



本科生毕业论文（设计）

题目： 基于无监督领域自适应的医学图像分割

姓 名 谢 华 俊

学 号 18343148

院 系 数学学院

专 业 数学与应用数学

指导教师 任传贤 (教授)

2022 年 3 月 16 日

基于无监督领域自适应的医学图像分割

Unsupervised Domain Adaptation
for Medical Image Segmentation

| | |
|------|----------|
| 姓 名 | 谢 华 俊 |
| 学 号 | 18343148 |
| 院 系 | 数学学院 |
| 专 业 | 数学与应用数学 |
| 指导教师 | 任传贤 (教授) |

2022 年 3 月 16 日

学术诚信声明

本人郑重声明：所呈交的毕业论文（设计），是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文（设计）不包含任何其他个人或集体已经发表或撰写过的作品成果。对本论文（设计）的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本论文（设计）的知识产权归属于培养单位。本人完全意识到本声明的法律结果由本人承担。

作者签名：

日 期： 年 月 日

【摘 要】

无监督领域自适应原理在医学图像领域中越来越受到关注，旨在解决医学图像中常见的域位移现象导致的分割模型性能下降问题。这篇文章提出了一个新的无监督域自适应框架用于跨模态的心脏图像分割。该框架基于解耦表示学习来提取心脏图像中的解剖学特征和模态特征，进而利用该解剖学特征来完成分割任务。同时使用了循环一致损失和对抗学习来进行图像层面上的领域自适应，实现跨模态的图像转换。我们在不成对的多模态 CT-MRI 心脏数据集上验证了该方法，实验结果表明此框架改进域位移问题的有效性。

关键词： 无监督领域自适应，医学图像分割，解耦表示学习，图像转换，对抗学习

[ABSTRACT]

Unsupervised domain adaptation has become an important and hot topic in recent studies on medical image computing, aiming to recover performance degradation when applying the neural networks to new testing domains. This paper proposes a novel unsupervised domain adaptation framework to effectively tackle the severe problem of domain shift. Our proposed method extracts anatomy features and modality features via disentangled representation learning and enhances the anatomy features towards the segmentation task. To handle unpaired training data, we fuse the cross-cycle consistency loss and adversarial learning to achieve image-to-image translation between different domains. We have extensively validated our method with a challenging application of cross-modality medical image segmentation of cardiac structures. Experimental results demonstrate the effectiveness of our method.

Keywords: Unsupervised domain adaptation, medical image segmentation, disentangled representation learning, image-to-image translation, adversarial learning

目录

| | | |
|-----|------------------|----|
| 1 | 绪论 | 1 |
| 1.1 | 选题背景与意义 | 1 |
| 1.2 | 本文方法概述 | 3 |
| 1.3 | 本文的论文结构与章节安排 | 4 |
| 2 | 相关工作 | 5 |
| 2.1 | 基于无监督域自适应的医学图像分割 | 5 |
| 2.2 | 解耦表示学习 | 6 |
| 3 | 方法 | 7 |
| 3.1 | 利用解耦表示来进行图像自适应 | 7 |
| 3.2 | 框架各组成成分的网络结构 | 8 |
| 3.3 | 其他损失函数 | 11 |
| 3.4 | 实现细节 | 13 |
| 4 | 实验与结果 | 14 |
| 4.1 | 数据集与评价指标 | 14 |
| 4.2 | 与相关方法的比较 | 14 |
| 4.3 | 解耦表示 | 16 |
| 4.4 | 跨模态图像生成 | 16 |
| 5 | 总结与展望 | 17 |
| | 参考文献 | 18 |
| | 附录 A 相关损失函数公式 | 22 |
| A.1 | 交叉熵公式 | 22 |
| A.2 | Dice 损失 | 22 |
| A.3 | 平均表面距离 | 22 |

| | |
|--------------|----|
| 致谢 | 23 |
|--------------|----|

插图目录

| | | |
|-----|------------------------------------|----|
| 1.1 | 2D 卷积神经网络卷积运算过程, 图源 ^[1] | 1 |
| 1.2 | U-Net 网络结构, 图源 ^[2] | 2 |
| 1.3 | 心脏的 CT 图, MRI 图及其对应的灰度直方图 | 3 |
| 3.1 | 整体网络结构 | 8 |
| 3.2 | 解剖学特征编码器 | 9 |
| 3.3 | 模态特征编码器 | 9 |
| 3.4 | SPADE 网络结构 | 10 |
| 3.5 | 由 SPADE ResBlock 组成的解码器结构 | 11 |
| 3.6 | 在 MMWHS 数据集上进行训练的损失 | 13 |
| 4.1 | 与相关方法的视觉分割效果比较 | 15 |
| 4.2 | MRI 图和 CT 图解剖学特征 | 16 |
| 4.3 | 跨模态图像生成: CT \rightarrow MRI | 16 |

表格目录

| | | |
|-----|--------------------------|----|
| 4.1 | 与相关方法的数值分割效果比较 | 15 |
|-----|--------------------------|----|

1 绪论

1.1 选题背景与意义

随着医疗技术的进步,各种新型的医学影像设备已广泛应用在临床诊断中,其中包括计算机断层扫描(CT)、磁共振成像(MRI)、超声成像(UI)等。医学图像中含有非常有用的信息,医生可利用CT及其他医学图像来诊断患者病情,医学图像已逐渐成为临床诊断的主要依据。因此,对医学图像处理的研究具有重要意义。其中,医学图像分割是该领域的研究热点,属于语义分割的一种,图像分割将图像划分成多个解剖学意义的区域,并在此基础上可以计算相应区域的相关定量指标。

一般地,医学图像分割可以用集合论的术语描述^[1]:给定一张医学图像 I 以及相似性约束集合 $C_i (i = 1, 2, \dots)$,对图像 I 的分割即得到其分划:

$$\bigcup_{x=1}^N R_x = I, \quad R_x \cap R_y = \emptyset, \quad \forall x \neq y, x, y \in [1, N] \quad (1.1)$$

其中每个 R_x 中的所有像素都满足相似性约束 $C_i (i = 1, 2, \dots)$,代表不同病理学意义的区域。利用传统的图像处理方法和机器学习方法来进行医学图像分割往往需要繁琐的特征工程或先验知识,而深度学习能够自动提取出复杂有效的特征,基于卷积神经网络的方法已经广泛应用在医学图像处理当中。

卷积神经网络已成功地应用于许多图像分类、目标检测和分割任务。假设输入图像的大小为 $H \times W \times 3$,大小为 (h, w, c) 的卷积核在图像的空间维度进行滑动,在每个通道进行相应位置的卷积运算,具体的2D卷积神经网络计算过程如

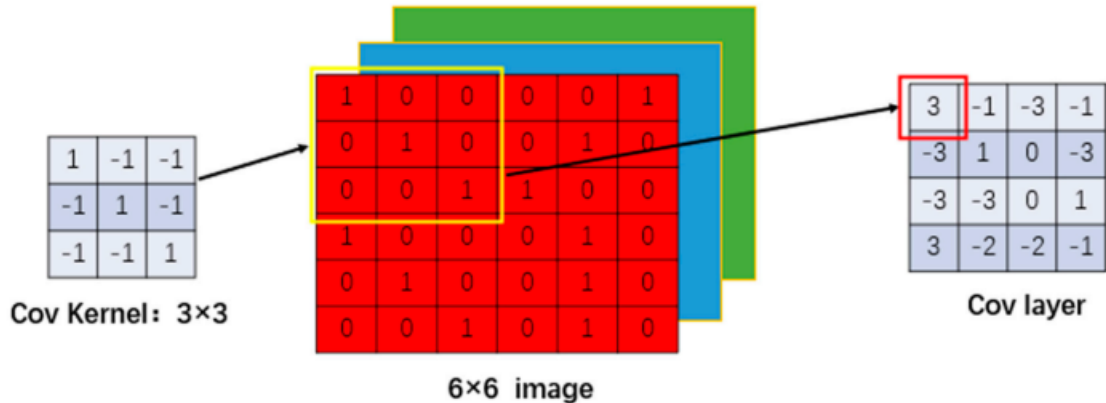


图 1.1 2D 卷积神经网络卷积运算过程, 图源^[1]

