

中图分类号: TP391

单位代码: 10425

学 号: S22070005



中國石油大學

# 硕士学位论文

**Master Degree Thesis of China University of Petroleum**

换装行人重识别方法研究

**Research on Cloth-Changing Person Re-Identification**

学科专业: 计算机科学与技术

研究方向: 图形图像与可视化

作者姓名: 李明越

指导教师: 刘玉杰 副教授

二〇二五年五月

# **Research on Cloth-Changing Person Re-Identification**

A Thesis Submitted for the Degree of Master

**Candidate: Mingyue Li**

**Supervisor: A.P. Yujie Liu**

Qingdao Institute of Software, College of Computer Science and  
Technology

China University of Petroleum (East China)

学位论文答辩委员会

答辩时间	2025 年 5 月 17 日		
答辩地点	综合实验楼 A402		
答辩委员会组成			
组成	姓名	职称	工作单位
主席	彭冲	教授	中国海洋大学
委员	齐连永	教授	中国石油大学（华东）
	徐九韵	教授	中国石油大学（华东）
	孙歧峰	副教授	中国石油大学（华东）
	王雷全	高级实验师	中国石油大学（华东）

## 关于学位论文的独创性声明

本人郑重声明：所呈交的论文是本人在指导教师指导下独立进行研究工作所取得的成果，论文中有关资料和数据是实事求是的。尽我所知，除文中已经加以标注和致谢外，本论文不包含其他人已经发表或撰写的研究成果，也不包含本人或他人为获得中国石油大学（华东）或其它教育机构的学位或学历证书而使用过的材料。与我一同工作的同志对研究所做的任何贡献均已在论文中作出了明确的说明。

若有不实之处，本人愿意承担相关法律责任。

学位论文作者签名：\_\_\_\_\_ 日期： 年 月 日

## 学位论文使用授权书

本学位论文作者完全了解中国石油大学（华东）有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交学位论文的复印件和电子版，允许本学位论文被查阅和借阅。本人授权中国石油大学（华东）可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用复印、缩印、扫描或其它手段保存和汇编本学位论文。（保密的学位论文在解密后适用本授权书）

学位论文作者签名：\_\_\_\_\_ 日期： 年 月 日

指导教师签名：\_\_\_\_\_ 日期： 年 月 日

## 摘 要

随着城市化进程的加速，城市中广泛部署的摄像头系统在保障公共安全方面发挥着越来越重要的作用。然而，传统的人工监控方法在处理海量视频数据时存在诸多局限性，例如监控人员容易疲劳、监控效率低下等。这些问题促使智能监控技术应运而生，以提高监控系统的效率和准确性。其中行人重识别技术的不断发展，使得跨摄像头追踪和犯罪预防变得更加高效，显著增强了城市的安全管理能力。

尽管如此，在现实场景中，行人服装的频繁变化仍然是一个尚未得到有效解决的问题，这在很大程度上限制了行人重识别技术的广泛应用。本文总结了目前面临的两类挑战：

(1) 服装颜色纹理以及款型变化是根源问题。服装作为行人外观的重要组成部分，其颜色和纹理的多样性及复杂性直接影响了图像中行人特征的提取。

(2) 辅助模型和服装标签的不准确性是目前技术受限的症结所在。辅助模型无法保证在不同换装数据数据集中的有效性，而错误的服装标签无法充分发挥模型的性能。

为了解决这些挑战，本文设计了两种基于深度学习的换装行人重识别方法，旨在提高模型在面对服装变化时的鲁棒性和准确性。具体研究内容如下：

(1) 基于图像分割的双流融合网络。本文设计一种双流融合网络，旨在将局部特征与高级语义特征进行有效融合。在局部特征流中，以卷积神经网络作为特征提取器，并引入服装标签来设计空间解纠缠模块。该模块能够对服装特征和身份特征进行解耦，从而提取出更加丰富且与身份高度相关的行人局部特征表示。在高级语义特征流中，采用视觉自注意力模型，深入挖掘图像的上下文信息，以提供更为鲁棒的行人语义特征。此外，对高级语义特征流的输入进行数据增强，这不仅充分利用了先验知识，还有效减少了先验模型噪声对模型的干扰。实验结果表明，所提出的双流融合网络在 PRCC 和 NKUP+数据集上取得了显著的效果，显著提高了识别身份的准确率。

(2) 基于主动学习的服装采样方法。从数据驱动的角度出发，本文设计一种基于主动学习的服装采样器。在采样阶段，通过正确的服装标签保证相同身份的服装多样性和相似服装的身份多样性，提供更有价值的数据，便于模型学习更多与服装无关的特征。为进一步应对服装颜色纹理变化以及人体轮廓变化，设计一种多服装实例一致性损失，引导模型挖掘更加鲁棒的行人表示。在 PRCC 和 NKUP+数据集上进行了大量

的实验，实验结果充分证明了本章方法的有效性，相较于目前最先进的方法，该方法在换装场景下具有更强的鲁棒性和更高的识别准确率。

针对服装变化引发的挑战，本文提出的两种解决方法在模型设计、训练策略和度量学习等方面发挥了重要作用，设计一套有效的双流融合网络，提供了新颖高效的服装采样策略，并针对换装场景设计更贴切的度量损失函数。通过将本文所提出的方法与当前主流方法进行对比，以及对本文方法进行消融实验，从多个角度充分验证了本文所提出方法的有效性。这些方法不仅在理论上具有创新性，而且在实际应用中也具有较高的实用价值，为换装行人重识别领域的发展提供了新的思路和方法。

**关键词：**换装行人重识别；卷积神经网络；视觉自注意力模型；主动学习；采样策略

## Abstract

With the acceleration of urbanization, widely deployed camera systems in cities are playing an increasingly important role in ensuring public safety. However, traditional manual monitoring methods have many limitations when dealing with massive video data, such as easy fatigue of monitoring personnel and low monitoring efficiency. These issues have prompted the emergence of intelligent monitoring technology to improve the efficiency and accuracy of monitoring systems. The continuous development of Person Re-Identification has made cross camera tracking and crime prevention more efficient, significantly enhancing the city's safety management capabilities.

However, in real-world scenarios, the frequent changes in person clothing remain an unresolved issue, which greatly limits the widespread application of Person Re-Identification. This thesis summarizes two types of challenges currently faced:

(1) The fundamental issue is the variation in clothing color, texture, and style. As an important component of pedestrian appearance, clothing's diversity and complexity in color and texture directly affect the extraction of pedestrian features in images.

(2) The inaccuracy of auxiliary models and clothing labels is currently the crux of technological limitations. The auxiliary model cannot guarantee its effectiveness in different dressing datasets, and incorrect clothing labels cannot fully utilize the performance of the model.

To address this challenge, this thesis proposes two deep learning based methods for Cloth-Changing Person Re-Identification, aiming to improve the robustness and accuracy of the model in the face of clothing changes. The specific research content is as follows:

(1) Dual-stream fusion network based on image segmentation. This thesis designs a dual stream fusion network aimed at effectively fusing local features with advanced semantic features. In the local feature flow, a convolutional neural network is used as the feature extractor, and clothing labels are introduced to design a spatial disentanglement module. This module can decouple clothing features and identity features, thereby extracting richer and highly identity related pedestrian local feature representations. In the advanced semantic feature flow, a vision transformer is adopted to deeply explore the contextual information of images, in order to provide more robust person semantic features. In addition, data augmentation is applied to the input of advanced semantic feature streams, which not only fully utilizes prior knowledge, but also effectively reduces the interference of prior model noise on the model. The experi-

mental results show that the proposed dual stream fusion network has achieved significant results on PRCC and NKUP+ datasets, significantly improving the accuracy of identity recognition

(2) Active learning-based clothing sampling method. From a data-driven perspective, this thesis designs a clothing sampler based on active learning. In the sampling stage, the diversity of clothing with the same identity and the diversity of clothing with similar identities are ensured through correct clothing labels, providing more valuable data for the model to learn more features unrelated to clothing. To further address changes in clothing color and texture, as well as variations in human body contours, a multi clothing instance consistency loss is designed to guide the model in mining more robust pedestrian representations. A large number of experiments have been conducted on the PRCC and NKUP+ datasets, and the experimental results fully demonstrate the effectiveness of our proposed method. Compared with the most advanced methods currently available, this method has stronger robustness and higher recognition accuracy in clothing change scenarios.

The two solutions proposed in this thesis for clothing-related challenges have played important roles in model design, training strategies, and metric learning. An effective dual-stream fusion network was designed, a novel and efficient clothing sampling strategy was provided, and a more suitable metric loss function was designed for clothing change scenarios. By comparing the proposed methods with mainstream methods and conducting ablation experiments on the proposed methods, the effectiveness of the proposed methods was fully validated from multiple perspectives. These methods are not only innovative in theory but also highly practical in real-world applications, providing new ideas and methods for the development of person re-identification in clothing change scenarios.

**Key words:** Cloth-Changing Person Re-identification; Convolutional Neural Network; Vision Transformer; Active Learning; Sampling Strategy



# 目 录

摘要 .....	i
Abstract .....	iii
第 1 章 绪论 .....	1
1.1 研究背景和意义 .....	1
1.2 国内外研究现状 .....	2
1.3 本文主要研究内容 .....	6
1.4 本文结构组织 .....	7
第 2 章 相关背景知识 .....	8
2.1 卷积神经网络 .....	8
2.2 视觉注意力 .....	10
2.3 行人重识别 .....	11
2.4 换装行人重识别数据集 .....	13
2.5 换装行人重识别评价指标 .....	16
2.6 换装行人重识别损失函数 .....	17
第 3 章 基于行人图像分割的换装行人重识别方法 .....	19
3.1 引言 .....	19
3.2 方法设计 .....	19
3.2.1 空间解纠缠模块 .....	20
3.2.2 数据增强模块 .....	21
3.2.3 双流融合模块 .....	23
3.2.4 损失函数 .....	24
3.3 实验设置 .....	25
3.4 对比实验 .....	25
3.5 消融实验 .....	27
3.5.1 各个模块的有效性 .....	27
3.5.2 ViT 模型的有效性 .....	28
3.5.3 多模态方法分析 .....	29

3.6 本章小结 .....	29
第 4 章 基于主动学习的换装行人重识别方法 .....	30
4.1 引言 .....	30
4.2 方法设计 .....	31
4.2.1 CS 采样器 .....	32
4.2.2 损失函数 .....	34
4.3 实验设置 .....	35
4.4 对比实验 .....	37
4.5 消融实验 .....	38
4.5.1 CS 采样器的有效性 .....	38
4.5.2 损失函数的有效性 .....	38
4.5.3 RAS 采样器跟 CS 采样器的比较 .....	39
4.5.4 不同采样器训练效率的比较 .....	40
4.6 可视化分析 .....	40
4.6.1 特征分布可视化 .....	41
4.6.2 激活特征图的可视化 .....	41
4.7 本章小结 .....	42
第 5 章 总结与展望 .....	43
5.1 总结 .....	43
5.2 展望 .....	43
参考文献 .....	45
攻读硕士学位期间取得的研究成果 .....	53

## 第 1 章 绪论

### 1.1 研究背景和意义

在过去，大街小巷尚未安装足够多的监控摄像头时，公共安全领域曾长期遭受着严峻的挑战，犯罪活动难以追踪，使得社会安定和民众安全受到严重侵害。近年来随着科技的进步，监控摄像头逐渐普及至城市的每一个角落，成为维护公共安全的重要防线。通过实时监控可以有效降低犯罪率，为城市的安全稳定提供了有力保障。

然而，在监控系统持续拓展覆盖范围的大背景下，数据呈海量爆发式增长态势。如此规模的数据量下，给单纯依靠人工实施监控的方式带来许多的问题，高昂的人力成本、长时间的注意力集中需求以及不可避免的疏忽，都使得人工监控的效率低下且容易出错。因此，借助机器辅助进行智能监控成为了建设智慧城市的必然选择。

在这一背景下，人脸识别技术<sup>[1][2]</sup>应运而生，并迅速在公共安全领域展现出巨大的应用潜力。通过精准识别人脸特征，该技术能够迅速锁定目标人物，为警方提供关键线索。然而，在实际应用中，由于遮挡、光线变化、姿态差异等多种因素的影响，高质量的人脸图像往往难以获取，从而降低了人脸识别的准确性和可靠性。

为了弥补这一不足，行人重识别技术<sup>[3][4][5][6]</sup>，尤其是换装行人重识别（Cloth-Changing Person Re-identification, CC Re-ID）技术，进一步为智慧监控系统提供了强有力的支持。该技术不仅能够跨越不同摄像头和时间点，对行人的全身特征进行持续追踪，还能在行人换装或姿态变化的情况下进行身份识别。这一特性使得换装行人重识别技术在追踪犯罪嫌疑人、实现跨摄像头连续追踪等方面具有独特的优势，为智慧城市建设注入了新的活力。通过行人重识别及相关技术，本文可以构建一个更加全面、高效、智能的公共安全监控体系，为城市的和谐稳定保驾护航。

然而换装行人重识别任务在实际应用中因服装的多样性和变化性而面临着巨大的挑战，这些挑战主要可以归结为以下两个方面：

（1）服装变化及其引发的人体轮廓改变是换装行人重识别的一大难题。服装作为行人外观的重要组成部分，其颜色和纹理的多样性及复杂性直接影响了图像中行人特征的提取。如图 1-1（a）所示，当行人更换服装时，这些颜色与纹理的显著变化会占据行人图像区域的很大一部分，极大地干扰了网络学习到的行人特征。这种干扰不仅使得特征表示变得不稳定，还可能导致检索结果的准确性大幅下降，使得系统难以准

确地在不同摄像头之间追踪同一行人。其次，人体轮廓同样是行人识别中的重要线索，但在实际场景中，如图 1-1 (b)，服装的多样性和季节性变化会极大地影响人体轮廓的呈现。例如，紧身或轻薄的服装对人体轮廓的影响相对较小，但宽身或厚重的衣服则会显著改变人体轮廓特征，使得同一行人在不同着装下的轮廓差异显著。这种轮廓变化不仅增加了特征匹配的难度，还可能导致检索算法在处理类似轮廓但不同身份的行人时产生混淆，进一步影响检索效果。

(2) 辅助模型和服装标签的不准确性同样是换装行人重识别需要面对的问题。在借助辅助模型直接获得行人特有特征作为辅助特征时，由于训练域和测试域的巨大差异，在换装数据集中效果难以保证，容易带来新的干扰。而在构建训练数据集时，对行人服装的标注往往是一个复杂且耗时的过程。虽然对同一人的不同服装进行标注相对容易，但要对不同行人的相同或相似服装进行准确标注则需要付出巨大的成本。在实际操作中，由于标注者的主观性和疏忽，可能会出现对不同的人穿着相同或相似服装时标签不一致的情况。这种标签的不一致性会直接影响网络训练的准确性，使得模型难以学习到真正有效的特征表示，进而降低换装行人重识别的性能。



图 1-1 服装变化所带来的两类挑战  
Figure 1-1 Two types of challenges brought about by clothing changes

## 1.2 国内外研究现状

换装行人重识别任务是在通用行人重识别研究的基础上，结合现实应用场景的需求而逐渐发展起来的。现有的大多数行人重识别模型在很大程度上依赖于颜色纹理特征，假设行人在不同摄像头下不会更换衣服。然而，当行人发生服装更换时，如短期内服装颜色纹理的变化或长期内服装类型的改变，利用服装信息反而会诱导模型做出

错误的判断。例如，在追捕犯罪嫌疑人时，嫌疑人可能会通过不断更换上衣和裤子来干扰办案人员的判断。为了解决这一问题，近年来越来越多的研究者开始关注换装行人重识别任务。

解决换装行人重识别的关键在于抑制服装信息的干扰并提取服装不变的特征。基于辅助模型的方法是当前最流行的方法之一，通过辅助模型获取与服装无关的特征，直接利用或与其他模态特征进行融合，从而提供可靠的先验知识帮助识别。例如，一些研究利用人体姿态估计模型来提取姿态特征，关注对服装和姿势变化不敏感的体型信息。而基于解纠缠的方法是换装行人重识别中的另一个热门研究方向，主要是在 RGB 图像上利用辅助模型或辅助标签来引导模型获取丰富的人体独特特征。此外，基于 Transformer 架构的方法在通用行人重识别任务中发挥着重要作用，因此也有相关工作在此基础上进行研究。这些方法在近年来取得了一定的进展，为换装行人重识别领域的发展提供了新的思路和解决方案。

#### （1）基于辅助模型的方法

在原始 RGB 图像中，服装的颜色和纹理特征较为明显，对换装行人重识别模型的干扰较大。为了解决这一问题，基于辅助模型的方法通过获取与服装关联较小的模态信息来提取相对鲁棒的特征。这些模态信息包括轮廓草图、人体解析图、人体姿态和步态等，可以直接利用或与其他模态特征进行融合，以获取更多与服装无关的特征。

Yang 等<sup>[7]</sup>在 2019 年的 TPAMI 上发表的工作认为在行人服装颜色纹理变化的情况时，衣服的厚度保持不变，即人的形状特征不会发生太大的变化，这时可以在轮廓图上引入空间极坐标变换来提取和聚合多角度下的人体形状特征，从而提取可靠且具有判别性的特征以识别行人。Liu 等<sup>[8]</sup>在 2022 年的 ACM MM 会议上提出一个名为 M2Net-F 的网络，利用轮廓图和人体解析图来丰富 RGB 图像特征。M2Net-F 集成了 RGB 图像、轮廓图像和人体解析图像的多模态特征，忽略 RGB 图像中不相关的误导性信息，从而进行换装情况下的身份识别，学习对服装变化具有鲁棒性的特征。Zhang 等<sup>[15]</sup>在 2022 年的 ICME 会议上提出了一种利用服装无关线索的多生物特征统一网络，引入了多生物特征分支，提取头部、颈部、肩部等多种生物特征，以抵御服装变化。这些研究为换装行人重识别提供了有效的解决方案，通过多模态特征的融合和创新的特征提取方法，显著提高了模型在面对服装变化时的识别准确率。

