

密级:_____



中国科学院大学
University of Chinese Academy of Sciences

硕士学位论文

基于迁移学习的人脸识别

作者姓名: _____ 吴俊婷 _____

指导教师: _____ 山世光 _____ 研究员 _____

中国科学院计算技术研究所

学位类别: _____ 工学硕士 _____

学科专业: _____ 计算机应用技术 _____

研究所: _____ 中国科学院计算技术研究所 _____

2013 年 5 月

Transfer Learning for Face Recognition

By

Junting Wu

A Thesis Submitted to

University of Chinese Academy of Sciences

In partial fulfillment of the requirement

For the degree of

Master of Computer Applied Technology

Institute of Computing Technology, Chinese Academy of Sciences

May, 2013

声 明

我声明本论文是我本人在导师指导下进行的研究工作及取得的
研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，本论
文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同
志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了
谢意。

作者签名：

日期：

论文版权使用授权书

本人授权中国科学院计算技术研究所可以保留并向国家有关部
门或机构送交本论文的复印件和电子文档，允许本论文被查阅和借
阅，可以将本论文的全部或部分内容编入有关数据库进行检索，可以
采用影印、缩印或扫描等复制手段保存、汇编本论文。

（保密论文在解密后适用本授权书。）

作者签名：

导师签名：

日期：

摘 要

虽然人脸识别技术现已广泛应用到许多场景中,比如安全监控、辅助相机聚焦等,但人脸识别技术仍然面临着许多挑战。其中一个实际应用中经常会遇到的问题是:训练与目标的人脸图像存在显著差异,从而导致训练得到的模型性能下降。该问题的根源在于传统的机器学习方法一般要求训练数据是对目标数据的良好描述,此时模型在训练数据上得到的性能在一定程度上代表其在目标数据上的性能,即训练误差可以是测试误差的良好估计。但是在实际应用当中,目标数据很可能来自与训练数据不同的分布,即目标数据对于模型来说是新的数据,因此模型无法有效应对这些新数据,从而导致人脸识别模型在目标数据上性能的退化。

解决上述问题的方法之一是迁移学习方法。迁移学习是通过将已有的知识迁移到新的数据场景下从而提升模型对新数据识别能力的方法。鉴于目前将迁移学习应用在人脸识别上的前人工作还很少,本文调研了机器学习等领域提出的迁移学习方法,在此基础上,针对目标域仅有无监督数据的情况,从三个不同的角度分析提出了基于迁移学习的解决方法:

1) 提出了基于目标域相似度矩阵估计的判别性特征提取方法

源域上的监督信息虽然并不与目标域上的完全一致,但目标域与源域都是基于身份信息分类的,因此可利用源域上可靠的监督信息来指导目标域上类别信息的估计。目标域上的类别信息可采用样本间的相似度矩阵进行描述,即要求属于同类的样本具有非零且非常接近的相似度,而不同类样本间相似度为 0。这样的相似度矩阵具有高度的判别性,且具有低秩的特性,其秩可用于衡量其域的判别性。因此,本方法旨在利用源域上的监督信息来指导目标域上具有高度判别性且低秩的相似度矩阵的估计。为了防止求解得到退化的相似度矩阵(比如秩为 1),则要求该相似度矩阵所描述的类的中心与属于该类的样本间距离较小。这个约束本方法将其形式化为目标域样本结合该相似度矩阵重构得到的样本与目标域样本间的误差应较小。由于估计目标域与源域遵循同一分类准则,则源域与目标域可共享同一判别性特征子空间,这要求源域与目标域在该子空间上能同时具有较好的判别性。其中目标域上采用相似度矩阵的秩、源域上采用 Fisher 准则定义的统计量来度量两个域的判别性。该方法提出可同时估计该特征子空间与目标域相似度矩阵的目标函数。由于低秩的相似度矩阵具有较好的判别性,因此可直接基于该相似度矩阵提取判别性的特征,而无需再利用无监督聚类进行类别估计。

2) 提出了基于源域样本目标化的人脸识别方法

同样针对目标域无监督数据问题,该方法试图将源域上带类别标签的样本迁移至目标域,从而获得一批可用于训练目标域模型的监督样本。具体来说,该方法将源域与目标域视为两个流形,源域样本的目标化其本质为确定源域样本在目标域流形上的位置。该方法采用目标域样本对源域样本进行稀疏编码来学习源域样本在目标域流形上的局部空间位置关系。但是在原始空间上由于源域与目标域存在明显差异,直接对源域样本进行稀疏编码时会涉及部分无用甚至有害的信息,实际上无法学习得到可靠的迁移系数。因此本方法借助源域与目标域无明显差异的公共子空间来学习可靠的迁移稀疏。这是因为源域与目标域之间包含丰富可共享的信息,在该空间上源域样本相对于目标域样本的局部空间位置关系很大程度上体现了在原始空间上目标化后源域样本相对于目标域样本的空间位置关系。本方法采用稀疏重构误差度量域间差异,并采用拉普拉斯矩阵来保持局部结构以避免域上流形有用信息的丢失,结合这两个约束来优化公共子空间同时学习迁移系数。最后将优化得到的稀疏编码用于在原始空间上对源域样本进行目标化。

3) 提出了融合公共与特有子空间上判别性特征提取的方法

当前基于公共子空间的迁移学习方法只利用到源域落在公共子空间上的监督信息,而忽略了目标域样本落在目标域特有子空间上的信息,而该信息亦可能对目标域上的分类问题有用。基于这一思路,该方法提出充分提取目标域整个空间上有用信息的融合方法。其中公共子空间上利用带标注源域所蕴含的可靠监督信息;目标域特有子空间上则利用无监督聚类的方法(比如 **Kmeans**)来提取目标域样本空间位置关系中所蕴含的类别信息,在此基础上利用有监督的特征提取方法(例如 **LDA**)来分别进行判别性的特征提取,最后融合这两部分判别性特征得到可对目标域整个特征子空间进行判别性信息提取的投影矩阵。在 **XM2VTS** 到 **FRGC** 等跨迁移实验上的结果表明该方法较只利用公共子空间上信息的迁移学习方法有更好的性能。

关键字: 迁移学习; 人脸识别; 低秩表示; 稀疏编码

ABSTRACT

Junting Wu (Computer Applied Technology)

Directed by Professor Shiguang Shan

Face recognition still confronts a great many problems despite that face recognition has been widely applied in many applications, i.e. security, assisting camera focusing. One of these challenges occurs that the model performance degenerates due to the significant disparity of training data and target data. The underlying reason is that most machine learning method obey the rules that training data describe the test data very well, so that the performance of model trained on the training data is with the high potential to attain the similar performance when applied on the target data, that is train error is a good estimate of test error under this condition. However, in the realistic application, target data are often different with the train data, that is the target data are almost new to the model, so that the model cannot efficiently deal with the target data, resulting in the degeneration of model performance on target data.

One way of solving this problem is transfer learning. Transfer learning can improve the model performance of recognizing new data through transferring the old knowledge to the application under the new data. Since there are few works on transfer learning in the field of face recognition, we review the proposed transfer learning methods in the field of machine learning etc., and then as to the scenario that target domain only consists of unlabeled data we propose three methods based on transfer learning for face recognition:

- 1) Discriminative feature extraction based on target domain affinity matrix estimation.

Even though the supervision information on the source domain isn't consistent with target domain's completely, the classifications both on target domain and source domain are based on identification, so that the source domain supervision information can be utilized to estimate the classification of target domain. The classification of target domain can be described by the affinity matrix, which has non-zero but similar affinities among the samples of the same object, and has zero affinities among the samples of different objects. This kind of affinity matrix is highly discriminative and

has the property of low-rank, and the rank of affinity matrix can be used to evaluate the discriminant level of domain. Therefore, this method aims at utilizing the source domain supervision information to direct the estimation of high-discriminative and low-rank affinity matrix of target domain. In order to avoid getting degenerated solution (i.e. rank is 1), the centers of clusters described by the affinity matrix should be close to the samples of the same cluster. The requirement is formulated as the small error between the target domain samples and the samples constructed by the target domain samples matrix and the affinity matrix. Since source domain and target domain obey the same rule of classification, so that source domain and target domain could share the same discriminative feature subspace, which requires the high discriminability of both target domain and source domain on this subspace. The discriminative information on the two domains is described by the rank of target affinity matrix and the variable defined by the Fisher rules on the source domain respectively. This method aims at utilizing the source domain supervision to direct the estimation of affinity matrix which reflects consistent discriminative-information on the common subspace. Since the low-rank affinity matrix is highly discriminative, so that it can be used directly to extract the discriminative features, without the need of clustering on it to estimate the classifications.

2) Unsupervised transfer learning via targeting source samples for face recognition

As to unsupervised target domain, this method attempts to transfer the labeled source domain samples to target domain, so that these targetized source domain samples can be used to train a model more suitable for target domain task. The method utilizes the sparse coding of source domain samples based on the dictionary of target domain samples to estimate the local special-positions of targetized source domain samples on the manifold of target domain. However the source domain and target domain have significant disparity on the original subspace, the sparse coding on it are not reliable since source domain samples contains some useless even harmful information. Therefore, this method proposes to learn the reliable sparse coding on the common subspace where source domain and target domain have no significant disparity. The feasibility lies on that source domain and target domain have rich information to be shared, the local special-positions of source domain samples relative to target domain samples on the common subspace highly reflects the local spatial-positions of targetized source domain samples on the manifold of target domain. This method utilizes the sparse reconstruction error to evaluate the disparity

between domains, the Laplacian matrix to preserve the locality structure for avoiding the information losing on the domains, and then combine them to optimize the common subspace as well as the sparse coding. Finally, this method utilizes the optimized sparse coding to transfer the source domain to target domain on the original subspace.

3) Fuse common and specific discriminative feature model

The recent transfer learning methods based on common subspace only utilize the supervision information of source domain on the common subspace, but ignore the information embedded in the target domain specific subspace, which may be useful for the classification task on the target domain. Therefore this method proposes the pipeline of how to completely extract the useful information on the whole target domain subspace. In detail, on the common subspace, there are supervision information of source domain; on the target-specific subspace, there are discriminative information embedded in the target samples locations, which can be extracted by unsupervised clustering (i.e. Kmeans). Based on these, supervised learning method (i.e. LDA) is used to extract the discriminative features respectively, and finally get the discriminative features of the whole target domain subspace by fused the extracted discriminative features on common and target specific subspace. From the results of transferring experiments such as XM2VTS to FRGC etc., this method attains better performance compared to other transfer learning methods based on common subspace.

Keywords: Transfer Learning, Face Recognition, Low-rank Representation, Sparse Coding

目 录

摘 要	I
ABSTRACT.....	III
目 录	VII
图目录	XI
表目录	XIII
第 1 章 绪论	1
1.1 研究背景及其意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究问题及其意义.....	5
1.2 国内外研究现状.....	9
1.2.1 迁移学习介绍.....	9
1.2.2 迁移学习方法.....	11
1.3 本文的研究内容及贡献.....	19
1.3.1 研究内容.....	19
1.3.2 研究贡献.....	20
1.4 论文的组织.....	21
第 2 章 基于目标域相似度矩阵估计的判别性特征提取方法	23
2.1 相关背景知识.....	23
2.1.1 线性判别分析.....	23
2.1.2 低秩表示.....	24
2.1.3 The Singular Value Shrinkage Operator	26
2.2 基本思路与方法框架.....	26
2.2.1 基本思路.....	26
2.2.2 方法框架.....	28
2.2.3 目标函数.....	28
2.3 优化方法与算法流程.....	30
2.4 实验及性能比较.....	33
2.4.1 实验数据库及协议.....	33

2.4.2 不同迁移问题下利用 TAME 迁移的人脸识别实验.....	37
2.4.3 参数实验分析.....	40
2.5 总结与分析.....	41
第 3 章 基于源域样本目标化的人脸识别	43
3.1 相关背景知识.....	43
3.1.1 局部保持映射.....	43
3.1.2 稀疏编码 (Sparse Coding)	44
3.2 基本思路与方法框架.....	47
3.2.1 基本思路.....	47
3.2.2 方法框架.....	49
3.2.3 目标函数.....	49
3.3 优化方法与算法流程.....	52
3.3.1 优化方法.....	52
3.3.2 算法流程.....	54
3.4 实验及性能比较.....	55
3.4.1 不同迁移问题下利用 TSC 迁移的人脸识别实验	55
3.4.2 实验分析.....	57
3.5 总结与分析.....	59
第 4 章 融合公共与特有子空间上判别性特征的迁移学习方法	61
4.1 相关背景知识.....	61
4.1.1 主成份分析.....	61
4.2 基本思路与方法框架.....	63
4.2.1 基本思路.....	63
4.2.2 方法框架.....	64
4.2.3 目标函数.....	66
4.3 优化方法与算法流程.....	67
4.3.1 优化方法.....	67
4.3.2 算法流程.....	68
4.4 实验及性能比较.....	69
4.4.1 不同迁移问题下利用 FCSD 迁移的人脸识别实验.....	69
4.4.2 实验分析.....	71
4.5 总结与分析.....	74
第 5 章 结束语	77