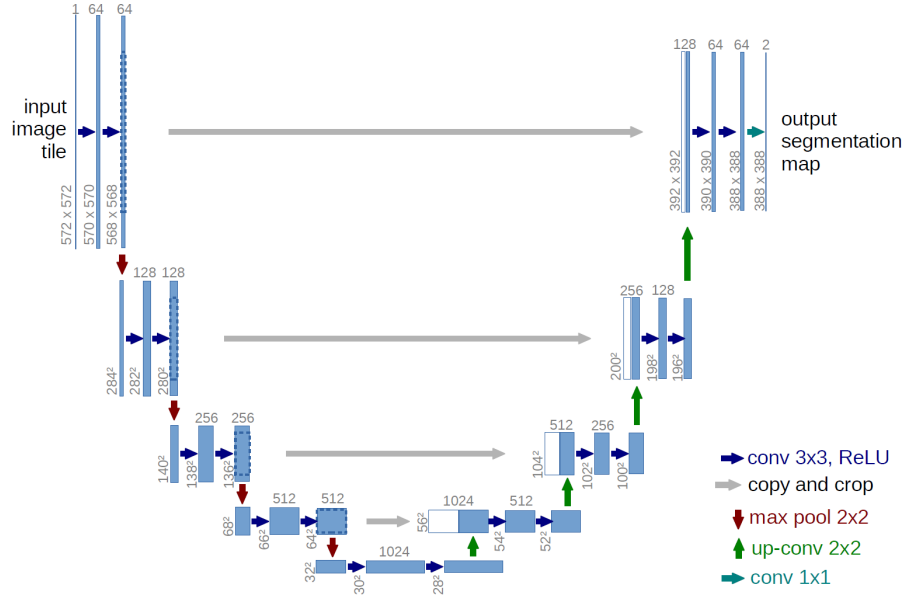


图 1.2 U-Net 网络结构，图源^[2]

图1.1所示。由卷积神经网络，下采样层，激活函数组成的深层神经网络在医学图像分割中已取得了良好的效果^[3,2,4]，其中 U-Net^[2] 的使用最为广泛，对其结构的拓展仍为该领域的热门方向。

U-Net 网络由一个 U 型网络和跳跃连接组成，如图1.2。该 U 型网络类似一种编码器解码器的结构。编码器包含四个子模块，每个子模块包含两层卷积神经网络及用于下采样的最大池化层，对称地，解码器的四个子模块也包含两层神经网络及上采样层。同时，解码器每个子模块的输入由上个子模块的输出以及编码器中对称的子模块的输出拼接而成，二者具有相同的分辨率，为的是结合低层和高层中的语义信息，帮助提取出图像中更复杂的结构，提升分割的准确率。

在临床诊断中，多模态的医学图像常被用于辅助诊断，而医学图像的标注成本比较昂贵，专家标注一张完整的分割图像大约需要 8 小时^[5]，解决这个问题的一种不失为有效的方法是基于学习的方法对一种模态的已有标注数据进行建模并用于另一种模态的图像分割。然而，机器学习方法通常假设训练集和测试集同属于一个数据分布，这个假设在该实际的临床场景中通常不成立，在训练集上训练好的模型在测试集上不能得到很好的泛化。先前的研究表明，测试误差会随着训练集和测试集的分布差异而增加^[6]，这就是域位移问题，训练集（源域）和测试集（目标域）的分布具有一定的差异。在医学图像领域，由于各种成像模式具有不同的物理原理，因此异构域位移的情况更加常见和严重，如多中心，跨模态等，见图1.3。在不需要目标域的带标注数据的情况下，无监督域自适应（Unsupervised Domain Adaptation, UDA）是解决由域位移带来的性能下降的一种方法。由于不需要额外的标注数据，UDA 在医学图像领域中越来越受到关注。

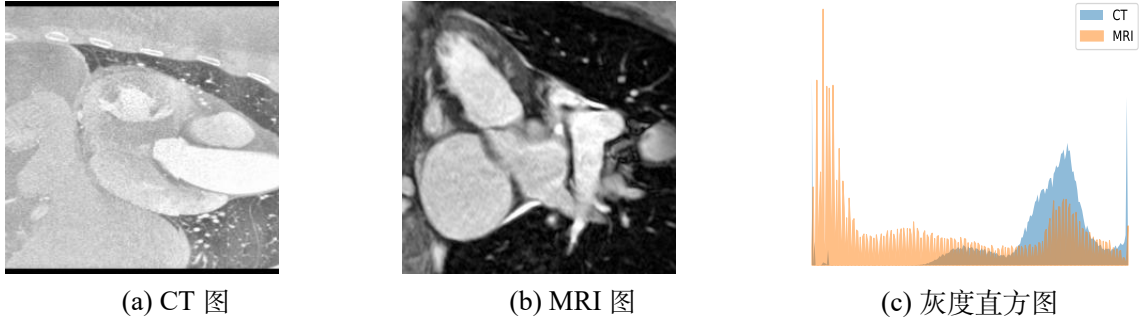


图 1.3 心脏的 CT 图，MRI 图及其对应的灰度直方图

在 UDA 的框架中,我们称带标注数据为源域,不带标注的数据为目标域,UDA 的目的对齐源域和目标域数据的分布,一种方法是特征自适应^[7-13],将不同模态的图像映射到一个模态不变的特征空间,另一种方法是图像自适应^[14],将图像从一种模态直接转换为有着相同解剖学结构的另一种模态。对于特征自适应,可以通过距离度量来显式地减小源域和目标域在特征空间上的分布差异,此外,也可以通过对抗学习来隐式地减小源域和目标域的差异。近期的研究^[15-16]提出特征自适应和图像自适应从互补的角度来缓解域位移问题,二者有效地相结合能够改进域自适应的表现。此外,多模态的医学图像具有丰富的层次性特征,一些工作^[17-19]利用解耦表示学习来提取医学图像中的域不变特征以及域特定特征。

1.2 本文方法概述

这篇文章提出了一个新的 UDA 框架用于跨模态心脏图像分割。该框架基于解耦表示学习来提取出不同模态的心脏图像的模态特征以及解剖学特征,结合图像自适应和特征自适应,进而利用不同模态的医学图像之间具有相似的解剖学特征来实现跨模态的心脏图像分割。主要贡献如下:

- 利用解耦表示学习提取出跨模态医学图像的模态特征和解剖学特征,进而实现非成对跨模态图像之间的转换。
- 在 CT-MRI 多模态心脏数据集进行相关实验,验证了方法的有效性。

1.3 本文的论文结构与章节安排

本文共分为六章，各章节内容安排如下：

第一章绪论。简单说明了本文章的选题背景与意义。

第二章为该研究方向上国内外的相关工作。

第三章为本文方法的详细介绍。

第四章为实验与结果。

第五章为总结与展望。

2 相关工作

2.1 基于无监督领域自适应的医学图像分割

最近, UDA 方法已广泛应用在医学图像领域的研究中。现有的 UDA 方法通常从三个角度对齐源域和目标域的分布。第一类是特征自适应, 通常是为了提取域不变特征, 最小化域之间的特征分布差异。第二类是图像自适应, 通过基于生成对抗网络的图像到图像转换实现跨域图像外观和风格的转变。第三类关注的是二者的结合。

2.1.1 特征自适应

传统的 UDA 方法大多通过距离度量来显式地减小源域和目标域的特征空间的分布差异。其中, MMD^[7] 是一种常用的距离度量。相应的一些扩展利用了特征分布的一些统计量, 如协方差^[8], 特征函数^[9] 来进行特征对齐。但由于用于图像分割的特征空间应该包括各种视觉线索, 如外观, 可能过于复杂和高维, 显式的距离度量如 KL 散度, Wasserstein 距离在考虑批量数据的情况下通常没有解析式, 而使用其他的一些度量如 MMD, 减小分布差异效果不明显, 而 CFD^[9] 在高维数据空间进行反向梯度传播会造成不可忽视的误差^[10]。因此, 另外一种主流的用于图像分割的域自适应方法是对抗学习, 一些工作通过对抗学习隐式地对齐特征空间, 从而学习到域不变的特征表示。对于存在严重域位移问题的跨模态分割, Dou et al.^[11] 通过微调具体的特征层并对监督特征采用对抗损失学习。Tsai et al.^[12] 过对齐输出空间来整合空间和几何结构信息。Wang et al.^[20] 提出了一种方法对眼底图像的熵和边界进行对抗自适应。Vesal et al.^[13] 通过对齐点云来对齐形状特征。

