

分类号 _____

编号 _____

U D C _____

密级 _____



南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

本科生毕业设计（论文）

题 目： 基于跨域数据的白内障手术场景

分割算法研究

姓 名： 区明阳

学 号： 11913001

系 别： 计算机科学与工程系

专 业： 计算机科学与技术

指导教师： 刘江 教授

2023 年 5 月 1 日

CLC _____

Number _____

UDC _____

Available for reference ☐Yes ☐No



SUSTech Southern University
of Science and
Technology

Undergraduate Thesis

Thesis Title: Semantic Segmentation of Cataract
Surgical Scene on Cross Domain Data

Student Name: Ou Mingyang

Student ID: 11913001

Department: Department of Computer Science
and Engineering

Program: Computer Science and Technology

Thesis Advisor: Professor Liu Jiang

Date: May 1, 2023

诚信承诺书

1. 本人郑重承诺所呈交的毕业设计（论文），是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料均真实可靠。

2. 除文中已经注明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的作品或成果。对本论文的研究作出重要贡献的个人和集体，均已在文中以明确的方式标明。

3. 本人承诺在毕业论文（设计）选题和研究内容过程中没有抄袭他人研究成果和伪造相关数据等行为。

4. 在毕业论文（设计）中对侵犯任何方面知识产权的行为，由本人承担相应的法律责任。

作者签名:

_____年____月____日

COMMITMENT OF HONESTY

1. I solemnly promise that the paper presented comes from my independent research work under my supervisor's supervision. All statistics and images are real and reliable.
2. Except for the annotated reference, the paper contents no other published work or achievement by person or group. All people making important contributions to the study of the paper have been indicated clearly in the paper.
3. I promise that I did not plagiarize other people's research achievement or forge related data in the process of designing topic and research content.
4. If there is violation of any intellectual property right, I will take legal responsibility myself.

Signature:

Date:

基于跨域数据的白内障手术场景 分割算法研究

区明阳

(计算机科学与工程系 指导教师: 刘江)

[摘要]: 白内障手术是治疗白内障的主要手段, 而开发手术系统有助于为手术医师在术中提供手术环境感知, 提供技术支持。其中, 手术场景的语义分割人物是基础性任务。

基于监督训练的深度神经网络在手术场景的语义分割任务上有较好的表现。然而, 1) 像素级别的手术场景训练标签的获取是极其昂贵的。2) 在单一数据集中训练的深度神经网络在临床应用中由于数据分布偏移问题常出现分割性能下降的情况 3) 尽管领域自适应方法的提出能有效改善数据分布偏移所带来的负面影响。近年来, 由于对数据隐私保护的重视程度提升, 在实际场景中我们往往无法直接利用训练域数据。

基于以上难点与挑战, 本文提出了一种基于知识蒸馏框架下的无源域领域自适应算法。该算法首先利用多视图学习与特征一致性约束, 在带训练标签的源域中训练出鲁棒性强的白内障手术场景语义分割网络。随后结合知识蒸馏框架与熵最小化的方法, 我们提出一套仅仅使用目标域无标签数据和源域模型的领域自适应方法, 提升模型在目标域的分割性能。

为了验证方法的可行性, 我们在两个白内障手术数据集上展开了一系列的对比实验, 实验结果表明提出的方法具有一定的有效性。

[关键词]: 语义分割; 白内障手术; 无源域领域自适应

[ABSTRACT]: Cataract surgery is the main treatment for cataracts, and developing surgical systems can provide surgeons with surgical environment perception and technical support. Among them, semantic segmentation of surgical scenes is a fundamental task.

Supervised deep neural networks have shown good performance in semantic segmentation tasks of surgical scenes. However, 1) obtaining pixel-level training labels for surgical scene is extremely expensive. 2) Deep neural networks trained on a single dataset often show a decrease in segmentation performance due to the problem of data distribution shift in clinical applications. 3) Although the proposed domain adaptation methods can effectively improve the negative impact of data distribution shift, in practical scenarios, we often cannot directly use the training domain data due to the increasing importance of data privacy protection.

Based on the above challenges, this paper proposes a source-free domain adaptation algorithm based on the knowledge distillation framework. The algorithm first uses multi-view learning and feature consistency constraints to train a robust cataract surgery semantic segmentation network in the source domain with labeled training data. Then, combining the knowledge distillation framework and the entropy minimization method, we propose a domain adaptation method that only uses unlabeled target domain data and source domain model to improve the segmentation performance of the model in the target domain.

To verify the feasibility of the proposed method, we conducted a series of comparative experiments on two cataract surgery datasets, and the experimental results show that the proposed method has certain effectiveness.

[Key words]: Semantic Segmentation; Cataract surgery; Source-free domain adaptation

目录

1. 引言	1
1.1 研究背景和意义	1
1.2 存在问题与挑战	2
1.3 论文创新点与贡献	3
2. 国内外研究相关研究现状	3
2.1 图像语义分割算法	3
2.2 计算机视觉领域自适应算法	4
3. 研究方法	5
3.1 问题阐述	5
3.2 方法概述	5
3.3 基于风格差异性与语义一致性约束的特征提取器	5
3.4 基于多视图学习的白内障语义分割网络	7
3.5 基于均值教师知识蒸馏的领域自适应方法	8
4. 实验	10
4.1 实验数据集	10
4.1.1 CaDIS 数据集	10
4.1.2 CatInstSeg 数据集	11
4.2 实现细节	12
4.2.1 软硬件环境与超参数	12
4.2.2 实验目的	13
4.2.3 对比方法	13
4.2.4 评价指标	14
4.3 实验结果	14
5. 结论	17

参考文献	18
附录	21
致谢	22

1. 引言

1.1 研究背景和意义

眼球接受光线的照射并在视网膜上成像，是人体形成视觉，对外界环境感知最重要的器官之一。眼的结构如图1所示，大致可以分为外部结构和内部结构。外部结构包含外层、中层和内层三层。外层包括在前半部分的透明近球状结构角膜（Cornea）以及在后半部分的稠密的白色纤维状的结构巩膜（Sclera）。中层包括由虹膜（Iris）、睫状体（Ciliary body）和脉络膜（Choroid）组成的葡萄膜。其中虹膜中有圆形开口，称为瞳孔（Pupil）。后层主要是视网膜，通过视神经将成像转化为电信号传输给大脑。眼的内部结构可以分为前房（Anterior chamber）、后房（Posterior chamber）以及（Vitreous chamber）玻璃房。在虹膜后房，晶状体（Lens）通过睫状小带（Zonules）和睫状体相连。

在成像的过程中，外界物体的光线首先透过角膜折射照进眼球，虹膜通过调节瞳孔的大小来控制进入光束的大小。随后晶状体对进入光线进行二次折射，使其聚焦在视网膜上，形成图像。此外，通过改变晶状体的形状，人可以聚焦不同距离的物体。因此在视觉的形成过程中，晶状体起着非常重要的作用^[1]。

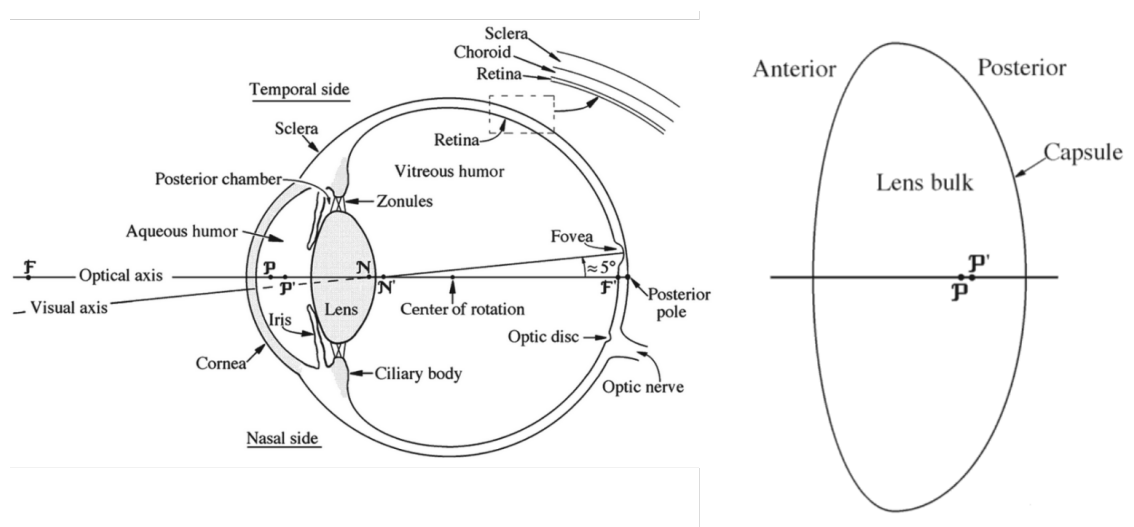


图 1 眼球结构图，图片来源于 Goossens^[1]