#### （2）基于解纠缠的方法

实际上，原始的 RGB 模态包含了丰富的与衣服无关的信息，这些信息在很大程度上

上未被现有方法充分利用。尽管一些换装行人重识别方法使用了强大的主干网络从原始图像中提取特征，但没有设计合理的损失函数，导致学习后的特征仅关注一些简单的与衣服无关的信息，如人脸，而其他重要的与服装无关的信息则被忽略。因此，近年来基于 RGB 图像进行解纠缠学习的方法得到了越来越多的关注，提取高语义的服装无关特征成为其主要研究方向，例如体型等信息。基于解纠缠的方法可以分为两类：数据驱动解纠缠和特征驱动解纠缠。

### ①数据驱动解纠缠

基于数据驱动的方法主要通过数据增强来模拟服装变化，迫使模型学习与服装无关的特征。该方法可分为基于图像空间重构和基于特征空间重构两种方式。

基于图像空间重构。Shu 等<sup>[9]</sup>在 2021 年的 SPL 期刊上提出的语义引导像素采样工作中采用随机化衣服颜色的数据增强方法，使用身体解析器收集服装区域，并在小批量中重新分配其像素值，模拟服装颜色纹理变化以增加训练数据。Guo 等<sup>[16]</sup>在 2022 年的 TMM 会议上提出的语义感知一致性网络中采用语义感知的数据增强补丁策略，随机组合衣服补丁，模拟同一个人穿不同衣服来进行数据增强。

基于特征空间重构。Han 等<sup>[17]</sup>在 2023 年的 CVPR 会议上提出的服装变化特征增强工作，首次在特征空间进行数据增强，自动建模特征分布扩展，反映一个人的服装颜色和纹理变化，以增强模型训练。该方法首先估计在相同身份不同服装场景下的服装变化正态分布，然后通过用于数据增强的生成器学习这个分布，来增加合理的服装变化用例。通过对抗学习的方法，保证增强特征最大限度地增加服装的变化，最大限度地减少身份属性的变化，以保证增强特征的有效性。

### ②特征驱动解纠缠

基于特征驱动的方法主要利用附加特征监督引导模型从 RGB 图像中获取与服装无关的特征，可分为基于辅助特征引导和基于服装特征对抗两种方式。

基于辅助特征引导。Qian 等<sup>[11]</sup>在 2020 年的 ACCV 会议上发表的工作中提出学习高语义人体特征如体型信息，通过设计形状嵌入模块和消除服装形状蒸馏模块，消除目前不可靠的服装特征。此方法利用现成的姿态估计模型编码关键点形状信息，利用注意机制对特征进行解纠缠提取身份相关特征，完全删除与服装外观相关的信息，只关注对透视和姿势变化不敏感的体型信息。Hong 等<sup>[18]</sup>在 2021 年的 CVPR 会议上的工作中指出预先训练好的辅助模型存在很大的误差，导致模型提取与服装无关的体型信

息效果不理想。为了解决这个问题，此工作提出了一种细粒度的 Shape-Appearance 互学习框架用于换装行人重识别任务，通过利用密集交互学习的方法，将细粒度的体型知识从形状流传递到身份流。Jin 等<sup>[13]</sup>在 2022 年的 CVPR 会议上做的工作首次引入步态识别作为一项辅助任务来驱动换装行人重识别模型，通过利用个人独特的与服装无关的步态信息来学习与服装无关的特征表示，步态特征流作为一个调节器，鼓励身份特征流在训练过程中捕捉服装不变的生物特征。

基于服装特征对抗：Gu 等<sup>[14]</sup>在 2022 年的 CVPR 会议上的服装对抗（CAL）工作首次提出只利用 RGB 图像挖掘体型等与服装无关的特征。此工作通过惩罚模型对于服装的识别能力来从原始 RGB 图像中挖掘衣服无关特征，利用对抗学习，将同一身份的所有服装分类认为属于正类，着重提取身体部分信息，在此基础上抑制服装信息以达到提取体型特征。Yang 等<sup>[19]</sup>在 2023 年的 CVPR 会议上所做的工作中提出使用因果蒸馏来提取与服装无关的特征。此工作指出服装与身份的纠缠表示导致基于统计的神经网络向易于区分的浅层服装特征收敛，这种统计关联让 ReID 模型产生了一种错误的看法，即服装和身份之间存在着很强的联系，这将破坏合理结果的最终预测。对此，该文提出的自动干预模型分析了服装对模型推理的影响，并采用双分支模型模拟因果干预，并通过模型训练逐步消除这种服装偏见。

在日常生活中，每个人的服装搭配跟他们的个性一样往往都是独一无二的，一个人可以穿很多套衣服，却很少会有人穿一模一样的衣服。所以在换装数据集中，不同行人即使穿相同衣服，其服装也会被标记为不同标签。因此，基于单模态的方法在借助服装标签解纠缠特征的同时也会引入新的错误关联。

### （3）基于 Transformer 架构的方法

He 等<sup>[20]</sup>在 2021 年 ICCV 会议上提出了在行人重识别领域第一篇基于视觉注意力机制（Vision Transformer, ViT）的工作 TransReID，为行人重识别模型设计了两个模块：边信息编码和拼图补丁模块。其中边信息编码将摄像机或视点之类的非可视信息编码为矢量嵌入插入到 ViT 中以消除各种相机或视点引起的偏差。拼图补丁模块通过对图像编码块进行位移和打乱操作，对其进行重排列，获得鲁棒的人体特征，提高模型识别能力和增强输入的多样性。Bansal 等<sup>[21]</sup>在 2022 年 WACV 会议上提出了一个基于自注意力机制的框架，直观地利用独特的基于软生物特征的识别信息，如步态特征，并将其与 ViT 特征表示配对，以允许模型生成长期结构和上下文关系，以应对服装变化。

Wang 等<sup>[22]</sup>在 2022 年的 ACCV 会议上首次提出了采用 Transformer 来处理 CC Re-ID 的多模态交互方法。首次设计了一种新型的形状语义嵌入模块，对行人的体型语义信息进行编码，并提出了一种共同注意对齐的相互交叉注意框架，将来自多种形态的特征进行对齐，然后在图像空间和体型空间之间有效地交互和传递具有身份意识但与服装无关的知识，从而获得更鲁棒的特征表示。

然而，上述各类方法不可避免的存在各种缺陷。借助辅助模型的方法，容易受辅助模型噪声的干扰；借助服装标签的方法，无法忽略部分标签错误关联的影响；纯 Transformer 方法，难于捕捉局部细粒度信息。

### 1.3 本文主要研究内容

本文的主要研究内容和贡献如下：

(1) 针对换装行人重识别任务中需要面对的两类主要挑战：服装变化及其引发人体轮廓改变、辅助模型和服装标签的不准确性，系统地分析了这些问题的由来以及对行人重识别的影响。详细来看，这些挑战均与服装直接关联，在服装中等变化情况下，即服装的纹理颜色发生变化，但人体轮廓保持不变，模型易受显著的服装区域影响导致网络推理阶段干扰人体局部特征向量的生成，从而影响识别结果的准确性；随着服装款型的巨大变化，人体轮廓等被破坏，在特征提取阶段的高级语义特征被严重影响，从而进一步增加识别难度；为了降低服装与身份之间的关联，许多研究需要借助服装标签或辅助模型，而这部分标签或先验模型的不准确性给换装任务带来新的挑战。

(2) 设计了一种基于图像分割的双流融合网络。其首先借助预训练的人体分割模型在图像层面进行数据增强，模拟服装颜色纹理变化从而抑制服装特征的干扰，但随之而来的先验模型的噪声给卷积神经网络特征提取器带来不小的干扰，为此本文使用视觉注意力模型进行特征提取，降低局部噪声干扰的同时提取更加鲁棒的高级语义信息，并与人体局部特征进行融合。其次利用空间解纠缠模块进一步在图像空间不同尺度上将服装与身份特征进行解纠缠。

(3) 设计了一种基于主动学习的服装采样方法。之前的方法在借助辅助模型或者服装标签的同时，不可避免会遇到由噪声带来的新的挑战。对此本文提出一种在不借助辅助模型和错误服装标签情况下，从训练策略出发，提出一种适合换装任务的服装采样器，从数据驱动角度利用主动学习的方法筛选有价值数据以训练模型。并设计相关实例一致性损失来应对服装剧烈变化带来的人体轮廓改变挑战。



(4) 通过大量的实验验证了所提出方法的有效性。在多个公开数据集上的测试结果表明，本文提出的方法在换装行人重识别任务中取得了显著的性能提升，充分证明了所提出方法的有效性和优越性。

## 1.4 本文结构组织

本文针对换装行人重识别的研究难题，系统地分为以下五个章节进行阐述：

第一章，绪论。首先介绍了换装行人重识别的研究背景和研究意义，指出了该领域在实际应用中的重要性和挑战性。随后，对换装行人重识别的发展历史和研究现状进行了全面概述，为后续章节的深入研究奠定了基础。最后，简要陈述了本文的主要研究内容和贡献。

第二章，相关背景知识介绍。本章详细阐述了卷积神经网络模型和视觉自注意力模型的基本原理，为后续章节中网络结构的设计提供了理论支撑。同时，针对换装行人重识别任务，介绍了相关的数据集、损失函数以及评价指标，为后续实验的设计和性能评估提供了依据。

第三章，基于图像分割的双流融合网络。本章首先详细介绍了双流融合网络的结构及其组成，包括整体结构、空间解纠缠模块、数据增强模块、双流融合模块等关键部分。随后，通过实验结果的对比和分析展示了该方法在换装行人重识别任务上的有效性。最后，对整个方法进行了总结。

第四章，基于主动学习的服装采样方法。本章提出了一种针对换装任务的服装采样方法，并详细介绍了其步骤。并设计了相应的实例一致性损失，在服装显著变化情况下亦能取得很好的效果。通过实验结果的对比和分析，验证了该方法在应对服装变化方面的优越性。最后，对这一方法进行了总结和评价。

第五章，总结与展望。本章首先对本文的主要研究内容和创新点进行了简要概括，并分析了本文方法的不足之处。随后，对未来可能的研究方法进行了展望，并提出了本文方法可能的改进思路。通过这一章节，为换装行人重识别领域的后续研究提供了有益的参考。

## 第 2 章 相关背景知识

### 2.1 卷积神经网络

卷积神经网络（Convolutional Neural Network, CNN）的发展历史是深度学习和计算机视觉领域中一个引人注目的篇章。CNN 的起源可以追溯到 20 世纪 50-60 年代，当时 Hubel 和 Wiesel 通过对猫的视觉皮层<sup>[23]</sup>的研究，发现了感受野的概念，为后来的 CNN 结构提供了重要的生物学灵感。1980 年代，日本学者福岛邦彦提出的 Neocognitron<sup>[24]</sup>是早期的一种卷积网络，用于手写字符识别，虽然没有使用反向传播算法进行训练，但它是 CNN 发展史上的一个重要里程碑。

1989 年，Yann LeCun 在手写字符识别中首次成功地将反向传播<sup>[25]</sup>算法应用于多层卷积网络，这一研究不仅验证了反向传播算法在多层网络中的有效性，还展示了卷积神经网络在图像处理中的潜力。此后，LeCun 等人继续改进这一技术，最终在 1998 年提出了 LeNet-5<sup>[26]</sup>的架构和应用，标志着卷积神经网络在图像识别领域的重大突破。然而，CNN 的发展在之后一段时间内受到了限制，直到 2012 年，Alex Krizhevsky、Ilya Sutskever 和 Geoffrey Hinton 设计的 AlexNet<sup>[27]</sup>在 ImageNet 竞赛中取得了历史性的胜利，这一成就不仅标志着深度学习在计算机视觉领域的复兴，也引发了全球对深度学习技术的广泛关注。

随后，深度学习社区见证了一系列创新模型的诞生，它们在性能和效率上不断突破极限。2015 年，GoogLeNet<sup>[28]</sup>以其创新的 Inception 模块引领了深度和宽度的双重扩展，而 VGGNet<sup>[29]</sup>则以其深度网络结构在图像识别任务中取得了显著的成绩。紧接着，何恺明等人在 2015 年提出的 ResNet<sup>[30]</sup>通过引入残差连接解决了深层网络训练中的梯度消失问题，使得网络能够构建得更深，为图像分类和目标检测等任务提供了强大的基础。

CNN 的发展不仅仅局限于图像识别领域。在医学领域<sup>[31]</sup>，能够帮助医生分析 X 光和 MRI 图像，辅助诊断疾病；在自动驾驶技术中<sup>[32]</sup>，用于识别道路标志和障碍物，保障行车安全；在金融领域<sup>[33]</sup>，也能通过分析交易图表来预测股市走势。尽管存在挑战，如对训练数据的大量需求<sup>[34]</sup>和模型解释性不强等问题，CNN 的发展仍在不断推进。研究人员正在努力提高其效率和准确性，同时也在探索如何减少对大量标注数据的依赖。

随着技术的不断进步，CNN 在人工智能发展中的角色越来越重要，其应用范围也在不断扩大，涉及图像、语音识别以及自然语言处理等众多领域。CNN 以其强大的特征提取能力和对图像干扰因素的鲁棒性，已经成为解决多种复杂问题的关键技术之一。

CNN 的基本结构由输入层、卷积层、激活函数、池化层、多层堆叠、全连接层和输出层构成。

**输入层：**输入层接收原始图像数据。图像通常由三个颜色通道（红、绿、蓝）组成，每个通道由一个二维矩阵构成，其中的每一个值表示像素值。

**卷积和激活：**卷积层将输入图像与卷积核进行卷积操作，如图 2-1 所示。然后，通过应用激活函数来引入非线性。这一步使网络能够学习复杂的特征。

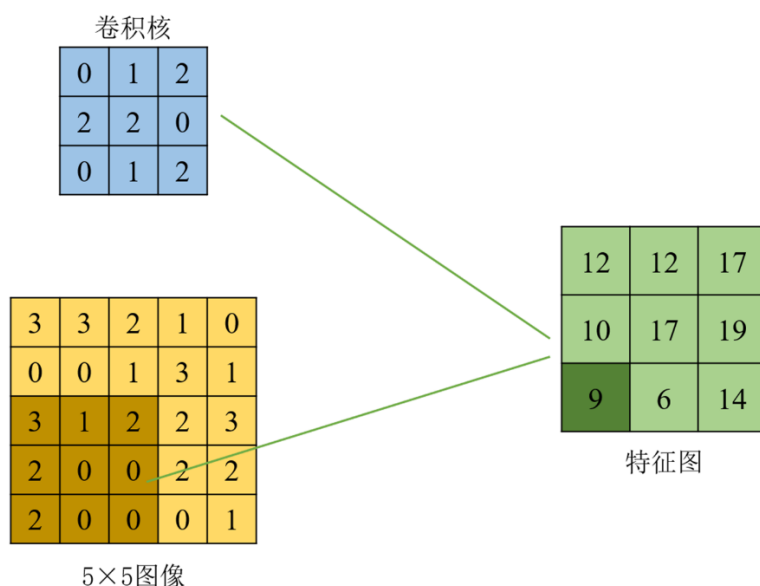


图 2-1 卷积操作  
Figure 2-1 Convolution operation

**池化层：**池化层通过减小特征图的大小来减少计算复杂性。它通过选择池化窗口内的最大值或平均值来实现，最大池化如图 2-2 所示。这有助于提取最重要的特征。

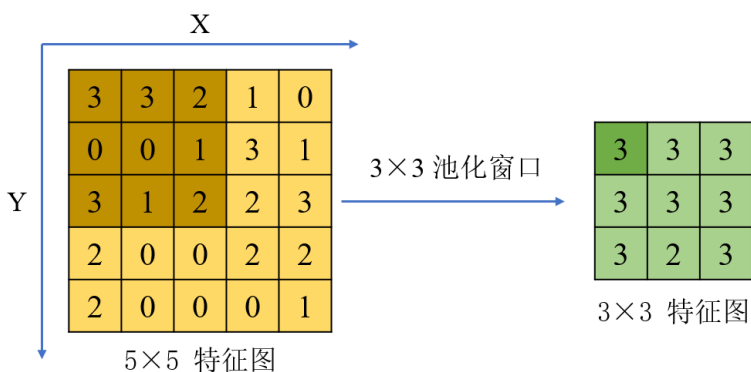


图 2-2 最大池化操作  
Figure 2-2 Maximum Pooling Operation

多层堆叠：CNN 通常由多个卷积和池化层的堆叠组成，扩大感受野以逐渐提取更高级别的特征。深层次的特征可以表示更复杂的模式。

全连接和输出：最后，全连接层将提取的特征映射转化为网络的最终输出。这可以是一个分类标签、回归值或其他任务的结果。

## 2.2 视觉 Transformer 模型

视觉注意力模型<sup>[35]</sup>（Vision Transformer, ViT）是首次将 Transformer 架构应用于图像分类的模型，最早由 Google 团队提出。因其结构简单、效果好且可扩展性强等因素，成为 Transformer 在 CV 领域应用的里程碑。与传统的卷积神经网络依赖卷积层获取局部特征不同，ViT 有强大的全局感知能力，能进行直接的全局关系建模，获取更多的上下文信息，其具体的网络结构如图 2-3 所示。