2.1.2 图像自适应

近年来一些图像到图像转换的方法如 CycleGAN^[14] 利用生成对抗网络来把源域的图像转换为与目标域风格类似的图像。然后, 这些生成的图像继承了源域图像的分割图, 可以用于目标域分割网络的监督学习。Zhang et al.^[21] 利用循环和形状一致性对抗网络进行多模态大脑 MRI 分割。在 Liu et al.^[22] 中, 作者结合基于注意力的神经网络生成目标域图像。Yang et al.^[17] 启发于解耦表示的思想, 训练变分自编码器将图像模态特征和内容特征进行解耦, 从而用来生成不同域的图像。

2.1.3 特征和图像自适应

近年来,一些研究提出将特征自适应和图像自适应结合起来,能更好地减轻域位移的影响^[15-16]。CyCADA^[15]将UDA作为一种风格迁移方法减小源域和目标域之间的外观差距,同时独立地对齐两个域的潜在特征空间。Chen et al.^[16]提出了SIFA框架,不同于^[15],其协同利用特征自适应和图像自适应成统一的网络,从互补的角度来对齐心脏图像MRI和CT两种模态的分布。他们还扩展SIFA为Bidirectional-SIFA^[23],添加深度监督特征的对齐,以及从双向的角度(MRI \leftrightarrow CT)来探索领域自适应在多模态医学图像分割的应用。Han et al.^[24]提出了一个对称结构的域自适应网络。Tomar et al.^[25]利用自注意空间自适应标准化的结构来使得网络关注医学图像中具有解剖学意义的区域。Ye et al.^[26]利用对比学习和原型相似度比较来显式地对齐源域和生成图像的特征。

2.2 解耦表示学习

对图像特征进行解耦已广泛应用于图像转换和风格迁移等计算机视觉应用任务中^[27-28]。InfoGAN^[29]将特征解耦成可解释的潜变量与噪声,同时最大化生成数据与潜变量之间的互信息,使得生成数据与潜变量更相关。Yang et al.^[17]通过训练变分自编码器将图像特征进行解耦,得到8维的风格特征以及高维表示的内容特征,利用自适应实例归一化(Adaptive Instance Normalization, AdaIN)和风格特征对内容特征进行仿射变换,进而实现不同域之间图像的风格迁移。最近的研究也开始将解耦表示学习用于医学影像分析当中^[18,30,19]。Pei et al.^[19]利用自注意力机制和归零损失函数来进一步对医学图像特征进行解耦,进而完成跨模态的医学图像分割。

3 方法

记来自源域的带标注数据集为 $\mathbb{D}_s = \{x_i^s, y_i^s\}_{i=1}^{m_s}$, 其中 $x_i^s \in \mathbb{R}^{w \times h \times 3}$, $y_i^s \in \mathbb{R}^{w \times h \times c}$, c 是图像中具有不同语义结构的数目, m_s 表示源域数据集的大小。记来自目标域的无标注数据集 $\mathbb{D}_t = \{x_i^t\}_{i=1}^{m_t}$, m_t 表示目标域数据集的大小。我们旨在利用带标注的源域数据集来对目标域的医学图像进行语义分割。这篇文章将图像的特征进行解耦, 从而学习到两个域的解剖学特征和模态特征, 文章提出的框架如图3.1所示。对于每个模态的图像, 首先利用编码器得到对应域的解剖学特征以及模态特征, 然后交换各自的模态特征, 将相应的解剖学特征和模态特征输入到解码器得到另一个模态的图像。参考 CycleGAN, 重复编码解码的操作对原图像进行重构。^①

3.1 利用解耦表示来进行图像自适应

不同模态的医学图像通常具有相似的解剖学结构, 即域不变特征, 二者的差异往往来自于其不同的模态特征, 即域特定的特征。将这些特征进行解耦, 进而实现跨模态图像之间的转换。如图所示, 编码器 $E_d^{anatomy}$ 和 $E_d^{modality}$ (下文将解剖学特征 anatomy 简记为 a , 模态特征 modality 简记为 m) 分别提取图像的解剖学特征以及模态特征, 数学形式的表达即 $z_d^c = E_d^c x^d$, 其中 $c \in \{a, m\}$, $d \in \{s, t\}$ 。为了实现解耦表示, 常用的方法是通过权重共享将跨模态图像的解剖学特征映射到同一个空间, 然而, 权重共享很难保证编码出相同的解剖学特征表示。这里引入判别器 D_a 来隐式地对齐跨域的域不变特征, 其对抗损失为

$$\begin{aligned} L_{adv}^a = & \mathbb{E}_{x_s} \left[\frac{1}{2} \log D_a(z_s^a) + \frac{1}{2} \log (1 - D_a(z_s^a)) \right] \\ & + \mathbb{E}_{x_t} \left[\frac{1}{2} \log D_a(z_t^a) + \frac{1}{2} \log (1 - D_a(z_t^a)) \right]. \end{aligned} \quad (3.1)$$

得到解耦表示后, 两个域特定的解码器, 记为 G_s, G_t , 用于生成两个模态的伪图。如图所示, G_s 可以结合源域图像 x_s 的模态特征和目标域图像 x_t 的解剖学特征来生产伪图 $x_s^* = G_s(z_s^m, z_t^a)$, 同理, 我们可以利用 D_t 得到 $x_t^* = G_t(z_t^m, z_s^a)$ 。启发于 CycleGAN, 重复该过程对原图像进行重构, 得到 x_s^{rec} 和 x_t^{rec} , 即 $x_s^{rec} = G_s(E_s^m(x_s^*), E_t^a(x_t^*))$, $x_t^{rec} = G_t(E_t^m(x_t^*), E_s^a(x_s^*))$ 。这些重构的图像应与原图像一致,

^① 开源代码: <https://github.com/Lawliet-Xie/undergraduate-thesis>

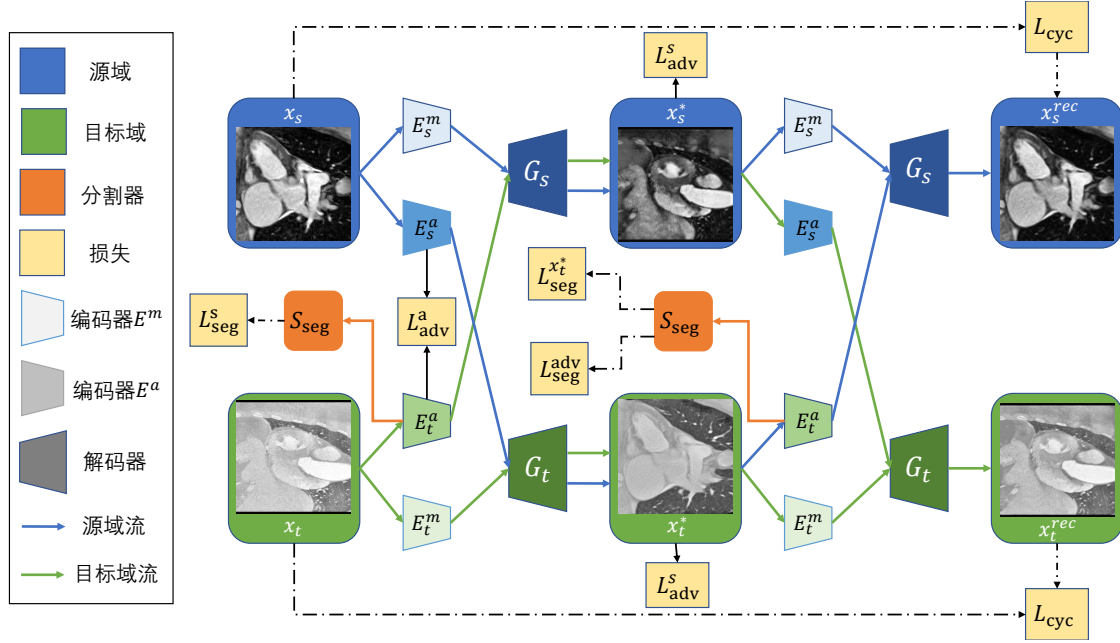


图 3.1 整体网络结构

故有循环一致性损失（cycle consistency loss）：

$$\mathcal{L}_{cyc} = \mathbb{E}_{(x_s, x_t) \in P(\mathbb{D}_s, \mathbb{D}_t)} \|x_s^{rec} - x_s\|_1 + \mathbb{E}_{(x_s, x_t) \in P(\mathbb{D}_s, \mathbb{D}_t)} \|x_t^{rec} - x_t\|_1. \quad (3.2)$$