晶状体的主要构成成份是晶体蛋白。白内障患者由于各种病理性的因素晶状体中晶体蛋白变性，使得晶状体的形态和折射率因此而改变，眼底成像因此变得模糊。在严重的情况下，光线无法透过晶状体，最终导致失明。具体来说，年龄的增长、先天性遗传、青光眼手术等其他眼科手术中的创伤、其他眼科疾病的影响、药物的使用以及系统性疾病都可能是罹患白内障的原因^[2]。根据 Steinmetz 等人的研究，白内障是全球中重度视力损伤的主要原因之一，也是失明最主要的原因。在 2020 年，全球 50 岁以上的白内障患病老年人人数高达 152 万^[3]。通过手术

将人工晶体替换浑浊的晶状体是最主要的治疗手段。然而，1) 在发展中国家，大量白内障患者缺乏获得手术治疗的机会，2) 白内障手术医师的培训速度远不及每年白内障患者的新增速度，3) 随着人口老龄化的家具，每年新增的老年白内障患者在不断增加以及 4) 目前在医学上仍然没有有效的预防或延缓白内障病变的方法和途径^[2]，以上四点导致了白内障亟需更高效的治疗方法。

近年来，手术机器人系统有了长足的发展，在包括白内障在内的眼科手术上有着良好的应用前景。根据人工介入的程度，手术机器人系统可以分为全人工、机器人器械辅助系统、远程操控系统、手术协作系统以及半自动或全自动手术机器人系统^[4]。在远程操控、手术协作和半自动或全自动手术机器人系统中，手术医师都需要通过视觉反馈来对复杂的手术环境进行感知，以作出下一步的手术操作决策。语义分割通过对输入图像进行像素级别的分类，能够帮助对白内障手术场景进行划分，增强对手术场景的感知，有助于更好地开发实现白内障手术机器人系统，提高白内障手术效率。同时，开发针对白内障语义分割的算法并与白内障手术培训系统结合，提供增强的视觉反馈，对于加快白内障手术医师的培训也有着重要的临床意义。

1.2 存在问题与挑战

对于自然图像的语义分割任务，在街景分割、遥感图像分割等领域都已经有了大量高效精准的语义分割算法。与此同时，针对医学图像的语义分割也有相应的突出研究成果。然而，在白内障手术场景分割的这个子课题下，仍有许多待未能解决的问题：

1. 由于白内障手术属于显微操作手术，手术器械大多形态较为纤细，在手术场景中的占比较小。此外，手术过程中可能出现出血，手术强光照射等不确定因素，导致手术器械外观形态发生改变。综上两点需要语义分割模型在白内障手术场景图像中需要更高的分割精度。
2. 此外，在医学领域中，待标注的数据集通常数据量较少，特别是白内障语义分割相关的医学图像数据集，需要大量的时间和精力对手术器械学习、识别以及手术场景的标注，使得良好的白内障手术语义分割数据集非常难以获得。与此同时，存在大量未标注的白内障手术数据。但是由于拍摄的场景、设备、图像的质量等诸多因素，导致不同数据集之间数据分布存在较大差异，也就是域偏移现象。临床运用中输入的数据也与训练数据有较大差异，导致模型性能的下降。
3. 尽管针对领域偏移的问题有许多领域自适应的研究，这些方法往往都需要对

训练域数据进行获取。而由于数据隐私保护、数据存储和传输的限制以及训练计算开销的因素，我们无法对训练域数据进行获取，而只能得到在训练域数据上训练好的语义分割模型。如何利用该模型与无标签的目标域数据进行领域自适应也是一值得研究的难题。

1.3 论文创新点与贡献

基于以上难点，我们提出一种提出一种基于知识蒸馏框架下的无源多视图领域自适应白内障手术场景语义分割算法。我们通过在训练域上利用特征一致性和差异性约束，结合多视图学习的方法训练出鲁棒的白内障语义分割网络；利用均值教师知识蒸馏的方法，实现无源域的领域自适应。具体来说，我们的创新点体现在：

1. 利用风格差异性和语义一致性的约束，结合扰动方程，训练出一个鲁棒的特征提取器，提升模型对白内障手术场景语义信息的提取能力。使用多视图学习的思路，利用两个差异化的特征学习器训练出一个鲁棒的白内障语义分割网络，提升在训练域分割精度的同时提高模型的泛化能力。
2. 传统均值教师模型^[5]的问题在于由于初始化的时候教师和学生模型都使用训练域的模型进行初始化，教师模型和学生模型耦合度高，参数更新方向相似，因此可能会导致性能上出现瓶颈。而我们提出的模型相由于在初始化的时候两个特征学习器天然的就有较大差异性，在领域自适应的过程中朝着不同的方向进行更新，因此理论上能打破传统教师学生模型的性能瓶颈

2. 国内外研究相关研究现状

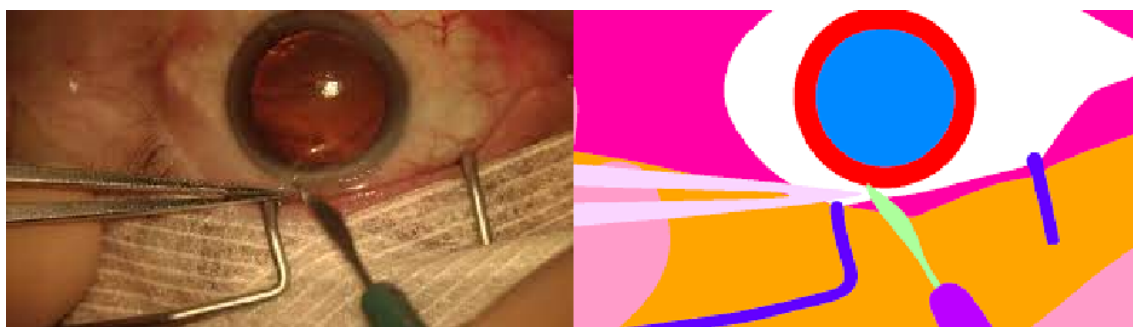


图 2 白内障语义分割图，数据来源于 Grammatikopoulou^[6]

2.1 图像语义分割算法

图像分割自计算机视觉领域诞生以来作为最基本的任务之一，在医学图像分析，自动驾驶，机器人感知等应用中发挥重要作用。传统的图像分割算法主要包括

阈值分割方法, 区域合并方法, k-means 聚类方法, 分水岭方法, Markov 随机场及活动轮廓等算法。2011 年, Lalys 等^[7]通过人工构筑形状, 颜色, 表面纹理相关特征, 并结合 HMM 算法实现对白内障手术场景的分割。2014 年, Quéllec 等^[8]利用条件随机场算法 (CRF) 根据帧级信息完成手术场景分割。2015 年, Zhao 等^[9]人用带有 L1 正则化的总变分模型实现了对瞳孔和虹膜的分割。2016 年, Abdullah^[10]等提出一种结合活跃轮廓和非循环正则化框架的方法, 在瞳孔虹膜分割任务中取得良好的结果。