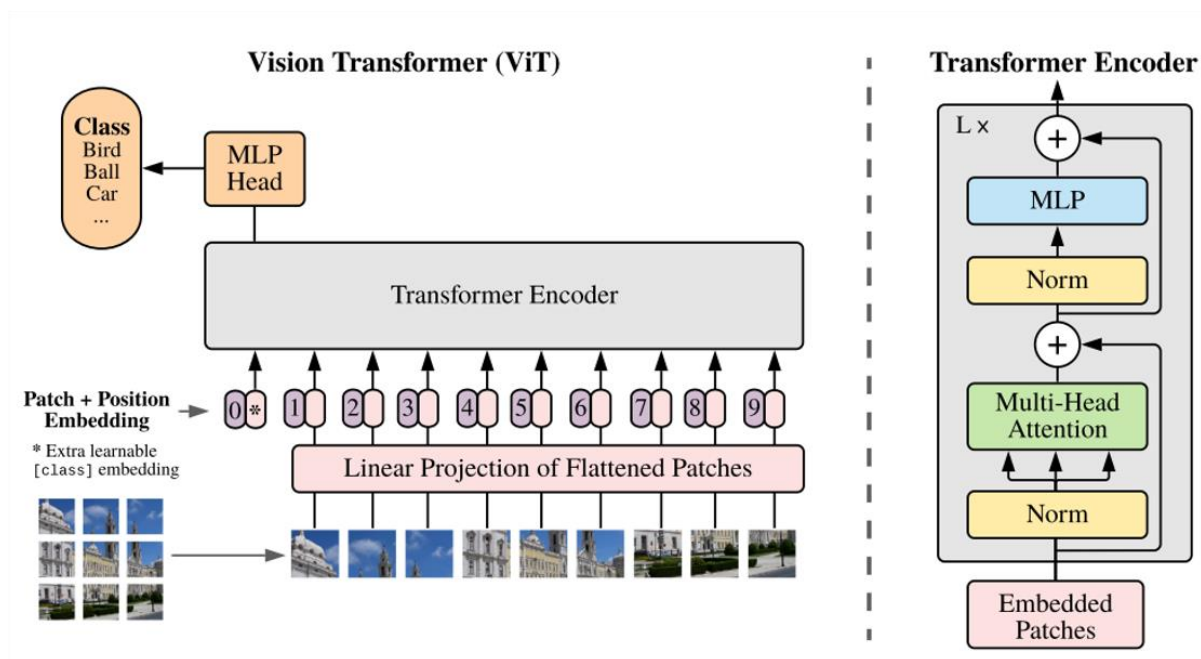


图 2-3 ViT 网络结构<sup>[35]</sup>  
Figure 2-3 ViT Network Structure<sup>[35]</sup>

其工作流程主要有以下几个步骤：

（1）图像划分：输入图片被分成大小相同数量固定的图像块，并将这些图像块进行铺平，这些图像块通常被称之为“补丁”（patch），每个 patch 相当于自然语言处理任务中的单词，代表图像的每一个局部区域。

（2）形成序列：将铺平的 patches 经线性映射获得其各自特征向量（token）。这些向量构成了模型的输入序列。

(3) 位置编码 (Position Embedding): 为输入的 token 序列额外添加一个可学习的 CLS token, 并为每个 token 添加位置信息, 保持其图像的空间结构。

(4) 编码器 (Transformer Encoder): 将 token 送入编码器中去, 利用多头自注意力机制关注特征之间的关系, 从而捕捉全局上下文信息。

(5) 分类头 (MLP Head): 在图像分类任务中, 最后由一个分类头对 CLS token 进行分类。

(6) 预训练和微调: 基于 Transformer 架构的模型基本需要先较大的数据集上进行预训练, 学习高质量的中间表示, 然后在下游任务中进行微调去学习特定的特征。

ViT 在 CV 领域成功的关键因素是通过自注意力机制实现输入自适应 (处理不同分辨率图像更加灵活)、远程依赖 (获取图像的全局理解) 和高级空间交互, 以及他们相互竞争的建模能力。但其无法利用图像本身的尺度、平移不变性和特征局部性等先验知识, 需要大规模数据集学习高质量的中间表示, 否则在数据集较小的任务上难以展示其优越的性能。

## 2.3 行人重识别

行人重识别技术最早起源于跟踪算法研究中, 旨在匹配同一个人不同场景中的图像, 被广泛应用于城市监控系统。在早期的方法研究中, 研究者致力于解决姿态、视角变化导致的识别不一致问题, 为此形成了一套比较成熟的行人重识别系统, 根据其设计流程, 将其分为表征学习、度量学习以及训练策略三个部分。

### (1) 表征学习

针对行人重识别任务, 现有数据集可分为图像和视频两类。针对图像数据, 主要有三类学习策略, 分别为全局特征、局部特征、辅助特征; 针对视频数据, 有时序特征学习策略。下面分别介绍四类学习策略:

(a) 全局特征: 为每个人物图像提取一个全局特征表示向量。在深度学习时代, 深度神经网络最初应用于图像分类任务, 主要是基于全局特征进行分类。而早期 Re-ID 任务也可看作分类任务, 将每个身份视为一个不同的类, 广泛使用的身份判别嵌入模型<sup>[36]</sup>开始将训练过程构建为一个多类分类问题, 目前在 Re-ID 社区广泛使用<sup>[37][38][39]</sup>。为了获取多尺度特征, Qian 等<sup>[40]</sup>开发了一种多尺度深度表征学习模型来捕获不同尺度上的判别性线索。为了进一步增强表征学习能力, 注意力机制<sup>[41]</sup>被广泛应用于 Re-ID 任务, 典型的策略如: 通道注意力<sup>[42][43][44]</sup>、空间注意力<sup>[45]</sup>和交叉注意力<sup>[46]</sup>。而随着

ViT<sup>[35]</sup>在图像分类任务的广泛应用，TransReID<sup>[20]</sup>也证明了纯 Transformer 架构在 Re-ID 任务也展示极其不错的性能。

(b) 局部特征：聚合各种局部特征，为每个人物图像制定组合表示。局部特征学习部分/区域聚合特征，面对不对齐场景更具有鲁棒性。为了获得人体的各个部位，可以使用先验模型根据人工解析图像或者姿态估计生成，也可以使用直观的方法进行水平分割获得脑部、上躯干和下躯干部分。

在基于先验模型的方法中，为了获取丰富的人体特征，常用的解决方案是将全身表示和局部部位特征结合起来。具体来说，设计了多通道聚合<sup>[47]</sup>、多尺度上下文感知卷积<sup>[48]</sup>、多阶段特征分解<sup>[49]</sup>和双线性池<sup>[50]</sup>来改善局部特征学习。另一种流行的解决方案是增强对背景杂波的鲁棒性，使用姿势驱动匹配<sup>[51]</sup>，姿势引导部分注意模块<sup>[52]</sup>，语义部分对齐<sup>[53][54]</sup>。

对于水平划分的区域特征，在基于部分的卷积基线（PCB）<sup>[39]</sup>中学习了多个部件分类器，是当前最先进的局部特征学习基线。

(c) 辅助特征：使用辅助信息改进特征表示学习，辅助信息包括属性、GAN 生成的图像等。Tay 等<sup>[55]</sup>将行人属性信息应用到分类框架中，提出一种属性注意网络，该网络将身份类别、行人属性统一到一个框架中来共同学习一个具有高判别度的特征。对于 Re-ID 任务，随机调整大小、裁剪和水平翻转<sup>[56]</sup>是常见的数据增强方法。此外，常用的数据增强方法还有利用生成对抗方法<sup>[57]</sup>模拟遮挡场景以增加可训练的数据。Huang 等<sup>[58]</sup>中提出了一种随机擦除策略，在输入图像中加入随机噪声模拟数据变化。Bak 等<sup>[59]</sup>生成在不同光照条件下渲染的虚拟人。这些方法通过扩充样本丰富了监督，提高了测试集的泛化能力。

(d) 时序特征：使用多个图像帧和时间信息来学习基于视频的时序表示，如人物走路时所具备的步态特征。基于视频的 Re-ID 同样是一个热门话题，其中每个人都由具有多帧的视频序列表示。由于其丰富的外观和时间信息，给视频特征表示学习带来了额外的挑战。主要的挑战是如何准确地捕捉时间信息。Yan 等<sup>[60]</sup>提出了一种递进/顺序融合框架来聚合连续多帧的人体区域表示。Jin 等<sup>[13]</sup>提出一种利用步态特征聚合人体特征以应对换装任务。

## (2) 度量学习

在深度学习时代之前，度量学习已经通过学习马氏距离函数<sup>[61]</sup>或投影矩阵<sup>[62]</sup>进行了广泛的研究。现如今度量学习的作用已被损失函数设计所取代，以指导特征表示学

习。对于 Re-ID 任务，常用的损失函数及其变种主要有三类：身份损失<sup>[63]</sup>、验证损失<sup>[64]</sup>和三元组损失<sup>[65]</sup>。

当把 Re-ID 任务看作分类任务时，即每个身份都是一个不同的类，主要使用身份损失（即交叉熵损失）进行分类监督；当作为匹配任务时，通过对比损失或二元验证损失来优化两两关系；当作为检索任务时，根据相似度进行排序，通过三元组损失来优化锚点与正样本之间的距离小于与负样本之间的距离，使得同一身份的图像特征更加相似以达到检索效果。

### （3）训练策略

采样策略在 Re-ID 模型学习中起着重要的作用。因为训练集中每个身份的图像数量差异很大，常用的采样策略无法应对这种情况，给训练过程带来极大的挑战。同时，正、负样本对的严重不平衡<sup>[66]</sup>也增加了训练策略设计的难度。处理不平衡问题最常用的训练策略是身份采样<sup>[65]</sup>。对于每个训练批量，随机选择一定数量的身份，然后从每个选择的身份中采样若干图像。这种批量采样策略保证了信息的正、负挖掘。为了处理正负样本之间的不平衡问题，自适应抽样是调整正负样本的常用方法，如样本率学习<sup>[44]</sup>、课程抽样<sup>[42]</sup>。另一种方法是样本重加权，利用样本分布<sup>[42]</sup>或相似差<sup>[67]</sup>来调整样本权重。Zhou 等<sup>[68]</sup>设计了一个有效的参考约束，将成对/三元组的相似性转换为样本到参考的相似性，解决了不平衡问题并增强了可判别性，同时对异常值具有鲁棒的。

为了高效地组合多个损失函数，多损失动态训练策略<sup>[69]</sup>可以自适应地对身份损失和三元组损失重新加权，合理组合各个组件。通过这种多损失训练策略，在行人重识别模型训练过程中可以获得持续的性能增益。

综上所述，在行人重识别任务中，可以从训练的三个阶段开展研究：模型设计、度量学习、训练策略。针对换装任务而言，本文同样从这三点出发，设计一套有效的双流融合网络，提供新颖高效的服装采样策略，并针对换装场景设计更贴切的度量损失函数。

## 2.4 换装行人重识别数据集

在换装行人重识别研究中，研究者们积极构建了在服装变化场景下的换装行人重识别数据集，如 PRCC<sup>[7]</sup>、NKUP+<sup>[8]</sup>、Celeb-reID<sup>[70]</sup>、Celeb-reID-light<sup>[71]</sup>、LTCC<sup>[11]</sup>、VC-Clothes<sup>[72]</sup>、COCAS<sup>[73]</sup>、LsAT<sup>[74]</sup>，下面分别对其进行介绍。

(1) PRCC<sup>[7]</sup>是一个广泛使用的服装颜色纹理变化的换装数据集,由 221 个身份的 33698 张图片组成。在训练集中包含 150 个身份,测试集中包含剩余的 71 个身份,没有重复身份出现。此外,该数据集提供额外的反应人体轮廓信息的轮廓草图图像,供后续研究者们使用。此数据集中的图片分别在 3 个不同的摄像机(A、B、C)视图下捕获的,其中摄像机 B 收集到的与 A 同一身份的服装未曾发生变化,而在摄像机 C 视角下其服装发生改变。

在测试阶段,该数据集可利用视角不同分别开展相同服装识别(Same Cloth)、跨服装识别(Cross Cloth)。另外,自 PRCC 数据集公开以来,越来越多的研究者针对此数据集开展换装行人重识别方法研究,并因该数据集是严格意义上的换装数据集,而被选为模型性能测试的基准之一。

(2) NKUP+<sup>[8]</sup>数据集是最新发布的换装数据集,由南开大学的 29 个室内外摄像头拍摄,包含 361 个身份的 40,217 张图像。与现有数据集相比,NKUP+ 提供了更加真实和多样化的场景信息,涵盖了室内和室外的多个场景、多种视角以及丰富多样的服装款式。数据集的拍摄周期长达 10 个月,覆盖了不同季节的服装变化,使其成为研究换装行人重识别的理想基准。

NKUP+ 中的测试集根据服装变化程度分为多个子集,包括相同服装、中等外观变化以及剧烈服装变化。这种设计为评估模型在不同难度下的性能提供了全面的测试场景。此外,NKUP+ 数据集的多样性和长期跨度使其能够更好地模拟现实监控环境中的行人服装变化,推动了行人重识别技术在实际应用中的发展。

(3) Celeb-reID<sup>[70]</sup>数据集是一个复杂且多样化的数据集,图像来源于互联网上的名人街拍(如谷歌图片、百度图片等),涵盖了不同的相机视角、场景、年龄、性别和国籍。该数据集包含 1,052 个身份的 34,189 张图像,其中训练集有 20,208 张图像(632 个身份),测试集有 13,978 张图像(420 个身份),测试集中约 30% 的图像为查询图像,其余 70% 作为图库集供查询。

Celeb-reID 数据集同时包含换装和不换装的场景,为换装任务提供了大规模的数据支持,推动了相关领域的发展。然而,由于其图像来源与传统行人重识别任务中监控摄像头所采集的图像存在差异,导致基于该数据集训练的模型在实际监控场景中的泛化能力有限。

(4) Celeb-reID-light<sup>[71]</sup>数据集是一个专注于换装场景的行人重识别数据集,是 Celeb-reID 的轻量级版本。该数据集包含 590 个行人身份的 10,842 张图像,所有图像



均来源于互联网上的名人街拍。与 Celeb-reID 一样, Celeb-reID-light 的训练集和测试集仅包含换装情况的数据, 且每个人在数据集中不会穿两次相同的衣服。这种设计使其成为一个纯粹的换装行人重识别数据集, 专注于模拟真实场景中行人更换服装的挑战。

(5) LTCC<sup>[11]</sup>数据集是一个室内换装数据集, 包含 152 个身份的 17,119 张图像, 由 12 个室内摄像头在超过两个月的时间跨度内捕获。数据集涵盖了多种照明、遮挡、相机视角、携带物品和姿势变化。它被划分为两个子集: 换装子集包含 91 个行人身份和 416 套不同服装的 14,783 张图像, 平均每人有 5 件不同衣服; 服装不变子集包含 61 个行人身份的 2,336 张图像, 其中行人的服装保持一致。LTCC 数据集通过模拟室内真实场景中的长期服装变化, 为行人重识别算法在换装场景下的性能评估提供了重要的基准, 但受限于数据集规模比较小, 在此数据集上的工作进展较为缓慢。

(6) VC-Clothes<sup>[72]</sup>数据集是一个基于高清逼真的游戏画面渲染的虚拟换装数据集, 包含 512 个身份, 共 19,060 张图像, 其中训练集有 9,449 张图像 (256 个身份), 测试集有 9,611 张图像 (256 个身份), 并模拟了现实世界中的光照变化和姿态变化等信息。该数据集包含 1 个室内场景和 3 个室外场景, 分别由 4 个摄像头拍摄。其中, 行人在摄像机 2 和摄像机 3 之间穿着相同的衣服, 而在摄像机 3 和摄像机 4 之间穿着不同的衣服。每个身份有 1~3 套衣服。VC-Clothes 数据集为研究算法在虚拟现实中的应用提供了基础。

(7) COCAS<sup>[73]</sup>数据集是一个大规模的实拍数据集, 包含 5,266 个身份的 62,382 张图像, 每个身份平均有 12 张图像, 涵盖了多种真实场景 (30 台摄像机拍摄), 包括不同的光照条件 (室内和室外) 和遮挡情况。COCAS 数据集的图像采集于多个商品交易市场, 通过复杂的人物关联和隐私保护策略构建而成。此外, 该数据集还引入了一种新的换装行人重识别设置, 查询时需要提供一张人物图像和对应的服装模板。为了保护隐私, 所有行人图像的面部区域均被模糊处理。但 COCAS 数据集目前尚未开源。

(8) LaST<sup>[74]</sup>数据集是目前规模最大的换装行人重识别数据集, 具有极高的时空跨度和多样性。该数据集包含超过 228,000 张行人图像, 涵盖了广泛的场景和时间范围, 模拟了行人可能在不同地点和时间更换服装的真实情况。图像采集自电影或电视剧截图, 涉及不同国家、年龄、场景、天气以及白天和夜晚等多种条件。这种丰富的多样性使得 LaST 数据集在研究行人重识别任务时更具挑战性, 同时也为模型学习提供了丰

富的数据支持。然而，由于其采集方式的限制，LaST 数据集的背景较为复杂，但其庞大的数据量和广泛的时空覆盖使其成为研究换装行人重识别的重要资源。

上述各数据集在来源和内容上各有特点，其具体信息如表 2-1 所示。例如，PRCC、NKUP+、LTCC 和 COCAS 数据集主要基于真实监控摄像头采集图像，而 Celeb-reID 和 Celeb-reID-light 数据集则源自互联网上的图片资源。VC-Clothes 数据集通过高清游戏画面渲染生成，LaST 数据集则是从电影或电视剧截图中整理而来。考虑到本文主要研究换装任务，所以实验主要在 PRCC 和 NKUP+上进行。PRCC 提供丰富的室内场景，拍摄周期较短，主要包含服装颜色纹理变化，而 NKUP+包含更加真实的室外场景，时间跨度大，包含各种程度上的服装变化。

表 2-1 换装行人重识别数据集  
Table 2-1 Person Re-identification Dataset for Cloth-Changing

方法	换装形容重识别数据集 (身份/数量)		相机数量	介绍
	训练集	测试集		
PRCC	150/17896	71/6927	3	实拍、每个人均有两套服装
NKUP+	181/22590	181/17627	29	实拍、时间跨度大
Celeb-reID	632/20208	420/13978	-	名人的街拍照片、背景杂乱
Celeb-reID-light	490/9021	100/1821	-	
LTCC	77/9576	75/7543	12	实拍、视角多、规模小
VC-Clothes	256/9449	256/9611	4	规模大、虚拟数据
COCAS	2800/34019	2466/28306	-	实拍、未开源
LaST	5000/ 71248	5806/135529	-	规模巨大、背景杂乱