其中 $\|\cdot\|_1$ 代表 L1 范数。

此外，为了使得伪图 x_s^* 和 x_t^* 接近于真实的图像，引入两个判别器 D_s 和 D_t ，用于对抗学习。基于生成对抗网络的思想，判别器 D_s 能够分别出 x_s 是真图，而 x_s^* 是伪图，同时以上的由编码器解码器组成的生成器结构能够生成出足够真实的样本来干扰判别器， D_t 同理。故目标函数为：

$$\min_{(E_s^a, E_t^m, G_t)} \max_{D_t} \mathcal{L}_{adv}^t = \mathbb{E}_{x_t \in P(\mathbb{D}_t)} [\log D_t(x_t)] + \mathbb{E}_{(x_s, x_t) \in P(\mathbb{D}_s, \mathbb{D}_t)} [\log(1 - D_t(x_t^*))], \quad (3.3)$$

$$\min_{(E_t^a, E_s^m, G_s)} \max_{D_s} \mathcal{L}_{adv}^s = \mathbb{E}_{x_s \in P(\mathbb{D}_s)} [\log D_s(x_s)] + \mathbb{E}_{(x_s, x_t) \in P(\mathbb{D}_s, \mathbb{D}_t)} [\log(1 - D_s(x_s^*))]. \quad (3.4)$$

3.2 框架各组成成分的网络结构

3.2.1 编码解剖学特征

解耦解剖学特征的编码器由卷积神经网络组成，将 2D 的医学图像映射到其空间表示，即 $E^a : x \rightarrow z^a$ ，其中 $z^a \in \mathbb{R}^{w \times h \times k}$ ， k 表示所需要的解剖学特征数。编

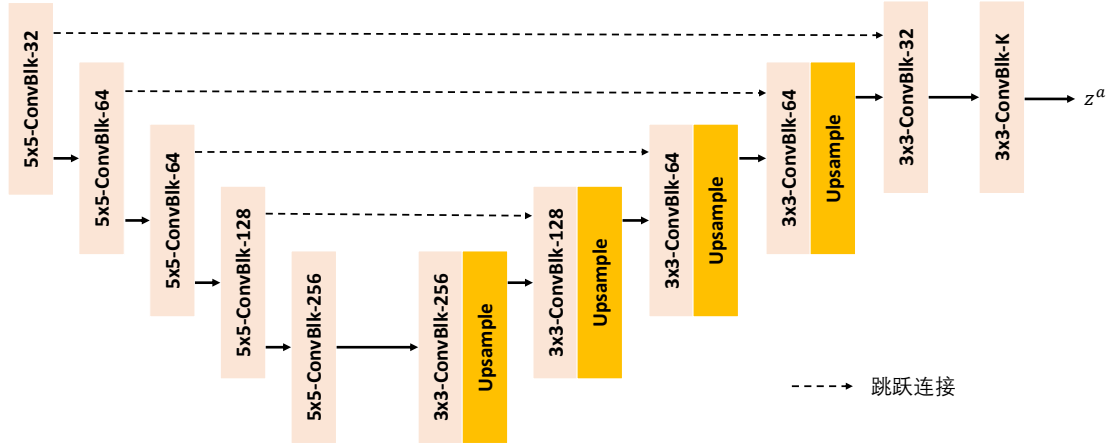


图 3.2 解剖学特征编码器

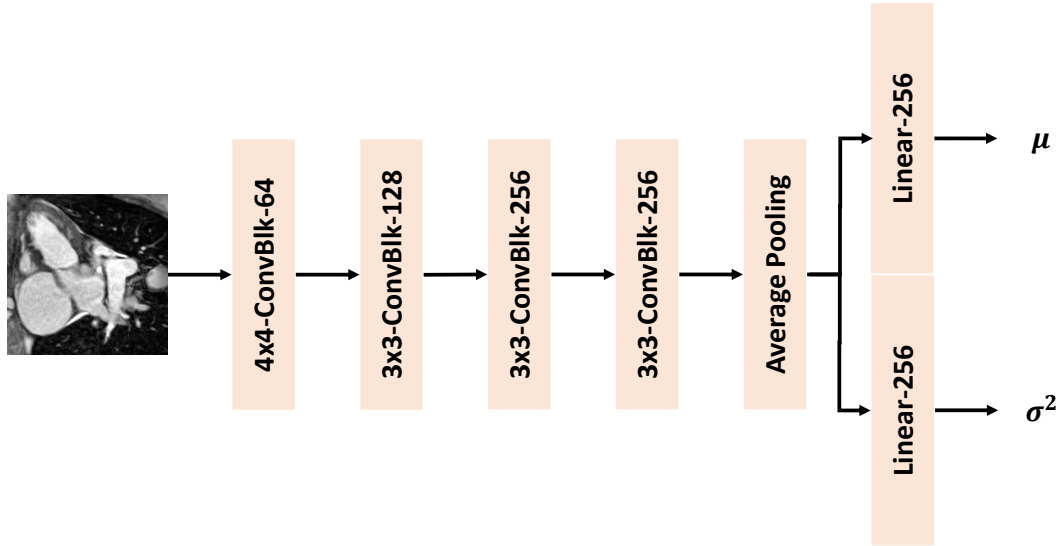


图 3.3 模态特征编码器

编码器整体采用类似 U-Net^[2] 的网络结构，如图3.2所示，包含了下采样，上采样以及跳跃连接，能够有效地融合重要的局部和非局部信息，提取出丰富的语义信息，得到特征中的每个通道表示心脏中一些具有解剖学意义的子结构。

3.2.2 编码模态特征

编码器 E^m 为了学习到医学图像模态特征 z^m 的后验分布 $q(z^m|x)$ ，采用变分自编码器^[31](Variational Autoencoder, VAE)。简单地说，VAE 学习一个低维的潜空间，该特征空间中的潜表示能够匹配一个先验分布 $p(z) = \mathcal{N}(\mathbf{0}, I)$ ，从风格迁移和解耦表示的角度来说，从该空间中进行模态的采样，能够生成各种模态的图像，整体结构如图3.3所示。

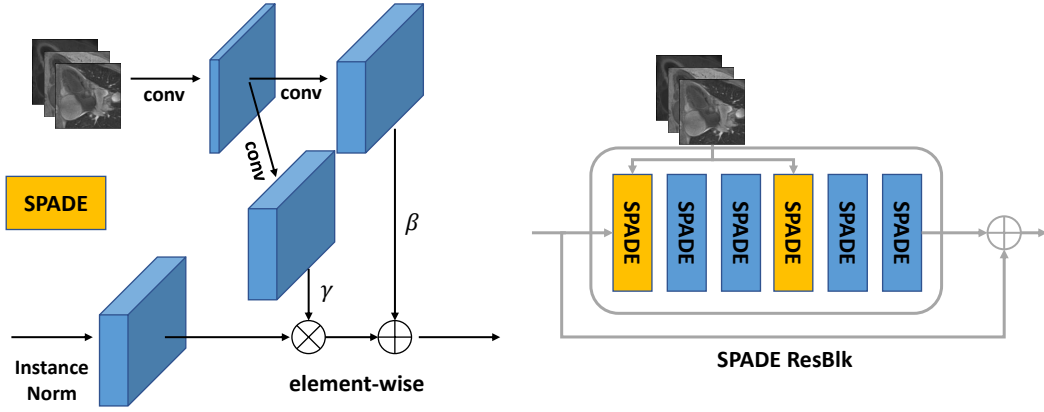


图 3.4 SPADE 网络结构

3.2.3 解码器

解码器 G 以模态特征和解剖学特征作为输入，用于进行图像的生成与重构。SPADE^[32] (Spatially Adaptive Denormalization) 是一种能有效利用语义信息以及风格信息进行图像生成的一种结构。与普通的卷积神经网络对比，其通过空间自适应归一化来更好地保留语义信息，具体公式如下：

$$\begin{aligned}\hat{x}_{c,i,j}(z^a) &= \gamma_{c,i,j}(z^a) * \frac{z_{c,i,j}^m - \mu_c}{\sigma_c} + \beta_{c,i,j}(z^a) \\ \mu_c &= \frac{1}{HW} \sum_{i,j} z_{c,i,j}^m, \quad \sigma_c^2 = \frac{1}{HW} \sum_{i,j} (z_{c,i,j}^m - \mu_c)^2\end{aligned}\quad (3.5)$$