近年来随着深度学习技术的快速发展, 一系列基于该技术的算法在图像语义分割领域取得瞩目的成果。在诸多方法中, 全卷积网络 (FCN)^[11]是最为经典的图像语义分割解决方案之一。FCN 是一个端到端, 采用编码器-解码器架构的卷积神经网络。相比于 VGG-16^[12]和 Alexnet^[13], FCN 用反卷积操作代替全连接来完成上采样的工作, 从而生成像素层级的分割图像。U-Net^[14]采用了完全对称的编码器-解码器的结构。同时 U-Net 增加了下采样特征图和上采样特征图之间的跳连, 使得低分辨率的信息和高分辨率的信息进行融合, 提供更精细的分割预测。PSPNet^[15]基于 ResNet^[16]的 backbone, 引入了池化金字塔模块。该模块并联四个不同规格的池化层, 生成不同尺度的特征图, 最后再将这些不同的特征图和原始特征图相加, 有效地解决了 FCN 模型上下文信息丢失的问题。

2.2 计算机视觉领域自适应算法

在计算机视觉领域中, 来自两个不同域的图片之间往往拥有不同的特征, 譬如纹理, 明暗等存在差异性。因此, 当以某个特定域的图片作为训练数据的模型倘若直接作用于另一个域时, 常有一定程度性能的下降。针对该挑战, 领域自适应方法尝试通过将两个不同域的数据映射到相同的表征空间中, 通过施加约束尽量拉近两个域在这个表征空间中特征的距离, 从而最大限度维持模型原有的性能。在医学图像分支中, 受限于数据采集的成像设备的不同, 采集场景光线等客观条件的不同以及受试者的差异性, 不同数据集之间存在域偏移问题, 因此领域自适应方法同样适用。领域自适应方法的实现主要通过深度学习技术, 其中 GAN^[17]的出现使得基于对抗的领域自适应方法成为潮流。GAN 主要包含一个生成网络和一个判别网络, 生成网络负责完成图像分析的主要任务, 而判别器对生成器的输出进行真伪判断, 两者相互拮抗, 一方随着另一方的优化而优化。DCGAN^[18]采用步长卷积替代上采样层, 在每一层加入 batch normalization 层, 从而提升了 GAN 训练的稳定性问题。CyCADA^[19]结合了 CycleGAN^[20]的一致性约束, 对两个域之间的输入图像进行像素级的风格迁移, 同时对迁移后的图像之间以及提取的特征之间进行对抗学习, 实现了具有领域自适应能力的语义分割模型。Luo^[21]等提出基于

协同训练和对抗学习思路的 CLAN 领域自适应模型。传统对抗学习仅在表征空间对两个域之间的特征进行全局的对齐，忽视了特征与类别的联合分布之间的对齐，导致部分特征语义的一致性遭到破坏。CLAN 提出的解决方案主要采用了协同训练的思路，使用具有差异性的解码器对相同的特征空间进行学习，生成本地特征评估图来匹配局部特征和语义类别，从而减轻全局特征对齐所带来的负面影响。此外，也存在一些通过简单图像预处理的方式降低源域和目标域输入图像差异的尝试，如 2020 年 Yang^[22]等提出结合傅里叶变换和反傅里叶变换，交换源域和目标域图像的低频信息，得到具有目标域风格的源域图片，从而降低域之间的差异问题。

3. 研究方法

3.1 问题阐述

在本小节我们对要解决的问题进行数学化阐述。源域上的白内障数据集 S 存在有 N 个匹配的白内障手术场景图像和语义分割标签 $\{(X_s, Y_s)\}_{s=1}^N$ 。通过监督学习，源域数据集 S 可以训练出一个白内障语义分割模型 ϕ_S 。然而，给定目标域 T ，数据集上只有 M 个无标签的白内障手术场景图像 $\{X_t\}_{t=1}^M$ 。由于在临床应用上，源域 S 和目标域 T 在数据分布上有较大的差异，导致源域模型 ϕ_S 直接在目标域 T 上进行白内障手术场景图像分割会出现性能下降的现象。在无源域的问题设置下，本论文的任务是研究如何只利用缺乏标签的目标域数据 $\{X_t\}$ ，将源域模型 ϕ_S 优化成目标域模型 ϕ_T ，使其在目标域上有较好的分割表现。

3.2 方法概述

为了解决白内障手术场景语义分割在无源域设置下的领域自适应任务，本论文提出一种基于知识蒸馏框架下的无源多视图领域自适应算法。该算法主要有三部分构成。首先算法对输入的源域图像 X_s 利用风格特征差异化约束和内容特征一致性约束训练出鲁棒的特征提取器 f (3.3 小节)。随后在第一步的基础上，算法基于多视图思想训练两个差异化的特征学习器 g_α, g_β ，在源域上得到具有较强泛化性的白内障手术场景语义分割模型 $\phi_S = f \circ (g_\alpha + g_\beta)$ (3.4 小节)。在领域自适应阶段，由于无法获取源域数据，算法通过均值教师知识蒸馏的领域自适应方法，将源域模型 ϕ_S 更新为在目标上有良好的分割性能的目标域模型 ϕ_T (3.5 小节)。

3.3 基于风格差异性与语义一致性约束的特征提取器

在计算机视觉领域中，图像包含的信息可以大致分为颜色风格信息，纹理材质信息以及结构语义信息。将图像输入全卷积网络后，不同层的卷积层对应不同信息的特征的提取。一般来说，颜色风格对应特征更多地是在浅层的卷积层进行提