5.1 本文总结.....	77
5.2 展望.....	78
参考文献	79
致 谢	i
作者简介	iii

图目录

图 1-1 人脸识别的流程图	2
图 1-2 传统机器学习流程	5
图 1-3 传统机器学习中的源域与目标域，其刻画的信息较为接近。	6
图 1-4 实际应用中的源域与目标域，其包含的信息存在显著的差异。	6
图 1-5 无监督迁移学习下的人脸识别问题。训练数据包含带标注的源域与无标注的目标域，要解决的是目标域上的人脸识别问题。	6
图 1-6 采用目标数据与非目标数据训练得到的模型（Fisherfaces+NN）分别在目标域为 OFD、FRGC 上的识别率。	8
图 1-7 采用目标数据与非目标数据训练得到的模型（Fisherfaces+NN）分别目标域为 MultiPIE 不同姿态子数据库上的识别率。	8
图 1-8 SGF 方法原文的说明图	15
图 2-1 诱导相似度矩阵估计（TAME）框架图	28
图 2-2 OFD 数据库的示例图	34
图 2-3 XM2VTS 数据库的示例图	35
图 2-4 FRGC 数据库上的示例图	35
图 2-5 MultiPIE 数据库上同一个对象在不同姿态下的示例图	36
图 2-6 TAME 方法在以 XM2VTS 数据库为源域，FRGC 数据库为目标域的实验上，其性能随参数变化的曲线。其中横轴为参数 τ ，曲线的不同颜色表示取不同的 λ 值。	41
图 2-7 TAME 方法在 MultiPIE 数据库上源域为 0° ，目标域为 -45° 的实验上，其性能随参数变化的曲线。其中横轴为参数 τ ，曲线的不同颜色表示取不同的 λ 值。	41
图 3-1 基于源域样本目标化的人脸识别方法（TSC）的方法框架	49
图 3-2 基于源域样本目标化（TSC）原理示意图	52
图 3-3 TSC 方法在源域为 45° 、目标域为 -45° 下，随 λ 值变化的曲线，以及随机初始化 W_s 、 W_t 矩阵时的标准差图。	57
图 4-1 主成份分析的前 8 维本征脸	62
图 4-2 FCSD 方法的原理示意图	64
图 4-3 FCSD 方法框架图	65
图 4-4 公共子空间模型在不同的 λ 下对公共子空间能量的敏感性。纵轴为人脸	

识别率，横轴为公共子空间保留的能量，曲线的不同颜色表示取不同的参数 λ 值	72
图 4-5 公共子空间特征模型性能在不同的 τ 下对公共子空间能量的敏感性。纵轴为人脸识别率，横轴为公共子空间保留的能量。曲线的不同颜色表示取不同的参数 τ 值。	73
图 4-6 MultiPIE 不同迁移问题下，公共子空间模型对于公共子空间保留的能量敏感性。	74
图 4-7 MultiPIE 两个不同的迁移问题下，FCSD 模型随参数 λ 和 τ 变化的性能曲线。	74

表目录

表格 1-1 常用的公共人脸数据库	3
表格 1-2 迁移学习的分类。其中“ \odot ”表示存在监督信息，“ \ominus ”表示无监督信息。	11
表格 2-1 用于模拟迁移学习的数据库：OFD、XM2VTS、FRGC.....	36
表格 2-2 用于模型迁移学习的数据库：MultiPIE	37
表格 2-3 跨采集条件的人脸识别问题，由 XM2VTS 迁移至 FRGC.....	37
表格 2-4 测试 TAME 方法在跨种族问题下的人脸识别性能	39
表格 2-5 测试 TAME 方法在跨采集条件问题下的人脸识别性能	39
表格 2-6 测试 TAME 方法在跨姿态问题下的人脸识别性能	40
表格 3-1 测试 TSC 在跨种族问题下的性能及与其他方法的对比.....	56
表格 3-2 测试 TSC 在跨采集条件问题下的性能及与其他方法的对比.....	56
表格 3-3 测试 TSC 在跨姿态问题下的性能及与其他方法的对比.....	56
表格 3-4 TSC 迁移方法的有效性。分别对比了直接在原始空间上进行稀疏重构 迁移的性能（SC），公共子空间模型（LCS）的性能，以及 TSC 方法的性能。	58
表格 4-1 测试 FCSD 在跨种族问题下的性能及与其他方法的对比	70
表格 4-2 测试 FCSD 在跨采集条件问题下的性能及与其他方法的对比	71
表格 4-3 测试 FCSD 在跨姿态问题下的性能及与其他方法的对比	71

第1章 绪论

1.1 研究背景及其意义

1.1.1 研究背景

人脸识别技术是一种利用计算机对包含人脸的图像、视频等进行分析，旨在判断图像中人脸所对应的身份信息的技术。人脸识别是计算机视觉领域的一个重要研究课题。

人脸识别的优势在于：1) 非接触性采集，相较于指纹识别或者 DNA 序列分析技术，人脸识别只需通过摄像设备采集人脸图像即可，而不需要用户接触仪器或者提供细胞等；2) 采集方式快速便捷。特别是由于目前摄影设备的普及，能采集到更多的人脸图像，使得人脸识别技术有着更加广泛地应用；3) 采集设备成本低，相较于 DNA 序列分析等技术，人脸识别分析技术快速便捷，在一般的家用电脑甚至手机设备上都可使用。人脸识别技术要求的设备成本低，可以广泛应用在实际应用中。

过去很长一段时间来，对人脸识别技术的需求主要来自公共安全领域。例如应用在辅助出入境管理上，利用人脸识别技术对用户提供的证件图像与现场采集的人脸图像进行分析，判断证件照与现场对象是否一致；此外还可与嫌疑犯或逃犯的图像进行比对，可在短时间内利用机器做批量分析，对和可疑分子长相类似的对象，返回相关的可疑分子图像供审核人员进一步参考，可增强审查的严密性，同时大大降低审核人员的工作负担；并且由于是机器分析，可以持续提供稳定的服务。此外人脸识别技术在视频监控、门禁设备和智能考勤等方面有广泛的需求与应用。

近年来，由于摄像设备价格更趋平民化，大部分个人都拥有几件带拍摄功能的机器，比如手机、便携式电脑等，这些使得人脸识别技术在个人应用方面也有随之有了更多的需求。例如 Google 公司的图像管理产品——Picasa 软件，内嵌人脸检测和人脸识别技术，具有根据身份信息进行相册管理的功能，其中人脸识别技术为该功能的核心算法之一；网易公司正尝试开发的利用人脸识别技术代替口令密码来登录邮箱的功能。另外一个有着巨大发展潜力的是目前备受青睐的智能手机应用程序开发领域，比如利用人脸识别进行加密与解锁等功能。

虽然目前人脸识别技术相关产品已经进入应用市场，但由于受到实际应用数

据与训练数据之间差异的影响，常常无法得到理想的性能。比如以下这种情况：用于训练模型的人脸图像通常为实验室控制条件下采集的，光照、表情变化有限；但是实际应用数据为现实条件下采集的，可能包含强烈的光照或者表情变化。由于训练数据总是有限的，故模型适用的人脸识别问题也可能局限在某些类似的情景下。为了使得人脸识别技术能够在面对不同的目标任务时具有一定的自适应性，即当带标签的训练数据与目标应用数据之间存在差异的问题下，如何充分提取训练数据中对于目标问题有益的信息，从而提高人脸识别模型在目标数据上的性能是人脸识别技术能否更鲁棒地应对实际问题的一个重要研究问题。

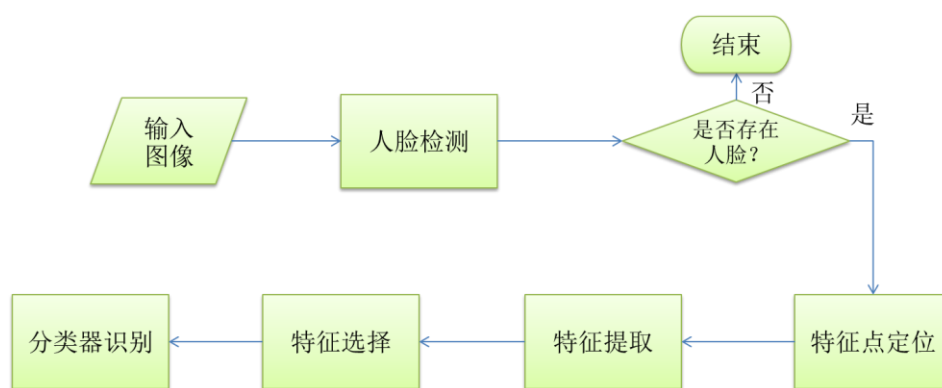


图 1-1 人脸识别的流程图

1.1.1.1 人脸识别流程介绍

人脸识别技术作为人机交互、人工智能等领域不可或缺的一个技术，其应用流程通常如图 1-1 所示。其中流程中的不同的过程可选择不同的方法，最后可训练得到不同的人脸识别模型。

1. 人脸检测

人脸检测技术的主要任务为判断给定的图像（静态图像或视频序列中的帧）中是否包含人脸，若包含人脸则返回人脸部分相对原始图像所在的坐标信息。将人脸检测推进到实时检测的开创性工作是由 2001 年的这篇文章[1]中提出的。

2. 特征点定位

特征点定位是指在人脸检测返回的区域上，对人脸的特征点进行定位，比如眼角、嘴角、轮廓特征点等[2]。特征点定位较为经典的算法有主动形状模型[3]（Active Shape Model, 简称 ASM）、主动表观模型[4]（Active Appearance Model, 简称 AAM）等。

3. 特征提取

特征提取技术是指从定位好的特征点上或者整个人脸区域上通过一些设计好的描述子进行特征的提取,提取得到的特征较之灰度或彩色图像具有更好的鲁棒性、判别性。常用的人脸特征描述子有 LBP[5](Local Binary Pattern), Gabor[6], SIFT[7], HOG[8]特征等。

4. 特征选择

由于上述提取的特征可能维数较高,且包含大量冗余信息,故一般采用特征选择技术选取包含较多信息量或对分类更为有用的部分特征,从而降低算法的复杂性。特征选择方法根据是否利用到监督信息,可分成无监督的特征提取方法,比如 PCA[9, 10], 以及有监督的特征提取方法,比如 LDA[11]。

5. 分类器设计

在面对已提取、选择好的特征时,人脸识别模型中通过训练分类器来进行身份信息的确认为或识别。在人脸确认(Face Verification)问题中,分类器需能够判断给定的一对人脸图像其身份信息是否相同;在人脸识别(Face Recognition)问题中,分类器要根据已知身份信息的注册图像(通常称之为 Gallery)来判断待识别图像的身份信息。常用的分类器有最近邻分类器(Nearest Neighbor, 简称 NN)、支持向量机[12](Support Vector Machines, 简称 SVM)等。

1.1.1.2 常用的公共人脸数据库

为了推进人脸识别技术的发展,一些研究所或大学等机构发布了一些人脸数据库。本节介绍部分常用的公共人脸数据库,这些数据库的获取地址整理如表格 1-1 所示。

表格 1-1 常用的公共人脸数据库

数据库	链接网址
AR	http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html
CAS-PEAL	http://www.jdl.ac.cn/peal/index.html
FERET	http://www.itl.nist.gov/iad/humanid/feret/feret_master.html
FRGC	http://www.nist.gov/itl/iad/ig/frgc.cfm
LFW	http://vis-www.cs.umass.edu/lfw/
OFD	http://www.aiar.xjtu.edu.cn/dfrlsjk5.htm
XM2VTS	http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/
YaleB	http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html

- AR 人脸数据库[13]: 该数据库包含在实验室控制条件下分两次(间隔 2 周)对 126 个对象(70 个男人、56 个女人)采集的超过 4,000 张的彩色图像。收

集时对采集对象无发型、服饰等方面的要求。该数据库采集的人脸图像中包含了不同的表情、光照条件以及遮挡情况，常用于不同表情、人脸遮挡问题下的人脸识别模型训练与评测数据库。

- **CAS-PEAL 人脸数据库[14]**: 由中科院计算所 JDL 实验室设计采集, 旨在收集实验室控制条件下受不同因素影响的人脸图像。该数据库对 1,040 名中国人 (595 个男性和 445 个女性) 采集了共 99,594 幅图像, 包含表情、年龄、饰物、光照和姿态等变化, 且划分了标准的训练数据集和测试数据集。
- **FERET 数据库[15, 16]**: 该数据库受美国国防部资助, 其全称为 **FacE REcognition technology Test**。其创建目的是为了发展自动人脸识别技术, 用于评测人脸识别技术投入实际应用 (比如辅助安全监管、人工智能等) 的可行性。为了模拟不同的人脸识别任务, 该数据集严格划分了训练集、不同的 Probe 集合与 Gallery。该数据库包含了年龄、光照、表情等变化。
- **FRGC 人脸数据库[17]**: 由美国国家标准局 (NIST) 策划采集的, 是一个包含超过 50, 000 张高分辨率人脸图像的数据库, 根据公布的协议划分成训练集和测试集。该数据库上的测试协议又将测试集划分成 6 个不同的子测试集, 用于评价人脸识别方法在不同情景下的性能。该数据库创建的初衷主要是为了评测人脸识别技术在高分辨率图像、3D 图像以及不同的预处理方法下的性能, 推动人脸识别技术在这几个方向上的发展。
- **LFW 数据库[18]**: 该数据库全称为 **The Labeled Faces in the Wild**, 旨在模拟非实验室控制条件下的人脸识别问题。LFW 数据库包含超过 13,000 幅从网络上收集而来的图像, 共 5749 个对象, 其中 1680 个对象包含两幅或两幅以上的图像。该数据库一般用于评测人脸识别模型在实际环境下的性能。
- **Multi-PIE 人脸数据库[19]**: 由美国卡内基梅隆大学策划采集的, 起初仅对 68 个对象采集了共 41368 幅人脸图像, 其中包括不同姿态、表情以及光照变化。目前广泛采用的是拓展之后的数据库, 拓展数据库中人数增加至 337 人。该数据库创建目的主要是为了模拟实验室控制条件下, 姿态、表情、光照等因素对于人脸识别的影响。该数据库在四个不同的时期 (**Four Sessions**) 对总共 337 个对象采集了超过 750,000 幅图像, 包含了 15 个不同的视角、19 种不同的光照。该数据库常用于姿态变化下人脸识别模型的训练与评测数据库。
- **OFD 人脸数据库[20]**: 该数据库由中国西安交通大学 **IAIR** 实验室采集, 旨在建立大规模、包含丰富变化的东方人数据库。OFD 数据库对 1247 名志愿者采集了共 33,669 幅图像, 主要包含姿态、光照、表情等变化。
- **Yale Face Database B 数据库 (简称 YaleB 数据库) [21]**: 该数据库是由耶鲁