在该模块中，解剖学特征 z^a 被映射到一个嵌入空间来产生仿射变换的参数 γ 和 β ，不同于 AdaIN 等其他条件归一化方法，SPADE 产生的 γ 和 β 包含了空间和相应的语义信息，接着利用 γ 和 β 对已经实例归一化的模态特征 z^m 来实现多模态的图像生成^[32]，如图3.4所示。解码器（图3.5）由多个 SPADE 块组成，能够充分融合解剖学特征和模态特征来进行伪图的生成以及原图的重建。

3.2.4 分割器

解剖学特征已经包含了医学图像中丰富的语义信息，在此基础上增加一层卷积神经网络 S_{seg} 来完成语义分割。由于源域图像带标注 y_s ，生成的伪图 x_t^* 与 x_s 共享相似的解剖学特征 z_s^a ，故 S_{seg} 对二者的分割预测 \hat{y}_{t^*} 和 \hat{y}_s 能够与 y_s 进行监督学

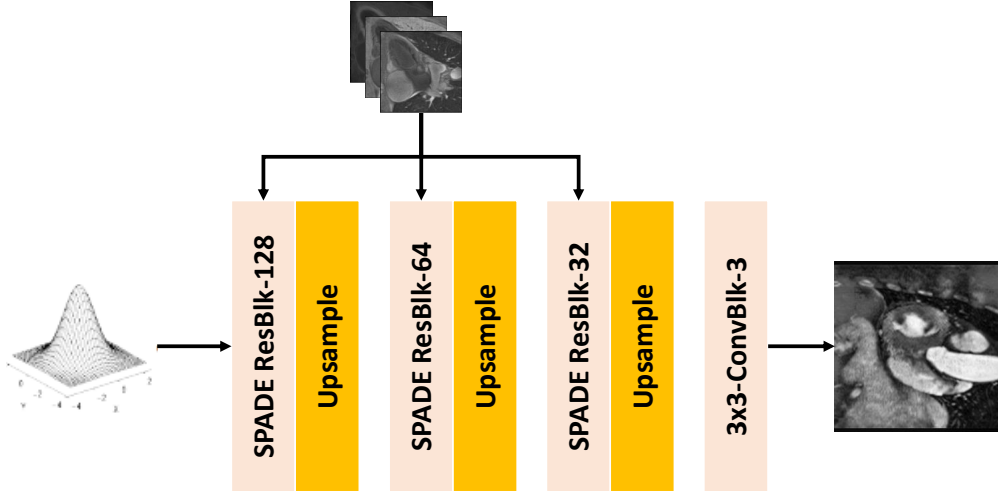


图 3.5 由 SPADE ResBlock 组成的解码器结构

习，有如下损失：

$$\mathcal{L}_{\text{seg}}^{x_s} = \mathcal{H}(y_s, \hat{y}_s) + \alpha \cdot \text{Dice}(y_s, \hat{y}_s), \quad (3.6)$$

$$\mathcal{L}_{\text{seg}}^{x_t^*} = \mathcal{H}(y_s, \hat{y}_{t^*}) + \alpha \cdot \text{Dice}(y_s, \hat{y}_{t^*}). \quad (3.7)$$

其中 H 为交叉熵损失，Dice 损失用于衡量语义分割效果，具体计算公式见附录， α 为超参数。总的分割损失可表示为 $\mathcal{L}_{\text{seg}} = \mathcal{L}_{\text{seg}}^{x_s} + \mathcal{L}_{\text{seg}}^{x_t^*}$ 。此外，引入判别器 D_{seg} ，进一步从语义空间上缩小源域和目标域图像的分布差异，

$$\begin{aligned} \min_{(E_t^a, S_{\text{seg}})} \max_{D_{\text{seg}}} \mathcal{L}_{\text{adv}}^{\text{seg}} &= \mathbb{E}_{x_t^*} [\log D_{\text{seg}}(S_{\text{seg}}(E_t^a(x_t^*)))] \\ &+ \mathbb{E}_{x_t} [\log (1 - D_{\text{seg}}(S_{\text{seg}}(E_t^a(x_t))))]. \end{aligned} \quad (3.8)$$

3.3 其他损失函数

- 正交损失

对得到的解剖学特征进行正交性约束，使得每个通道能学习到不同病理意义的子结构，减少特征的冗余性，更好地解耦解剖学特征和模态特征。

$$\mathcal{L}_{\text{ortho}} = \mathbb{E}_{x_s} [\|E_s^a(x_s)E_s^a(x_s)^T - I\|_F] + \mathbb{E}_{x_t} [\|E_t^a(x_t)E_t^a(x_t)^T - I\|_F]. \quad (3.9)$$

其中 I 是单位矩阵， $\|\cdot\|_F$ 代表 Frobenius 范数。

- KL 散度损失

该损失函数目的是使模态表示 $q(z^m|x) = \mathcal{N}(\mu(x), \sigma(x)^2)$ 与先验高斯分布 $p(z) = \mathcal{N}(\mathbf{0}, I)$ 一致，从图像生成的角度，我们从该先验高斯分布中随机采样，能得到不同模态的图像。

$$\begin{aligned}\mathcal{L}_{\text{KL}} &= \mathcal{D}_{\text{KL}}(q(z^m|x) \| p(z)) \\ &= - \int q(z^m|x) \log \frac{q(z^m|x)}{p(z)} dz \\ &= \frac{1}{2} \sum_{j=1}^d (\sigma_j + u_j^2 - \log \sigma_j - 1).\end{aligned}\tag{3.10}$$

- 重构损失

除了交叉循环重构损失，以 E_s^a 和 E_s^m 为输入，进入解码器 G_s 理应能重构出原来的图像 x_s ，对于目标域图像 x_t 同理，即如下重构损失：

$$\mathcal{L}_{\text{id}} = \mathbb{E}_{x_s} [\|G_s(E_s^a(x_s), E_s^m(x_s)) - x_s\|_1] + \mathbb{E}_{x_t} [\|G_t(E_t^a(x_t), E_t^m(x_t)) - x_t\|_1].\tag{3.11}$$

结合以上提到的损失函数，总的目标函数为

$$\begin{aligned}\mathcal{L} &= \lambda_1 \mathcal{L}_{adv}^a + \lambda_2 \mathcal{L}_{cyc} + \lambda_3 \mathcal{L}_{adv}^s + \lambda_4 \mathcal{L}_{adv}^t \\ &\quad + \lambda_5 \mathcal{L}_{\text{seg}} + \lambda_6 \mathcal{L}_{adv}^{\text{seg}} + \lambda_7 \mathcal{L}_{\text{ortho}} + \lambda_8 \mathcal{L}_{\text{KL}} + \lambda_9 \mathcal{L}_{\text{id}}.\end{aligned}\tag{3.12}$$

其中， $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7, \lambda_8, \lambda_9$ 为超参数，用于平衡各个模块的重要性。第一项损失旨在使得源域和目标域图像能够提取到域不变特征，第二项和最后一项损失旨在重构回原来的图像，第三第四项损失旨在生成足够真实的伪图从而能够干扰判别器，第五项为利用监督学习进行语义分割产生的损失，其余项的提出是为了促进不同模态图像之间域自适应，缓解域位移问题。