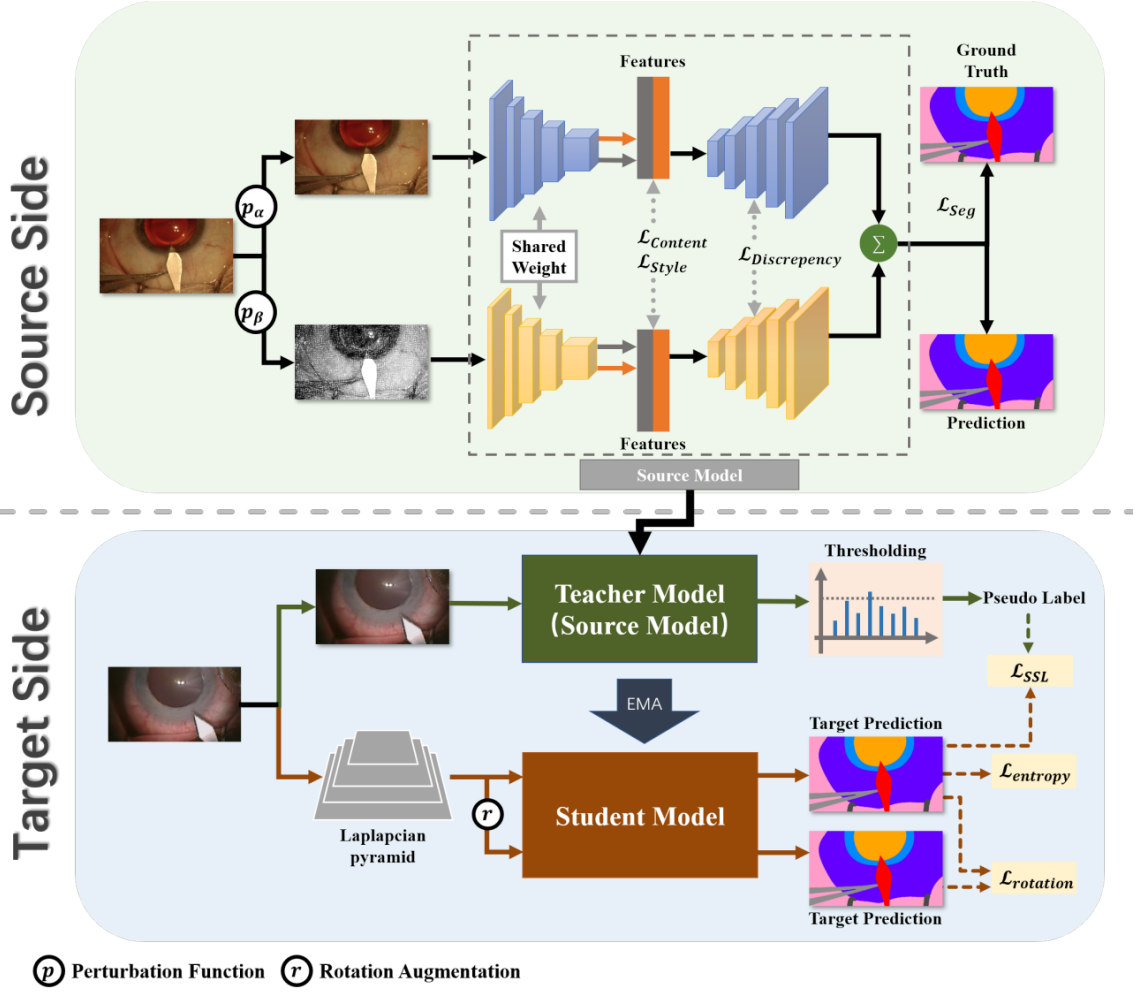


图 3 算法流程图

取，随后是纹理材质对应的特征，而最深层次的结构语义特征需要经过多层的卷积获得^[23]。设 $\phi_i(X)$ 为给定输入图像 S 经过 i 层卷积后得到的特征图，则对于 L 层的特征提取器来说，我们可以得到 $F = \{\phi_1(X), \phi_2(X), \dots, \phi_L(X)\}$ 的激活特征图集合。在这里，我们选择 $\phi_1(X)$ 和 $\phi_2(X)$ 作为对输入图像颜色风格信息的表述，通过线性双插值上采样方法分别将特征图采样至输入图大小，随后沿着特征维度进行拼接，得到风格特征 F_{sty} 。另一方面，我们选择最后一层的激活特征图 $\phi_L(X)$ 作为对语义信息的表示 F_{cont} 。

由于源域和目标域白内障手术场景图像数据分布的差异性，我们需要特征提取器 f 对不同风格的输入图像的语义特征提取具有鲁棒性。因此，给定一张源域的白内障手术场景图像 X_s ，我们首先选择两种不同的扰动函数 p_α 和 p_β ，对 X_s 进行扰动来模拟风格的变化，得到两张不同风格的图像 X_s^α 和 X_s^β 。我们分别地对他们进行特征提取，并希望他们各自得到的风格特征 $(F_{sty}^\alpha, F_{sty}^\beta)$ 是具有差异性的，

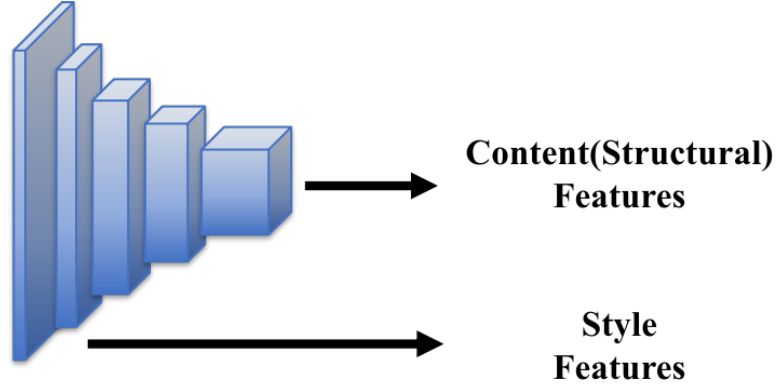


图 4 特征提取图

因此我们添加一个风格差异化的损失 \mathcal{L}_{sty} 对特征提取器 f 进行约束，定义如下：

$$\mathcal{L}_{sty} = \max\left(\frac{1}{n} \sum_i^n G(F_{sty}^\alpha, F_{sty}^\alpha), \frac{1}{m} \sum_j^m G(F_{sty}^\beta, F_{sty}^\beta)\right) \quad (1)$$

其中 $G(F_{sty}^\alpha, F_{sty}^\beta)$ 为 X_s^α 和 X_s^β 的 Gram 矩阵，计算方法为：

$$\begin{aligned} G(F_{sty}^\alpha, F_{sty}^\beta) &= 1 - D_{\cosine}(F_{sty}^\alpha, F_{sty}^\beta) \\ &= \frac{F_{sty}^\alpha, F_{sty}^\beta}{\|F_{sty}^\alpha\| \|F_{sty}^\beta\|} \end{aligned} \quad (2)$$

同时，我们希望 X_s^α 和 X_s^β 提取出来的语义特征 $F_{cont}^\alpha, F_{cont}^\beta$ 尽可能的相似，因此我们对特征提取器施加语义一致性损失 \mathcal{L}_{cont} ：

$$\mathcal{L}_{cont} = \left\| G(F_{cont}^\alpha, F_{cont}^\alpha) - G(F_{cont}^\beta, F_{cont}^\beta) \right\|_1 \quad (3)$$

最终我们得到的特征提取器的训练损失 $\mathcal{L}_{enc} = \mathcal{L}_{sty} + \mathcal{L}_{cont}$ 。

3.4 基于多视图学习的白内障语义分割网络

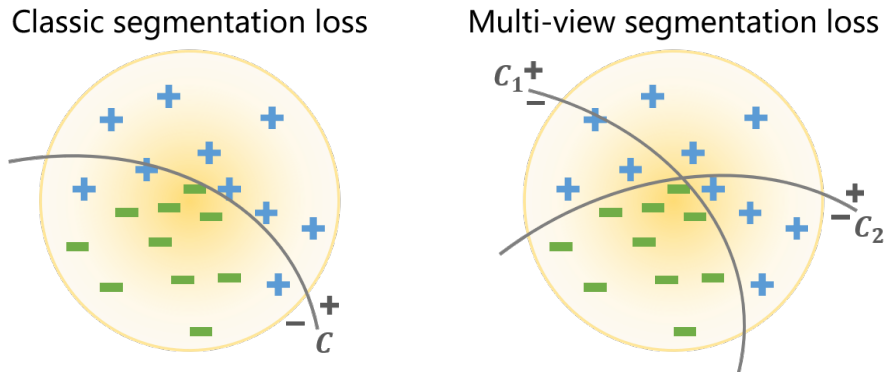


图 5 多视图学习，图片来源于 Ou^[24]

深度学习在现实世界的应用时，往往会面临来源于不同采集方式的数据，即多视图数据。举例来说，在对网页分类的任务中，网页的文字内容和超链接信息分别可以作为对同一个网页描述的两组不同视图的数据。多视图学习针对这类数据集，对每个视图都分别学习一个函数实现样本到标签的映射，通过函数之间的联合优化提升模型的泛化性能。与此同时，单视图的数据通过人为地创造出多视图图并与用多视图学习方法也能训练出泛化性强的模型^[25]。

基于以上研究，不同风格的源域图像 X_s^α 和 X_s^β 作为不同视图的数据进行特征提取，紧接着两个差异化的特征学习器 g_α, g_β 分别对特征进行学习，输出像素级别的语义分割图像 $\hat{Y}_s^\alpha, \hat{Y}_s^\beta$ 并根据真实标签 Y_s 进行监督训练。监督训练的分割损失为：

$$\mathcal{L}_{Seg} = - \sum_{k=1}^K Y_{s,k} \log(\hat{Y}_s^\alpha + \hat{Y}_s^\beta) \quad (4)$$

为了保证特征学习器的差异性，我们对 g_α 和 g_β 的参数 ($param$) 进行约束 $\mathcal{L}_{Discrepancy}$ ：

$$\begin{aligned} \mathcal{L}_{Discrepancy} &= 1 - D_{cosine}(param(g_\alpha), param(g_\beta)) \\ &= \frac{param(g_\alpha) \cdot param(g_\beta)}{|param(g_\alpha)| |param(g_\beta)|} \end{aligned} \quad (5)$$

通过该约束，我们希望学习器 g_α 和 g_β 对于相同的特征有不同的激活程度，即选择不同的特征进行学习。同时由于两个学习器都在监督训练的损失 \mathcal{L}_{seg} 约束下优化，各自都能保证较好的分割能力。最终，结合特征提取器的损失函数 \mathcal{L}_{enc} ，我们得到源域训练阶段的损失函数 \mathcal{L}_S ：

$$\mathcal{L}_S = \lambda_1 \mathcal{L}_{enc} + \lambda_2 \mathcal{L}_{Discrepancy} + \mathcal{L}_{Seg} \quad (6)$$