大学 (Yale University) 采集创建, 仅采集了 10 个对象。该数据库中的每个对象各采集了 596 不同采集条件 (9 种姿态 \times 64 种正面光照) 下的图像, 此外在各个姿态下又采集一幅包含背景漫反射的图像, 故构成总数为 5850 幅的数据库。该数据库包含丰富细致的姿态、光照变化, 可用于受姿态、光照影响下人脸识别的模型训练与性能评测数据库。

- XM2VTS 多模态人脸数据库[22]: 该数据库采集时, 对 295 个对象约间隔 1 个月采集一次, 共采集了 4 次。每次采集时都采用高分辨率的设备记录下人脸以及语音数据, 主要包含讲话时的脸部变化 (Speech Shot) 以及头部旋转时的变化 (Head Rotation Shot)。该数据库也公布了标准的人脸确认评测协议。

1.1.2 研究问题及其意义

1.1.2.1 研究问题

传统的人脸识别技术遵循着和大部分机器学习方法相同的流程, 如图 1-2 所示。传统的机器学习方法大都包含一个隐式的假设, 即源数据 (传统机器学习中称之为训练数据) 与目标数据来自相同的分布 (如图 1-3 所示), 因此在源数据上训练的模型在源数据与目标域数据上的性能较为接近。但是在实际应用当中, 源数据与目标数据常常可能来自不同的分布 (如图 1-4 所示)。这种应用场景下, 即使在源数据上训练得到一个较为满意的模型 (未过拟合), 若将其直接应用在目标数据上, 一般会得到较低的性能 (如图 1-7、图 1-6 所示)。

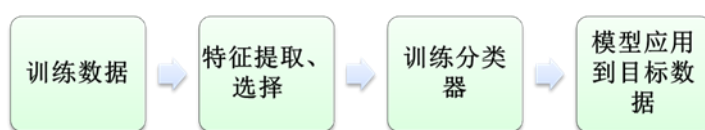


图 1-2 传统机器学习流程中分类器训练与测试流程

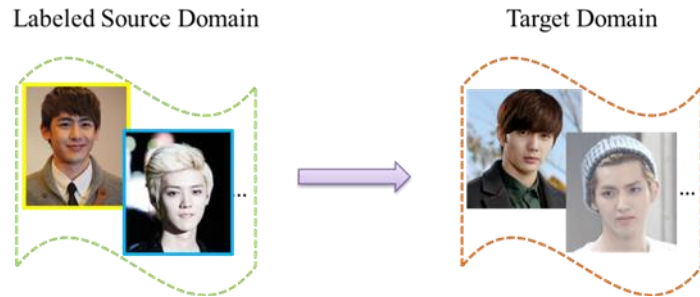


图 1-3 传统机器学习中的源域与目标域，其刻画的信息较为接近。

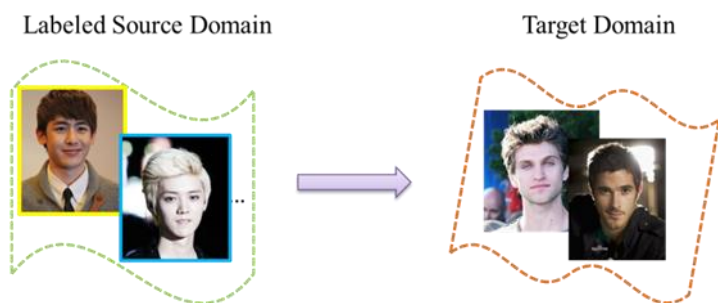


图 1-4 实际应用中的源域与目标域，其包含的信息存在显著差异。

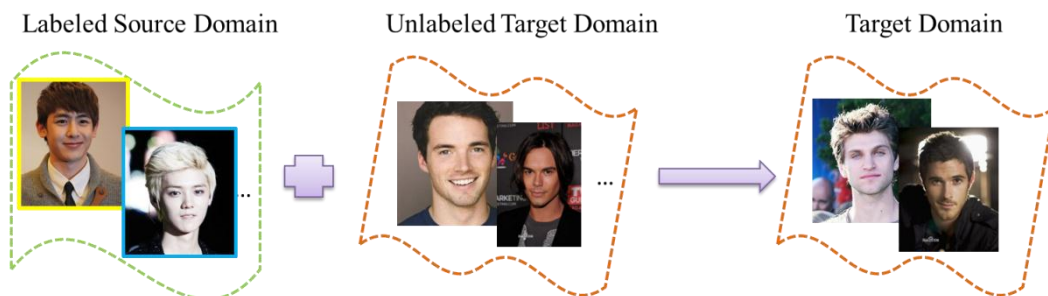


图 1-5 无监督迁移学习下的人脸识别问题。训练数据包含带标注的源域与无标注的目标域，要解决的是目标域上的人脸识别问题。

源域与目标域间之间的差异会导致相同的人脸识别方法分别在源域与目标域上训练得到的模型，应用在目标域上后得到的性能存在较大的差异。为了证明这种现象的存在，我们在几个公共的人脸数据库上进行了模拟实验。图 1-6 为在 OFD、FRGC、XM2VTS 这三个人脸数据库上展开模拟实验的性能图，纵轴为人脸识别性能，横轴为目标域数据库。图上的红色柱条表示采用目标域数据库上的

人脸图像作为训练数据训练得到的模型 (Fisherfaces+NN)，在目标域测试数据上应用时得到的人脸识别性能，其中训练数据与测试数据不重叠；图上的绿色柱条表示采用来自非目标数据库的数据库（这里采用 XM2VTS 数据库）上的训练数据训练得到的模型 (Fisherfaces+NN)，应用在目标域测试数据上得到的人脸识别性能。由图 1-6 可见，红色柱条代表的模型性能远远高于绿色柱条所代表的模型。由于这两种颜色柱条所采用的人脸识别方法相同，仅是训练数据不同，红色柱条的训练数据与目标域来自相同分布，而绿色柱条的训练数据与目标域来自不同分布，故可见训练数据与目标数据之间的差异会导致人脸识别模型其性能的下降。类似地，在另外一个数据库 (MultiPIE) 上的模拟实验同样支持该结论，实验结果如图 1-7 所示，纵轴为模型的人脸识别率，横轴为目标数据库，即 MultiPIE 数据库按照姿态的不同而划分的子数据库。其中红色柱条表示采用各姿态子数据库 (-45° 、 -30° 、 -15° 、 0° 、 15° 、 30°) 上的训练数据训练得到的模型 (Fisherfaces+NN)，应用在自身数据库的测试数据上得到的人脸识别性能；绿色柱条表示采用 MultiPIE 数据库上 45° 姿态下的人脸图像作为训练数据训练得到的模型 (Fisherfaces+NN)，应用在不同姿态下的性能。在图 1-7 所示的系列实验中，红色柱条所代表的模型性能优于采用绿色柱条所代表的性能。这个实验同样说明训练数据与测试数据之间的差异会导致模型性能的下降。

上文的模拟实验已经表明，训练数据与目标数据之间的差异会导致人脸识别模型性能下降，因此要学习得到适合目标域上人脸识别的模型，尽量应采用与目标域来自同分布的数据作为训练数据，最直接的作法是重新采集并标注一批目标域数据。但在实际应用中，目标的应用数据变化不定，如若每次遇到新一批的应用数据就要采集并标注大量的目标域数据，这是非常耗费人力、物力的，因此并不太可行。训练时需要一批这样的数据，是因为该数据中包含丰富的目标域判别性信息，从而使得训练得到的模型可很好地应对目标域上的人脸识别问题。但其实无标注的目标域样本已包含丰富的目标域信息，只是苦于无任何监督信息。另一方面虽然可用于监督训练的人脸数据库通常仅包含有限的人脸变化，并不能全面地涵盖目标域上所有的人脸变化，但这些人脸数据库上包含可靠的标注信息，内含人脸图像依据身份信息分类的知识。因此，如若可以充分利用这两部分数据上各自所携带的信息，提取得到可逼近理想情况下带标注目标域样本所刻画的目标域人脸分类信息，则可更好地应对目标域上的人脸识别模型。

这种结合利用源域与目标域的学习方法，在学术上称之为迁移学习 (Transfer Learning)，第 1.2 节中将具体介绍该方法。根据源域与目标域上学习任务是否相同、是否携带监督信息，可细分成多种的子问题。本研究只着眼于如图 1-5 (带方框的图像表示已知标注信息，无边框的图像无标注信息) 所示的人脸识别问题，

即训练数据为带标注的源域与无标注的目标域样本，而待解决的问题为目标域上的人脸识别问题。这种情景的问题中，由于目标域上无任何的监督信息，故在迁移学习中一般称之为无监督的迁移学习问题。

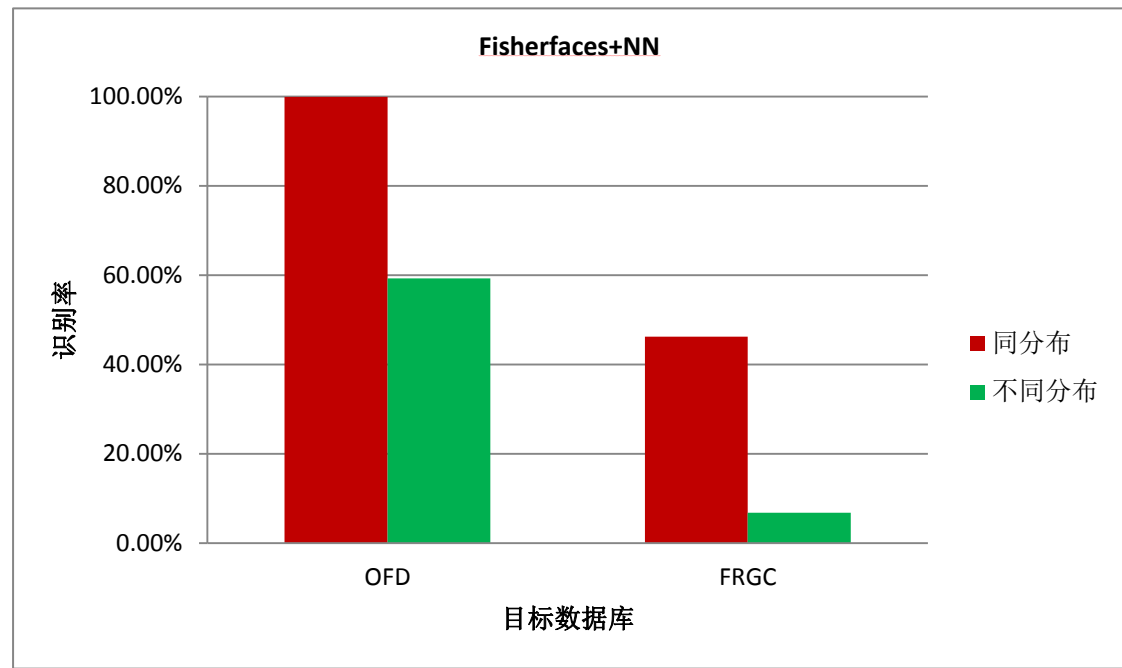


图 1-6 采用目标数据与非目标数据训练得到的模型（Fisherfaces+NN）分别在目标域为 OFD、FRGC 上的识别率。

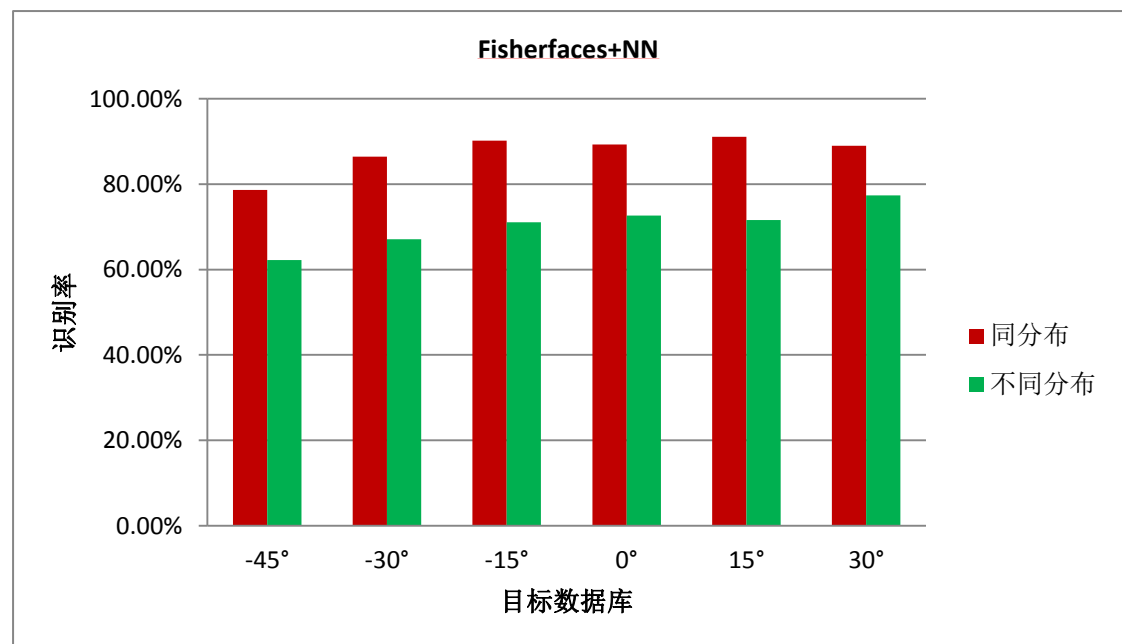


图 1-7 采用目标数据与非目标数据训练得到的模型（Fisherfaces+NN）分别目标域为 MultiPIE 不同姿态子数据库上的识别率。

1.1.2.2 研究难点

利用迁移学习的思路来解决当带标注的源域与目标域存在差异人脸识别问题时，存在下述的几个难点。

- 1) 迁移什么。由于因为源域中对目标域有利的信息通常并不是显而易见的，如何全面地发现源域中 useful 信息是首先需要解决的问题。
- 2) 如何迁移。即使已经知道源域中对于目标域有帮助的信息，如何将这些信息尽可能准确、无损失的迁移至目标域也是迁移学习方法需要解决的难题。
- 3) 迁移的有效性。源域中可能包含对目标域上的人脸识别问题有害的信息，并不是迁移越多源域知识越好，故如何在全面迁移源域有用信息的基础上，杜绝源域中对目标域上的人脸识别任务有害的信息，保证迁移信息的有效性亦是迁移过程中必须考虑的难题。

