法默认包含了动量（momentum）参数，其值设置为0.9，同时加入了权重衰减（weight decay）以正则化模型，其值设置为。初始学习率被设定为，并且配合了指数衰减的学习率调度器，其衰减率（gamma）为0.9，这意味着学习率会在每个epoch后乘以0.9，以此来逐渐降低学习率，寻找更精确的最小值。

## 3.测试结果

### 3.1 定量测试结果

分别在MM-WHS 2017数据集和Abdominal数据集上进行测试。

表 2 MM-WHS 2017数据集的测试结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| MMWHS | | | | | |
| MR | | | | | |
|  | Dice | | | | |
| classes | AA | LAC | LVC | MYO | Avg |
| unet2D | 0.79553 | 0.862217 | 0.903405 | 0.821521 | 0.845668 |
| CT | | | | | |
|  | Dice | | | | |
| classes | AA | LAC | LVC | MYO | Avg |
| unet2D | 0.873125 | 0.862938 | 0.897506 | 0.760832 | 0.8486 |

表 3 Abdominal数据集的测试结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Abdomainal | | | | | |
| MR | | | | | |
|  | Dice | | | | |
| classes | Liver | R.kidney | L.kidney | Spleen | Avg |
| unet2D | 0.929368 | 0.902777 | 0.749538 | 0.766687 | 0.837092 |
| CT | | | | | |
|  | Dice | | | | |
| classes | Liver | R.kidney | L.kidney | Spleen | Avg |
| unet2D | 0.907819 | 0.852985 | 0.842988 | 0.621654 | 0.806362 |

### 3.2 对比测试结果

为了测试unet2D网络的域适应性能，增加unet2Dw/o对比实验对其进行验证，为进一步设计域适应网络做准备。

表 4 MM-WHS 2017数据集上的对比实验结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| MMWHS | | | | | |
| CT->MR | | | | | |
|  | Dice | | | | |
| classes | AA | LAC | LVC | MYO | Avg |
| unet2D | 0.79553 | 0.862217 | 0.903405 | 0.821521 | 0.845668 |
| unet2Dw/o | 0.172665 | 0.318049 | 0.236207 | 0.022715 | 0.187409 |
| MR->CT | | | | | |
|  | Dice | | | | |
| classes | AA | LAC | LVC | MYO | Avg |
| unet2D | 0.873125 | 0.862938 | 0.897506 | 0.760832 | 0.8486 |
| unet2Dw/o | 0.215117 | 0.338228 | 0.466618 | 0.23598 | 0.313986 |

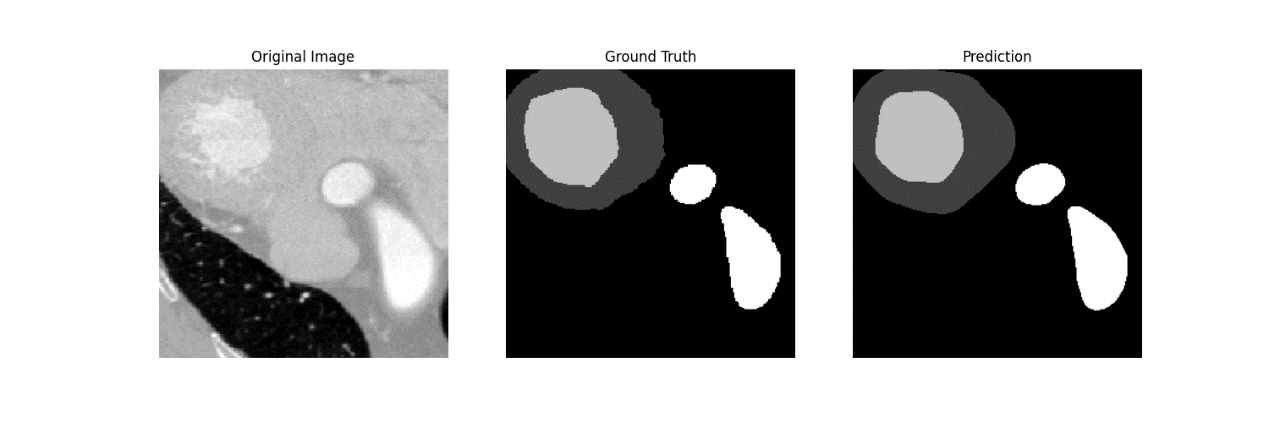
注：CT->MR表示CT作为源域数据，MR作为目标域数据；MR->CT表示MR作为源域数据，CT作为目标域数据。