3.5 基于均值教师知识蒸馏的领域自适应方法

在领域自适应阶段，我们无法获取源域数据集 S 进行训练，只能利用无标签的目标域数据集 T 对源域模型 ϕ_S 进行优化。在这种问题设置下，均值教师模型^[5,26]提出利用知识蒸馏框架进行无源域的领域自适应。其思想是利用源域训练好的模型 ϕ_S 初始化教师模型 $\phi_{teacher}$ 和学生模型 $\phi_{student}$ ，将目标域的图像 X_t 分别输入两个模型，得到输出 $\hat{Y}_{t,teacher}, \hat{Y}_{t,student}$ ，并计算一致性损失 $\mathcal{L}_{Consistency}$ 。由于教师模型更新速度更慢，保留更多源域的分割能力。另一方面，学生模型更新速度更快，因此对目标域数据的特征学习更为积极，结合一致性损失从而将源域模型 ϕ_S 迁移到目标域模型 ϕ_T 。基于这种思想，本论文结合源域训练好的鲁棒白内障语义分割网络进行领域自适应。

如图3下半部分所示，首先设置教师模型和学生模型，两者都与源域训练模型

具有相同的结构。在源域上测试得到的泛化性能较高的源域模型参数初始化教师模型，使得教师模型保留源域的分割能力。同时，考虑到源域模型参数会引入域偏见（Domain Bias），影响学生模型在对目标域的学习和更新，使用随机参数对其进行初始化。

与传统知识蒸馏框架^[5]类似，教师模型和学生模型分别对输入的目标与图像进行分割，通过教师模型的分割结果引导学生模型的更新。由于源域和目标域的数据分布不一致，教师模型的分割结果可能包含较多的噪声，因此本论文引入自监督学习的思路通过设置阈值构建伪标签：

$$\hat{y}_{teacher}^n = \begin{cases} \hat{y}_{teacher}^{n,*} & \hat{y}_{teacher}^{n,*} > \tau \\ 0 & else \end{cases} \quad (7)$$

其中， $\hat{y}_{teacher}^n$ 对应教师模型分割结果 $\hat{Y}_{t,teacher}$ 的像素点，而 $\hat{y}_{teacher}^{n,*} = \text{argmax}(\phi_{teacher}(X_t))$ 代表着该像素点上模型的预测结果， τ 则为选择的阈值。通过阈值选择，教师模型的输出噪声能够得到一定程度的过滤，提升对学生模型引导的准确程度。

对于学生模型，为了使得学生模型对图像边缘进行更好的学习，我们首先对根据目标域图像构建拉普拉斯金字塔（Laplacian Pyramid）对目标域图像进行增强。随后输入到学生模型中 $\phi_{student}$ 得到分割图像 $\hat{Y}_{t,student}$ 。将学生模型的输出和教师模型的输出进行自监督训练，得到损失 L_{SSL} ：

$$\mathcal{L}_{SSL} = - \sum_{k=1}^K \hat{Y}_{t,teacher} \log(\hat{Y}_{t,student}) \quad (8)$$

在学生模型每经历 k 次迭代更新后，本文使用指数移动平均（Exponential Moving Average, EMA）的方法将学生模型的知识迁移到教师模型当中。对于第 k 次迭代时教师模型的参数 $\theta(\phi_{teacher})$ 和学生模型的参数 $\theta(\phi_{student})$ ，本文的更新策略如下：

$$\theta_t(\phi_{teacher}) = \psi \theta_{t-k}(\phi_{teacher}) + (1 - \psi) \theta_t(\phi_{student}) \quad (9)$$

其中 ψ 为常量，表示教师模型参数在更新后保留源域知识的程度。

模型在更新的过程中，单凭自监督损失模型无法对包含噪声的区域进行处理，因此本文在提出的框架之上引入了最大平方损失（Maximum Square Loss）^[27]，通过降低噪声区域的不确定度提升模型的输出效果：

$$\mathcal{L}_{Entropy} = - \frac{1}{2HxW} \sum_{i=1}^H \sum_{j=1}^W (p_{student}^{i,j})^2 \quad (10)$$

公式中， $p_{student}^{i,j}$ 代表着学生模型在某一个像素点上输出的概率分布。

为了进一步进一步提高学生模型对目标域的数据的学习，方法额外加入了一致性的约束。具体的做法是对拉普拉斯金字塔增强后的图像进行旋转的增广，输入到学生模型得到输出 $\hat{Y}'_{t,student}$ ，并于无增广的分割结果计算一致性损失：

$$\mathcal{L}_{Rotation} = \left\| \hat{Y}'_{t,student} - \hat{Y}_{t,student} \right\|_1 \quad (11)$$

最终我们在目标与进行领域自适应的总损失如下：

$$\mathcal{L}_T = \mathcal{L}_{SSL} + \mathcal{L}_{Entropy} + \mathcal{L}_{Rotation}$$

4. 实验

在本章中，我们将对上一章所提出的无源域领域自适应算法进行实验与展示。我们将会首先介绍实验当中使用到的数据集。随后我们将会对实验细节，包括软硬件环境与超参数，使用的对比方法，实验任务以及评估指标进行详细的阐述。最后我们将对实验结果进行展示。

4.1 实验数据集

首先我们将对实验用到的数据集进行介绍。我们将会用到两个白内障手术场景语义分割的数据集，分别是 CaDIS^[6]数据集和 CatInstSeg 数据集^[28]

4.1.1 CaDIS 数据集

CaDIS^[6]数据集全称 Cataract dataset for surgical RGB-image segmentation，是一个白内障手术场景语义分割数据集（图6）。该数据集是根据 Al Hajj 等人^[29]于 2019 年发布的 CATARACTS 白内障手术视频数据集得到的。

CATARACTS^[29]手术视频记录于 2015 年 1 月至 9 月的布雷斯特大学医院，共计 50 段白内障超声乳化手术视频。患者年龄最小 23 岁，最大 83 岁，平均年年龄 61 岁。患者的患病情况包括老年白内障，外伤性白内障和屈光不正。手术总共由三位外科医生进行清晰度为 1920 x 1080 像素。

CaDIS 数据集从 CATARACTS 数据集中其中 25 个训练视频中进行抽帧操作，并将图像清晰度从 1920 x 1080 像素调整为 960 x 540 的大小，共计 4670 张手术图片，每一张样本图片都有一张对应的像素级别的分割标签，两者皆为 png 格式。数据集由 4 名标注者对白内障手术场景进行像素级别的语义标注。其中，标注类别共 36 类，包含 29 类手术器械类别和 4 类解剖学构造以及 3 个杂项类别。

在处理白内障手术数据集时，我们首先需要对样本图片进行大小归一化，为了确保所有图片在尺寸上保持一致，我们需要将图片的大小进行归一化处理。这

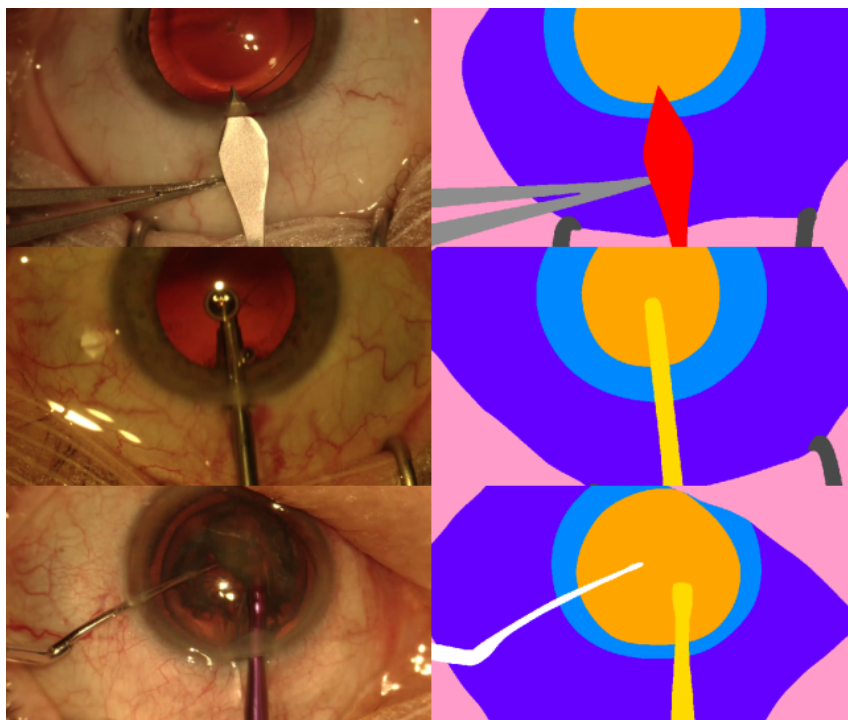


图 6 CaDIS 数据集^[6]