1.2 国内外研究现状

由于当前人脸识别领域还未有关于迁移学习的相关工作，故本节将介绍当前在机器学习、物体识别等领域中提出的关于迁移学习的工作。这些方法分别从公共子空间、信息论等角度来迁移源域上的有用信息。

1.2.1 迁移学习介绍

人类在面临新任务时，通过接触部分新任务问题，对之前相关任务中的旧知识进行调整，可更快地学好或者应对新任务。迁移学习正是受到这种学习方式的启发，希望在对新任务或者目标数据预先有了一定了解的情况下，充分结合相关的旧任务中所包含的知识，进而指导模型的学习，使其能比只利用了旧知识或有限目标域知识的模型，能更好地应对新任务。

在迁移学习中存在两个概念，即域 \mathcal{D} (Domain) 和学习任务 \mathcal{T} (Task)。具体来说，域上又包含两部分概念——样本所在的特征子空间 \mathcal{X} 和域上样本所构成的分布 $P(X)$ ，即 $\mathcal{D} = \{\mathcal{X}, P(X)\}$ 。若两个域所在的特征子空间 (Feature Subspace) 或者其上的样本分布 (Distribution) 不同，则这两个域存在差异，即具有以下关系：若 $\mathcal{X}_S \neq \mathcal{X}_T$ 或者 $P(\mathcal{X}_S) \neq P(\mathcal{X}_T)$ ，则 $\mathcal{D}_S \neq \mathcal{D}_T$ 。类似地，学习任务亦包含两部分——类别空间 \mathcal{Y} (Label Subspace) 和目标预测函数 (Objective Predictive Function) $f(\cdot)$ ，即 $\mathcal{T} = (\mathcal{Y}, f(\cdot))$ 。若两个学习任务中标签所在的类别空间或者目标预测函数不同，那么这两个学习任务是不同的，即具有以下关系：若 $\mathcal{Y}_S \neq \mathcal{Y}_T$ 或者 $f_S(\cdot) \neq f_T(\cdot)$ ，则 $\mathcal{T}_S \neq \mathcal{T}_T$ 。

学术上对迁移学习的定义为，给定源域 \mathcal{D}_S 和在其上的学习任务 \mathcal{T}_S ，以及目标域 \mathcal{D}_T 和在其上的学习任务 \mathcal{T}_T ，迁移学习的目标在于利用 \mathcal{D}_S 和 \mathcal{D}_T 的知识来提升在 \mathcal{D}_T 上目标预测函数的学习，其中 $\mathcal{D}_S \neq \mathcal{D}_T$ ，或者 $\mathcal{T}_S \neq \mathcal{T}_T$ [23]。

机器学习领域认为迁移学习过程中需要回答三个问题：1) 迁移什么知识？在应用迁移学习时，需要分析可以迁移的知识，比如域间公共的信息。目前有关迁移学习的文章大部分是以提取域间公共的信息作为目标。这是因为在公共子空间上，源域与目标域之间无明显差异，故目标域可直接借用源域上的知识；2) 是否迁移。在应用迁移学习来提升模型性能之前，需要判断源域的知识对于目标域的任务学习是促进作用还是抑制作用。研究上，把对目标域上学习任务有利的源域称之为正源域 (Positive Source Domain)；反之，起抑制作用的源域称之为负源域 (Negative Source Domain)。在利用源域时，并不是源域越多越好，需要警惕负源域可能带来的负面影响。由于一般选取先验上相关的域作为源域，故当前大部分的迁移学习只关注如何迁移相关源域上的有用知识，省略了对域是否相关有利的判断 3) 如何迁移？大部分研究工作围绕在“如何迁移”这个问题上，提出各种不同的迁移学习方法，以使得目标域上的学习任务能够借鉴到更多有用的源域信息。

常见的迁移学习方法根据目标域和源域上学习任务是否相同，主要可分成两大类——归纳迁移学习 (Inductive Transfer Learning) 和直推迁移学习 (Transductive Transfer Learning) (如表格 1-2 所示)。归纳迁移学习指当源域与目标域上有不同的学习任务时的迁移问题，学习任务的不同可能是由于类别空间不同也可能是目标预测函数不同引起的。更具体地，再根据源域和目标域上是否包含监督信息，又可进一步细分成以下的子问题：1) 源域和目标域都有监督信息的多任务学习[24-28] (Multi-task Learning)；2) 源域无监督信息，目标域有监督信息的自主学习[29] (Self-taught Learning)。直推迁移学习指的是当源域和目标域上有相同学习任务时的迁移问题。目前的研究主要集中在当源域有监督信息，而目标域无监督信息的这类问题上。其中域自适应[30-38] (Domain Adaptation)、方差迁移[39-41] (Covariate Shift) 以及样本选择偏差[42-45] (Sample Selection Bias) 等问题都属于直推迁移学习。一般来说，归纳迁移学习的难度要高于直推迁移学习。这是因为由于源域与目标域上的学习任务不同，归纳迁移学习方法需要提取能在不同域之间起类似作用的高层知识，通常较难约束；而直推迁移学习常基于同类的样本应该被投影到较近的区域这种一般性的先验约束，对齐源域与目标域间的知识。目前研究比较多的归纳迁移学习，一般假设目标域上包含有一定的带监督信息的数据。这是因为如果目标域上没有监督信息，这种情况下的迁移学习难度较大，因此较少有文章涉及。但是实际应用中，目标域通常为变化的。面对不同的目标域就要标注一批目标域样本是不太实际的，更合理的情形是目标域是无监督的。故研究源域是带标注而目标域为无标注情形下的迁移学习更具有实际应用价值。

表格 1-2 迁移学习的分类。其中“☺”表示存在监督信息，“☹”表示无监督信息。

类型类型	$T_s ? T_t$	相关问题	源域监督信息	目标域监督信息
归纳迁移学习 Inductive Transfer Learning	=	Multi-task Learning	☺	☺
		Self-taught Learning	☹	☺
直推迁移学习 Transductive Transfer Learning	\neq	Domain Adaptation, Co-variate Shift, Sample Selection Bias	☺	☹

1.2.2 迁移学习方法

当前的迁移学习方法根据迁移时所对应的机器学习步骤不同（如图 1-2 所示），主要可分成基于两种不同层面上的迁移方法：基于分类器层面[46]和特征层面[31, 34]上的迁移。基于分类器的迁移方法旨在对源域上训练得到的分类器，根据提升其在目标域上的可分性等准则对其进行迁移。基于特征的迁移，一般以寻找到一个能够降低源域与目标域间差异的特征特征子空间为目标，从而使得在该空间上目标域可直接利用源域上的知识。一般可采用将源域和目标域投影到第三个公共子空间或者直接将源域投影到目标域的方式进行迁移。

1.2.2.1 基于分类器的迁移学习

基于分类器的迁移学习方法一般假设源域与目标域具有接近类似的判别分界面，通过对分类器在目标域上判别性能的估计，指导源域上训练得到的分类器进行迁移。

自适应的支持向量机

自适应的支持向量机 [46]（Adaptive Support Vector Machine, 简称 Adaptive-SVM）假设目标分类函数与源域分类函数相差一个“Delta”函数，通过一个类似 SVM 的目标函数估计得到该“Delta”函数，将其加和到源域分类函数上，则可获得目标分类函数，从而完成对源域分类器的迁移。Adaptive-SVM 主要用于处理源域有丰富监督信息而目标域只含有限监督信息情形下的迁移问题。这种情况下，采用迁移学习的必要性是因为：1）目标域的监督信息是有限的，直接利用这批数据无法训练得到可靠的分类器；2）源域虽然有丰富的监督信息，可

训练得到较为可靠的分类器，但不一定能适合目标域上的分类任务。故 Adaptive-SVM 希望借助有丰富监督信息的源域，在目标域上有限的带监督数据的指导下，迁移源域上训练得到的分类器，使其既具有可靠性又更适用于目标域上的分类问题。

具体来说，该方法先在源域上通过标准的 SVM 目标函数，如公式(1-1)所示，优化得到源域分类器 $f^a(x)$ 。

$$\begin{aligned} \min_w \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & \xi_i \geq 0, \end{aligned} \quad (1-1)$$

$$y_i w^T \phi(x_i) \geq 1 - \xi_i, \forall (x_i, y_i) \in \mathcal{D}_S$$

其中， $\sum_{i=1}^N \xi_i$ 用于衡量样本的分类误差， $\|w\|^2$ 是一个与类边缘（margin）成反比的正则项。该目标函数旨在寻找到一个同时具有最小分类错误与最大类边缘这两个性质的分界面 w 。Adaptive-SVM 假设目标域上的分类函数可由源域的分类函数加上一个“Delta”函数得到，即如公式(1-2)所示。

$$f(x) = f^S(x) + \Delta f(x) = f^a(x) + w^T \phi(x) \quad (1-2)$$

目标分类函数应该尽量降低目标域中样本的分类误差并且具有较大的类边缘，Adaptive-SVM 将其形式化成如下所示的目标函数。

$$\begin{aligned} \min_w \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & \xi_i \geq 0 \\ & y_i f^a(x_i) + y_i w^T \phi(x_i) \geq 1 - \xi_i, \forall (x_i, y_i) \in D^T \end{aligned} \quad (1-3)$$

由于源域与目标域间存在的相关性，源域分类函数会是目标分类函数很好的初始化函数，故在有限的带监督目标域样本下借助 Adaptive-SVM 可较快地收敛得到可靠的 $\Delta f(x)$ 函数，最后得到可靠的目标分类函数。

1.2.2.2 基于特征的迁移学习

基于特征的迁移学习旨在通过特征提取、选择等方式，使得在某一个特征空间上，源域与目标域之间的差异足够小，目标域可直接共享源域上的有用信息。常见的这种子空间是源域与目标域共享的部分特征子空间，即公共子空间。基于特征的迁移学习方法关键在于如何度量域间的差异，在此基础上寻找降低域间差异的特征子空间。

迁移成分分析

迁移成分分析（Transfer Component Analysis, 简称 TCA[34]）将源域与目标域投影至再生核希尔伯特空间（Reproducing Kernel Hilbert Space, 简称 RKHS）上，借用最大化中心差异准则（Maximum Mean Discrepancy, 简称 MMD）的推论来定义域间的距离，进而以降低该域间距离为目标来迁移跨域不变的成分。文章[47]中利用 MMD 准则的推论得到在 RKHS 空间两个分布间非参数化的距离估计，即在 RKHS 空间上域间距离等价于域上中心点间的距离。具体地说，域样本为 $X = \{x_1 \ x_2 \ \cdots \ x_{n_1}\}$ 与域样本为 $Y = \{y_1 \ y_2 \ \cdots \ y_{n_2}\}$ 的两个域，其在 RKHS 空间上的距离为，

$$\text{MMD}(X, Y) = \left\| \frac{1}{n_1} \sum_{i=1}^{n_1} \phi(x_i) - \frac{1}{n_2} \sum_{i=1}^{n_2} \phi(y_i) \right\|_{\mathcal{H}}^2 \quad (1-4)$$

其中 $\|\cdot\|_{\mathcal{H}}$ 为 RKHS 空间的范数， $\phi(\cdot)$ 为将样本投影至 RKHS 空间的函数。TCA 的目标为寻找一个投影函数 $\phi(\cdot)$ ，使得源域与目标域间的距离最小化。在 TCA 方法之前，TCA 作者发表在另一篇更早文章[48]中提出的目标函数为，

$$\text{argmin}_{K \succeq 0} \text{tr}(KL) - \lambda \text{tr}(K) \quad (1-5)$$

其中 $\text{tr}(\cdot)$ 为矩阵迹函数，上述公式中的其他项计算方式如下，

$$K = \begin{bmatrix} K_{S,S} & K_{S,T} \\ K_{T,S} & K_{T,T} \end{bmatrix} \subseteq \mathbb{R}^{(n_S+n_T) \times (n_S+n_T)}, \quad (1-6)$$

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

$$L = \begin{bmatrix} L_{S,S} & L_{S,T} \\ L_{T,S} & L_{T,T} \end{bmatrix},$$

$$L_{ij} = \frac{1}{n_S^2}, \text{ if } x_i, x_j \subseteq X_S \quad (1-7)$$

$$L_{ij} = \frac{1}{n_T^2}, \text{ if } x_i, x_j \subseteq X_T; \text{ otherwise } L_{ij} = -\frac{1}{n_S n_T}$$

公式(1-5)第一项旨在最小化两个域间的距离， $\text{tr}(KL)$ 即为公式(1-4)所定义的域间距离；第二项旨在最大化域上方差，以保留域上的信息； λ 为调节参数。目标函数(1-5)可转成 SDP（Semi-definite Program）问题求解。

但该方法由于并不直接估计 $\phi(\cdot)$ ，而是估计源域与目标域样本在 RKHS 空间的核矩阵，故无法处理 out-of-sample 的问题。因此作者对该工作进行改进，提出公式(1-8)所示的目标函数。

$$\begin{aligned} \arg \min_w \quad & \text{tr}(W^T K L K W) + \mu \text{tr}(W^T W) \\ \text{s.t.} \quad & W^T K H K W = I_m \end{aligned} \quad (1-8)$$