## 2.5 换装行人重识别评价指标

在测试阶段，数据集主要分为查询集和图库集。query 表示对查询集中每一张待检索的图片，包含  $n$  张图片， $q_i$  表示第  $i$  个 query，则查询集可记作： $Q = \{q_1, \dots, q_i, \dots, q_n\}$ 。gallery 表示图库集中的每一张图片，包含  $m$  张图片， $g_i$  表示第  $i$  个 gallery，则图库集可记作： $G = \{g_1, \dots, g_i, \dots, g_n\}$ 。对于每个 query 的查询过程： $q_i$  提取特征变量后依次跟  $G$  中每张图片的特征向量进行比对，计算出距离，按照距离大小对 gallery 进行升序排列，排序后的数据集记作  $G_{q_i}$ 。

(1) 平均准确率平均值 (mAP)

将 $q_i$ 命中的图片组成新的数据集，记为 $\bar{G}_{q_i} = \{\bar{g}_1, \dots, \bar{g}_j, \dots, \bar{g}_{k_i}\}$ ， $k_i$ 代表命中的数量。假设 $\bar{g}_j$ 在 $G_{q_i}$ 中的排位记为 $r_j$ ，在 $\bar{G}_{q_i}$ 中的排位记为 $\bar{r}_j$ ，则 $q_i$ 命中的准确率为：

$$AP = \frac{1}{k_i} \sum_{j=1}^{k_i} \frac{\bar{r}_j}{r_j} \quad (2-1)$$

对所有 query 重复这个过程，得到平均准确率的平均值如下：

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (2-2)$$

## （2）rank-1 准确度（R1）

对于 $G_{q_i}$ 中， $q_i$ 跟位于第一位的 gallery 身份相同则称 $q_i$ 满足首位命中（rank-1 击中）， $Q$ 中所有满足首位命中的 query 数量记为 $\bar{n}$ ，则 rank-1 准确率记为：

$$R1 = \frac{\bar{n}}{n} \quad (2-3)$$

## 2.6 换装行人重识别损失函数

在换装行人重识别中，主要使用以下几个损失函数：

（1）交叉熵损失（Cross-Entropy Loss）：对于分类任务而言，该损失可以加强行人图像特征的类内紧凑性，保证相同身份特征在嵌入空间的距离较小，公式如下：

$$L_{CE} = -\sum_{i=1}^n y_i \log(p_i) \quad (2-4)$$

其中  $n$  表示数据集中的  $n$  个不同身份， $p_i$  是图像特征向量经分类头输出的概率分布， $y_i$  是 one-hot 编码后的身份标签。

（2）三元组损失（Triplet Loss）：最早由谷歌团队提出，常用在人脸识别任务中，对 Re-ID 而言可以进一步保证行人图像特征的类间分离性，做到不同身份相似样本的区分，其公式如下：

$$L_{Triplet} = \sum_{i=1}^n \max\left(0, \text{margin} + \text{Dist}(\text{anchor}^i, p^i) - \text{Dist}(\text{anchor}^i, n^i)\right) \quad (2-5)$$

其中  $\text{Dist}$  表示距离度量函数， $\text{margin}$  是一个阈值超参， $\text{anchor}^i$  锚点， $p^i$  是跟锚点身份相同的正样本， $n^i$  是跟锚点身份不同的负样本。

（3）中心损失（Center Loss）：为每一个类维护一个类中心，使得同一类图像更加靠近类中心，其公式如下所示：

$$L_{Center} = \frac{1}{2} \sum_{i=1}^n \left( Y_i - center^{(y_i)} \right)^2 \quad (2-6)$$

其中  $Y_i$  表示图像  $i$  最后的特征表示,  $y_i$  表示图像  $i$  的身份类别,  $center^{(y_i)}$  表示该类的类中心。

(4) 均方差损失 (MSE Loss): 常用于回归任务, 计算模型预测值与真实值的差异。在 Re-ID 任务中, 可以用来监督数据增强前后的特征表示, 来获得更鲁棒的行人特征, 其公式如下:

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^n \left( Y_i - \bar{Y}_i \right)^2 \quad (2-7)$$

其中  $Y_i$  和  $\bar{Y}_i$  分别代表数据增强前后的特征表示。

以上损失函数通常会结合使用, 通过加权组合以达到对判别性身份特征的综合优化。根据具体数据集和场景, 对上述超参进行赋值以及对损失函数进行组合。

## 2.7 本章小结

本章详细介绍了卷积神经网络和视觉 Transformer 模型这两类骨干网络、行人重识别系统构成、以及换装领域常用的数据集、评价指标、损失函数。卷积神经网络通过卷积层和池化层可以获取丰富的局部特征。视觉 Transformer 模型利用自注意力机制处理图像块序列, 能够捕捉长距离依赖关系, 对全局特征建模更为有效。在计算机视觉领域, 行人重识别模型主要由表征学习、度量学习、训练策略三部分构成。

## 第3章 基于行人图像分割的换装行人重识别方法

### 3.1 引言

当采集到的行人图像中的服装发生中等程度的变化，即款式保持不变但颜色和纹理发生显著变化时，传统的行人重识别方法往往难以有效消除服装的干扰，导致识别准确率相对较低。为了解决这一问题，近年来的研究重点在于如何更好地利用辅助模型提供的先验知识来辅助识别。最初的研究思路是直接借助辅助模型提取与服装无关的特征，例如，Yang 等<sup>[7]</sup>通过提取人体轮廓图进行识别，Liu 等<sup>[8]</sup>则提出将 RGB 图像、人体轮廓图和人体解析图等特征进行融合。然而，直接依赖辅助模型在真实场景中的可靠性难以保证，且泛化性能较低。此外，RGB 图像本身已包含丰富的与服装无关的信息，因此最前沿的研究更关注于如何从原始图像中提取更多的人体特征，以减少服装的干扰，即实现服装与身份特征的解纠缠。在基于解纠缠的方法中，通常会利用辅助模型的特征来引导网络学习，例如，Qian 等<sup>[11]</sup>利用先验姿态估计模型引导学习姿态特征，获取对服装和姿势变化不敏感的体型信息；Jin 等<sup>[13]</sup>则利用先验解析模型获取人体解析图像，并针对图像提取步态特征进行引导。此外，利用辅助模型的先验知识进行数据重构<sup>[9]</sup>，模拟服装变化，也能有效帮助模型实现解纠缠。

上述研究工作主要集中在模型设计上，常用的特征提取网络主要是卷积神经网络，如 DenseNet-121<sup>[8]</sup>和 ResNet-50<sup>[14][17][19]</sup>。然而，当 CNN 作为特征提取器时，一个不可避免的问题是，在借助辅助模型抑制服装干扰的同时，可能会因先验知识中的噪声而导致与服装有纠缠关系的人体体型等高级语义信息的较大损失。针对这一难点，本章提出了一种双流融合网络，分别包括局部特征流和高级语义流。在局部特征流中，使用 CNN 提取显著人体局部特征，如人脸；在高级语义流中，采用视觉注意力模型提取人体高级语义信息，如体型。为了降低服装的影响，在局部特征流中，基于解纠缠方法设计了双分支解纠缠模块，利用服装标签引导解纠缠；在高级语义流中，利用先验人体解析图像进行数据增强，模拟服装颜色变化，从而减少服装区域对人体特征的影响。

### 3.2 方法设计

为了解决行人服装颜色纹理变化影响识别检索效果，首先考虑获取丰富的人体局

部特征，在此基础上融合人体高级语义特征，能更好地应对实际换装任务。基于此，本章提出了基于服装信息的双流融合网络（Dual Stream Fusion Network, DSFN），分为局部特征流和高级语义流，在局部特征流，设计空间解纠缠模块（Spatial Decoupling Module, SDM）来对服装和身份信息进行对抗蒸馏从而实现解纠缠；在高级语义流，设计了数据增强模块（Data Augmentation Module, DE）从图像空间进行重构，利用直接的人体先验知识来模拟服装变化，将增强前后的图像经 ViT 获取其高级语义特征，比较前后的实例一致性去引导模型学习与服装纹理颜色无关的人体独特特征。最后经双流融合模块（Dual Stream Fusion Module, DSFM）将局部特征和高级语义特征进行融合。图 3-1 展示本章提出的双流融合网络的网络结构。

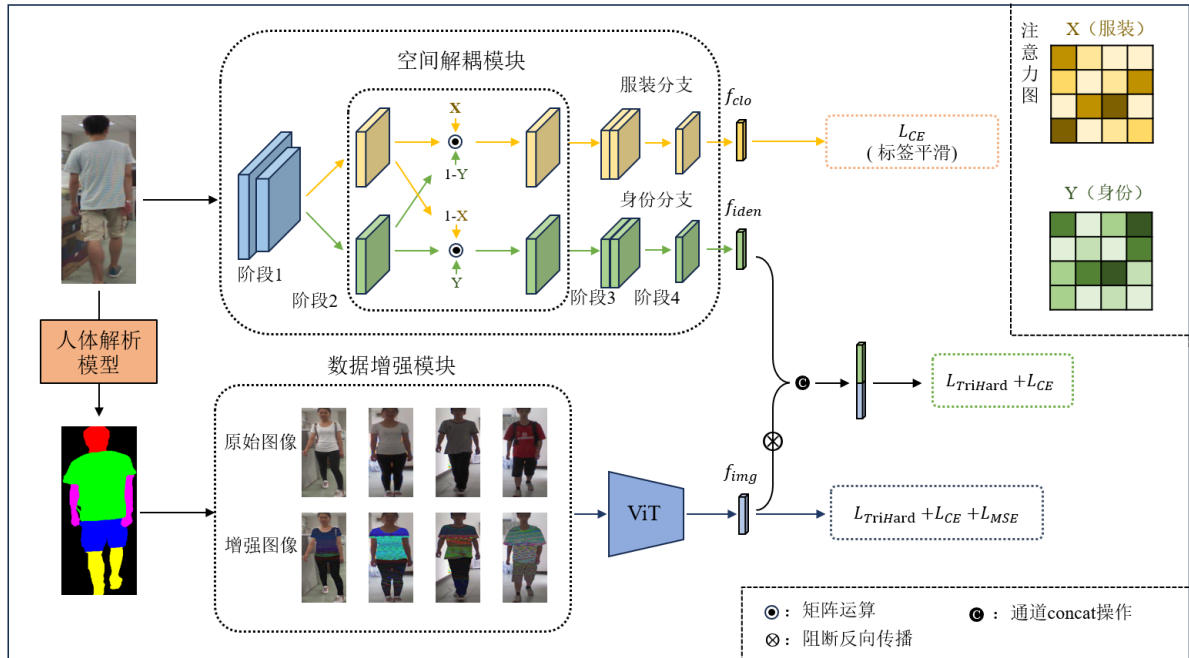


图 3-1 双流融合网络  
Figure 3-1 Dual Stream Fusion Network

### 3.2.1 空间解纠缠模块

在局部特征流中为获取丰富的人体特征，本章设计了一个双分支的服装-身份解纠缠网络，如图 3-2 所示。在传统 ResNet-50 基线模型的基础上增加一个服装分支，并设计空间解耦模块对服装分支和身份分支的特征在特征空间进一步解纠缠。在训练过程中，RGB 图像首先输入到一个全卷积网络中，之后通过阶段 1 的共享权重获取共享特征图，在阶段 2 之后进入双分支分别生成服装特征图  $m_{clo}$  和身份特征图  $m_{iden}$ ，而为了进一步对两者进行解纠缠，紧跟其后利用 SDM 将其在特征空间进行分离，通过注意力机制引导两种特征聚焦于不同的空间区域，聚焦过程如公式 3-1 所示。

$$m'_{iden} = SA_{iden} \otimes (1 - SA_{clo}) \otimes m_{iden} \quad (3-1)$$

$$m'_{clo} = SA_{clo} \otimes (1 - SA_{iden}) \otimes m_{clo} \quad (3-2)$$

其中  $SA_{iden}$  和  $SA_{clo}$  代表  $m_{iden}$  和  $m_{clo}$  分别通过一个独立的空间注意力模块后生成的与输入特征图空间尺寸相同的单通道注意力图。这个注意力图的每个位置的值表示该位置的重要性，值越高表示越重要。在四肢图像区域上服装信息和行人信息是互补的存在，由于两者的紧密相连，其特征往往存在很大程度的耦合，导致其关注各自的特征的同时，也会受彼此的影响。对此通过对抗蒸馏的方式来进行解纠缠，上述公式中将每种特征图乘以自己的注意力图和另外一种注意力图的逆图，通过关注彼此不关注的区域从而进一步增强自己的特征表示。

为了获得更好的效果，在阶段 2 和阶段 3 后均加入 SDM，从不同尺度上进行解纠缠，使不同类型的特征映射覆盖不同的语义信息。在得到两种特征图后利用平均池化模块对两种特征图进行处理，对空间信息进行聚合，得到输出向量  $f_{iden}$  和  $f_{clo}$ 。

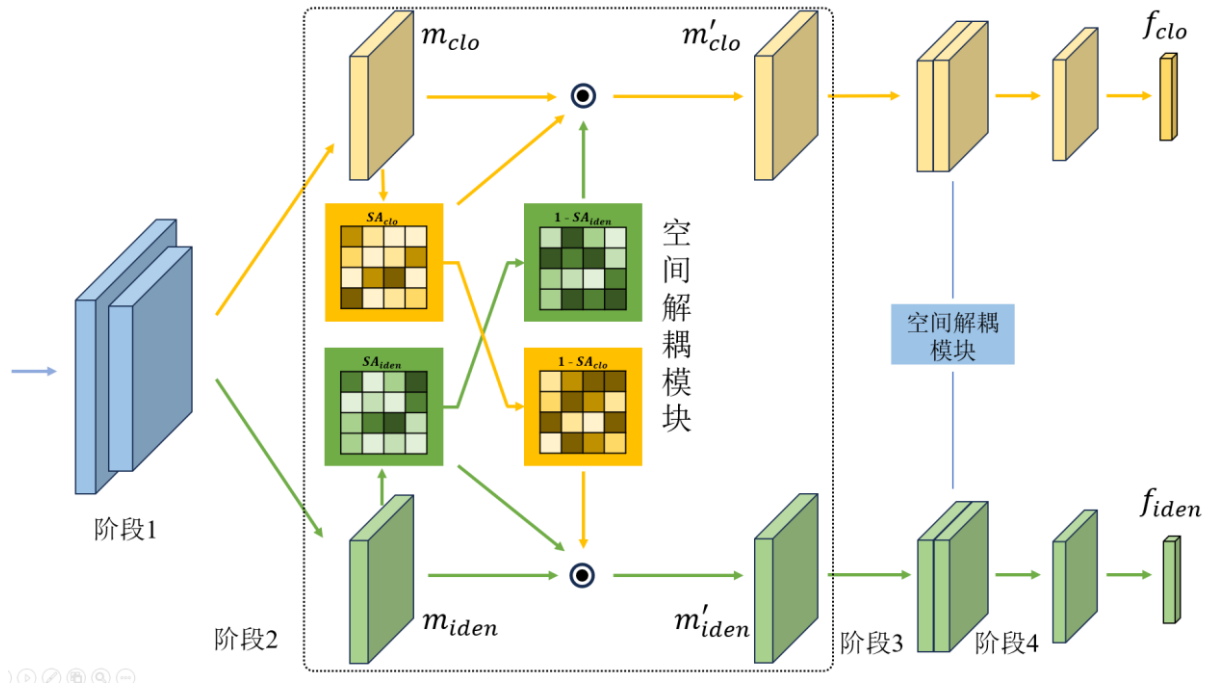


图 3-2 服装-身份解纠缠网络  
Figure 3-2 Clothing-Identity Disentanglement Network

### 3.2.2 数据增强模块

在行人重识别任务中，对图像进行数据增强是一种常见手段，在提高模型识别准确率的同时可以显著降低训练域的过拟合。常见的数据增强手段有随机裁剪<sup>[56]</sup>和随机擦除<sup>[58]</sup>等。随机裁剪通过从原始图像中随机选择一个区域进行裁剪，以增加数据集的

多样性，从而提高模型的泛化能力和鲁棒性。为了保证输入图像的规格一致，通常情况下需要先对图像进行 padding 操作，然后根据规格大小进行随机裁剪。随机擦除则是通过在训练过程中随机选择图像中的一个矩形区域，并用随机值擦除该区域的像素，从而生成具有不同遮挡级别的训练图像。这种方法可以减少过拟合的风险，并使模型对一定程度上的遮挡具有鲁棒性。

针对换装场景，则有一些独特的数据增强方法，比如利用先验模型对姿态特征、步态特征等进行增强。本章所采用的是利用先验分割模型获取的人体解析图像来进行数据增强。它通过将人体的上衣和裤子颜色纹理将其打乱，以降低服装变化对模型的干扰，如图 3-3 所示。