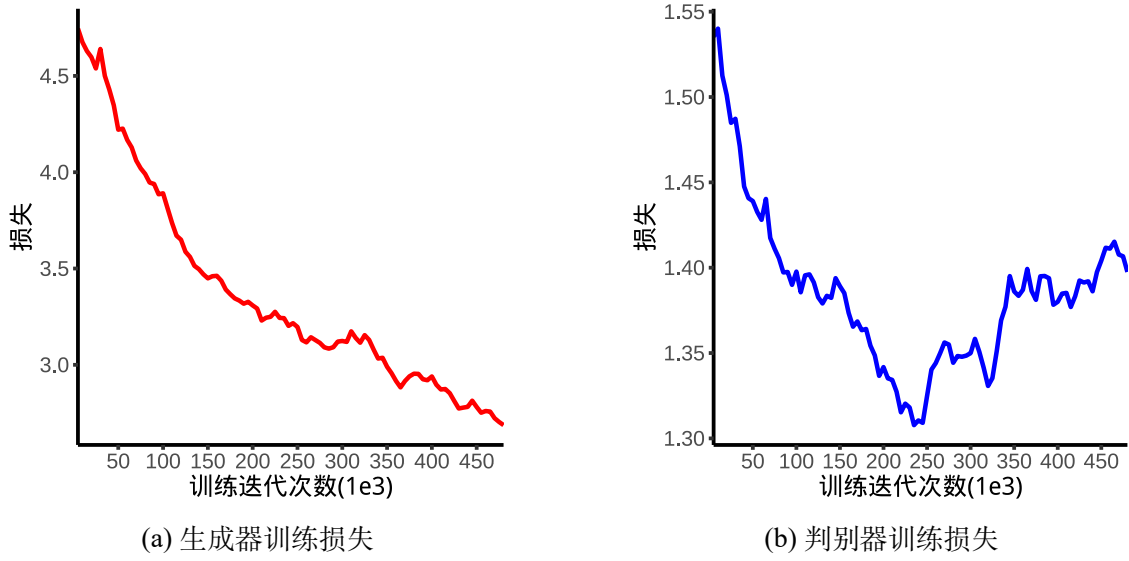


图 3.6 在 MMWHS 数据集上进行训练的损失

3.4 实现细节

这一小节我们补充训练过程中各种参数的设定以及相应模块的网络结构设置。

在训练过程当中，对分割器模块使用初始学习率为 1×10^{-3} 的 Adam 优化器，学习率以每两个 epoch 以 0.9 的倍率进行衰减，对于其他模块则以恒定的 2×10^{-4} 的学习率进行随机梯度下降更新参数，总训练 epoch 数为 50，数据批量大小为 8。

每个模块基于卷积神经网络块 (ConvBlock)，由 3×3 卷积核的 CNN，实例归一化以及参数为 0.2 的 leaky RELU 组成。为了生成对抗网络训练的稳定性，参照 CycleGAN 的设定，我们使用了基于均方误差的对抗损失。此外，对引入的四个判别器，采用 PatchGAN^[33] 的设置对 70×70 的补丁进行判别。

关于超参数的设定， $\alpha = 1, \lambda_1 = 2, \lambda_2 = 10, \lambda_3 = 2, \lambda_4 = 2, \lambda_5 = 0.5, \lambda_6 = 2, \lambda_7 = 1, \lambda_8 = 0.01, \lambda_9 = 2.5$ ，解剖学特征数 $K = 8$ 。在 MMWHS 数据集上训练的误差如图 3.6 所示，图 3.6a 展示了生成器的训练损失，即式 3.12 中总损失减去判别器损失，图 3.6b 则展示了判别器在训练过程中的损失。可以看到，生成器损失随着迭代次数不断下降，意味着框架能生成足够真实的伪图以及解剖学特征编码器能编码出域不变的特征，而判别器损失则先降后增，说明训练阶段初期判别器能够分辨出真图和伪图，而到了后期，由于生成器能够生成足够真实的图像来干扰判别器，使得判别器的损失回升。

4 实验与结果

4.1 数据集与评价指标

我们在 Multi-Modality Whole Heart Segmentation Challenge 2017 数据集^[34] 上验证文章所提出的 UDA 方法的有效性。该数据集用于 MRI 图和 CT 图的心脏分割，包含从不同中心收集的不对应的 20 个病人的 MRI 图和 CT 图。数据带标注分割图，包括升主动脉 (Ascending Aorta, AA)、左心房血腔 (Left Atrium Blood Cavity, LAC)、左心室血腔 (Left Ventricle Blood Cavity, LVC) 以及左心室心肌 (Myocardium, MYO)。

实验采用 MRI 图作为源域，CT 图作为目标域，通过 UDA 方法来进行跨模态的医学图像分割。每个模态随机挑选 80% 的数据用于训练，剩余 20% 的数据用于测试。目标域 CT 图的标注仅在测试中使用，不参与到训练过程当中。所有数据均进行均值为 0，方差为 1 的标准化。由于原始数据是 3D 图像，为了减小模型参数量，我们对图像进行切片，并裁剪至 256×256 的大小，用 2D 图像进行模型训练。此外，进行旋转，翻转，仿射变换等数据增强来防止过拟合。

在测试过程中，实验采用两组常用的评价指标来量化分割模型的表现。其中之一是 Dice 系数 ([%])，计算预测的分割图与真实的分割图的重叠体积，越高代表效果越好，另一个是平均表面距离 ASD([voxel])，衡量分割模型在边界处的表现，ASD 值越低越好。

4.2 与相关方法的比较

我们将文章提出的方法与几种常用的 UDA 方法进行比较，包括 DANN, ADDA, CycleGAN, CyCADA 等以及该数据集上的 SOTA 方法 SIFA。其中，DANN 和 ADDA 仅进行了特征自适应，CycleGAN 仅进行了图像自适应，CyCADA, SIFA 则对二者进行结合。此外，实验还衡量了模型效果的上界和下界，上界即用带标注的 CT 图进行监督学习 (Supervised)，下界即不进行域自适应 (W/o adaptation)，直接将在 MRI 数据上进行训练的模型用于 CT 图像的分割。

表4.1统计了各方法的结果，可以看到我们的方法与没进行域自适应 (W/o adaptation) 相比有了大幅的提升，且与其他 UDA 方法相比也有一定的竞争力。无域自适应的模型在四个心脏结构上仅获得了 2.1% 的平均 Dice，从中可以知道 MRI 图

像和 CT 图像之间严重的域位移问题。经过域自适应调整, 平均 Dice 提升至 64.4%, 平均 ASD 减小至 6.5%, 而且在升主动脉这一结构上的 Dice 分数超过了所有方法, 在平均表面距离 ASD 指标上, 文章提出的框架取得了 6.5 个体素的最小值, 意味着解耦表示学习提取出的解剖学特征用于分割能得到较为平滑的表面, 使得与真实分割图的表面距离较小, 从而表现出我们方法的有效性。

表 4.1 与相关方法的数值分割效果比较

| Method | Volumetric Dice \uparrow | | | | | Volumetric ASD \downarrow | | | | |
|----------------------------|----------------------------|-------------|-------------|-------------|-------------|-----------------------------|------------|------------|------------|------------|
| | AA | LAC | LVC | MYO | Average | AA | LAC | LVC | MYO | Average |
| Supervised | 90.7 | 92.4 | 89.2 | 85.2 | 89.4 | 0.9 | 2.3 | 3.1 | 2.1 | 2.1 |
| W/o adaptation | 6.1 | 0 | 1.3 | 1.2 | 2.1 | 64.4 | 78.3 | 68.5 | 38.3 | 62.4 |
| DANN ^[35] | 39.0 | 45.1 | 28.3 | 25.7 | 34.5 | 16.2 | 9.2 | 12.1 | 10.1 | 11.9 |
| ADDA ^[36] | 47.6 | 60.9 | 11.2 | 29.2 | 37.2 | 13.8 | 10.2 | N/A | 13.4 | N/A |
| PnP-AdaNet ^[11] | 74.0 | 68.9 | 61.9 | 50.8 | 63.9 | 12.8 | 6.3 | 17.4 | 14.7 | 12.8 |
| SynSeg-Net ^[37] | 71.6 | 69.0 | 51.6 | 40.8 | 58.2 | 11.7 | 7.8 | 7.0 | 9.2 | 8.9 |
| AdaOutput ^[12] | 65.2 | 76.4 | 54.4 | 43.6 | 59.9 | 17.9 | 5.5 | 5.9 | 8.9 | 9.6 |
| CycleGAN ^[14] | 73.8 | 75.7 | 52.3 | 28.7 | 57.6 | 11.5 | 13.6 | 9.2 | 8.8 | 10.8 |
| CyCADA ^[15] | 72.9 | 77.0 | 62.4 | 45.3 | 64.4 | 9.6 | 8.0 | 9.6 | 10.5 | 9.4 |
| SIFA v1 ^[16] | 81.1 | 76.4 | 75.7 | 58.7 | 73.0 | 10.6 | 7.4 | 6.7 | 7.8 | 8.1 |
| SIFA v2 ^[23] | 81.3 | 79.5 | 73.8 | 61.6 | 74.1 | 7.9 | 6.2 | 5.5 | 8.5 | 7.0 |
| Ours | 84.8 | 50.8 | 64.0 | 57.8 | 64.4 | 3.2 | 5.4 | 10.6 | 6.7 | 6.5 |