与公式(1-5)中的直接估计源域与目标域样本所构成的核矩阵不同, 公式(1-8)中的目标函数是在已知投影至 RKHS 空间核函数 $\phi(\cdot)$ (比如径向基函数等) 的基础上, 旨在优化该核空间上最小化源域与目标域间距离的投影矩阵 W 。故当遇到新样本时, 先通过已知的核函数 $\phi(\cdot)$ 将其投影到该 RKHS 空间, 再通过优化得到的投影矩阵 W 将源域样本投影至与目标域分布较近的子空间上。改进后的目标函数已知核函数 $\phi(\cdot)$, 可处理 out-of-sample 的问题。

Sampling Geodesic Flow

受到增量学习 (Incremental Learning) 的启发, Sampling Geodesic Flow (简称 SGF[31]) 的作者通过估计源域与目标域之间具有渐进式变化的中间子空间, 来提取更多描述目标域的信息。SGF[31] 作者认为这些中间子空间中亦包含了许多关于目标域的信息, 将源域样本投影至这些子空间即可获得更多关于目标域描述的特征, 在此基础上训练得到的模型中由于可利用更多的目标域信息, 故在目标域上可得到更好的性能。

为了能估计这些中间子空间, SGF 从 Grassmann 流形的角度来学习这些中间子空间。具体来说, SGF 将源域 \mathcal{D}_S 和目标域 \mathcal{D}_T 所在的子空间 (比如 PCA 空间 $W \in R^{N \times d}$) 视为 Grassmann 流形[49, 50] $G_{N,d}$ 上的两个点; 则这些中间子空间应位于 $G_{N,d}$ 上连接源域 \mathcal{D}_S 子空间与目标域 \mathcal{D}_T 子空间的几何路径 (Path) 上。在给定 Grassmann 流形上两个点的情况下, 如何估计该流形上两点间路径的方法见该篇文章[51]。在通过文献[51]中提出的方法确定的路径上, 均匀选取路径上的 n 个点 (包含起点与终点), 这些点即为 SGF 所要估计的“潜在的”中间子空间 $\mathbf{S} = \{S_1, S_2, \dots, S_n\}$ 。

在得到这些中间子空间的集合 \mathbf{S} 之后, SGF 通过将源域上的样本投影到这些中间子空间上, 扩充了源域样本的特征向量维数, 使其包含更多目标域的信息。由于扩充特征时并未改变源域上样本的类别信息, 故可在拓展特征后的带标注源域样本上进行模型的学习。投影至该空间的源域样本由于携带更多的目标域信息, 理论上训练得到的模型可更好的应对目标域问题。

该方法中对若采样 n 个中间子空间, 则特征维数扩增了 n 倍, 极大地增加了后续模型训练的计算复杂度。采用比如线性判别分析 (LDA) 这种特征提取方法, 就会涉及到对 $nd \times nd$ 大小的矩阵进行奇异值分解的操作, 其计算复杂度会非常高, 故 SGF 推荐采用偏最小二乘回归 (PLS) 的方法作为后续的判别性特征提取

方法。

形象来看，SGF 的原理如图 1-8 所示。其中绿色的“+”和“×”号分别表示源域上的正负样本；红色的“×”号表示目标域上无标注的样本。将源域与目标域的特征空间视为 Grassmann 流形上的两个点，这两点在该流形上的几何路径，如图中虚线所示。在这条路径上采样 n 个点（包含起点与终点），即得到能够刻画由源域变化到目标域的 n 个渐变式中间子空间。最后将源域样本投影到这些子空间上，即可对其特征进行扩增，使其包含更多的目标域信息。

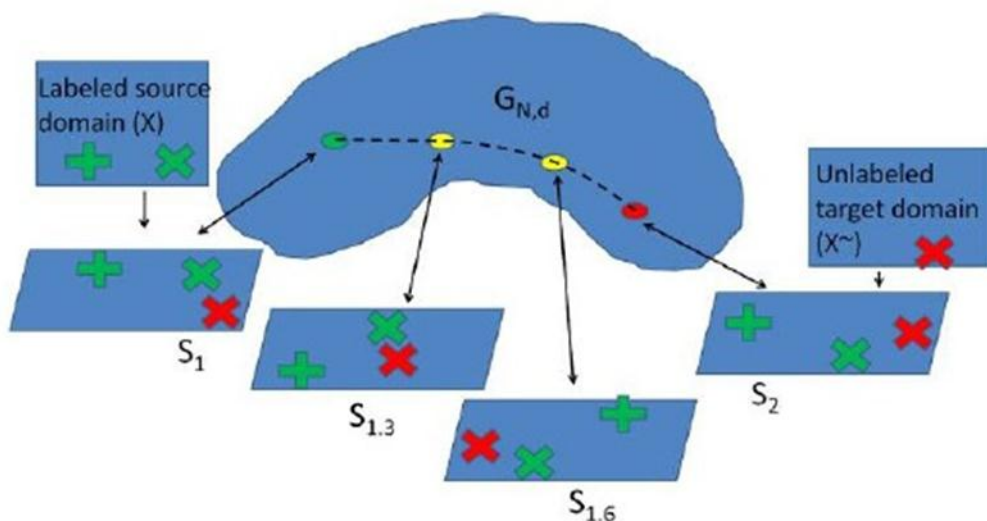


图 1-8 SGF 方法原文的说明图

无监督域适应问题下基于信息论的判别性聚类

2012 年 ICML¹ 会议的这篇文章 [52] 提出了基于信息论（Information-Theoretical Learning）来处理目标域为无标注情况下迁移学习问题的方法，简称 ITL。ITL 利用信息论中的互信息量[53]（Mutual Information）来定义两个变量之间的相关性，比如样本的特征向量与其类别标签之间的相关性、样本的特征向量与域标签之间的相关性。ITL 方法中提及的判别性聚类（Discriminative Clustering）是指在样本间空间位置关系具有一定的判别性情况

¹ ICML 为国际会议 International Conference on Machine Learning 的缩写

的域上进行的聚类。域上样本的空间位置关系具有判别性是指位于同一域上或分别位于不同域上的样本间距离关系都具有以下性质：同类的样本距离近，不同类样本距离远的性质。ITL 方法是在假设源域与目标域上样本间空间位置关系具有上述判别性的基础上提出的。此外，为了使得目标域上的样本可根据其与源域中样本的邻近关系，来估计其类别信息，ITL 进一步假设源域与目标域上的样本都属于相同的 C 类。在上述的两个条件下，ITL 提出利用互信息量来提取能增强域上样本空间位置关系的判别性特征投影的目标函数，使得不论在源域或者目标域上，同类样本非常近，而不同类样本尽量远，从而采用无监督聚类可得到更为可靠的分类。

具体来说 ITL 认为理想的判别性投影应该具备以下两个条件：

1) 目标域样本特征向量与其类别标签之间的互信息量 $I_t(X;Y)$ 尽可能大。这是因为当 $I_t(X;Y)$ 较大时，表明基于该样本的特征向量可较为确定地估计该样本的类别信息，即样本属于某一类的可能最大，而属于其他类的可能非常小。在这种情况下，目标域上的样本空间位置关系极具判别性。具体来说， $I_t(X;Y)$ 的定义如下，

$$\begin{aligned} I_t(X;Y) &= H[p_0] - \frac{1}{M} \sum_t H[p_t] \\ p_0 &= \frac{1}{M} \sum_t p_t \\ p_t &= [p_{t1} \quad p_{t2} \quad \cdots \quad p_{tC}]^T \end{aligned} \quad (1-9)$$

其中 $H[\cdot]$ 为熵函数； p_t 为目标域样本 x_t 属于 C 类的概率向量， p_0 则是目标域上所有样本属于各个类的平均概率，先验上该概率应该较为均匀，即 p_0 的熵应较大。 p_t 中属于每一类的概率可通过以下公式得到。

$$p_{tc} = \sum_{j \neq t} p_{ij} \delta_{jc}, c=1,2,\dots,C. \quad (1-10)$$

其中当源域样本 x_j 属于第 c 类时， δ_{jc} 为 1；否则为 0，故 p_{tc} 为通过目标域样本 x_t 与属于第 c 类的所有源域样本的 p_{ij} 加和得到的。若源域样本 x_j 为目标域样本 x_t 的 k 个源域样本近邻，即 $x_j \in N_k(x_t)$ ， p_{ij} 的计算方式如下所示，

$$\begin{aligned} p_{ij} &= \frac{e^{-d_{ij}^2}}{\sum_{j \neq t(x_t)} e^{-d_{ij}^2}} \\ d_{ij} &= \|Lx_i - Lx_j\|_2^2 \end{aligned} \quad (1-11)$$

若源域样本 x_j 不属于目标域样本 x_t 的 k 个源域样本近邻，即 $x_j \notin N_k(x_t)$ ，

$p_{ij} = 0$ 。

2) 经过投影之后, 源域与目标域上的样本特征向量与其域标签之间的互信息量 $I_{st}(X, Q)$ 尽量小。这是因为 $I_{st}(X, Q)$ 较小时, 表明通过样本的特征向量很难判断出该样本来自哪个域, 即源域与目标域上的样本在该投影空间上交织混合, 无明显的域间差异。具体来说, $I_{st}(X, Q)$ 的计算方式如下所示,

$$I_{st}(X, Q) = H[q_0] - \frac{1}{M+N} \sum_i H[q_i] \quad (1-12)$$

$$q_0 = \frac{1}{M+N} \sum_i q_i$$

其中 q_i 的计算方式与上述计算 p_i 时类似, 只是此时其所属的类别仅为 2 类, 即 q_i 包含样本属于源域或目标域的概率, 其计算公式如下所示,

$$q_{il} = \sum_{j \neq i} p_{ij} \delta_{jl}, l = S, T. \quad (1-13)$$

其中当样本 x_j 的域标签为 l 时, δ_{jl} 为 1; 否则为 0。若样本 x_j 为样本 x_i 的 k 个近邻样本之一, 则 p_{ij} 计算方式如(1-11)所示; 否则为 0。 q_0 为域上所有样本属于源域或目标域的平均概率向量。

ITL 认为目标的判别性特征投影矩阵 L 应在上述的两个互信息量之间进行平衡, 故可得到如下所示的目标函数,

$$\begin{aligned} \min & -I_t(X; Y) + \lambda I_{st}(X; Q) \\ \text{s.t.} & \text{Trace}(L^T L) \leq d \end{aligned} \quad (1-14)$$

其中 λ 为调节参数, 用于平衡这两部分互信息量。通过该目标函数估计得到的投影矩阵既降低了目标域与源域之间的差异, 同时使得目标域样本具有判别性的空间位置关系, 故在此空间上, 对目标域进行无监督聚类得到虚拟标签后进而基于监督特征提取方法 (比如 LDA) 可获取有效的判别性特征, 从而更好地应用目标域上的分类问题。

迁移的特征降维

迁移的特征降维方法[54] (Transferred Dimensionality Reduction, 简称 TDR) 旨在结合源域与目标域上的信息来学习具有判别性的低维嵌入。该文章的作者思路到无标注目标域上样本的空间位置关系一定程度上反应了目标域特有的类别关系。基于这一思路, TDR 采用无监督聚类 (比如 Kmeans[55]) 对目标域上样本间空间位置关系所体现的类别信息进行提取, 同时结合源域上可靠的监督信息, 共同估计目标域上的类内类间散度矩阵。在此基础上 TDR 提出了迭代优化的判别性特征提取方法。

具体来说，TDR 认为一个有效的判别性特征空间应具有以下性质。

1) 最大化目标域上的判别性。TDR 的作者将不同的域视为流形，认为虽然源域与目标域中样本的近邻样本结构有所差异，但域上的样本仍是按照同一个概念进行分类，故判别性的信息可一定程度上在源域与目标域间共享。TDR 方法中认为这部分可共享的判别性信息体现在源域与目标域上不同类样本间的空间位置关系上，这一信息 TDR 采用类间散度矩阵来描述。这部分信息的共享，TDR 将其形式化为加和源域与目标域上的类间散度矩阵的方式来实现，如公式(1-15)所示。其中源域上的类间散度矩阵 S_{b_s} 可根据源域上带标注信息的样本来计算得到可靠的类间散度矩阵；目标域上类间散度矩阵虽然无法直接估计，但由于目标域上的样本间空间位置关系一定程度上也反映了目标域特有的类间关系，故可通过无监督聚类的方式来粗略地估计目标域上样本的类别信息，在此基础上估计得到目标域的类间散度矩阵 S_{b_t} 。根据公式(1-15)估计得到的类间散度矩阵 S_b ，既包含了源域上描述样本类间关系的可靠信息，又包含了目标域特有的类间信息。

$$S_b = S_{b_s} + S_{b_t} = \sum_{i=1}^{C_s} l_i m_{s_i} m_{s_i}^T + \sum_{j=1}^K l_j m_{t_j} m_{t_j}^T \quad (1-15)$$

由于目标域上估计的类间散度矩阵 S_{b_t} 是通过聚类（比如 Kmeans）估计类中心计算得到的，故可将 S_{b_t} 视为关于无监督聚类得到的类指示矩阵 H_u （Indicator Matrix）的函数，即 $S_{b_t}(H_u)$ 。在估计得到一个较为可靠且更为接近描述目标域上类间关系的类间散度基础上，TDR 认为理想的判别性投影应能尽量地最大化该类间散度矩阵。

2) 判别性投影能尽量地保存域上的样本间位置关系。域上样本间的位置关系可采用拉普拉斯矩阵来描述。这一部分信息与两个域上共同的协方差矩阵结合，共同刻画域上的样本结构，如下所示，

$$S_t + \lambda XLX^T = X(I + \lambda L)X^T. \quad (1-16)$$

其中 X 表示包含源域与目标域所有样本的矩阵， L 表示在该全部样本上的拉普拉斯矩阵，关于拉普拉斯矩阵的介绍详见算法 3-1。

基于上述两点对投影矩阵的要求，TDR 将其形式化成一个类似 Fisher 准则的目标函数，如下公式(1-17)所示。

$$\arg \max_{W, H_u} \frac{W^T (S_{b_s} + S_{b_t}(H_u)) W}{W^T (S_t + \lambda XLX^T) W} \quad (1-17)$$