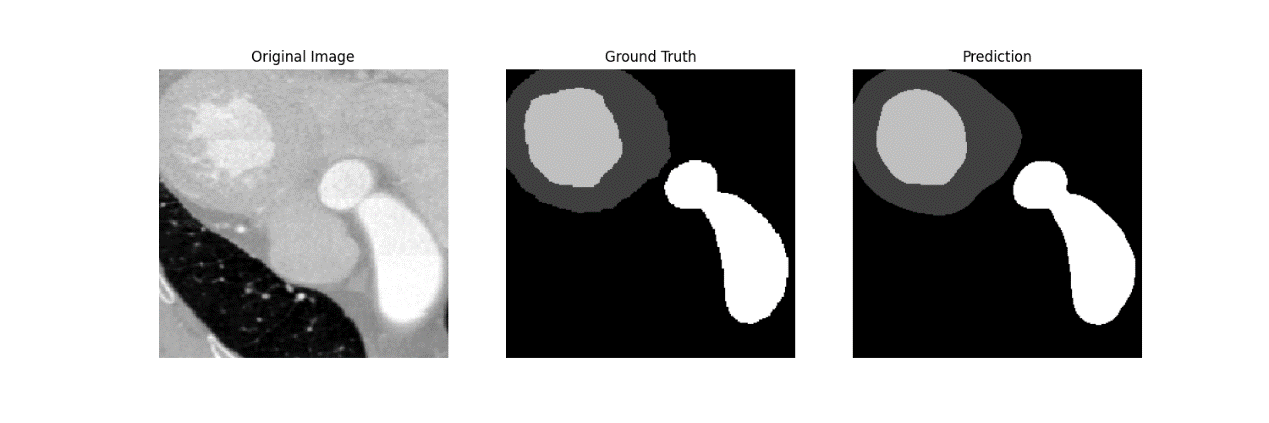
表 5 Abdominal数据集上的对比实验结果

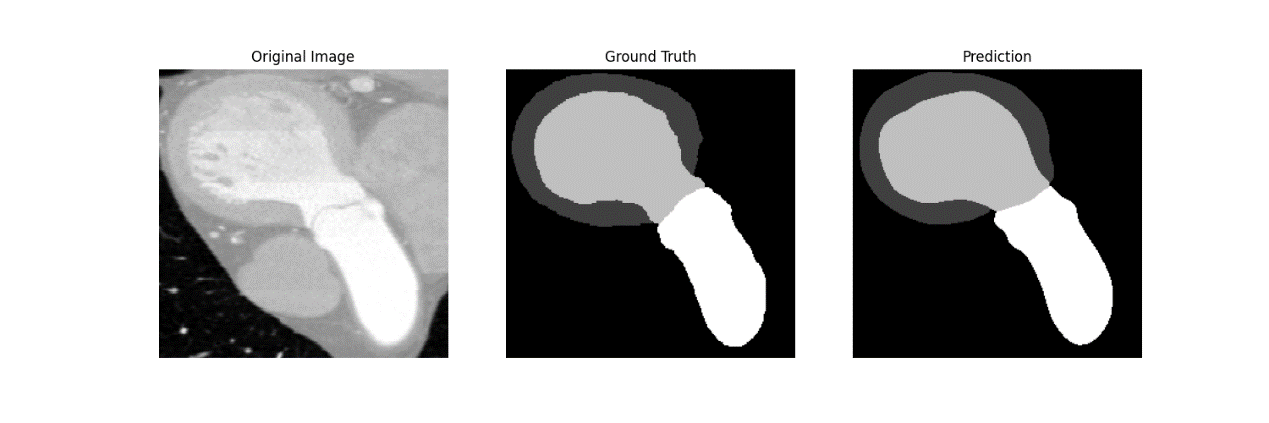
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Abdomainal | | | | | |
| CT->MR | | | | | |
|  | Dice | | | | |
| classes | Liver | R.kidney | L.kidney | Spleen | Avg |
| unet2D | 0.929368 | 0.902777 | 0.749538 | 0.766687 | 0.837092 |
| unet2Dw/o | 0.121348 | 0.056242 | 0.141547 | 0.106488 | 0.106406 |
| MR->CT | | | | | |
|  | Dice | | | | |
| classes | Liver | R.kidney | L.kidney | Spleen | Avg |
| unet2D | 0.907819 | 0.852985 | 0.842988 | 0.621654 | 0.806362 |
| unet2Dw/o | 0.538844 | 0.187134 | 0.306393 | 0.101269 | 0.28341 |