可以通过将所有图片调整为相同的分辨率 720 x 405 来实现。这将使得计算过程更加高效，并减少由于图片尺寸差异带来的计算误差。接下来，原始数据集中可能包含许多细分的手术标签。为了简化模型的训练过程并提高准确性，我们需要将这些细分标签合并成 13 个大类。这可以通过为每个原始标签分配一个新的类别编号来实现。合并后的标签类别应该具有较高的区分度，以便于模型更好地识别不同类型的手术。最后，删除实验不使用的标签类别：在某些情况下，我们可能不需要使用所有的标签类别来进行实验。为了提高实验效率，我们可以删除实验中不需要使用的标签类别。这可以通过筛选出与实验目标相关的标签类别，然后从数据集中移除其他不相关的类别来实现。这将有助于减少计算资源的浪费，并加快模型训练速度。通过以上步骤，我们可以为白内障手术数据集进行有效的预处理，为后续的实验和分析打下坚实的基础。

4.1.2 CatInstSeg 数据集

此数据集的内容是从公开的白内障手术视频集中精选而来的特定白内障手术帧，这些视频经过 COCO Annotator 手动进行了详细的标注。在挑选过程中，相关机构专注于筛选具有最佳视觉质量的帧，如避免出现运动模糊等影响观察的因素，并确保涵盖各种不同的白内障手术场景，以便为每个类别提供丰富的样本。

最终发布的这个白内障手术数据集包含了广泛应用于 COCO-Format 的 9 种关键手术器械的详细标注信息。数据集中共收录了 281 张具有不同白内障手术特点

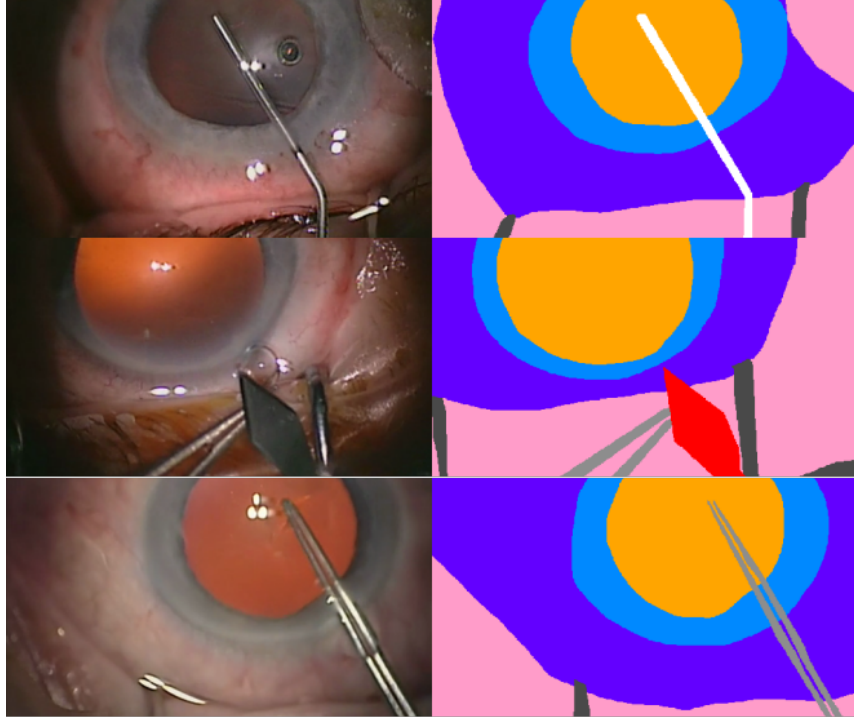


图 7 CatInstSeg 数据集^[28]

的已标注图像。这些图像不仅有助于研究人员了解手术过程中所需的各种器械，还有助于开发和优化计算机视觉算法在白内障手术领域的应用。

在处理白内障手术数据集的过程中，我们首先要对样本图片的尺寸进行统一化，从而确保所有图片具有一致的大小，便于后续计算和分析。然后，我们将对图片手术标签进行整合，将其归为 13 个类别，这样做有助于简化模型训练流程并提高模型的预测精度。最后，为了优化实验效果和精确度，我们将移除实验过程中不必要的标签类别。

4.2 实现细节

4.2.1 软硬件环境与超参数

此次试验的硬件与软件设施如下表所示。表2阐述了实验所需的硬件条件。表1描述了实验涉及的软件配置。

表 1 软件环境配置

OS	Ubuntu 22.04.2 LTS
Compiler	GNU Compiler v7.5.0, CUDA Compiler v10.0
MPI	Open MPI 4.0.4
Python version	Python 3.10.9

表 2 硬件环境配置

CPU	12th Gen Intel® Core™ i7-12700 × 20
GPU	NVIDIA GeForce RTX 3070
Memory	32.0 GiB, DDR4 2133Mhz
Hard Disk	3.0 TB SSD

4.2.2 实验目的

我们的实验目标是验证所提出的基于知识蒸馏框架下的无源多视图领域自适应白内障手术场景语义分割算法在应对白内障手术场景分割任务时所展现的有效性和鲁棒性。为实现此目标，我们将通过以下几个方面的实验验证和评估：

首先，我们将评估所提出算法在训练域上的分割精度，通过对比其他领域自适应方法和传统语义分割方法，以展示我们的方法在提取白内障手术场景的语义信息方面的优势。

其次，我们将探讨所提出算法的泛化能力。通过将算法应用于具有不同数据分布的测试域，我们将评估该方法在处理不同来源和质量数据时的稳定性和鲁棒性。

进一步，我们将分析所提出的多视图学习策略以及知识蒸馏框架在提升模型性能方面的贡献。通过对比仅使用单一视图和不采用知识蒸馏的方法，我们将揭示这些创新点在改善分割性能和泛化能力方面的作用。

最后，我们将关注所提出的无源领域自适应策略在处理未标注目标域数据时的有效性。通过在不同的目标域进行实验，我们将评估该方法在缺乏标注数据的情况下是否仍能实现领域自适应。

综合以上各方面的实验评估，我们希望充分验证所提出的基于知识蒸馏框架下的无源多视图领域自适应白内障手术场景语义分割算法在解决白内障手术场景分割任务中的有效性和鲁棒性。

4.2.3 对比方法

在本小节中，我们对将要进行对比的方法逐一进行介绍。为了验证我们提出的基于多视图学习的白内障手术场景语义分割模型在目标域上的鲁棒性，我们选择了 U-net^[14]、Deeplabv3^[30]、FCN8s、FCN16s、FCN32s^[11]。我们希望多视图语义分割模型通过与上述模型的对比实现更优的性能表现，以此来验证模型的泛化性能。

同时，为了验证我们提出的基于均值教师知识蒸馏的领域自适应方法在无源域领域自适应白内障语义分割任务上的有效性，我们将提出的框架与 Maximum

Square Loss^[27], Tent^[31], STROTSS^[32], SFDA-CellSeg^[33]进行比较。

4.2.4 评价指标

语义分割任务是计算机视觉领域中的一个重要任务，其目的是将图像中的每个像素分配给一个特定的类别。为了评估语义分割模型的性能，我们通常使用一些指标来衡量模型的准确性。其中一个常用的指标就是并交比（Intersection over Union, IoU）^[34]。