该目标函数中包含两个未知变量 W 与 H_u ，TDR 通过迭代优化的方式进行求解，具体的优化流程如算法 1-1 所述。

算法 1-1 迁移的特征降维（Transferred Dimensionality Reduction）算法

1. 初始化：通过 Kmeans 为目标域上的样本赋予标签，得到 H_u ；计算两个域上所有样本的拉普拉斯矩阵 L 。
2. 监督地降维： 在已知 H_u 的情况下，优化目标函数 $\arg \max_w \frac{W^T (S_{b_s} + S_{b_t}) W}{W^T (S_t + \lambda X L X^T) W}$ 取其前 $m = C + K - 1$ 维的特征向量作为降维投影矩阵。
3. 紧致目标域样本的聚类： 在第 2 步优化得到低维空间中进行无监督的 Kmeans 聚类，更新样本类中心的指示矩阵 H_u ，Kmeans 聚类的目标函数可形式化成如下所示的公式，即求解可使样本在该空间下具有最大类间散度矩阵的指示矩阵。 $\arg \max_{H_u} S_{b_t} (H_u)$
4. 若第 3 步更新的指示矩阵 H_u 与第 2 步的指示矩阵变化不大，则停止迭代优化，进入第 5 步；否则回到第 2 步。
5. TDR 低维嵌入 TDR 最后得到的低维嵌入为上述第 2 步优化得到的 m 维投影矩阵 W 。将目标域的样本投影至该空间可得到具有判别性的样本分布，即， $x \rightarrow z = W^T x。$ 在此基础上采用类似最近邻分类器等方式可将目标域样本进行有效的分类。

1.3 本文的研究内容及贡献

1.3.1 研究内容

在实际应用中，人脸识别技术通常会由于训练数据与目标数据之间存在显著的差异，而导致模型的性能下降。针对这样的问题，本研究采用迁移学习的方式来应对该问题。不同情况下的迁移学习问题，可采用不同的迁移学习方法。

当源域与目标域间有较大的重叠时，即源域有许多样本可被目标域上的任务重复利用，那么可采用对源域样本进行加权（Instance Reweighting）[56]的方式进行迁移学习。一般是对源域中较为靠近目标域的样本赋予较高的权重，对不像目标域的样本赋予较低的权重，从而降低源域中对目标域上学习任务具有不利

作用的样本的影响。这一类的迁移问题可采用样本采样偏差（Sample Selection Bias）[42, 43, 45]或者协方差矩阵偏移（Covariate Shift）[39-41]的方法进行迁移学习。但是当源域与目标域无法直接通过对源域样本进行加权从而降低样本间的差异时，这一类的迁移学习方法就无法适用。在如图 1-3 所示的迁移问题中，并无法保证源域中一定包含可被重用的样本，故这一类的方法并不适用于这类迁移设置下的人脸识别问题。

进一步来说，若源域与目标域之间存在些浅显易捕捉的先验，比如源域与目标域的模型具有相同的函数形式，并且共享某些参数或者分界面，则可利用这些先验，借助目标域样本来对源域模型进行修正或更新，使其能更好的应对目标域任务。针对这一类的迁移学习问题有以下的工作[37, 54, 57]。但在实际应用中，源域与目标域通常相差较大，比如位于不同的子空间且域上样本分布还不太相同。这种情况下，域间可共享的信息可能无法浅显地捕获或形式化成有用先验。在这种情况下，需要对域间可共享的信息进行深度提取。本研究所着眼的问题中在域与目标域极度相关，都是人脸图像，但仍然具有显著差异的情形上，比如不同的种族、采集条件等。当前针对该问题的研究主要集中在寻找源域与目标域间公共的特征子空间方法上。这是因为公共子空间上源域与目标域具有非常接近，该空间上对源域上学习任务有利的信息，通常来说对于目标域任务也是可靠、有利的。但是公共子空间毕竟只包含了目标域上部分的信息，缺少目标域特有的部分信息，而这部分信息对于解决目标域上的人脸识别问题亦可能有其贡献。

在分析与吸收前人工作的基础上，本研究从两个不同的角度提出不同的迁移学习方法来应对如图 1-5 所示的人脸识别问题：1）从利用源域的监督信息来指导目标域上类别信息估计的角度；2）从迁移源域中可靠有用信息的角度。

1.3.2 研究贡献

从解决由于源域与目标域存在差异而导致模型性能下降的问题出发，本文首次在人脸识别领域引入迁移学习的思路，并提出可行的方法来提升人脸识别模型在目标域上的性能。

在分析了当前迁移学习工作的基础上，本研究分别从两个角度出发，提出了更为有效的迁移学习方法来提升人脸识别性能。

与一般的迁移学习都只从提取源域上有用的信息不同，本研究考虑了以源域作为以目标域无监督聚类的验证域角度出发，提出了基于目标域相似度矩阵估计的判别性特征提取方法。

针对迁移学习方法一般只利用到源域与目标域间公共的子空间的情况，本文从两个角度提出了如何全面目标域整个特征子空间上有用信息的方法：1）基于

源域样本目标化的人脸识别方法；2) 融合公共与特有子空间判别性特征提取的迁移方法。

1.4 论文的组织

第 1 章 为绪论，主要介绍了本研究的研究背景，国内外目前的研究现状以及本文的研究内容及贡献。

第 2 章 介绍提出的基于目标域相似度矩阵估计的判别性特征提取方法。首先介绍相关的线性判别分析、低秩表示等背景知识，随后分析迁移思路，以及该方法的应用框架，然后具体介绍了该方法的目标函数、优化方法与实现算法，最后介绍了相关的实验部分。

第 3 章 介绍提出的基于源域样本目标化的人脸识别方法。首先介绍相关的局部保持投影、稀疏编码这两个背景知识，然后分析迁移思路以及方法框架，再具体介绍该方法中的核心目标函数；随后介绍该目标函数的优化方法与算法流程，并且汇报了该方法在不同迁移问题下的性能。

第 4 章 介绍提出的融合公共与特有子空间判别性特征提取的迁移方法。首先介绍相关的主成份分析这一背景知识，然后依序介绍该方法的基本思路、方法框架、目标函数、优化方法与算法流程，再汇报了相关的实验部分。

第 5 章 为结束语。对本文进行全文总结，并且阐述了下一步的工作展望。

第2章 基于目标域相似度矩阵估计的判别性特征提取方法

由于目标域上样本的空间位置关系未必能强烈地反映身份信息的异同，故直接采用无监督聚类得到的聚类结果未必是基于身份信息的分类。另一方面源域上所包含的标注信息却是基于身份信息的。这表明在某一能反映身份信息异同的特征子空间上，源域与目标域上的样本能同时在该空间上具有判别性的分布。基于这一思路，本章提出的方法中将源域上带有身份信息的样本作为评价准则，用于指导目标域上相似度矩阵的估计，目标的相似度矩阵能反映目标域上样本身份信息的异同，在此基础上再利用 LDA 提取判别性的特征。

与一般的迁移学习方法都是基于提取源域中有效信息的角度不同，本章提出的方法中，源域上的监督信息是用于评价目标域上估计得到的相似度矩阵反映的类别信息是否与其一致，并不直接利用源域上的信息。

2.1 相关背景知识

2.1.1 线性判别分析

线性判别分析（Linear Discriminative Analysis，简称 LDA）是一种有监督的特征提取方法，广泛应用在统计、模式识别和机器学习等领域。LDA 因为目标函数简单、具有解析解，得到的性能通常与 state-of-the-art 的方法具有可比性。因此，LDA 一般作为有监督特征提取方法的基线（Baseline）方法。在人脸识别领域，LDA 是非常常用的对比方法。

LDA 方法的核心思想是，希望经过特征提取后，样本的空间位置关系具有判别性，即同类样本尽量近，不同类样本尽量远。类内类间的样本空间位置关系可采用类内与类间散度矩阵来描述，因此，LDA 方法的目标函数为，寻找一个特征投影矩阵 W ，经过 W 投影后，在投影空间上类间散度矩阵尽量大，类内散度矩阵尽量小，这一准则学术上称之为 Fisher 准则，具体目标函数如下公式所述。

$$W = \arg\max \frac{W^T S_b W}{W^T S_w W} \quad (2-1)$$

其中 w 即为所求的线性判别投影方向； S_w 为类内散度矩阵， S_b 为类间散度矩阵，求解方式如公式(2-2)所示。其中 $m_i \Big|_{i=1}^{NC}$ 为第 i 类的中心点， m_o 为全部样本的中心点。

$$S_w = \sum_{i=1}^{NC} \sum_{j=1}^{c_i} (x_{ij} - m_i)(x_{ij} - m_i)^T, \quad (2-2)$$

$$S_b = \sum_{i=1}^{NC} n_i (m_i - m_o)(m_i - m_o)^T. \quad (2-3)$$

LDA 的目标函数(2-1)的求解问题可转换成求解如公式(2-4)所示的广义特征向量问题。

$$S_b w_i = \lambda S_w w_i \quad (2-4)$$

由于 LDA 方法其目标函数为 Fisher 准则，故该方法又称 Fisher's LDA，简称 FLD。理论上表明特征值越大，其所对应的特征向量判别性越大。因此应用 LDA 选取投影矩阵时，都会按特征值从大到小对特征向量进行排序；若要投影至 d 维空间，则选取前 d 个特征值所对应的特征向量，构成投影矩阵。一般实际应用中，原始特征维数远大于类别数 NC ，这种情况下 LDA 有效的特征向量维数为 $NC-1$ 维，其中 NC 为所有样本的类别数。

在人脸识别领域 Belhumer[58]等人将先利用 PCA 降维后再基于 LDA 提取的判别性特征称之为 Fisherfaces。LDA 作为一种简单而有效的有监督特征提取方法，在人脸识别领域被广泛应用。

2.1.2 低秩表示

给定数据矩阵 $X = [x_1 \ x_2 \ \cdots \ x_n]$ ，其中 $x_i \in R^p \Big|_{i=1}^n$ 为单个样本的特征列向量，则由词典 D 对数据矩阵 X 进行表示，其目标函数如下，

$$X = DZ, \quad (2-5)$$

其中 $Z = [z_1 \ z_2 \ \cdots \ z_n]$ 为系数矩阵， z_i 为在词典 D 上 x_i 的表示系数。若对系数矩阵 Z 加上低秩的约束，则可形式化成如下所示的目标函数，

$$\begin{aligned} \min_Z \text{rank}(Z) \\ \text{s.t. } X = DZ \end{aligned} \quad (2-6)$$

包含对矩阵求秩的目标函数是一个非凸函数，求解较为困难。这篇文章[59]中证明表示矩阵的核范数 $\|\cdot\|_*$ 是矩阵秩的良好估计，且该范数的计算是凸的，可更好地优化求解。故目标函数(2-6)的解可由下述公式的解良好逼近。

$$\begin{aligned} \min_Z \|Z\|_* \\ \text{s.t. } X = DZ \end{aligned} \quad (2-7)$$

对样本矩阵 X 进行表示的词典 D 一般是通过迭代优化或者无监督聚类得到

的。根据文章[60]中的推导表明，若直接采用矩阵 X 作为词典，对自身进行低秩表示，如公式(2-8)所示，得到的系数矩阵 Z 一定程度上可表示 X 矩阵中样本间的亲疏关系。

$$\begin{aligned} \min_Z \|Z\|_* \\ \text{s.t. } X = XZ \end{aligned} \quad (2-8)$$

矩阵 X 中可能包含些噪声样本，为了增强系数矩阵 Z 对噪声的鲁棒性，则可允许对 X 的某些样本进行低秩表示时存在一定的重构误差，可形式化成如下所示的目标函数，

$$\begin{aligned} \min_{Z,E} \|Z\|_* + \lambda \|E\|_{2,1} \\ \text{s.t. } X = XZ + E \end{aligned} \quad (2-9)$$

其中 λ 为调节参数，用于平衡这两部分范数所起的作用，当 λ 取相对较大值时，目标函数(2-9)等价于目标函数(2-8)[60]；另外 $\|\bullet\|_{2,1}$ 称之为 $l_{2,1}$ -范数，其计算方式如下，

$$\|E\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^n (E_{ij})^2} \quad (2-10)$$

$\|E\|_{2,1}$ 控制系数矩阵 Z 对样本矩阵 X 的表示，希望大部分的样本都能被完美表示，而对某些噪声样本则允许存在误差。

根据文章[60]中介绍，在优化目标函数(2-9)时，为了求解方便，可将其转化成如下所示的目标函数，

$$\begin{aligned} \min_{Z,E,J} \|J\|_* + \lambda \|E\|_{2,1} \\ \text{s.t. } X = XZ + E, \\ Z = J \end{aligned} \quad (2-11)$$

该目标函数可由增强的拉格朗日方法[61, 62]（Augmented Lagrange Multiplier，简称 ALM）求解，具体的求解步骤可参见文章[60]。该具体求解过程中，涉及到求解一个包含弗罗贝尼乌斯范数（Frobenius 范数）与 $l_{2,1}$ 范数的目标函数，如(2-12)所示。

$$W^* = \min_W \lambda \|W\|_{2,1} + \frac{1}{2} \|W - Q\|_F^2 \quad (2-12)$$

该优化问题的最优解为，

$$W^*(:,1) = \begin{cases} \frac{\|q_i\| - \lambda}{\|q_i\|} * q_i & \text{若 } \lambda < \|q_i\| \\ 0 & \text{其它} \end{cases} \quad (2-13)$$

其中 q_i 为矩阵 Q 的第 i 个列向量。

2.1.3 The Singular Value Shrinkage Operator

奇异值减缩算子 (The Singular Value Shrinkage Operator) 是对给定矩阵的奇异值进行缩减的一种操作。具体来说, 给定矩阵 $X \in \mathbf{R}^{m \times n}$, 其秩为 r , 首先对 X 进行奇异值分解, 得到如下等式,

$$X = U \Sigma V^T, \Sigma = \text{diag}(\{\sigma_i\}_{1 \leq i \leq r}). \quad (2-14)$$