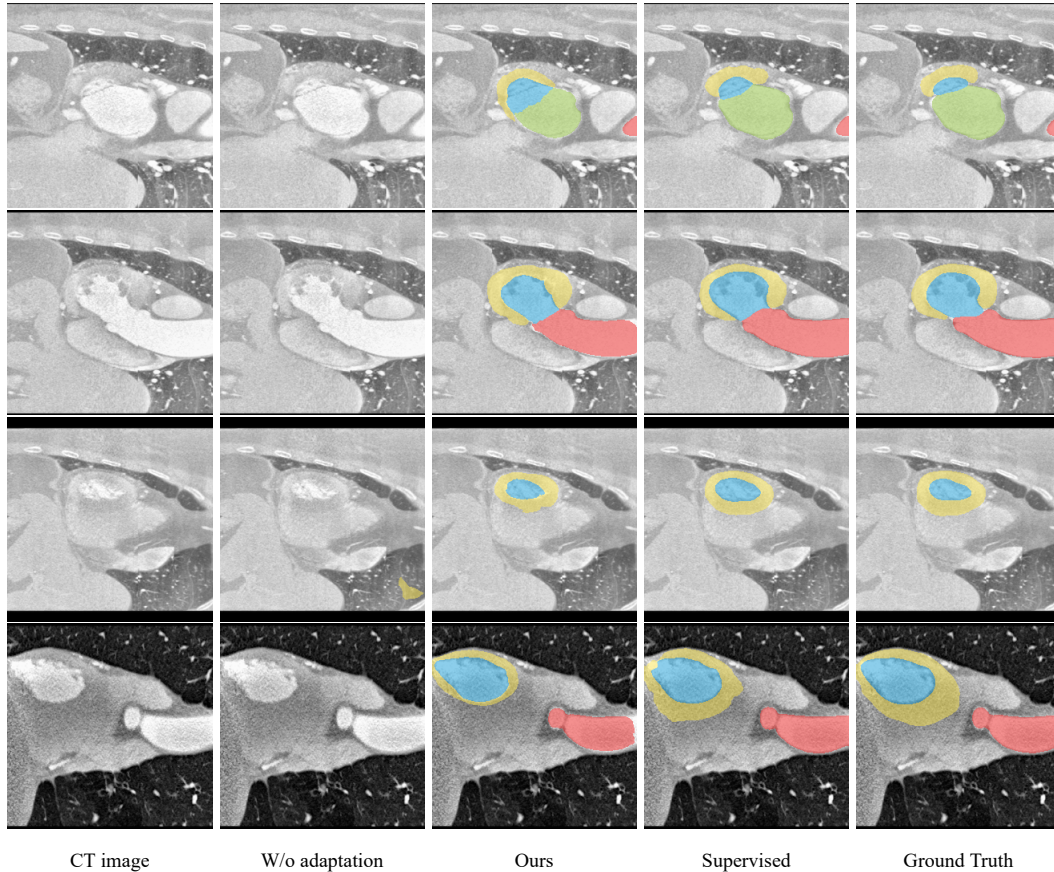
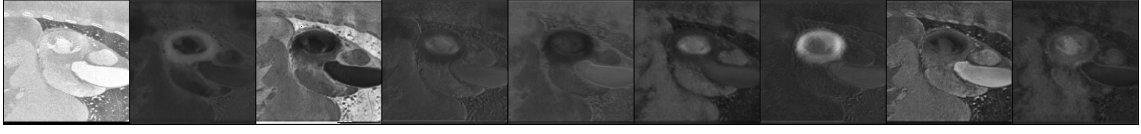


图 4.1 与相关方法的视觉分割效果比较



(a) E_s^a 编码器对 MRI 图编码出的解剖学特征



(b) E_t^a 编码器对 CT 图编码出的解剖学特征

图 4.2 MRI 图和 CT 图解剖学特征

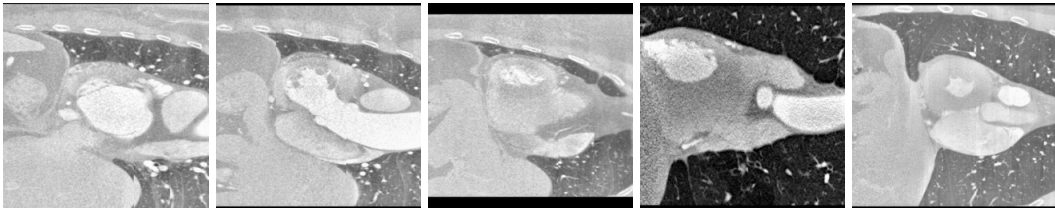
视觉比较结果如图4.1所示。可以看到，在不进行域自适应的情况下，网络很难对心脏结构进行有效的分割。而在进行域自适应之后，缓解了域位移问题带来的模型性能下降问题，分割器可以生成有解剖学意义的分割图，

4.3 解耦表示

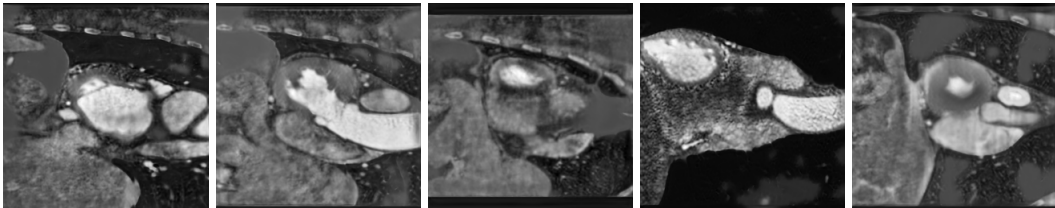
如图4.2所示，模型的解剖学特征编码器能够得到图像中有解剖学病理学意义的相关子结构，且在正交约束下，每个通道能提取到不同意义的病理区域，有效地促进了分割器的效果以及跨模态图像的生成。

4.4 跨模态图像生成

图4.3中展示了跨模态图像的生成，从 CT 图转换到 MRI 图，可以看到，图像的外观转换为相应模态，同时保留了有效的语义区域用于分割，从中可以看出该框架有效地解耦模态特征和解剖学特征用于图像的生成。



(a) CT 图



(b) 生成的 MRI 图

图 4.3 跨模态图像生成：CT \rightarrow MRI

5 总结与展望

这篇文章提出了一个新的 UDA 框架用于跨模态心脏图像分割。该框架利用解耦表示学习来提取多模态图像的解剖学特征和模态特征，同时利用对抗学习来进行图像自适应和特征自适应的融合。我们在不成对的多模态心脏数据集 MMWHS 上验证了该方法的效果。实验结果表明该框架改进域位移问题的有效性。

对于未来的工作，可以把该框架从 2D 图像扩展到与临床诊断更相关的 3D 图像，同时应用到不同的多模态数据集上来验证方法的鲁棒性以及泛化性，此外，可以利用生成对抗网络来进行多模态医学图像的生成进行自监督训练来缓解缺乏带标注医学图像数据的问题。