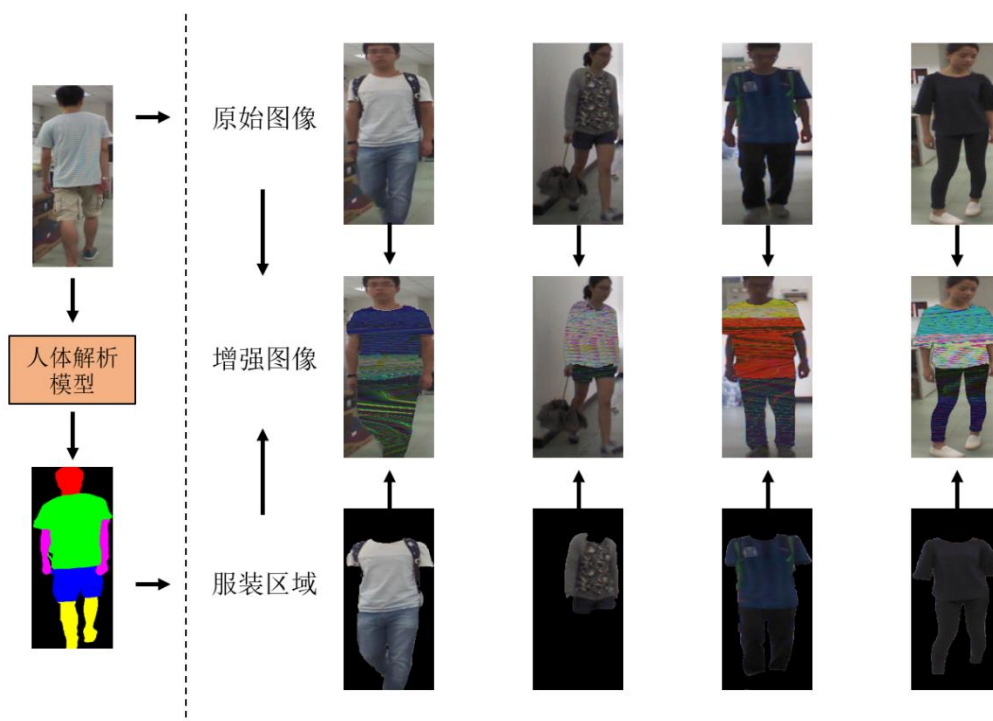


图 3-3 原始图像及增强后的图像  
Figure 3-3 Original Image and Enhanced Image

在换装场景中，将原始图像与打乱服装区域像素的图像输入网络，可以引导模型更关注服装不变的人体特征，减少服装的干扰。然而，由于辅助模型的训练场景与换装数据集的训练场景存在较大的不同，导致在分割时准确率难以保证，可能会误将人体其他区域识别为服装区域，从而在增强后的图像中引入明显的噪声，如图 3-4 所示。如果继续使用传统的 CNN 模型作为特征提取器，这部分噪声会通过卷积层的滑动窗口被不断放大，严重干扰模型的优化，降低方法的有效性。

针对这一问题，本章提出使用 ViT 模型代替 CNN 模型进行特征提取。ViT 模型的



核心在于自注意力机制，能够直接捕捉图像中不同区域之间的全局依赖关系，从而更好地关注人体的结构特征。此外，ViT 通过为图像中的每个区域分配注意力权重，减少对噪声区域的权重，以降低噪声区域对整体特征表示的影响。这种全局信息捕捉能力和对噪声的鲁棒性，使得 ViT 在处理换装场景中的复杂变化时具有明显优势，能够有效提高模型性能。



图 3-4 模型在部分图片存在的噪声干扰  
Figure 3-4 Noise interference in some images of the model

### 3.2.3 双流融合模块

本章的双流网络由基于 CNN 的双分支解纠缠网络和基于 ViT 的特征增强网络组成。与基于 CNN 的骨干网络相比，ViT 具有更强大的全局注意力机制，可以用来提取行人的高级语义特征。

具体来说，CNN 流在 stage4 阶段生成最终的行人特征表示，而 ViT 流在获取其特征表示之前，需要先将输入图像划分为若干个固定大小的图像块（patch）。接着，将每个图像块通过线性投影转换为一维向量，并将这些向量输入到标准的 Transformer 编码器中进行处理。最终，将 Transformer 编码器最后阶段的 CLS token 作为行人的特征表示。通过将 CNN 流和 ViT 流得到的两类特征进行通道拼接（concat），获得更为丰富和全面的行人特征表示，如公式 3-3 所示。

$$f_{person} = f_{iden} \odot f_{img} \quad (3-3)$$

其中 $\odot$ 表示将两种特征向量进行通道级拼接， $f_{iden}$ 和 $f_{img}$ 分别表示 CNN 流特征和 ViT 流特征，将其拼接形成最终的行人特征 $f_{person}$ 。

这种双流融合网络的设计充分利用了 CNN 和 ViT 各自的优势。CNN 在提取局部特征方面具有强大的能力，能够捕捉到丰富的行人局部特征，如人脸和局部纹理等，这些特征对于区分不同个体具有重要意义。而 ViT 则擅长捕捉全局特征，能够更好地理解行人的整体结构和语义信息，如体型和姿态等，这些信息在面对服装变化时具有较高的鲁棒性。通过将两种特征进行融合，不仅能够充分利用原始图像中的丰富信息，还能有效减少服装变化对行人识别的干扰，从而提高模型在复杂场景下的识别准确率。此外，这种融合方式还能够增强模型的泛化能力，使其在面对不同类型的服装变化时，依然能够准确地识别出行人的身份特征。

### 3.2.4 损失函数

在模型训练阶段，针对 CNN 流和 ViT 流分别设计了相应的损失优化策略。为了确保模型具备基础的身份识别能力，对两个流中的身份分类器均采用三元组损失和交叉熵损失进行联合监督。此外，为了加速模型的收敛过程，进一步引入了在线硬样本挖掘（TriHard Loss）机制，通过动态选择最具挑战性的样本对来优化模型的训练效果。

在 ViT 流中，为了进一步提取与人体相关的特征并减少服装区域的干扰，额外引入了均方差损失来对图像增强前后的特征进行一致性监督。损失函数如公式 3-4 所示。

$$L_{ViT} = L_{CE} + L_{TriHard} + L_{MSE} \quad (3-4)$$

其中  $L_{CE}$  和  $L_{TriHard}$  分别是交叉熵和三元组损失。 $L_{MSE}$  是均方差损失，这一损失函数确保了在数据增强过程中，模型能够学习到更鲁棒的特征表示，从而在保持人体特征的同时，降低服装变化对识别结果的影响。

在 CNN 流中，针对服装分支，采用了软标签交叉熵损失进行监督。对于身份分支，结合了交叉熵损失和难样本三元组损失对拼接后的特征进行联合监督。这种联合监督策略不仅能够确保模型学习到准确的身份特征，还能够通过难样本挖掘进一步提升模型的判别能力。损失函数如公式 3-5 所示。

$$L_{CNN} = L_{CE} + L_{TriHard} + L_{CE}^{label} \quad (3-5)$$

其中  $L_{CE}^{label}$  表示软标签交叉熵损失，这种损失函数通过引入软标签，能够有效减少错误服装标签对模型训练的误导，从而提高模型对服装特征的鲁棒性。

最终的总损失函数综合了上述各个损失项，通过加权求和的方式进行计算，具体的表达式如公式 3-6 所示。这一总损失函数的设计旨在平衡不同损失项对模型训练的贡献，确保模型在学习过程中能够同时优化多个关键目标，从而在服装变化的行人重识

别任务中取得更优的性能。

$$L_{Total} = L_{ViT} + L_{CNN} \quad (3-6)$$

### 3.3 实验设置

在本研究中，CNN 流采用 ResNet-50 作为特征提取器，并利用在 ImageNet 数据集上预训练的模型参数进行初始化。为了进一步丰富特征的粒度，将 ResNet-50 架构中最后一个卷积层的步幅设置为 1。在 ViT 流中，选用 ViT-Base 模型作为特征提取器，输入图像的尺寸被调整为  $256 \times 128$  像素，每个图像块的大小设定为  $16 \times 16$  像素，从而整个输入图像被划分为  $16 \times 8$  个图像块。

在 CNN 流中，为了提升模型的识别能力，采用两阶段训练策略。在训练的初始阶段，服装分支和身份分支分别依据各自特定的优化策略进行独立训练。这一过程旨在使模型初步聚焦于服装区域和身份区域，从而建立起基础的服装和身份识别能力，为后续的特征解纠缠工作奠定坚实基础。随后，通过引入空间解纠缠模块（SDM），模型进一步在空间维度上实现服装特征与身份特征的解耦。这种从简单到复杂的训练策略，使模型首先在较为基础的任务中学习关键特征，随后逐步过渡到更具挑战性的特征解纠缠任务。这一方法不仅有助于模型更精准地理解和区分服装与身份特征，还能显著增强其在实际应用中的鲁棒性和泛化能力。

在具体的实验设置中，整个训练过程共计 60 个 epoch。前 10 个 epoch 为第一阶段，专注于服装分支和身份分支的独立优化。后 50 个 epoch 为第二阶段，通过 SDM 模块来促进服装特征和身份特征在空间上的有效解耦。

### 3.4 对比实验

考虑到本文的主要研究方向是换装任务，因此实验主要在 PRCC<sup>[7]</sup>和 NKUP+<sup>[8]</sup>数据集上进行。PRCC 数据集提供了丰富的室内场景，其拍摄周期相对较短，主要涵盖了服装颜色和纹理的变化，是换装任务中常用的基准数据集之一，便于与其他方法进行公平的对比分析。NKUP+数据集则包含了更加真实的室外场景，具有较长的时间跨度，并且包含了各种程度的服装变化，这使得它能够进一步验证本文方法在真实复杂场景中的有效性和鲁棒性。在测试阶段，分别进行了相同服装识别（Same Cloth, SC）和跨服装识别（Cross Cloth, CC）两种场景的实验。

表 3-1 本章方法和其他算法的比较  
Table 3-1 Comparison of the methods in this chapter with other algorithms

方法	来源	PRCC				NKUP+			
		SC		CC		SC		CC	
		R1	mAP	R1	mAP	R1	mAP	R1	mAP
PCB <sup>[39]</sup>	ECCV 18	99.8	97.0	41.8	38.7	76.0	60.9	9.3	5.1
MGN <sup>[75]</sup>	MM 18	99.5	98.4	33.8	35.9	81.6	69.3	15.4	9.0
IANet <sup>[76]</sup>	CVPR 19	99.4	98.3	46.3	46.9	-	-	-	-
TransReID <sup>[20]</sup>	ICCV 21	99.6	98.7	44.6	46.5	81.1	69.6	19.7	10.9
SPT <sup>[7]</sup>	TPAMI 21	64.2	-	34.4	-	-	-	-	-
FSAM <sup>[18]</sup>	CVPR 21	98.8	-	54.5	-	-	-	-	-
GI-ReID <sup>[13]</sup>	CVPR 22	86.0	-	33.3	-	-	-	-	-
M2Net <sup>[8]</sup>	MM 22	<u>99.5</u>	<u>99.1</u>	<u>59.3</u>	<b>57.7</b>	<u>86.2</u>	<u>72.3</u>	<u>24.0</u>	<u>11.0</u>
baseline		99.4	98.2	45.9	46.9	81.0	70.8	16.6	9.5
DSFN(ours)		<b>99.8</b>	<b>99.2</b>	<b>61.6</b>	<u>56.9</u>	<b>90.7</b>	<b>83.0</b>	<b>26.3</b>	<b>12.3</b>

具体而言，本小节对所提出的方法与四类通用行人重识别方法（PCB<sup>[39]</sup>、MGN<sup>[75]</sup>、IANet<sup>[76]</sup>、TransReID<sup>[20]</sup>）以及四类专门针对换装场景的方法（SPT<sup>[7]</sup>、FSAM<sup>[18]</sup>、GI-ReID<sup>[13]</sup>、M2Net<sup>[8]</sup>）进行了全面比较，相关结果详细记录于表 3-1 中。在通用行人重识别领域，由于行人服装颜色通常保持不变，这四种通用方法能够借助稳定的服装特征有效地识别行人身份，因而在相同服装（SC）模式下取得了较高的识别精度。然而，在跨服装（CC）模式下，服装特征的可靠性大幅下降，容易导致模型误判，进而使得识别精度显著降低。

相比之下，专门针对换装场景的行人重识别方法通过降低服装特征的干扰，能够提取出更具鲁棒性的行人特征。这些方法不仅在 CC 模式下实现了精度的显著提升，而且在 SC 模式下依然保持了较高的性能表现。在这些换装方法中，主要与多模态方法进行对比。这些方法通过引入额外的辅助模态信息，在面对服装变化时取得了显著的性能提升。例如，其中 SPT<sup>[7]</sup> 方法仅使用人体轮廓图提取轮廓特征进行识别，而 FSAM<sup>[18]</sup>和 M2Net<sup>[8]</sup> 方法则在 RGB 图像的基础上，进一步融合了两种不同的模态信息，从而增强了模型对服装变化的适应能力。

本章提出的方法通过利用先验分割模型的先验知识进行数据增强，并融合人体局部特征与高级语义信息，性能较之前的多模态方法有明显的提升。在 CC 模式下，针对

PRCC 数据集，本文方法 DSFN 在 R1 和 mAP 两种指标上的表现较基线分别提升了 15.7%和 10.0%，并且相较于次优方法 M2Net 在 R1 指标上提升了 2.3%。对于 NKUP+ 数据集，DSFN 方法相较于基线模型在 R1 和 mAP 两种指标上的表现分别提高了 9.7%和 2.8%，同时也比次优的 M2Net 方法分别提高了 2.3%和 1.3%。这些结果表明，即使在不依赖额外更多模态信息的情况下，本章提出的方法依然能够有效地应对换装场景中的挑战，显著提高行人重识别的准确率。

表 3-2 不同模块在 PRCC 上的消融实验  
Table 3-2 Ablation experiments of different modules on PRCC

方法	CNN		ViT			PRCC	
	$f_{CNN}$	SDM	$f_{ViT}$	DE	$L_{MSE}$	R1	mAP
1	✓	×	×	×	×	45.9	46.9
2	✓	✓	×	×	×	52.0	51.7
3	×	×	✓	×	×	44.7	45.8
4	×	×	✓	✓	×	57.1	53.1
5	×	×	✓	✓	✓	58.9	55.7
6	✓	✓	✓	✓	✓	61.6	56.9

### 3.5 消融实验

#### 3.5.1 各个模块的有效性

在本小节中，对本章提出的各个模块在 PRCC 数据集上进行了详细的消融实验，并展示了 CC 模式下的 R1 和 mAP 指标数据，具体结果如表 3-2 所示。本章提出的方法主要由两大模块构成：CNN 流和 ViT 流。

在 CNN 流中，沿用之前的基线模型，选用 ResNet-50 作为特征提取器，并采用交叉熵损失和三元组损失进行监督。基线模型在 R1 和 mAP 指标上分别取得了 45.9%和 46.9%的效果。在此基础上，引入 SDM 模块，进一步对特征进行解纠缠，从而组成了本章方法的 CNN 流。通过 SDM 模块的加入，模型能够更有效地分离服装特征和身份特征，使得 CNN 流在 R1 和 mAP 指标上分别提高了 6.1%和 4.8%。

在 ViT 流中，选用 ViT-Base 作为特征提取器，并引入数据增强模块，有效扩充了训练域，丰富了行人穿着不同服装的场景。这种数据驱动的方法能够引导模型学习到人体的独特特征，从而在 R1 和 mAP 指标上分别实现了 12.4%和 7.3%的显著提升。为

了进一步增强模型对服装变化的适应能力，本章还加入了均方差损失，以约束图像增强前后的特征一致性。这一改进使得 ViT 流在 R1 和 mAP 指标上分别进一步提高了 1.8% 和 2.6%。

最后，将 CNN 流和 ViT 流的特征进行融合，充分发挥了两个模块的优势。融合后的模型在 R1 和 mAP 指标上取得了 61.6% 和 56.9% 的优越效果，充分证明了本章提出方法的有效性。通过这些消融实验，可以清晰地看到各个模块对模型性能提升的贡献，以及双流融合策略在换装行人重识别任务中的显著优势。

**表 3-3 不同特征提取器在 PRCC 上的消融实验**  
**Table 3-3 Ablation experiments of different feature extractors on PRCC**

方法	高级语义流特征提取器	PRCC	
		R1	mAP
高级语义流	CNN	57.8	54.9
	ViT	58.9	55.7
高级语义流+局部特征流	CNN	57.9	54.8
	ViT	61.6	56.9

### 3.5.2 ViT 模型的有效性

针对先验分割模型在换装任务中存在明显噪声问题，本节对单独的高级语义流方法进行了实验，以验证不同特征提取器的性能优劣，实验结果如表 3-3 所示。通过在 PRCC 数据集上对不同特征提取器进行消融实验，可以清晰地看到，在高级语义流中，采用 CNN 作为特征提取器时，R1 和 mAP 指标分别达到了 57.8% 和 54.9%；而采用 ViT 作为特征提取器时，这两个指标分别提升至 58.9% 和 55.7%。这一结果表明，相较于 CNN，ViT 在高级语义流中表现出了明显的性能优势，R1 和 mAP 指标分别提升了 1.1% 和 0.8%。这不仅验证了之前提出的用 ViT 模型代替 CNN 模型以减少噪声干扰的构想，还突显了 ViT 在处理高级语义信息时的优越性。

ViT 模型凭借其强大的全局感知能力，能够自适应地建模全局依赖关系，从而在降低噪声干扰的同时学习到更加鲁棒的高级语义特征。基于这一优势，将不同特征提取器的高级语义流与局部特征流进行融合，并进一步进行实验。实验结果表明，将 ViT 模型学习到的特征与局部特征融合后，两种特征互为补充，能够进一步提升识别性能。相比之下，将 CNN 模型学习到的特征与局部特征融合后，性能提升并不明显。这一发

现从侧面验证了 ViT 模型能够补充 CNN 模型所缺乏的高级语义信息，而将两者进行融合则能有效提升识别精度，为换装行人重识别任务提供了更强大的特征表示。

**表 3-4 不同方法在各阶段所需辅助模型数量**  
**Table 3-4 The number of auxiliary models required for different methods in each stage**

方法	辅助模型数量		PRCC	
	训练阶段	测试阶段	R1	mAP
M2Net <sup>[8]</sup>	2	2	59.3	57.7
DSFN(ours)	1	0	61.6	56.9

### 3.5.3 多模态方法分析

如表 3-4 所示，多模态方法 M2Net<sup>[8]</sup>通过两种辅助模型获取人体轮廓图像特征和人体解析图像特征来补充 RGB 图像特征，从而取得了较为先进的效果。然而，这种基于特征融合的方法在测试阶段仍需依赖两种辅助模型来提取特征，这不仅限制了其泛化性能，还使其容易受到辅助模型计算质量的影响，难以保证模型在不同场景下的稳定性能。