其中 U 、 V 分别为 $m \times r$ 、 $n \times r$ 的正交矩阵, σ_i 为 X 的正数奇异值。记对 X 进行奇异值缩减算子的操作为 $S_\tau(X)$, 其计算方式如下,

$$\begin{aligned} S_\tau(X) &= U S_\tau(\Sigma) V^T \\ S_\tau(\Sigma) &= \text{diag}(\{\sigma_i - \tau\}_{+, 1 \leq i \leq r}) \end{aligned} \quad (2-15)$$

其中, $\{\sigma_i - \tau\}_+$ 操作表示取 $\max(\sigma_i - \tau, 0)$ 。由于该操作将奇异值中小于 τ 的部分置为 0, 其他大于 τ 的值也向 0 方向缩减, 故称之为奇异值缩减算子 (The Singular Value Shrinkage Operator)。

根据文章[63]中的推导, 奇异值缩减操作是如公式(2-16)所示的目标函数的最优解。

$$S_\tau(X) = \arg \min_X \left\{ \frac{1}{2} \|X - Y\|_F^2 + \tau \|X\|_* \right\} \quad (2-16)$$

2.2 基本思路与方法框架

2.2.1 基本思路

传统的机器学习方法只能单独地利用有监督的源域或无监督的目标域来训练模型, 得到的模型无法很好地应对目标域上的分类任务。具体来说, 单独利用有监督源域的方法是直接将源域上训练得到的有监督模型应用到目标域上。源域上训练得到的模型是适用于源域问题的, 但由于源域与目标域间存在差异, 该模型并不一定适用目标域上的问题, 从而导致该模型在目标域上的性能下降, 这一现象已由图 1-6、图 1-7 所示的实验结果证明。单独利用无监督目标域的方法一般是首先对目标域进行无监督的聚类, 为目标域样本赋值虚拟的“样本”, 将目

标域由起始的无监督域变成带“监督”的域，然后再利用监督的机器学习方法来提取具有判别性的特征或分类器。但是这种应用方法，由于目标域上的无监督聚类所提取的类别信息未必与身份信息相一致，例如可能得到按照光照、表情等的异同而分类，因此基于此估计的类别信息进行的监督学习方法提取得到的信息未必对目标域上的人脸识别问题有利，一般来说，无法得到可靠理想的性能。

由于身份信息是源域与目标域共同的分类准则，因此可利用源域样本上的监督信息来指导目标域上的类别估计，使其反映身份信息的异同。目标域上样本间的类别信息可采用样本间的相似度矩阵 Z 来表示，要求具有高度判别性的相似度矩阵只在属于同一类的样本间有非零相似度，而在不同类样本间相似度为 0。因此，可利用源域样本上的监督信息来指导目标域上具有高度判别性的相似度矩阵的估计，使其反映与源域标注信息一致的类别信息（身份信息）。本研究借助一个公共子空间来实现源域监督信息对目标域上相似度矩阵估计的指导。具体来说，目标域与源域上类别信息的一致性表明目标域与源域具有较为一致的判别性公共子空间，即在该公共子空间上源域与目标域同时具有较好的判别性。

其中源域样本在某一空间的判别性程度可利用源域上带标注信息的样本进行评价。本研究选用机器学习中常用的 Fisher 准则来度量监督样本在某一子空间上的判别性程度，该准则的介绍见第 2.1.1 节。目标域样本在该子空间上的判别性程度，可利用在该子空间上目标域样本间相似度矩阵的秩来体现。这是因为若该相似度矩阵只在同类样本间具有极高的相似度，都接近某一极大值，而在不同类样本间相似度都为 0，这样的相似度矩阵具有低秩的性质，其秩的大小反映了该域上判别性的程度。此外，该相似度矩阵所描绘的每个类的中心应与属于该类的样本距离较小，以防止求解得到退化的相似度矩阵（比如秩为 1）。这个约束可形式化为该目标域样本矩阵结合该相似度矩阵得到的样本矩阵与目标域样本间的误差应尽量较小。综上关于目标域相似度矩阵 Z 的性质，该相似度矩阵应为一个低秩矩阵，同时尽量满足 $U_t = U_t Z$ ，其中 U_t 为目标域样本矩阵。

综上分析，本章提出同时优化公共子空间与高度判别性的目标域相似度矩阵的目标函数。基于该目标函数估计的相似度矩阵虽然并未直接为目标域上的样本赋予标签，但能非常清晰地描绘目标域上样本间身份信息的异同，故基于该相似度矩阵可计算得到的能反映身份信息异同的类内类间散度矩阵。在此基础上可基于 Fisher 准则提取到适用与目标域上人脸分类的判别性特征。本研究将该方法称之为基于目标域相似度矩阵估计的判别性特征提取，即 Discriminative Feature Extraction based on Target Affinity Matrix Estimation，简称 TAME。

相较于源域上提取的判别性特征，本方法提取的判别性特征是直接在目标域上提取的，该空间上的特征可更好地反映目标域上样本间身份信息的异同。另一

方面，本方法提取的判别性特征是基于源域上监督信息的指导下学习得到的，相较于直接在目标域上基于无监督聚类提取特征，本方法提取得到的特征更加可靠。

2.2.2 方法框架

本章提出的判别性特征提取方法，其应用的方法框架如图 2-1 所示。首先通过 TAME 的目标函数(2-17)求解得到能反映身份信息的判别性相似度矩阵 Z ；在此基础上，将该相似度矩阵应用在目标域的原始空间上，估计能反映身份信息的类内类间散度矩阵；最后基于 Fisher 准则提取判别性的特征。在求解得到相似度矩阵 Z 后亦可采用比如无监督聚类的方式，对目标域上样本进行分类。但由于低秩表示估计得到的相似度矩阵具有较强的判别性，故可直接基于该矩阵估计类内类间散度矩阵，从而提取判别性的特征。

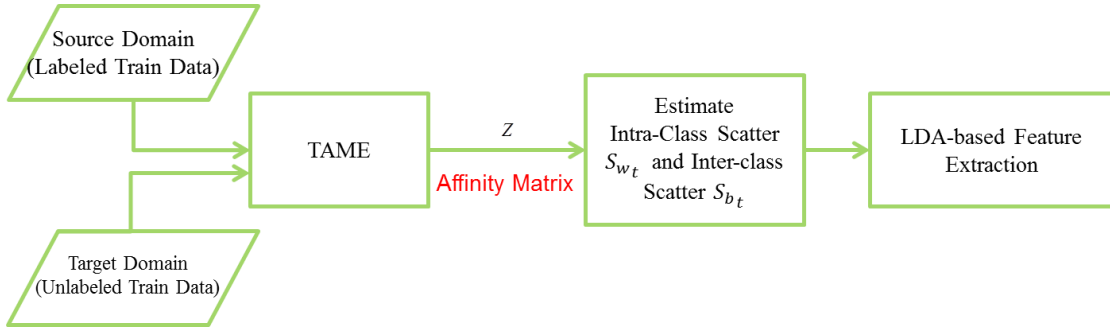


图 2-1 诱导相似度矩阵估计（TAME）框架图

2.2.3 目标函数

记目标域上所有样本构成的矩阵为 $U_t = [u_1 \ u_2 \ \cdots \ u_{n_t}]$ 上，其每一列为一个样本的特征向量，其中 n_t 为目标域上的总样本数。基于上述的分析，即寻找某一特征子空间，在该特征子空间上源域与目标域上样本间的空间位置关系都具有较好的判别性，记 W 为由原始空间投影至该特征子空间的投影矩阵。样本间空间位置关系的判别性，源域上采用 Fisher 准则来估计，即 $\text{tr}\{W^T (S_{w_s}^{-1}) S_{b_s} W\}$ ，其中 S_{w_s} 、 S_{b_s} 分别为源域上的类内、类间散度矩阵；目标域上采用如第 2.1.2 节中介绍的低秩表示系数矩阵的秩来估计。综上所述，则可形式化得到如(2-17)所示的目标函数。

$$\begin{aligned}
 W, Z = \arg \min_{W, Z} & \|Z\|_* + \lambda \|E\|_{2,1} - \tau \text{tr}\{W^T (S_{w_s}^{-1}) S_{b_s} W\} \\
 \text{s.t. } & WU_t = WU_t Z + E \\
 & \|w_i\| = 1, i = 1, \dots, c_s - 1
 \end{aligned} \tag{2-17}$$

其中 Z 即为所要求的目标域相似度矩阵。该目标函数中存在两个未知量 W 、

Z ，可采用迭代优化的方法进行求解，其中要求投影矩阵 W 列向量长度为 1 是为了防止求解得到退化的投影矩阵，如零矩阵。 τ 为调节参数，用于平衡源域与目标域上的判别性程度。 λ 是用于增强低秩表示鲁棒性的调节参数，较小的 λ 值使得低秩表示时允许对某些样本的重构存在误差。

该目标函数同时优化得到一个由投影矩阵 W 所确定的特征子空间以及能反映身份信息异同的目标域相似度矩阵 Z 。投影矩阵 W 所确定的特征子空间使得目标域与源域上的样本空间位置关系同时能反映一定的身份信息，但由于原始空间中源域与目标域上样本的近邻位置关系有较大的差异，该特征子空间是在源域与目标域之间进行了折中得到的结果，并不是最适合目标域上分类任务的特征子空间。另一方面，估计得到的相似度矩阵 Z 则是纯粹反映了目标域样本基于身份信息异同的相似度矩阵。故可基于该相似度矩阵，将其应用在原始空间，结合原始空间上目标域样本间的关系，提取得到更适合目标域上人脸识别的判别性特征。基于文章[60]中的分析，采用自身矩阵作为词典进行低秩表示的系数矩阵 Z 经过对样本位置的整理后，可得到如(2-18)所示的形式，

$$Z = \begin{bmatrix} Z_1 & 0 & 0 & 0 \\ 0 & Z_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & Z_r \end{bmatrix} \quad (2-18)$$

即属于同一类的样本间存在非零的相似度矩阵 Z_i ，而不同类的样本间其相似度为 0。该矩阵反映了某一空间上样本间的亲疏关系，同时极具判别性。其中 r 为系数矩阵 Z 的秩， Z_1, Z_2, \dots, Z_r 为 r 个非零子矩阵，同时有满足 $\text{rank}(Z_i) = 1$ 。为了说明方便，下文即在 Z 为如(2-18)所示的情况下进行分析。 $Z_i|_{i=1}^r$ 中非零系数所对应的样本非常相似，故在词典编码时才会互相表示；此外 $Z_i|_{i=1}^r$ 的秩为 1 是因为相似的样本，其与词典 U_i 中每个样本间的关系亦非常接近。具体来说，若相似度子矩阵 $Z_i|_{i=1}^r$ 中涉及 c_i 个样本，记为 U_{t_i} ， $Z_i|_{i=1}^r$ 本身可写成如下公式所示的形式，

$$Z_i = \begin{bmatrix} z_i & z_i & \cdots & z_i \end{bmatrix}_{c_i \times c_i} \quad (2-19)$$

$$z_i = \begin{bmatrix} z_{i1} & z_{i2} & \cdots & z_{ic_i} \end{bmatrix}^T$$

其中 z_i 为相似度子矩阵 Z_i 中的基本列向量， z_i 中的每一元素体现了待表示样本与词典 U_i 中所对应的样本间相似度。在既要保证重构误差尽量小，又保证相似度子矩阵 Z_i 的秩为 1 的情况下， $U_{t_i} z_i$ 应是 U_{t_i} 中样本加权后的中心点。故可通过公式(2-20)计算得到目标域上不同对象的中心点。

$$M = U_i Z = \begin{bmatrix} \underbrace{m_1 \cdots m_1}_{c_1} & \underbrace{m_2 \cdots m_2}_{c_2} & \cdots & \underbrace{m_r \cdots m_r}_{c_r} \end{bmatrix} \quad (2-20)$$

相应地，每个样本到其加权的类中心距离 B 计算如下，

$$B = U_t - M. \quad (2-21)$$

在已知类中心的情况下，目标域上的类内散度矩阵 S_{w_t} 与类间散度矩阵 S_{b_t} 可由下列公式计算，其中 u_{ij} 表示相似度子矩阵 Z_i 第 j 列元素所对应的目标域样本。

$$\begin{aligned} S_{b_t} &= M \times M^T = \sum_{i=1}^r c_i m_i m_i^T \\ S_{w_t} &= B \times B^T \\ &= (U_t - M)(U_t - M)^T \\ &= \sum_{i=1}^r \sum_{j=1}^{c_i} (u_{ij} - m_i)(u_{ij} - m_i)^T \end{aligned} \quad (2-22)$$

在估计得到能够反映身份信息的类内散度矩阵 S_{w_t} 与类间散度矩阵 S_{b_t} 基础上，可应用 Fisher 准则来提取具有判别性的特征，具体介绍详见第 2.1.1 节。

2.3 优化方法与算法流程

关于目标函数为(2-17)的求解方式，我们采用基于拉格朗日乘子的方式进行迭代优化求解。根据[60]文章中的分析，可将原目标函数改写成如(2-23)所示的目标函数进行求解。

$$\begin{aligned} \min_{Z, E, J, W, Y_1, Y_2} & \|J\|_* + \lambda \|E\|_{2,1} - \tau \operatorname{tr}\{W^T (S_{w_s}^{-1} S_{b_w}) W\} \\ \text{s.t.} & W^T U_t = W^T U_t Z + E \\ & Z = J \end{aligned} \quad (2-23)$$