参考文献

- [1] LIU X, SONG L, LIU S, et al. A review of deep-learning-based medical image segmentation methods[J]. Sustainability, 2021, 13(3): 1224.
- [2] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. [S.l.]: Springer, 2015: 234-241.
- [3] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv preprint arXiv:1412.7062, 2014.
- [4] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.
- [5] ZHUANG X. Challenges and methodologies of fully automatic whole heart segmentation: a review[J]. Journal of healthcare engineering, 2013, 4(3): 371-407.
- [6] BEN-DAVID S, BLITZER J, CRAMMER K, et al. Analysis of representations for domain adaptation[J]. Advances in neural information processing systems, 2007, 19: 137.
- [7] LONG M, ZHU H, WANG J, et al. Deep transfer learning with joint adaptation networks[C]//International conference on machine learning. [S.l.]: PMLR, 2017: 2208-2217.
- [8] SUN B, SAENKO K. Deep coral: Correlation alignment for deep domain adaptation[C]//European conference on computer vision. [S.l.]: Springer, 2016: 443-450.
- [9] WU F, ZHUANG X. Cf distance: A new domain discrepancy metric and application to explicit domain adaptation for cross-modality cardiac image segmentation[J]. IEEE Transactions on Medical Imaging, 2020, 39(12): 4274-4285.
- [10] WU F, ZHUANG X. Unsupervised domain adaptation with variational approximation for cardiac segmentation[J]. IEEE Transactions on Medical Imaging, 2021, 40(12): 3555-3567.
- [11] DOU Q, OUYANG C, CHEN C, et al. Pnp-adanet: Plug-and-play adversarial domain adaptation network with a benchmark at cross-modality cardiac segmentation[J]. arXiv preprint arXiv:1812.07907, 2018.
- [12] TSAI Y H, HUNG W C, SCHULTER S, et al. Learning to adapt structured output space for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2018: 7472-7481.

- [13] VESAL S, GU M, KOSTI R, et al. Adapt everywhere: Unsupervised adaptation of point-clouds and entropy minimization for multi-modal cardiac image segmentation[J]. IEEE Transactions on Medical Imaging, 2021, 40(7): 1838-1851.
- [14] 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. [S.l.: s.n.], 2017: 2223-2232.
- [15] HOFFMAN J, TZENG E, PARK T, et al. Cycada: Cycle-consistent adversarial domain adaptation[C]//International conference on machine learning. [S.l.]: PMLR, 2018: 1989-1998.
- [16] CHEN C, DOU Q, CHEN H, et al. Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation[C]//Proceedings of the AAAI conference on artificial intelligence: volume 33. [S.l.: s.n.], 2019: 865-872.
- [17] YANG J, DVORNEK N C, ZHANG F, et al. Unsupervised domain adaptation via disentangled representations: Application to cross-modality liver segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. [S.l.]: Springer, 2019: 255-263.
- [18] CHARTSIAS A, JOYCE T, PAPANASTASIOU G, et al. Disentangled representation learning in cardiac image analysis[J]. Medical image analysis, 2019, 58: 101535.
- [19] PEI C, WU F, HUANG L, et al. Disentangle domain features for cross-modality cardiac image segmentation[J]. Medical Image Analysis, 2021, 71: 102078.
- [20] WANG S, YU L, LI K, et al. Boundary and entropy-driven adversarial learning for fundus image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. [S.l.]: Springer, 2019: 102-110.
- [21] 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. [S.l.: s.n.], 2018: 9242-9251.
- [22] LIU X, WEI X, YU A, et al. Unpaired data based cross-domain synthesis and segmentation using attention neural network[C]//Asian Conference on Machine Learning. [S.l.]: PMLR, 2019: 987-1000.
- [23] 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.
- [24] HAN X, QI L, YU Q, et al. Deep symmetric adaptation network for cross-modality medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2021, 41(1): 121-132.

- [25] TOMAR D, LORTKIPANIDZE M, VRAY G, et al. Self-attentive spatial adaptive normalization for cross-modality domain adaptation[J]. IEEE Transactions on Medical Imaging, 2021, 40(10): 2926-2938.
- [26] YE Z, JU C, MA C, et al. Unsupervised domain adaption via similarity-based prototypes for cross-modality segmentation[M]//Domain Adaptation and Representation Transfer, and Affordable Healthcare and AI for Resource Diverse Global Health. [S.l.]: Springer, 2021: 133-143.
- [27] LEE H Y, TSENG H Y, HUANG J B, et al. Diverse image-to-image translation via disentangled representations[C]//European Conference on Computer Vision. [S.l.: s.n.], 2018.
- [28] LEE H Y, TSENG H Y, MAO Q, et al. Drit++: Diverse image-to-image translation via disentangled representations[J]. International Journal of Computer Vision, 2020: 1-16.
- [29] CHEN X, DUAN Y, HOUTHOOFT R, et al. Infogan: Interpretable representation learning by information maximizing generative adversarial nets[J]. Advances in neural information processing systems, 2016, 29.
- [30] CHEN C, DOU Q, JIN Y, et al. Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. [S.l.]: Springer, 2019: 447-456.
- [31] KINGMA D P, WELLING M. Auto-encoding variational bayes[J]. arXiv preprint arXiv:1312.6114, 2013.
- [32] PARK T, LIU M Y, WANG T C, et al. Semantic image synthesis with spatially-adaptive normalization[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. [S.l.: s.n.], 2019: 2337-2346.
- [33] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2017: 1125-1134.
- [34] ZHUANG X, SHEN J. Multi-scale patch and multi-modality atlases for whole heart segmentation of mri[J]. Medical image analysis, 2016, 31: 77-87.
- [35] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks[J]. The journal of machine learning research, 2016, 17(1): 2096-2030.
- [36] TZENG E, HOFFMAN J, SAENKO K, et al. Adversarial discriminative domain adaptation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2017: 7167-7176.

- [37] HUO Y, XU Z, MOON H, et al. Synseg-net: Synthetic segmentation without target modality ground truth[J]. IEEE transactions on medical imaging, 2018, 38(4): 1016-1025.

附录 A 相关损失函数公式

A.1 交叉熵公式

交叉熵用于衡量两个概率分布之间的差异，常用于分类问题，语义分割是像素级的一类分类问题。二元交叉熵公式定义为：

$$\mathcal{L}_{CE}(y, \hat{y}) = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})). \quad (\text{A.1})$$

其中 \hat{y} 为预测值， y 为真实值。而在语义分割中，不同语义的区域往往面积（体积）不一致，存在数据不平衡的问题，故常常使用平衡交叉熵 (Balanced Cross-Entropy)：

$$\mathcal{H}(y, \hat{y}) = -(\beta * y \log(\hat{y}) + (1 - \beta) * (1 - y) \log(1 - \hat{y})). \quad (\text{A.2})$$

其中， $\beta = 1 - \frac{y}{H * W}$ ， H 和 W 为图像的宽度和高度。

A.2 Dice 损失

给定两个集合 A 和 B ，二者的 Dice 分数定义为：

$$\text{Dice}(A, B) = \frac{2|A \cap B|}{|A| + |B|}. \quad (\text{A.3})$$

其范围为 $[0, 1]$ ，值越大表示两个集合重合度越高。

A.3 平均表面距离

对于 3D 物体的每个体素，如果在其 18 邻域范围内至少有一个像素不是物体，则认为该体素属于表面，依次遍历所有的体素便得到物体的表面。记两个 3 维图像的表面为 X 和 Y ，平均表面距离 (Average Surface Distance) 定义为：

$$\text{ASD}(X, Y) = \frac{\sum_{x \in X} \min_{y \in Y} d(x, y)}{|X|}. \quad (\text{A.4})$$

其中 $d(\cdot, \cdot)$ 代表欧氏距离。

致谢

四年时间转眼即逝，青涩而美好的本科生活快告一段落了。回首这段时间，我不仅学习到了很多知识和技能，而且提高了分析和解决问题的能力与养成了一定的科学素养。这一切与老师的教诲和同学们的帮助是分不开的，在此对他们表达诚挚的谢意。

然后要感谢的是我的指导老师任传贤教授。我作为一名本科生，缺少学术研究经验，不能很好地弄清所研究问题的重点、难点和热点，也很难分析自己的工作所能够达到的层次。任老师对整个研究领域有很好的理解，以其渊博的知识和敏锐的洞察力给了我非常有帮助的方向性指导。他严谨的治学态度与辛勤的工作方式也是我学习的榜样，在此向任老师致以崇高的敬意和衷心的感谢。

同时，我还要感谢大学四年和我一起走过的同学朋友对我的关心与支持，与他们一起学习、生活，让我在大学期间生活的很充实，给我留下了很多难忘的回忆。

最后我要感谢我的家人，正是他们的无私的奉献和支持，我才有了不断拼搏的信心和勇气，才能取得现在的成果。

谢华俊

2022年3月16日