相比之下，本文提出的基于数据增强的方法仅需在训练阶段借助辅助模型，而在测试阶段则完全依赖双流融合网络来提取特征，从而摆脱了对辅助模型的依赖。这种方法不仅更适合实际场景的应用，而且在识别效果上也展现出了明显的优势。通过数据增强策略，模型能够在训练过程中学习到更加鲁棒的特征表示，从而在测试阶段无需辅助模型的情况下，依然能够实现高精度的行人重识别。

## 3.6 本章小结

本章提出一种基于图像分割的双流融合网络。从模型设计出发，设计双流融合网络将 CNN 特征与 ViT 特征进行融合，兼顾人体高级语义信息和局部特征。利用先验模型提供的先验知识设计了数据增强模块，为避免先验模型噪声的干扰，仅在 ViT 分支进行数据增强，在利用先验知识的同时，最大限度的减少其噪声干扰对模型的影响。在 CNN 分支获取局部特征的同时，利用服装标签并设计空间解纠缠模块来进一步降低服装区域的影响。实验结果表明，所提出的双流融合网络在现有数据集上效果显著。

## 第4章 基于主动学习的换装行人重识别方法

### 4.1 引言

目前的换装行人重识别方法根据是否借助辅助模型可分为两大类：多模态方法和单模态方法。多模态方法利用辅助模型提取额外特征，如轮廓草图<sup>[7][8]</sup>、人体解析图<sup>[9][10]</sup>、人体姿势<sup>[11][12]</sup>和步态<sup>[13]</sup>。这些特征有助于模型在服装发生变化时做出更准确的判断，因为它们使用了除 RGB 图像之外的信息。然而，这些方法不可避免存在一些局限性。首先，它们并没有充分利用原始 RGB 图像中与服装无关的固有身份信息，而这些信息对于准确识别行人身份至关重要。其次，这些方法的性能可能受到辅助模型计算质量的影响，如果辅助模型的输出不准确，那么最终的重识别结果也可能受到影响。单模态方法则尝试通过对抗学习或因果蒸馏等技术，直接从 RGB 图像中提取与服装无关的特征。但这种方法也面临其他挑战，即服装标签与身份之间存在错误的关联。在现实生活中，每个人的服装组合往往和他们的个性一样独特，一个人可以穿多套不同的衣服，但很少有人会穿同样的衣服。因此，即使不同的人穿着相同的服装，他们也会被错误地贴上不同的服装标签，这会导致模型混淆不同身份。



图 4-1 由服装变化导致的两类识别失败场景

Figure 4-1 Two types of recognition failure scenarios caused by clothing changes

针对换装情况，本文总结了两种主要影响模型准确率的样本数据，如图 4-1 所示。当模型同时面对同一身份的服装变换以及不同身份穿着相似服装的情境时，很难准确识别出穿着相似服装的不同身份。对此，很多研究者从解纠缠角度设计方法，借助辅助模型或者服装标签以数据驱动方式将服装与身份之间的关系进行解纠缠。通过在训练过程中模拟服装变化来训练判别型模型，从而引导模型加强对这两种情况的辨别能



力，以获得更鲁棒的行人表示。例如，Shu 等<sup>[9]</sup>提出在图像空间进行重构，借助先验的人体解析模型模拟服装变化，在小批量样本中保证相同身份的服装多样性。而 Han 等<sup>[17]</sup>则认为在特征空间进行重构更为有效，他们利用服装标签模拟服装变化，并在小批量样本中同时保证相同身份的服装多样性和相似服装的身份多样性。这些方法为提高换装重识别的准确率提供了新的思路和解决方案。

但之前的解纠缠方法主要聚焦于模型设计层面，在一定程度上受到了辅助模型或服装标签的制约。对此，本文转而研究适合换装任务的训练策略，在不依靠辅助模型和错误服装标签的情境下进行解纠缠。借鉴主动学习<sup>[18][77]</sup>的理念，本文认为向模型输入一批具有更富有挑战性的数据，能够助力模型实现快速且高质量的训练成果。而在基于深度学习的 Re-ID 任务中，模型通常借助采样器选取小批量样本进行训练。常用的 PK 采样器<sup>[77]</sup>通过随机挑选  $P$  个不同身份的个体，并为每个个体随机抽取  $K$  张图像，以此构成模型训练的小批量样本。然而，PK 采样器所选取的样本往往缺乏针对性，很难为判别型模型提供真正有价值的数据。

针对 CC Re-ID 任务的特殊需求并结合两类主要失败现象，本文认为在采样阶段设计新的采样器，以确保相同身份的服装多样性和相似服装的身份多样性，在应对服装变化时是切实有效的。为此，本文创新性地提出了一种基于主动学习的服装采样器（Clothing Sampler, CS）。该采样器具备精准的样本挖掘能力，既能精准定位具有相似服装特征的不同身份图像，又能发现同一身份下着装各异的图像样本。从数据驱动的角度，模型将主动深入挖掘更为鲁棒的身份特征，同时摒弃那些可能干扰识别效果的服装特征。通过引入这种新型采样器，模型训练过程将更加高效，最终的识别性能也将得到显著提升。

从人脸识别损失函数<sup>[1]</sup>的研究中可以看出，强调难样本有助于提高模型的判别能力。通过采用 CS 方法，可以确保小批量样本中服装的多样性，对此我们专门设计了一个多实例一致性损失，通过将相同服装负样本与锚点靠近，限制了不同服装正样本的分布，进一步指导模型学习与服装无关的特征。

## 4.2 方法设计

在采样阶段，摒弃了传统的 PK 采样器，转而采用创新的 CS 采样器来获取用于训练的小批量样本。与 PK 采样器不同，CS 采样器在构建小批量样本时，特别注重服装的多样性。具体而言，它不仅确保了相同身份个体的服装多样性，即为每个身份选取

多张不同服装的图像，还兼顾了相似服装的身份多样性，即选取多个身份中穿着相似服装的图像。这种采样策略使得模型在训练过程中能够接触到更为丰富和具有代表性的服装变化情况，从而有助于提升模型对服装变化的适应能力和身份识别的准确性。整体模型方法的框架如图 4-2 所示，清晰地展示出 CS 采样器在整个模型训练流程中的关键作用和与其他模块的协同关系。

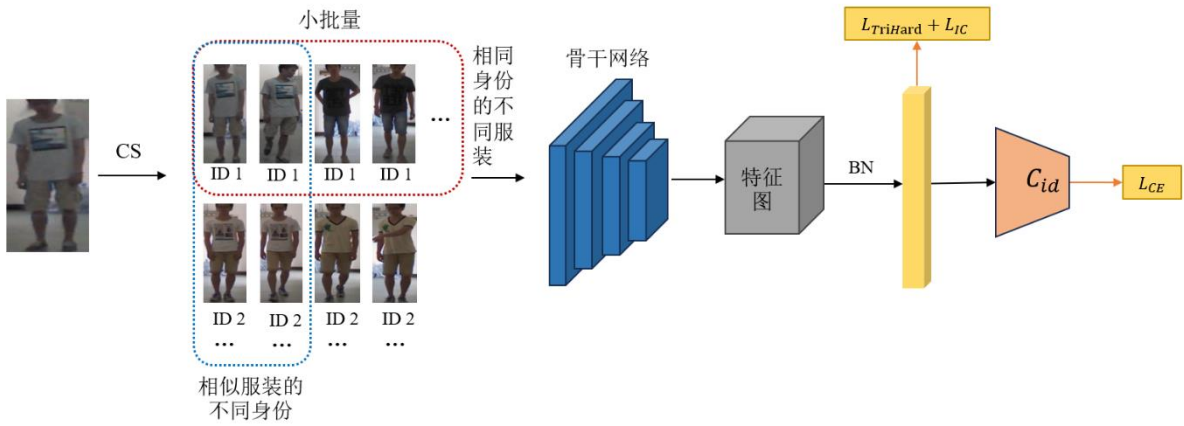


图 4-2 模型方法图

Figure 4-2 Model Method Diagram

#### 4.2.1 CS 采样器

在行人重识别任务中，常用的 PK 采样器主要用于度量学习以选择批量样本。然而，由于其随机性，PK 采样器无法在小批量样本中提供丰富的有价值 and 具有挑战性的例子，更不用说为判别学习提供足够的信息。因此，本文提出了一种基于主动学习的新采样器，称为服装采样器 CS，确保小批量样本中的“具有相同身份的服装多样性”和“具有相同服装的身份多样性”。这种采样策略迫使模型以数据驱动的方式将服装特征和身份特征进行解纠缠，从而获得鲁棒的服装无关特征。上述描述的两种采样器的采样结果如图 4-3 所示。

其中不同的颜色表示不同的批量，不同的形状表示不同身份。不同的数字表示不同的服装。PK 采样器随机选取 P 个人，每个人随机选取 K 张图像。CS 采样器主动选择具有相似服装的 P 个人并且每个人统一选择具有不同服装的 K 张图像。

受随机外观采样器<sup>[8]</sup> (Random Appearance Sampler, RAS) 的启发，依据服装标签能够找出相同身份的不同服装图像，因此在选择的小批量样本中相同服装的身份多样性可以得到保障。然而，获取不同身份间相似服装的图像却面临较大挑战。为攻克这一难题，本文借助一种客观规律（相似服装图像在特征空间中同样保持其相似性）来

设计方法。基于上述规律，通过分析输入图像在当前训练网络中的特征相似性来判断它们是否为相似服装图像。这一方法有效地解决了在采样过程中难以直接获取不同身份相似服装图像的问题，为模型提供了更为丰富和具有挑战性的训练样本，有助于提升模型对服装变化的适应能力和身份识别的准确性。

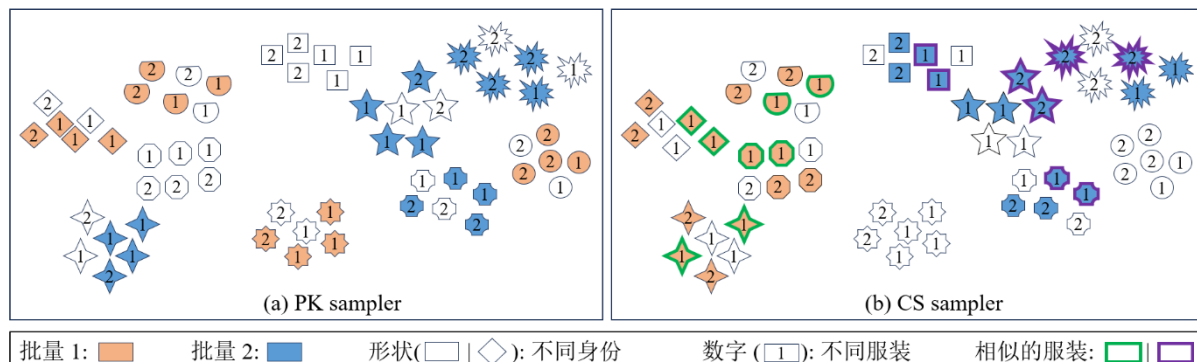


图 4-3 PK 采样器和 CS 采样器  
Figure 4-3 PK sampler and CS sampler

具体来说，CS 采样器算法的实现分为两个阶段：

第一阶段是在每个 epoch 开始时构建特征相似性矩阵，并过滤保留最近邻关系，具体步骤如下：

(1) 从每个身份的图像集中随机挑选一张图片，构建一个身份库，为后续的特征相似性分析奠定基础；

(2) 利用当前阶段训练有素的模型，为身份库中的每张图像精准提取特征，获取其在特征空间中的表示；

(3) 计算身份库中各特征之间的相似度，精心构建起相似度矩阵，并对每一行按降序排列，以便后续筛选出最具相似性的身份组合。

第二阶段是为每个小批量精心挑选  $P$  个穿着相似的身份，每个身份选取  $A$  类服装，每类服装再挑选  $N$  张图像：

(4) 依据相似度矩阵，为每个身份精准定位与其最相似的  $P-1$  个身份，巧妙构成小批量中的  $P$  个身份集合，如图 4-3 (b) 中绿色框所示的身份组合；

(5) 对于每个身份，精心挑选  $A$  类服装，确保服装的多样性和代表性，例如图 4-2 (b) 中选取的 2 类服装；

(6) 在每类服装中，细致挑选  $N$  张图像，进一步丰富小批量样本的服装变化情况，如图 4-3 (b) 中为每类服装选取的 2 张图像。

最终，在采样阶段成功获得了与数据集中身份数量相匹配的小批量样本，为模型训练提供了高质量且具有挑战性的数据支持。

#### 4.2.2 损失函数

为了赋予模型精准提取人物特征的能力，采用交叉熵损失和难样本三元组损失对模型进行联合训练，具体公式如公式 4-1 所示。交叉熵损失负责衡量模型预测标签与真实标签之间的差异，促使模型准确地对人物身份进行分类；而难样本三元组损失则专注于优化特征的度量学习，通过拉近同一身份样本特征的距离，同时推远不同身份样本特征的距离，进一步强化模型对人物特征的区分度。这两种损失函数的协同作用，使模型在训练过程中能够综合考虑分类准确性和特征相似性，从而有效提升识别的性能。

$$L_{ID} = L_{CE} + L_{TriHard} \quad (4-1)$$

其中  $L_{CE}$  和  $L_{TriHard}$  分别是交叉熵和三元组损失。值得注意的是，在采样过程中，我们发现不同的人可能会穿着相似的衣服，这给特征提取带来了挑战。为了解决这一问题，本章在训练中添加了在线困难样本挖掘技术。通过在线挖掘难分的正负样本对，扩大类间距离并减少类内距离，模型能够更有效地提取与衣服无关的特征，从而提升人物重识别的准确性和鲁棒性。

为了缓解因个人换装导致的特征变化显著差异问题，精心设计了一种多实例一致性损失 IC Loss，以进一步抑制服装特征的学习。该损失由三元组损失的变体 MA Loss 和 MC Loss 共同构成。在 MA Loss 中，针对锚点样本，选取与锚点样本身份相同、服装也相同的最远样本作为负样本，而将与锚点样本身份相同但服装不同的最远样本视为正样本，具体计算公式如公式 4-2 所示。

$$L_{MA} = \sum_{p=1}^P \sum_{a=1}^A \sum_{n=1}^N \sum_{b=1, b \neq a}^A [\beta + \max_{n'=1, \dots, N'} \text{Dist}(f(x_a^n), f(x_b^{n'})) - \max_{n'=1, \dots, N, n' \neq n} \text{Dist}(f(x_a^n), f(x_a^{n'}))] \quad (4-2)$$

其中  $f(x_a^n)$  表示在一个小批量样本中第  $n$  个行人着装为  $a$  时图像  $x_a^n$  ( $1 \leq a \leq A, 1 \leq n \leq N$ ) 的特征， $f(x_a^{n''})$  代表同一个人相同衣服的其他图像， $f(x_b^{n'})$  表示同一个人的其他服装图像， $\text{Dist}$  表示距离度量函数， $\beta$  是一个设置为 0 的超参数。通过将锚点到这些正样本的距离限制为小于到负样本的距离，网络得以学习到交叉服装图像的鲁棒特征。然而，若仅依赖差异最大的不同服装样本作为正样本，则模型可能过度关注由姿势、角

度和服装款式引起的特征差异，而忽视服装颜色和纹理的变化。为此，本章设计了 MC Loss 与 MA Loss 协同使用，公式如 4-3 所示。

$$L_{MC} = \sum_{p=1}^P \sum_{a=1}^A \sum_{n=1}^N \sum_{b=1, b \neq a}^A [\beta + \min_{n'=1, \dots, N'} \text{Dist}(f(x_a^n), f(x_b^{n'})) - \max_{n^*=1, \dots, N, n^* \neq n} \text{Dist}(f(x_a^n), f(x_a^{n^*}))]_+ \quad (4-3)$$

在 MC Loss 中，采用最近的不同服装样本作为正样本，其中特征差异更多源自服装颜色纹理变化，从而有效解决了 MA Loss 在处理服装变化特征时的不鲁棒问题。

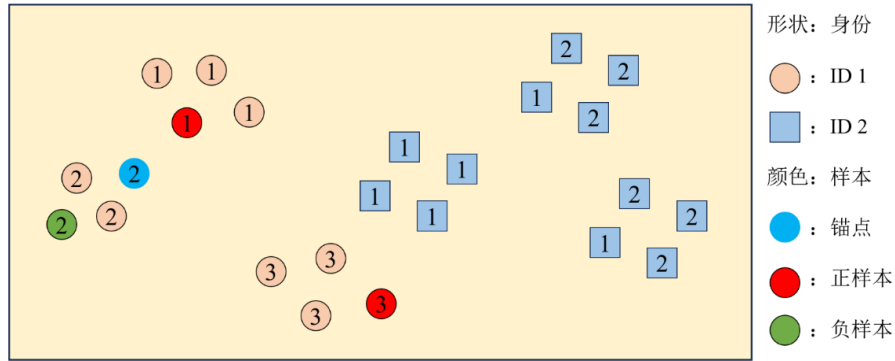


图 4-4 多服装实例一致性损失的可视化

Figure 4-4 Visualization of Consistency Loss in Multiple Clothing Instances

通过结合 MA Loss 和 MC Loss，本章成功地使具有相同身份的不同服装图像的特征更加接近，显著提升了模型在面对换装情况时的识别性能和鲁棒性。最终的实例一致性损失 IC Loss 如公式 4-4 所示。

$$L_{IC} = L_{MA} + L_{MC} \quad (4-4)$$

在图 4-4 中，在特征空间将损失进行可视化，在面对不同身份服装相似的情况下，多服装实例一致性损失会进一步引导模型学习更鲁棒的行人表征，保证类内特征的紧凑性。最后，总损失函数计算如公式 4-5 所示：

$$L_{Total} = L_{ID} + L_{IC} \quad (4-5)$$

### 4.3 实验设置

沿用传统的基线模型架构，选用 ResNet-50 作为骨干网络，并引入在 ImageNet 数据集上预训练的模型参数。为了统一输入图像的尺寸，将输入图像调整为 256×128 像素。在数据预处理阶段，采用了随机水平翻转、随机裁剪和随机擦除等数据增强技术，