并交比（IoU）是一种评估模型预测准确性的指标，它衡量了两个区域之间的重叠程度。在语义分割任务中，IoU 用于衡量预测的分割区域和真实分割区域之间的相似性。具体来说， $\text{IoU} = (\text{预测区域与真实区域的交集}) / (\text{预测区域与真实区域的并集})$ ，计算方法如下：

$$\text{IoU} = \frac{\sum_{i=1}^m \sum_{j=1}^n (P(i, j) \cap G(i, j))}{\sum_{i=1}^m \sum_{j=1}^n (P(i, j) \cup G(i, j))} \quad (12)$$

值得注意的是，IoU 的取值范围为 0 到 1。当 IoU 为 0 时，表示预测区域与真实区域完全不重叠；当 IoU 为 1 时，表示预测区域与真实区域完全重叠。因此，较高的 IoU 值意味着模型预测的分割效果较好。

为了评估整个语义分割模型的性能，我们通常会计算所有类别的平均 IoU（mean IoU, mIoU）。mIoU 是所有类别的 IoU 分数的平均值，它可以综合反映模型在不同类别上的预测性能。在实际应用中，我们会使用 mIoU 来比较不同语义分割模型的性能，并据此选择合适的模型

4.3 实验结果

表 3 以 CaDIS 数据集训练的模型在目标域 CatInstSeg 数据集上的测试结果 (%) 其中红色为该类上性能最佳指标，蓝色为次佳指标

Model	Iris	Pupil	Cornea	Skin	Knife	Cannula	Forceps	Handpiece	Eye Re.	mIoU
U-Net	28.57	49.96	42.63	51.76	15.93	11.91	17.63	5.61	15.74	26.64
Deeplabv3	40.48	68.77	54.70	58.31	23.82	19.66	14.11	13.54	27.12	35.61
FCN8s	23.08	36.29	19.85	45.21	6.38	10.71	8.87	6.50	11.90	18.76
FCN16s	22.12	37.15	21.70	47.37	7.53	13.20	8.19	8.80	17.43	20.39
FCN32s	24.02	40.48	30.77	50.98	2.15	16.93	6.50	13.25	16.43	22.39
Ours	28.23	75.08	60.86	58.21	13.81	15.38	14.43	14.58	28.33	34.32

根据提供的表格3数据，以 CaDIS 数据集训练的模型在目标域 CatInstSeg 数据集上的测试结果在 Iris 分割任务上，Deeplabv3 表现最好，准确率为 40.48%，远高于其他模型。在 Pupil 分割任务上，我们提出的模型表现最佳，准确率达到了 75.08%，

明显优于其他模型。在 Cornea 分割任务上，我们提出的模型同样优于其他模型，准确率为 60.86%，领先第二名 Deeplabv3 (54.70%)。在 Skin 分割任务上，Deeplabv3 和我们提出的模型表现相近，分别为 58.3% 和 58.21%。这两个模型在 Skin 任务上的表现均优于其他模型。在 Knife 分割任务上，Deeplabv3 的准确率为 23.82%，领先其他模型。在 Handpiece 分割任务上，我们提出的模型表现最好，准确率为 14.58%，略高于其他模型。在 Retractor 分割任务上，我们提出的模型再次取得最佳表现，准确率为 28.33%，领先第二名 Deeplabv3 (27.12%)。在 IoU (Intersection over Union) 评价指标上，Deeplabv3 的表现最好，达到了 35.61%，紧随其后的是我们提出的模型，准确率为 35.61%。

我们提出的模型在 Pupil、Cornea、Handpiece 和 Retractor 四个任务上表现最佳，这表明我们的模型在解决这些关键组件的分割问题上具有较强的能力。Deeplabv3 在 Iris、Skin 和 Knife 三个任务上表现最好。尽管在其他任务上略逊于我们提出的模型，但 Deeplabv3 在这些部分的表现仍然相当出色。U-Net 在 Iris 和 Knife 两个任务上表现较好，但在其他任务上的表现相对较差，尤其是在 Handpiece 和 Retractor 任务上。FCN8s、FCN16s 和 FCN32s 三个模型在所有任务上的表现均较弱。这可能表明，这些传统的语义分割方法在解决白内障手术场景分割任务时存在局限性。在 IoU (Intersection over Union) 评价指标上，Deeplabv3 的表现最好 (35.61%)，我们的模型紧随其后 (35.61%)。这意味着，虽然我们的模型在某些任务上的表现优于 Deeplabv3，但在整体性能上，Deeplabv3 略胜一筹。总体来说，我们提出的模型在多个分割任务上表现优异，尤其在 Pupil、Cornea、Handpiece 和 Retractor 任务上表现最佳。尽管在 IoU 指标上略逊于 Deeplabv3，但我们的模型仍展现出了良好的有效性和鲁棒性。这些结果为我们在进一步优化和改进模型时提供了有价值的参考。同时这些结果表明，我们提出的基于知识蒸馏框架下的无源多视图领域自适应白内障手术场景语义分割算法在解决白内障手术场景分割任务上具有潜力。

表 4 基于 CaDIS 数据集训练的源域模型在目标域 CatInstSeg 数据集上进行领域自适应的测试结果 (%) 其中红色为该类上性能最佳指标，蓝色为次佳指标

Model	D_S	D_T	Iris	Pupil	Cornea	Skin	Knife	Cannula	Forceps	Handpiece	Eye Re.	mIoU
Source only	✓		28.23	75.08	60.86	58.21	13.81	15.38	14.43	14.58	28.33	34.32
Max. Square Loss	✓	✓	63.27	85.72	62.73	69.96	36.12	15.73	13.71	22.02	14.23	42.61
STROTSS(DeepLabv3)	✓	✓	47.43	78.16	47.88	61.52	13.96	15.66	13.92	3.75	28.57	34.54
SFDA-CellSeg		✓	-	-	-	-	-	-	-	-	-	-
Tent		✓	54.47	88.05	67.83	66.56	0.00	0.00	1.34	1.35	0.24	31.09
Ours($\tau=0.75$)		✓	56.24	86.38	65.46	69.67	28.54	14.43	20.00	14.39	18.04	41.46
Ours($\tau=0.99$)		✓	54.36	77.59	56.53	64.57	43.61	12.47	14.27	21.07	0.00	38.28

在横向分析基于 CaDIS 数据集训练的源域模型在目标域 CatInstSeg 数据集上进行

领域自适应的测试结果时，我们可以从以下几个方面进行观察：MaxSquareLoss 模型在 Iris、Skin、Cannula、Handpiece 任务上表现较好，总体上的 mIoU 指标也为最佳，然而该模型在训练过程中使用到了源域数据，在实际临床应用中可能面临隐私保护等诸多短板。Tent 模型在 Pupil 和 Cornea 任务上表现优秀，分别取得 88.05% 和 67.83% 的准确率，但在 Knife 和 Forceps 分割上表现不佳，得分为 0.00%，总体 mIoU 为 31.09%。这是由于 Tent 模型使用的是熵最小化领域自适应损失函数，该种损失函数很容易由于类别不平衡问题导致手术器械等图像中占比较小，出现频率较少的模型在目标域上分割性能大幅下降。STROTSS 模型在各个任务上的表现均不太理想，总体 IoU 为 34.54%，在五种模型中排名次低。

我们的模型 ($\tau = 0.75$) 在 Iris, Pupil, Cornea 以及 Skin 上都有较高的性能表现任务上具有最高的准确率，相比于源域模型提升了 7.14、%，在总体上的性能也达到了 41.46%，总体排名第二。尽管 MaxSquareLoss 方法在总体上优于我们的模型，但是我们的模型在数值上与 MaxSquareLoss 差距较少，且我们的模型没有使用源域数据，在临床应用上更具优势。我们的模型 ($\tau = 0.99$) 在 Knife 任务上表现出最高的准确率，达到 43.61%，但在其他任务上表现一般，总体 IoU 为 38.28%。总体来看，我们的模型 ($\tau = 0.75$) 在整体性能上具有较高的 mIoU，表明在进行领域自适应时，我们的模型具有较好的有效性和鲁棒性。然而，在处理器械分割的类别不平衡问题上我们仍有较大的进步空间。