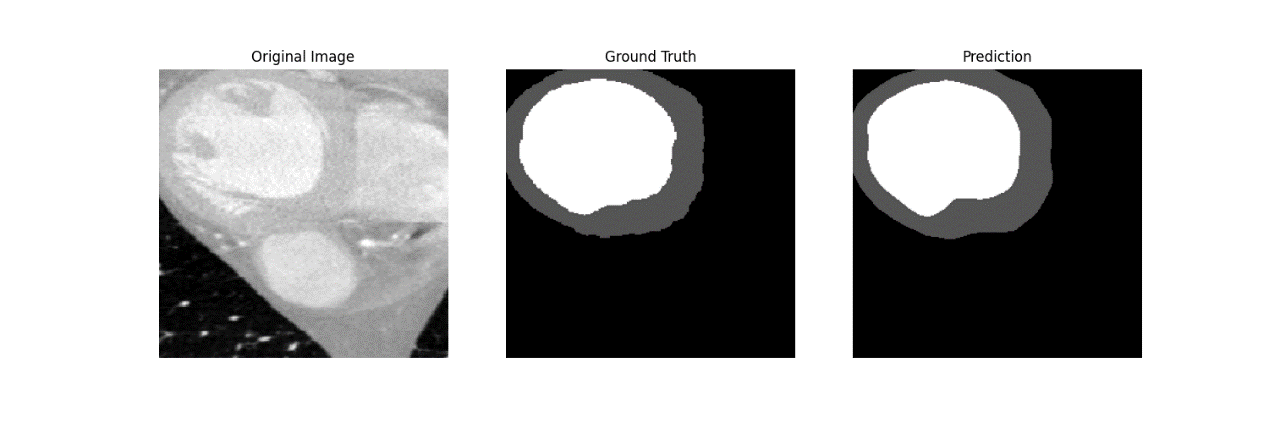
注：CT->MR表示CT作为源域数据，MR作为目标域数据；MR->CT表示MR作为源域数据，CT作为目标域数据。

### 3.3结果可视化









## 4.结果分析

实验结果表明，在Abdominal和MM-WHS 2017数据集上，UNet2D模型在图像分割任务中表现出较好的性能。无论是从CT到MR还是从MR到CT的域适应，UNet2D在平均Dice系数上均取得了较高的分数，这表明了模型在不同模态间的泛化能力。

在Abdominal数据集上，UNet2D在CT到MR转换中的平均Dice得分为0.837092，在MR到CT转换中平均得分为0.806362，这表明UNet2D能够有效地处理两种模态之间的图像分割任务。相比之下，unet2Dw/o模型在相同任务中的表现显著下降，CT到MR转换的平均Dice得分为0.106406，MR到CT转换为0.28341，这突显了UNet2D模型在分割精度上的优势。

在MM-WHS 2017数据集上，UNet2D同样展现出了良好的性能，CT到MR转换的平均Dice得分为0.845668，MR到CT转换为0.8486，进一步证实了模型的有效性。而unet2Dw/o模型在这一数据集上的表现同样不佳，CT到MR转换的平均Dice得分为0.187409，MR到CT转换为0.313986。

通过对比实验，我们发现UNet2D模型相比未进行优化的unet2Dw/o模型，在两个数据集上均有显著的性能提升。这归因于UNet2D模型在训练过程中采用了有效的正则化技术、优化的损失函数以及合适的域适应策略，这些因素共同作用使得模型在分割任务上更为精准。