以增强模型对图像变化的鲁棒性，防止过拟合现象的发生。为了获取更为细致的特征，将最后一个卷积层的 stride 参数设定为 1，从而丰富特征粒度。在特征提取环节，同时运用全局平均池化和全局最大池化两种策略，从不同角度捕捉图像的关键信息。随后，将这两种池化得到的特征进行连接操作，整合不同池化策略的优势。为了进一步优化特征的分布，使用 BatchNorm 对图像特征进行归一化处理，使其具有更稳定的分布特性，然后将处理后的特征输入到全连接层进行分类任务。在训练过程中，采用交叉熵损失和三元组损失进行监督，以确保模型能够准确地学习到区分不同身份的特征表示。同时，为了高效地优化模型参数，选择 Adam 优化器来进行梯度下降优化，实现模型的快速收敛和性能提升。

在 PRCC<sup>[7]</sup>数据集中，CS 各参数设置为，P=8，A=2，N=4，学习率初始化为  $3.5e-4$ ，每 30 个 epoch 后除以 10，共训练 90 个 epoch。在 NKUP+数据集中，P=8，A=4，N=3，学习率初始化为  $3e-4$ ，每 30 个 epoch 后除以 10，共训练 90 个 epoch。

表 4-1 CS 和其他算法在 PRCC 和 NKUP+数据集上的比较  
Table 4-1 Comparison of CS and other algorithms on PRCC and NKUP+datasets

方法	来源	大小	类型		PRCC				NKUP+			
					SC		CC		SC		CC	
			RGB	Other	R1	mAP	R1	mAP	R1	mAP	R1	mAP
PCB <sup>[39]</sup>	ECCV 18	384×192	●	-	99.8	97.0	41.8	38.7	76.0	60.9	9.3	5.1
MGN <sup>[75]</sup>	MM 18	-	●	-	99.5	98.4	33.8	35.9	81.6	69.3	15.4	9.0
IANet <sup>[76]</sup>	CVPR 19	384×192	●	-	99.4	98.3	46.3	46.9	-	-	-	-
TransReID <sub>[20]</sub>	ICCV 21	256×128	●	-	99.6	98.7	44.6	46.5	81.1	69.6	19.7	10.9
SPT <sup>[7]</sup>	TPAMI 21	256×128	-	ske	64.2	-	34.4*	-	-	-	-	-
FSAM <sup>[18]</sup>	CVPR 21	256×128	●	sli+pose	98.8	-	54.5*	-	-	-	-	-
GI-ReID <sup>[13]</sup>	CVPR 22	256×128	●	gait	86.0	-	33.3	-	-	-	-	-
M2Net <sup>[8]</sup>	MM 22	256×128	●	ske+par	99.5	99.1	59.3	57.7	86.2	72.3	24.0	11.0
CAL <sup>[14]</sup>	CVPR 22	384×192	●	-	<b>100</b>	<u>99.8</u>	55.2	55.8	<u>99.7*</u>	<b>92.4*</b>	<u>24.8*</u>	<u>13.2</u>
AIM <sup>[19]</sup>	CVPR 23	384×192	●	-	<b>100</b>	<b>99.9</b>	57.9	58.3	-	-	-	-
CCFA <sup>[17]</sup>	CVPR 23	384×128	●	-	99.6	98.7	<u>61.2</u>	<u>58.4</u>	-	-	-	-
baseline		384×192	●	-	99.5	97.0	57.0	52.3	99.7	80.8	26.8	10.9
CS(ours)		384×192	●	-	<u>99.9</u>	99.2	<b>64.8</b>	<b>61.5</b>	<b>99.9</b>	<u>88.6</u>	<b>29.0</b>	<b>13.3</b>

## 4.4 对比实验

为验证本章方法的可行性，本小节继续在 PRCC<sup>[7]</sup>和 NKUP+<sup>[8]</sup>数据集上进行对比实验。在上章对比方法的基础上，加入三种先进的单模态解纠缠方法（CAL<sup>[14]</sup>, AIM<sup>[19]</sup>, CCFA<sup>[17]</sup>）进行对比验证 CS 方法的先进性，如表 4-1 所示。近年来，采用只用 RGB 图像进行解纠缠学习的方法展示出很好的效果，如 CAL 通过惩罚模型对于服装的识别能力来从原始 RGB 图像中挖掘衣服无关特征帮助进行识别，AIM 使用因果蒸馏来消除服装干扰的同时提取更多利于区分不同行人的特征信息，CCFA 则是在特征空间进行数据增强，模拟人的服装颜色和纹理变化，从而减少模型对服装信息的依赖。但这些方法由于服装标签不正确而在服装之间产生了新的关联，身份识别从而会受到影响。

CS 方法在各项评估指标上均优于所有比较方法，取得了最先进的成果。基线模型使用 RAS 采样器，并通过 ID 损失和 MA 损失的联合监督进行训练，在 PRCC 数据集中表现出优秀的测试结果，CC 模式下 R1 和 mAP 可以达到 57.0%和 52.3%。本章 CS 方法采用 CS 采样器，并通过 ID 损失、MA 损失和 MC 损失的联合监督进行训练。在 CC 模式下，针对 PRCC 数据集进行测试，CS 方法在 R1 和 mAP 两种指标上的表现较基线分别提升了 7.8%和 9.2%，并且相较于次优方法 CCFA 分别提高了 3.6%和 3.1%。对于 NKUP+数据集，CS 方法相较于基线模型在 R1 和 mAP 两种指标上的表现分别提高了 2.2%和 2.4%，同时也比次优的 CAL 方法分别提高了 5.2%和 0.1%。实验结果充分表明，本章所提出的 CS 方法在采样阶段能够有针对性地提供在换装情况下更有价值的数据，引导模型学习到更为鲁棒的行人表示。在面对同一身份的服装变化以及不同身份的服装相似情况时，CS 方法能够实现更准确的识别，显著提升了行人重识别的性能和鲁棒性。

表 4-2 PRCC 上不同采样器的消融实验  
Table 4-2 Ablation experiments of different samplers on PRCC

Method	Sampler			CC	
	PK	RAS	CS	R1	mAP
1	✓	×	×	49.7	52.0
2	×	✓	×	53.8	53.1
3	×	×	✓	<b>54.2</b>	<b>54.5</b>

## 4.5 消融实验

### 4.5.1 CS 采样器的有效性

在 ID 损失的指导下，本章复现了 PK、RAS 和 CS 三种采样器在 PRCC 数据集上的实际性能，具体结果如表 4-2 所示。通过对比分析，可以清晰地观察到不同采样器在该数据集上的表现差异。PK 采样器作为传统的基线方法，在处理换装情况时存在一定的局限性，难以充分挖掘与服装无关的身份特征。RAS 采样器虽然在一定程度上考虑了服装的多样性，但其采样策略仍不够精细，无法完全适应复杂的换装场景。在 CC 模式下，本文提出的 CS 方法在性能上优于其他常用的采样技术。与传统的 PK 采样相比，CS 方法在 R1 准确率上实现了 4.5% 的提升，在 mAP 指标上也提高了 2.5%。

这些结果充分表明，CS 方法运用主动学习的理念，在构建批量样本时，精心挑选出既具有相似服装但身份不同的样本，又包含相同身份却着装各异的样本。这种策略使得模型能够更高效地学习到与服装无关的特征，从而在面对行人服装变化的检索任务时，展现出更强的鲁棒性和准确性。CS 方法的这一优势，对于提升行人重识别系统在复杂场景下的实际应用效果具有重要意义。利用主动学习的概念，在一个批量中提供具有相似服装但不同身份的样本，以及具有相同身份但不同服装的样本，帮助模型高效地学习与服装无关的特征，这对于检索服装变化的行人证明是更有帮助的。

表 4-3 PRCC 上损失函数的消融实验  
Table 4-3 Ablation experiment of Loss Function on PRCC

Method	Loss			SC		CC	
	ID	MA	MC	R1	mAP	R1	mAP
1	✓	×	×	100	99.7	54.2	54.5
2	✓	✓	×	99.8	99.0	58.7	57.3
3	✓	×	✓	99.9	99.5	55.5	55.0
4	✓	✓	✓	99.9	99.2	64.8	61.5

### 4.5.2 损失函数的有效性

表 4-3 展示了在 PRCC 数据集上对各个损失函数进行的消融研究。根据在 PRCC 数据集的实验结果可以看到，加入 MA 损失时，模型性能显著提升，CC 模式下 R1 和 mAP 指标分别提升了 4.5% 和 2.8%。这表明 MA 损失能够有效提取与服装无关的特征，减少类内距离，从而提高模型的准确性。然而，仅依赖 MA 损失时，模型可能过度关



注由姿势、角度和服装款式引起的特征差异，而忽视服装颜色和纹理的变化，这在处理服装变化特征时可能导致性能下降。

为了弥补这一不足，本章引入了 MC 损失与 MA 损失协同使用。在 MA 损失基础上，R1 和 mAP 指标又提升了 6.1%%和 4.2%。在 MC 损失中，采用最近的不同服装样本作为正样本，这些样本的特征差异更多源自服装颜色变化。通过这种方式，MC 损失能够有效解决 MA 损失在处理服装变化特征时的不鲁棒问题，进一步提升模型对服装变化的适应能力。

上述研究结果表明，通过结合两种损失，模型能够更全面地学习人体特征，有效应对服装变化，从而在行人重识别任务中实现更高的鲁棒性和准确性。

表 4-4 RAS 和 CS 策略在 PRCC 上的实验  
Table 4-4 Experimental study of RAS and CS strategies on PRCC

Loss	Sampling Strategy	SC		CC	
		R1	mAP	R1	mAP
MA	RAS	99.5	97.0	57.0	52.3
	CS	99.8	99.0	58.7	57.3
MA+MC	RAS	99.8	99.0	59.8	58.5
	CS	99.9	99.2	64.8	61.5

#### 4.5.3 RAS 采样器跟 CS 采样器的比较

本小节对三元组损失的有效性进行了相关实验，不同采样策略在采用不同三元组损失进行训练的实验结果如表 4-4 所示。在使用 PK 采样策略选择批量样本时，尽管三元组损失可以通过在线困难样本挖掘来获取不同身份的难样本图像，但由于选择身份的随机性，批量中难以保证存在相似服装的不同身份图像，且对同一身份图像选取时难以提供此身份的多种服装图像。之前的 RAS 采样方法在此基础上对图像选择加以干预，保证批量中包含同一身份的不同服装图像，从而更好地发挥三元组损失的性能。而本文提出的 CS 方法能够进一步从数据集中挖掘相似服装的不同身份图像供模型训练。从表 4-4 的数据可以观察到，在 CC 模式下，仅使用 MA 损失对模型进行训练时，CS 方法与 RAS 方法相比，R1 和 mAP 指标分别提高了 1.7%和 5.0%。当联合使用 MA 损失和 MC 损失时，R1 和 mAP 指标分别提高了 5.0%和 3.0%。通过上述实验可以看出，本文提出的 CS 方法在批量中提供相似服装的不同身份图像，可以进一步促进模型学习更加鲁棒的特征，有助于识别服装变化的身份。

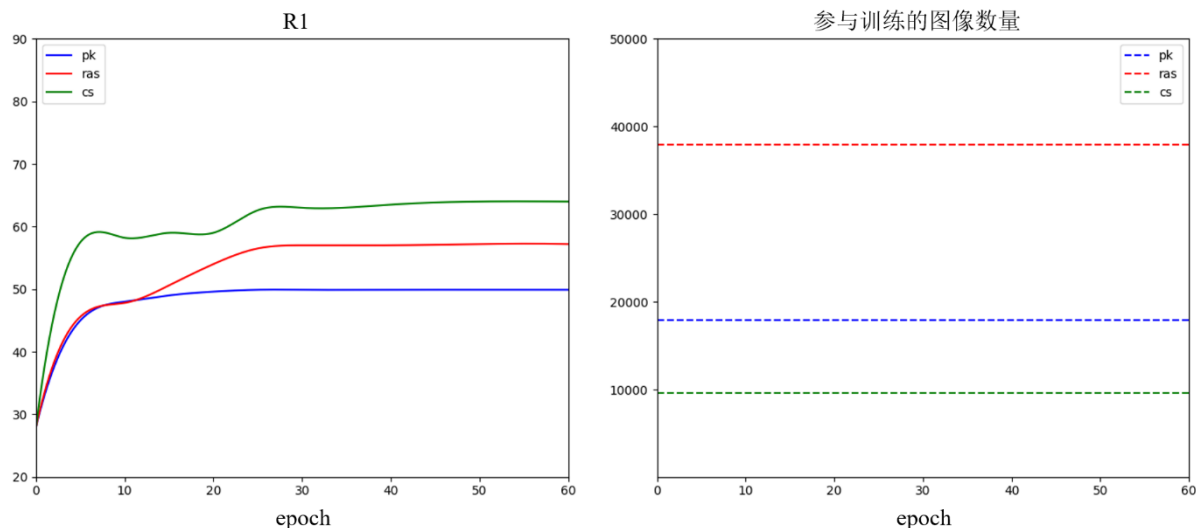


图 4-5 R1 和参与训练图像数量的变化曲线

Figure 4-5 The variation curve of R1 and the number of participating training images

#### 4.5.4 不同采样器训练效率的比较

在本节中，对不同采样器方法的训练效率进行了实验对比。实验结果如图 4-5 所示，其中实线代表不同方法的收敛速度和 R1 准确率变化曲线，虚线代表每种方法在一个 epoch 中所需参与训练的图像数量。

从图中可以直观地观察到，本文提出的 CS 方法相较于之前的 PK 和 RAS 方法，在更短的 epoch 内就能取得更具竞争力的效果。此外，根据图中的虚线所示，在每个 epoch 的训练过程中，CS 方法所需参与训练的图像数量远低于其他方法。这一结果进一步证明了 CS 方法通过选择更有价值的数据参与训练，能够更高效、更高质量地训练判别型模型。

#### 4.6 可视化分析

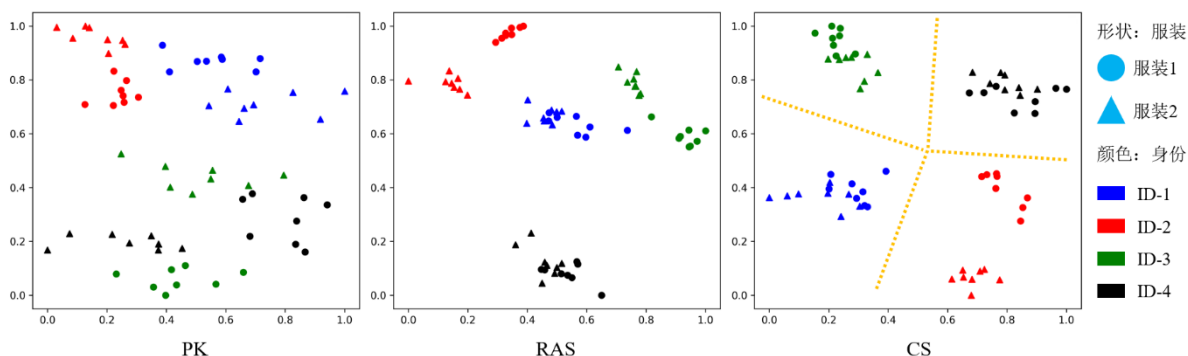


图 4-6 在 PRCC 数据集上，t-SNE 图像特征的可视化

Figure 4-6 Visualization of t-SNE Image Features on PRCC Dataset

#### 4.6.1 特征分布可视化

为了更直观地评估本文提出的 CS 方法，本文采用 t-SNE<sup>[78]</sup>方法来可视化网络输出的高维特征向量，选择了四个穿着相似的人进行展示，结果如图 4-6 所示。图中，不同的颜色代表不同的身份，而圆圈和三角形分别代表同一个人的第一套和第二套衣服，

从图中可以看出，与 PK 方法相比，RAS 方法生成的特征具有较为紧凑的类内特征，这使其能够有效地应对同一个人在不同着装情况下的变化。然而，RAS 方法在区分穿着相似服装的不同身份时遇到了困难，类间特征同样紧凑，这在面对不同身份但穿着相似服装的场景时，会使得特征难以区分。

相比之下，CS 方法能够同时妥善处理同一个人穿着不同衣服以及不同的人穿着相似衣服的情况，展现出更优越的特征区分能力。这表明 CS 方法在处理复杂场景时具有更高的鲁棒性和准确性，能够更好地应对服装变化带来的挑战。

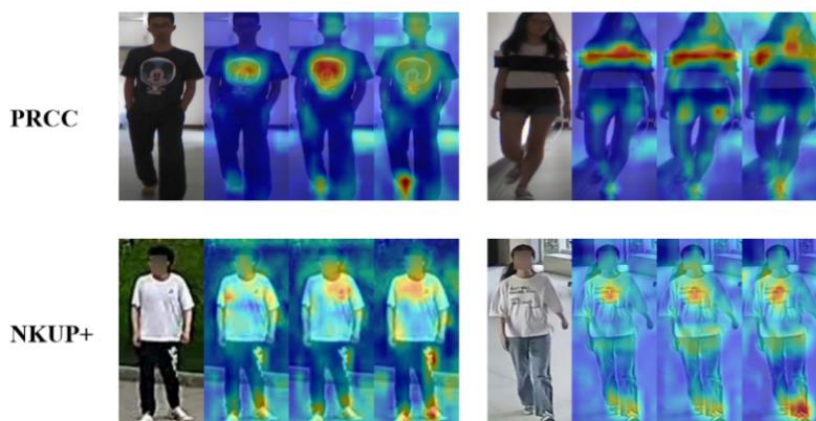


图 4-7 激活特征图的可视化  
Figure 4-7 Visualization of Activation Feature Map

#### 4.6.2 激活特征图的可视化

如图 4-7 所示，本章在 PRCC 和 NKUP+数据集上对三种方法的激活特征图进行了可视化。每组中，第二、第三和第四列分别展示了基于 PK 采样方法、基线模型和 CS 方法的激活图。从这些特征图中可以直观地观察到，CS 方法更侧重于与身份相关的区域，如头部、脚部和关节，而不是与服装相关的区域。这表明 CS 方法能够有效地引导模型关注人体的关键部位，而不是服装的细节。