通过观察和分析两个表格中的实验结果，我们可以得出以下结论：

在源域 CaDIS 数据集上训练的模型在目标域 CatInstSeg 数据集上的测试结果表明，Deeplabv3 模型和我们的模型在绝大多数任务上表现优越，尤其是在 Pupil、Cornea 和 Skin 任务上。然而，在 Knife 和 Handpiece 任务上，我们的模型具有更高的准确率。

在进行领域自适应的测试时，我们的模型 ($\tau = 0.75$) 在总体性能上表现最佳，具有较高高的 IoU，表明我们的无源多视图领域自适应白内障手术场景语义分割算法在处理白内障手术场景分割任务时具有较好的有效性和鲁棒性。

在某些特定任务上，例如 Pupil 和 Cornea，其他模型（如 Tent）在这些任务上的表现优于我们的模型，这意味着我们的模型在这些任务上仍有优化和改进的空间。

综上所述，我们的无源多视图领域自适应白内障手术场景语义分割算法在大部分任务上表现出优越的性能，并在总体 IoU 上优于其他模型。然而，在某些特定任务上，我们的模型仍有改进的空间，需要进一步优化和调整以提升整体性能。

5. 结论

白内障手术场景的语义分割对于开发白内障手术术中辅助系统具有重要意义。然而，由于 1) 像素级别的人工数据标注是及其昂贵的。2) 基于单一训练数据集的白内障手术场景语义分割模型在临床应用中由于输入数据分布的差异问题导致分割性能下降。3) 尽管领域自适应方法能够较好地解决前两点的问题，由于数据隐私，传输存储等限制导致在现实应用场景下只有训练域训练出来的分割模型，而无法用到训练数据进行领域自适应。基于以上三点挑战，一种基于知识蒸馏框架下的无源域领域自适应算法，该方法分为两个阶段。在源域训练阶段，结合图像扰动与特征一致性，方法约束源域模型对输入数据有较好的特征提取能力。同时结合多试图学习方法，两个差异性的解码器对特征进行不同视图的学习，通过互补提升最终模型的鲁棒性。在目标域领域自适应阶段，由于缺少源域训练数据的参与，本文提出利用均值教师的知识蒸馏框架对目标域数据进行自监督学习，同时加入熵最小化和旋转一致性约束提高模型对目标域数据的学习能力，实现无源域领域自适应的语义分割任务。本文为了验证提出方法的可行性，在 CaDIS 数据集和 CatInstSeg 数据集上进行白内障手术场景语义分割的领域自适应实验，并与其他领域自适应方法进行比较，实验结果表明本文提出的算法的能够在无源域训练数据的条件下提升模型的手术场景语义分割性能，具有一定程度的有效性。

参考文献

- [1] Goossens M, Mittelbach F, Samarin A. The LATEX companion[M]. Addison-Wesley Reading, 1994.
- [2] Lam D, Rao S K, Ratra V, et al. Cataract[J]. Nature reviews Disease primers, 2015, 1(1): 1-15.
- [3] Steinmetz J D, Bourne R R, Briant P S, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study[J]. The Lancet Global Health, 2021, 9(2): e144-e160.
- [4] Gerber M J, Pettenkofer M, Hubschman J P. Advanced robotic surgical systems in ophthalmology[J]. Eye, 2020, 34(9): 1554-1562.
- [5] Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[J]. Advances in neural information processing systems, 2017, 30.
- [6] Grammatikopoulou M, Flouty E, Kadkhodamohammadi A, et al. CaDIS: Cataract dataset for surgical RGB-image segmentation[J]. Medical Image Analysis, 2021, 71: 102053.
- [7] Lalys F, Riffaud L, Bouget D, et al. A framework for the recognition of high-level surgical tasks from video images for cataract surgeries[J]. IEEE Transactions on Biomedical Engineering, 2011, 59(4): 966-976.
- [8] Quellec G, Lamard M, Cochener B, et al. Real-time segmentation and recognition of surgical tasks in cataract surgery videos[J]. IEEE transactions on medical imaging, 2014, 33(12): 2352-2360.
- [9] Zhao Z, Ajay K. An accurate iris segmentation framework under relaxed imaging constraints using total variation model[C]. in: Proceedings of the IEEE international conference on computer vision. 2015: 3828-3836.
- [10] Abdullah M A, Dlay S S, Woo W L, et al. Robust iris segmentation method based on a new active contour force with a noncircular normalization[J]. IEEE transactions on systems, man, and cybernetics: Systems, 2016, 47(12): 3128-3141.
- [11] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. in: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. ArXiv preprint arXiv:1409.1556, 2014.

- [13] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[C/OL]. in: Pereira F, Burges C, Bottou L, et al. Advances in Neural Information Processing Systems: vol. 25. Curran Associates, Inc., 2012. https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.
- [14] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]. in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. 2015: 234-241.
- [15] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]. in: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.
- [16] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. in: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [17] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [18] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. ArXiv preprint arXiv:1511.06434, 2015.
- [19] Hoffman J, Tzeng E, Park T, et al. Cycada: Cycle-consistent adversarial domain adaptation[C]. in: International conference on machine learning. 2018: 1989-1998.
- [20] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. in: Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.
- [21] Luo Y, Zheng L, Guan T, et al. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation[C]. in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2507-2516.
- [22] Yang Y, Soatto S. Fda: Fourier domain adaptation for semantic segmentation[C]. in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 4085-4095.
- [23] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]. in: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13. 2014: 818-833.
- [24] Ou M, Li H, Liu H, et al. MVD-Net: Semantic Segmentation of Cataract Surgery Using Multi-View Learning[C]. in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2022: 5035-5038.

- [25] Zhao J, Xie X, Xu X, et al. Multi-view learning overview: Recent progress and new challenges[J]. *Information Fusion*, 2017, 38: 43-54.
- [26] Araslanov N, Roth S. Self-supervised augmentation consistency for adapting semantic segmentation[C]. in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021: 15384-15394.
- [27] Chen M, Xue H, Cai D. Domain adaptation for semantic segmentation with maximum squares loss[C]. in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019: 2090-2099.
- [28] Fox M, Taschwer M, Schoeffmann K. Pixel-Based Tool Segmentation in Cataract Surgery Videos with Mask R-CNN[C/OL]. in: de Herrera A G S, González A R, Santosh K C, et al. 33rd IEEE International Symposium on Computer-Based Medical Systems, CBMS 2020, Rochester, MN, USA, July 28-30, 2020. IEEE, 2020: 565-568. <https://doi.org/10.1109/CBMS49503.2020.00112>. DOI: 10.1109/CBMS49503.2020.00112.
- [29] Al Hajj H, Lamard M, Conze P H, et al. CATARACTS: Challenge on automatic tool annotation for cataRACT surgery[J]. *Medical image analysis*, 2019, 52: 24-41.
- [30] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. *ArXiv preprint arXiv:1706.05587*, 2017.
- [31] Wang D, Shelhamer E, Liu S, et al. Tent: Fully test-time adaptation by entropy minimization[J]. *ArXiv preprint arXiv:2006.10726*, 2020.
- [32] Kolkin N, Salavon J, Shakhnarovich G. Style transfer by relaxed optimal transport and self-similarity[C]. in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019: 10051-10060.
- [33] Li Z, Li C, Luo X, et al. Towards Source-Free Cross Tissues Histopathological Cell Segmentation via Target-Specific Finetuning[J]. *IEEE Transactions on Medical Imaging*, 2023.
- [34] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. in: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014: 580-587.

附录

致谢

大学入学以前，我曾对自己说要做一个优秀的人。如今大学即将毕业，我距离入学的目标仍有一段距离，但是无论如何在这四年中，我得到了一定程度的成长。

在这段四年的旅途中，要感谢的人太多，在背后默默支持我的家人，我遇到的每一位老师，每一位朋友，每一位只有一面之缘的人。可以说我是幸运的，这么多相遇的人都曾在某一个时刻给予我启迪，给我温暖，向我伸出援助之手。

在这里，我诚挚地感谢家人与每一位在大学四年里交谈的，结识的，相遇的老师、同学、朋友甚至是陌生人，你们或刻意或不经意的帮助都对我心智上的成长起了重要的作用。衷心祝愿各位在往后的生活中，平安喜乐，万事顺遂。