从特征图的角度来看，CS 方法通过在批量中模拟更多的服装变化，并利用难样本数据，能够引导模型学习更多与人体相关的特征，从而减少服装特征的干扰。这一结果进一步证明了 CS 方法在处理服装变化时的有效性和鲁棒性，为行人重识别任务提供了更高质量的特征表示。

## 4.7 本章小结

本章提出了一种基于主动学习的服装采样策略，可以在没有准确服装标签的情况下，提供有价值的数据供模型学习。利用有限正确的服装标签，可以通过服装采样器在小批量样本中同时保证相同身份的服装多样性和相似服装的身份多样性，以迫使模型去学习更多与服装无关的特征。为了进一步发挥方法的性能，本章设计了多实例一致性损失，来应对服装的各类变化导致的挑战。实验结果表明该方法显著提高了换装 Re-ID 模型在面对服装变化时的鲁棒性。

## 第5章 总结与展望

### 5.1 总结

本文首先介绍了换装行人重识别的研究背景和研究意义，指出了该领域在实际应用中的重要性和挑战性。在实际场景中经常会遇到两类导致模型识别失效的情况：同一身份穿着不同的衣服、不同的身份穿着相似的衣服。根据此现象，本文总结并明确了当前研究中存在的两大挑战：服装变化及其引发人体轮廓改变、辅助模型和服装标签的不准确性。

为了应对上述的挑战，本文深入研究现有方法，并针对其中的不足，提出了两种创新性的解决方案：基于图像分割的双流融合网络、基于主动学习的服装采样方法。并通过大量实验证明方法在实际场景的合理性和有效性。

（1）基于图像分割的双流融合网络。从模型设计出发，设计双流融合网络将 CNN 特征与 ViT 特征进行融合，兼顾人体高级语义信息和局部特征。为充分利用辅助模型提供的先验知识，设计了数据增强模块。为避免辅助模型噪声的干扰，仅对 ViT 高级语义流进行数据增强，获取人体高级语义信息。在 CNN 局部特征流中，为进一步降低服装区域的影响，设计空间解纠缠模块利用服装标签进行解纠缠。

（2）基于主动学习的服装采样方法。从训练策略出发，本文设计一种基于服装的采样器，不同于过去的随机采样，此采样器在供训练的小批量样本中保证相同身份的服装多样性和相似服装的身份多样性，基于主动学习的方式从数据驱动角度迫使模型去学习更多与服装无关的特征，并且仅在采样阶段使用部分正确的服装标签。同时本文设计了多实例一致性损失，配合服装采样器来应对服装的各类变化导致的挑战。

综上所述，本文从行人重识别模型的三个阶段：模型设计、训练策略、度量学习分别进行研究，有效应对服装引发的各类挑战。这些方法在实验中表现出色，为解决实际应用场景中的复杂服装变化问题提供了有力的解决方案。

### 5.2 展望

尽管本研究在解决换装行人重识别任务中取得了显著进展，但仍有许多潜在的研究方向和改进空间。以下是一些未来研究的主要工作：

(1) 数据增强：先验的人体解析模型受限于训练域，在换装行人重识别任务中效果不佳，在利用分割信息的同时存在部分程度的噪声干扰，影响模型的效果。而随着大模型的发展，通过海量数据中训练的分割大模型，如 SAM 等，在下游任务中同样表现相当出色，在未来的工作，我们将探索这些大模型在换装任务中的应用，利用其中丰富的中间表示来提高方法性能。

(2) 多模态融合：本文研究工作主要集中于图像场景，而在实际场景中往往有多种模态信息可供利用，所以融合来自不同传感器模态的信息可能对于提高换装行人重识别的性能非常关键。例如，结合视频序列信息、热红外图像等，可以提供更全面、鲁棒的特征表示。未来的工作可以在多模态融合方面进行深入研究。

(3) 弱监督学习及泛化能力：本文提出的 CS 方法虽然在不依赖辅助模型和错误服装标签的情况下实现了最先进的性能，但它确实需要正确的类内服装标签来进行采样和监督学习。在当前的数据获取过程中，可以通过摄像头或日期获得服装伪标签。在未来的工作中，我们将探索弱监督学习来应对服装变化带来的挑战，解决数据标注问题。通过这种方式，我们可以减少对精确标签的依赖，使得模型能够更好地适应服装变化的情况，提高在实际应用中的泛化能力。

通过这些方向上的进一步研究，可以为解决换装行人重识别问题提供更全面、有效的解决方案，并促进该领域的进一步发展。

## 参考文献

- [1] Deng J, Guo J, Xue N, et al. Arcface: Additive angular margin loss for deep face recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4690-4699.
- [2] Wen Y, Zhang K, Li Z, et al. A discriminative feature learning approach for deep face recognition[C]//Computer vision—ECCV 2016: 14th European conference, amsterdam, the netherlands, October 11–14, 2016, proceedings, part VII 14. Springer International Publishing, 2016: 499-515.
- [3] 刘玉杰, 周彩云, 李宗民, 等. 基于增强特征融合网络的行人重识别方法[J]. 计算机、设计与图形学学报, 2021, 33(2): 232-240.
- [4] 邵文斌, 刘玉杰, 孙晓瑞, 等. 基于残差增强注意力的跨模态行人重识别[J]. 图学学报, 2023, 44(1): 33.
- [5] Zhang Z, Lan C, Zeng W, et al. Relation-aware global attention for person re-identification[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 3186-3195.
- [6] Zhang J, Niu L, Zhang L. Person re-identification with reinforced attribute attention selection[J]. IEEE Transactions on Image Processing, 2020, 30: 603-616.
- [7] Yang Q, Wu A, Zheng W S. Person re-identification by contour sketch under moderate clothing change[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 43(6): 2029-2046.
- [8] Liu M, Ma Z, Li T, et al. Long-Term Person Re-identification with Dramatic Appearance Change: Algorithm and Benchmark[C]//Proceedings of the 30th ACM International Conference on Multimedia. 2022: 6406-6415.
- [9] Shu X, Li G, Wang X, et al. Semantic-guided pixel sampling for cloth-changing person re-identification[J]. IEEE Signal Processing Letters, 2021, 28: 1365-1369.
- [10] Chao H, He Y, Zhang J, et al. Gaitset: Regarding gait as a set for cross-view gait recognition[C]//Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 8126-8133.

- 
- [11] Qian X, Wang W, Zhang L, et al. Long-term cloth-changing person re-identification[C]//Proceedings of the Asian Conference on Computer Vision. 2020.
  - [12] Li Y J, Luo Z, Weng X, et al. Learning shape representations for clothing variations in person re-identification[J]. arXiv preprint arXiv:2003.07340, 2020.
  - [13] Jin X, He T, Zheng K, et al. Cloth-changing person re-identification from a single image with gait prediction and regularization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 14278-14287.
  - [14] Gu X, Chang H, Ma B, et al. Clothes-changing person re-identification with rgb modality only[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 1060-1069.
  - [15] Zhang G, Liu J, Chen Y, et al. Multi-biometric unified network for cloth-changing person re-identification[J]. IEEE Transactions on Image Processing, 2023, 32: 4555-4566.
  - [16] Guo P, Liu H, Wu J, et al. Semantic-aware Consistency Network for Cloth-changing Person Re-Identification[C]//Proceedings of the 31st ACM International Conference on Multimedia. 2023: 8730-8739.
  - [17] Han K, Gong S, Huang Y, et al. Clothing-Change Feature Augmentation for Person Re-Identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 22066-22075.
  - [18] Hong P, Wu T, Wu A, et al. Fine-grained shape-appearance mutual learning for cloth-changing person re-identification[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 10513-10522.
  - [19] Yang Z, Lin M, Zhong X, et al. Good Is Bad: Causality Inspired Cloth-Debiasing for Cloth-Changing Person Re-Identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 1472-1481.
  - [20] He S, Luo H, Wang P, et al. Transreid: Transformer-based object re-identification[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 15013-15022.
  - [21] Bansal V, Foresti G L, Martinel N. Cloth-changing person re-identification with self-attention[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Com-



- puter Vision. 2022: 602-610.
- [22] Wang Q, Qian X, Fu Y, et al. Co-attention Aligned Mutual Cross-Attention for Cloth-Changing Person Re-identification[C]//Proceedings of the Asian Conference on Computer Vision. 2022: 2270-2288.
- [23] Hubel D H, Wiesel T N. Receptive fields and functional architecture of monkey striate cortex[J]. The Journal of physiology, 1968, 195(1): 215-243.
- [24] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological cybernetics, 1980, 36(4): 193-202.
- [25] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4): 541-551.
- [26] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [27] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [28] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [29] Simonyan K. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [30] 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.
- [31] 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.
- [32] Chen Z, Huang X. End-to-end learning for lane keeping of self-driving cars[C]//2017 IEEE intelligent vehicles symposium (IV). IEEE, 2017: 1856-1860.
- [33] Ozbayoglu A M, Gudelek M U, Sezer O B. Deep learning for financial applications: A survey[J]. Applied soft computing, 2020, 93: 106384.
- [34] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image data-

- base[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.
- [35] Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [36] Zheng L, Zhang H, Sun S, et al. Person re-identification in the wild[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1367-1376.
- [37] Zheng Z, Zheng L, Yang Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro[C]//Proceedings of the IEEE international conference on computer vision. 2017: 3754-3762.
- [38] Zhong Z, Zheng L, Cao D, et al. Re-ranking person re-identification with k-reciprocal encoding[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1318-1327.
- [39] Sun Y, Zheng L, Yang Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 480-496.
- [40] Qian X, Fu Y, Jiang Y G, et al. Multi-scale deep learning architectures for person re-identification[C]//Proceedings of the IEEE international conference on computer vision. 2017: 5399-5408.
- [41] Yang F, Yan K, Lu S, et al. Attention driven person re-identification[J]. Pattern Recognition, 2019, 86: 143-155.
- [42] Wang C, Zhang Q, Huang C, et al. Manacs: A multi-task attentional network with curriculum sampling for person re-identification[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 365-381.
- [43] Shen Y, Xiao T, Li H, et al. End-to-end deep kronecker-product matching for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6886-6895.
- [44] Wang Y, Chen Z, Wu F, et al. Person re-identification with cascaded pairwise convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 1470-1478.

- [45] Li W, Zhu X, Gong S. Harmonious attention network for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 2285-2294.
- [46] Luo C, Chen Y, Wang N, et al. Spectral feature transformation for person re-identification[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 4976-4985.
- [47] Cheng D, Gong Y, Zhou S, et al. Person re-identification by multi-channel parts-based cnn with improved triplet loss function[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1335-1344.
- [48] Li D, Chen X, Zhang Z, et al. Learning deep context-aware features over body and latent parts for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 384-393.
- [49] Zhao H, Tian M, Sun S, et al. Spindle net: Person re-identification with human body region guided feature decomposition and fusion[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1077-1085.
- [50] Suh Y, Wang J, Tang S, et al. Part-aligned bilinear representations for person re-identification[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 402-419.
- [51] Su C, Li J, Zhang S, et al. Pose-driven deep convolutional model for person re-identification[C]//Proceedings of the IEEE international conference on computer vision. 2017: 3960-3969.
- [52] Xu J, Zhao R, Zhu F, et al. Attention-aware compositional network for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 2119-2128.
- [53] Zhang Z, Lan C, Zeng W, et al. Densely semantically aligned person re-identification[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 667-676.
- [54] Guo J, Yuan Y, Huang L, et al. Beyond human parts: Dual part-aligned representations for person re-identification[C]//Proceedings of the IEEE/CVF international conference

- on computer vision. 2019: 3642-3651.
- [55] Tay C P, Roy S, Yap K H. Aanet: Attribute attention network for person re-identifications[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7134-7143.
- [56] Luo H, Jiang W, Gu Y, et al. A strong baseline and batch normneuralization neck for deep person reidentification[J]. arXiv preprint arXiv:1906.08332, 2019.
- [57] Huang H, Li D, Zhang Z, et al. Adversarially occluded samples for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 5098-5107.
- [58] Zhong Z, Zheng L, Kang G, et al. Random erasing data augmentation[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 13001-13008.
- [59] Bak S, Carr P, Lalonde J F. Domain adaptation through synthesis for unsupervised person re-identification[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 189-205.
- [60] Yan Y, Ni B, Song Z, et al. Person re-identification via recurrent feature aggregation[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14. Springer International Publishing, 2016: 701-716.
- [61] Koestinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints[C]//2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012: 2288-2295.
- [62] Liao S, Li S Z. Efficient psd constrained asymmetric metric learning for person re-identification[C]//Proceedings of the IEEE international conference on computer vision. 2015: 3685-3693.
- [63] Dai Z, Chen M, Gu X, et al. Batch dropblock network for person re-identification and beyond[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 3691-3701.
- [64] Varior R R, Shuai B, Lu J, et al. A siamese long short-term memory architecture for human re-identification[C]//Computer Vision–ECCV 2016: 14th European Conference,

- Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14. Springer International Publishing, 2016: 135-153.
- [65] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification[J]. arXiv preprint arXiv:1703.07737, 2017.
- [66] Ning E, Wang C, Zhang H, et al. Occluded person re-identification with deep learning: a survey and perspectives[J]. Expert systems with applications, 2024, 239: 122419.
- [67] Ristani E, Tomasi C. Features for multi-target multi-camera tracking and re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6036-6046.
- [68] Zhou J, Su B, Wu Y. Easy identification from better constraints: Multi-shot person re-identification from reference constraints[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 5373-5381.
- [69] Zheng F, Deng C, Sun X, et al. Pyramidal person re-identification via multi-loss dynamic training[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 8514-8522.
- [70] Huang Y, Xu J, Wu Q, et al. Beyond scalar neuron: Adopting vector-neuron capsules for long-term person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(10): 3459-3471.
- [71] Huang Y, Wu Q, Xu J, et al. Celebrities-reid: A benchmark for clothes variation in long-term person re-identification[C]//2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019: 1-8.
- [72] Wan F, Wu Y, Qian X, et al. When person re-identification meets changing clothes[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 830-831.
- [73] Yu S, Li S, Chen D, et al. Cocas: A large-scale clothes changing person dataset for re-identification[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 3400-3409.
- [74] Shu X, Wang X, Zang X, et al. Large-scale spatio-temporal person re-identification: Algorithms and benchmark[J]. IEEE Transactions on Circuits and Systems for Video Tech-

- nology, 2021, 32(7): 4390-4403.
- [75] Wang G, Yuan Y, Chen X, et al. Learning discriminative features with multiple granularities for person re-identification[C]//Proceedings of the 26th ACM international conference on Multimedia. 2018: 274-282.
- [76] Hou R, Ma B, Chang H, et al. Interaction-and-aggregation network for person re-identification[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 9317-9326.
- [77] Liao S, Shao L. Graph sampling based deep metric learning for generalizable person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 7359-7368.
- [78] Hinton G, Van Der Maaten L. Visualizing data using t-sne journal of machine learning research[J]. Journal of Machine Learning Research, 2008, 9: 2579-2605.

## 攻读硕士学位期间取得的研究成果

### （1）攻读硕士学位期间发表的学术论文

- [1] Liu Y, Li M, Jing J, et al. Clothing Sampling Based on Active Learning For Cloth-Changing Person Re-identification[C]//2024 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2024: 1-6. (CCF-B 类)
- [2] Jing J, Liu Y, Li M, et al. Modeling Sketches both Semantically and Structurally for Zero-Shot Sketch-Based Image Retrieval is Better[C]// 2024 Pacific Graphics (PG). The Eurographics Association. (CCF-B 类)

### （2）攻读硕士学位期间参与的主要科研项目

- [1] 基于人工智能的水驱微观驱替实验分析方法测试，中国石油化工股份有限公司胜利油田分公司勘探开发研究院，项目编号: 302000018-20-FW2099-0130，主要完成人。
- [2] 2022 年冬季奥林匹克运动会典型冰上项目多源数据智能分析系统。 项目编号: 2019 YFF0301800, 2019-2022，参与。

## 致 谢

时时光荏苒，岁月如梭，转眼间我在中国石油大学（华东）的求学生涯已接近尾声。在此，我怀着无比感激之情，向所有在这段旅程中给予我支持与帮助的人们，致以最诚挚的谢意。

首先，我要感谢石油大学。本科四年与硕士三年的学习经历，让我积累了丰富的专业知识，锻炼了综合能力。学校严谨的学术氛围与多元的校园文化，为我的成长提供了优越的环境。

其次，我由衷感谢我的研究生导师刘玉杰老师。在学术上，刘老师悉心指导，助我攻克难关；在生活中，他无私关怀，给予我温暖与力量。刘老师的言传身教，让我领悟到科研的严谨态度与做人的高尚品德。

同时，我也要向图形与可视化课题组的李宗民老师表达深深的谢意。李老师在课题研究中给予我宝贵意见，拓展了我的专业视野，提升了专业技能与团队合作能力。

在学术探索的道路上，我的同门们也给予了我莫大的帮助与支持。我们共同讨论学术问题、分享研究心得，携手攻克技术难题，彼此审阅论文、提出修改建议。同门间的深厚情谊与默契合作，让我的研究生生活充满欢声笑语，学术之路不再孤单。

此外，我要感谢我的家人。他们是最坚实的后盾，无条件地给予爱与支持。在我考研、读研的征途中，家人默默承担家庭重担，为我创造良好的学习环境；在我迷茫、疲惫时，用鼓励的话语、温暖的拥抱，让我重新振作。

最后，我要感谢生活中遇到的每一个人。你们的陪伴与支持，让我的人生更加丰富多彩。每一份帮助、每一次鼓励，都如点点繁星，照亮我前行的道路，让我在成长的旅途中，不断汲取前行的力量。

再次感谢所有给予我支持与帮助的人们，是你们让我的求学生涯充满意义与温暖。未来，我定将怀揣感恩之心，继续努力前行，用所学知识回馈社会，为祖国的发展贡献自己的力量。