在此基础上，采用拉格朗日乘子方法，将带约束的目标函数(2-23)整理成一个如(2-24)所示的目标函数。

$$\begin{aligned} \min_{Z, E, J, W, Y_1, Y_2} & \|J\|_* + \lambda \|E\|_{2,1} - \tau \operatorname{tr}\{W^T (S_{w_s}^{-1} S_{b_w}) W\} \\ & + \langle Y_1, W^T U_t - W^T U_t Z - E \rangle + \langle Y_2, Z - J \rangle \\ & + \frac{\mu}{2} (\|W^T U_t - W^T U_t Z - E\|_F^2 + \|Z - J\|_F^2) \end{aligned} \quad (2-24)$$

该目标函数可采用迭代更新不同的对象的策略进行求解，直至迭代收敛。下文介绍不同对象的优化方式。

1. 固定其他，更新 J

当固定其他时，保留下原目标函数中关于 J 的项，则可得到如下所示的关于 J 的目标函数，

$$J = \arg \min_J \|J\|_* + \langle Y_2, Z - J \rangle + \frac{\mu}{2} \|Z - J\|_F^2. \quad (2-25)$$

对其进行整理，可得如公式所示的目标函数，

$$J = \arg \min_J \frac{1}{2} \left\| J - \left(Z + \frac{1}{\mu} Y_2 \right) \right\|_F^2 + \frac{1}{\mu} \|J\|_*. \quad (2-26)$$

该目标函数可根据 2.1.3 节中介绍的奇异值缩减算子求解，其解如公式(2-33)所示。

2. 固定其他，更新 Z

当固定其他项不变时，包含原目标函数中关于 Z 的项，得到如下所示的关于 Z 目标函数，

$$\begin{aligned} Z = \arg \min_Z & \langle Y_1, W^T U_t - W^T X U_t - E \rangle \\ & + \langle Y_2, Z - J \rangle + \frac{\mu}{2} \left(\|W^T U_t - W^T U_t Z - E\|_F^2 + \|Z - J\|_F^2 \right), \end{aligned} \quad (2-27)$$

对其进行整理可得如公式所示的目标函数，

$$Z = \arg \min_Z \left\| W^T U_t Z - \left(W^T U_t - E + \frac{1}{\mu} Y \right) \right\|_F^2 + \left\| Z - \left(J - \frac{1}{\mu} Y_2 \right) \right\|_F^2. \quad (2-28)$$

对再一步整理，可得到如下目标函数，

$$Z = \arg \min_Z \left\| \begin{bmatrix} W^T U_t \\ I \end{bmatrix} Z - \begin{bmatrix} W^T U_t - E + \frac{1}{\mu} Y_1 \\ J - \frac{1}{\mu} Y_2 \end{bmatrix} \right\|_F^2. \quad (2-29)$$

这是一个关于 Z 的最小二乘目标函数，其解如公式(2-35)所示，

3. 固定其他，更新 E

当固定其他不变时，包含原目标函数中关于 E 的项，得到如下所示的关于 E 目标函数，

$$\begin{aligned} E = \arg \min_E & \lambda \|E\|_{2,1} + \langle Y_1, W^T U_t - W^T U_t Z - E \rangle \\ & + \frac{\mu}{2} \|W^T U_t - W^T U_t Z - E\|_F^2. \end{aligned} \quad (2-30)$$

对其进行整理，可得到如所示的目标函数，

$$E = \arg \min_E \frac{1}{2} \left\| E - \left(W^T U_t - W^T U_t Z + \frac{1}{\mu} Y_1 \right) \right\|_F^2 + \frac{\lambda}{\mu} \|E\|_{2,1} \quad (2-31)$$

该目标函数可根据公式(2-12)(2-13)求解, 得到如公式(2-36)(2-37)所示的解。

4. 固定其他, 更新 W

当固定其他不变时, 包含原目标函数中关于 W 的项, 得到如下所示的关于 W 目标函数,

$$\begin{aligned} W = \arg \min_W & -\tau \operatorname{tr} \left\{ W^T \left(S_{w_s}^{-1} S_{b_w} \right) W \right\} + \langle Y_1, W^T U_t - W^T U_t Z - E \rangle \\ & + \frac{\mu}{2} \|W^T U_t - W^T U_t Z - E\|_F^2 \\ \text{s.t. } & w_i = \frac{w_i}{\|w_i\|}, i = 1, 2, \dots, c_s - 1 \end{aligned} \quad (2-32)$$

该目标函数是个在单位球上求解最优值的问题, 采用文章[64]中提供的方法进行求解。

综上关于目标函数(2-23)中各项的求解, 其完整的求解过程如算法 2-1 所示。

算法 2-1 基于目标域相似度矩阵估计的判别性特征提取 (TAME)

输入: 目标域样本矩阵 $U_t = [u_1 \ u_2 \ \cdots \ u_{n_t}]$, 其中每个列向量为目标域上的一个样本, 源域上的类内散度矩阵 S_{w_s} 、类间散度矩阵 S_{b_s} , 调节参数 λ 和 τ , 收敛精度 ε 。

初始化: $Z = J = \mathbf{0}_{n_t \times n_t}$, $E = \mathbf{0}_{d \times n_t}$, W 可为随机矩阵, 亦或取源域上的 FisherLDA 矩阵作为初始化 (列向量长度归一为 1), $Y_1 = 0$, $Y_2 = 0$, $\mu = 10^{-6}$, $\rho = 1.1$, $\varepsilon = 1$

循环直至收敛

1. 固定其他, 更新 J

根据公式(2-33)更新 J , 其中 $S_{1/\mu}$ 为奇异值缩减算子, 具体运算介绍见第 2.1.3 节。

$$\begin{aligned} J &= S_{1/\mu} \left(Z + \frac{1}{\mu} Y_2 \right) \\ &= S_{1/\mu} (U \Sigma V^T) \\ &= U \operatorname{diag} \left(\{ \sigma_i - \tau \}_{+, 1 \leq i \leq r} \right) V^T \end{aligned} \quad (2-33)$$

其中 U 、 Σ 、 V 为对 $Z + \frac{1}{\mu} Y_2$ 进行奇异值分解所得的矩阵, 即有,

$$Z + \frac{1}{\mu} Y_2 = U \Sigma V^T \quad (2-34)$$

2. 固定其他, 更新 Z

$$Z = (U_t^T W W^T U_t + I)^{-1} \left(U_t^T W W^T U_t - U_t^T W E + J + \frac{1}{\mu} (U_t^T W Y_1 - Y)_2 \right) \quad (2-35)$$

3. 固定其他，更新 E

$$E^*(:,1) = \begin{cases} \frac{\|q_i\| - \lambda}{\|q_i\|} * q_i & \text{若 } \lambda < \|q_i\| \\ 0 & \text{其它} \end{cases} \quad (2-36)$$

其中，

$$\begin{aligned} Q &= W^T U_t - W^T U_t Z + \frac{1}{\mu} Y_1 \\ &= [q_1 \quad q_1 \quad \cdots \quad q_{n_t}] \end{aligned} \quad (2-37)$$

4. 固定其他，更新 W

采用文章[64]中提出的算法求解目标函数(2-32)

5. 更新乘法子

$$\begin{aligned} Y_1 &= Y_1 + \mu (W^T U_t - W^T U_t Z - E) \\ Y_2 &= Y_2 + \mu (Z - J) \end{aligned} \quad (2-38)$$

6. 更新参数 μ

$$\mu = \min(\rho\mu, \max_{\mu}) \quad (2-39)$$

7. 检查收敛条件

若得到的 W 、 Z 、 J 和 E 满足如下条件，则收敛，结束循环；反之回到步骤 1。

$$\|W^T U_t - W^T U_t Z - E\|_{\infty} < \varepsilon \text{ and } \|Z - J\|_{\infty} < \varepsilon \quad (2-40)$$

结束循环

输出： Z

2.4 实验及性能比较

2.4.1 实验数据库及协议

不同的人脸数据库采集了来自不同对象的、包含不同脸部变化的人脸图像，比如不同的光照条件、表情、姿态等变化，可用于检测在不同条件下的人脸识别模型的性能。选取包含不同人脸识别的数据库，可用于模拟源域与目标域存在差异的人脸识别问题。本次实验预期将用到以下 4 个人脸数据库，为了使得数据库间的差异较为可控，我们在这些数据库上挑选了部分图像，分别构成训练集、

Probe 集合与 Gallery，其选取过程介绍如下所示。

OFD[20]: 该数据库上的人脸图像主要包含光照和姿态变化。本研究随机选取了 OFD 数据库中的 800 个对象，每个对象选取 4 幅正面、中性表情、略带光照变化的图像构成实验的 OFD 数据集。具体来说，在这批图像上选取 400 个对象的图像作为训练数据集，剩下的 400 人的图像作为测试数据集。其中在测试集上，任意选取 1 幅构成 Gallery，剩下的 3 幅构成 Probe 集合。OFD 数据库上的一些人脸图像如图 2-2 所示。

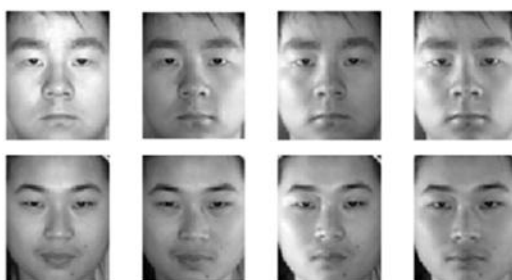


图 2-2 OFD 数据库的示例图

XM2VTS[22]: 该数据库上的人脸图像主要包括光照、姿态、表情等变化。本研究实验时选用每个对象 4 个时期中各两幅（即一人 8 幅）的正面、略带表情的图像作为本次实验的数据集。在挑选过后的数据集上，随机选取每个对象的 4 幅图像作为训练数据，剩下的 4 幅图像作为测试集。在测试集上，任意选取 1 幅构成 Gallery，余下 3 幅构成 Probe 集合。XM2VTS 数据库上的一些人脸图像如图 2-3 所示。



图 2-3 XM2VTS 数据库的示例图

FRGC[17]: 该数据库上的人脸图像主要反映非非控制条件下采集的人脸图像变化。本研究在 FRGC 提供的标准训练集上，每人选取 10 幅构成本次实验的训练集。相应的测试集，本研究选用 FRGC 原始数据库上公布的实验 4 测试集并对其进行改造，将其由一个人脸确认评测数据库改成一个人脸识别测试数据库。具体来说，在该实验室 4 测试集上，随机选取每个对象的 1 幅人脸图像构成 Gallery，在余下的测试集上随机每人选取 2-6 幅构成 Probe 集合。FRGC 数据库上的部分人脸图像如图 2-4 所示。



图 2-4 FRGC 数据库上的示例图

MultiPIE[19]:该数据上的人脸图像主要包含表情与姿态变化。本研究只选用该数据库上属于这 7 个姿态—— -45° 、 -30° 、 -15° 、 0° 、 15° 、 30° 、 45° ，和这 3 种表情——中性表情、微笑、厌恶，并且无明显光照变化的图像作为实验图像，构建实验数据集。在选中的人脸图像上，随机选取 201 个对象的图像用于构建训练集，余下 136 个对象的图像用于构建测试集。在此基础上，再根据姿态的不同划分成 7 个子训练集和 7 个对应姿态的测试集。MultiPIE 数据库上，同一对象在不同姿态下的示例图像如图 2-5 所示。

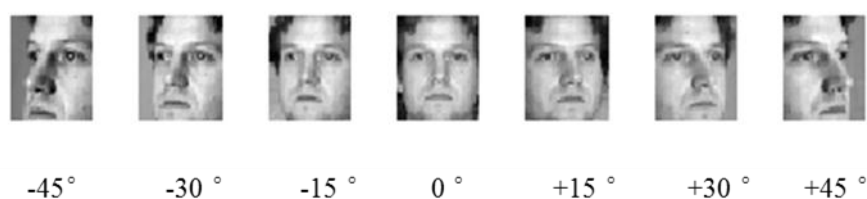


图 2-5 MultiPIE 数据库上同一个对象在不同姿态下的示例图

经过上述方式挑选数据后,这 4 个数据库上的训练集合、Probe 集合与 Gallery 中涉及的对象与图像数目等信息如表格 2-1 和表格 2-2 所示。根据图 1-6 可知,采用 FRGC 数据库上的人脸图像作为训练数据得到的模型 (Fisherfaces+NN),应用在 XM2VTS 数据库上得到性能,比不上由 XM2VTS 数据库上的人脸图像训练得到的模型性能。由表格 2-1 可知,这是因为 XM2VTS 数据库与 FRGC 数据库中图像的采集条件不同。本研究采用 XM2VTS 数据库与 FRGC 数据库之间的迁移实验,用于模拟跨采集条件的人脸识别问题。若 XM2VTS 为源域,FRGC 为目标域,则该方向的迁移学习问题中,其具体的训练集与测试集构成如表格 2-3 所示。

类似地,数据库 OFD 与 XM2VTS 数据库间最主要的差别为种族不同而导致的人脸表观差异,故可由这两个库上的迁移问题模拟跨种族情况下的人脸识别问题;在 MultiPIE 数据库上,不同姿态的人脸图像构成一个子集,这些子集之间的迁移学习问题可用于模拟跨姿态条件下的人脸识别问题。

表格 2-1 用于模拟迁移学习的数据库: OFD、XM2VTS、FRGC

Database	Collection Condition		Race	Train Set	Probe Set	Gallery
	Constrained	Wild				
OFD	√	-	Mongol	4×400	3×400	1×400
XM2VTS	√	-	Caucasian (Main)	4×295	3×295	1×295
FRGC	-	√	Caucasian (Main)	10×222	(2~6) ×466=2520	1×466

表格 2-2 用于模型迁移学习的数据库: MultiPIE

Database	Train	Probe	Gallery
MultiPIE View $i, i = 1, \dots, 7$	(2~9) $\times 201 = 1383$	(1~8) $\times 136 = 553$	1×136

表格 2-3 跨采集条件的人脸识别问题, 由 XM2VTS 迁移至 FRGC

Domain	Data		Subject Number	Image Number
XM2VTS (Source Domain)	Labeled Train Set		295	4×295
FRGC (Target Domain)	Unlabeled Train Set		222	10×222
	Test Set	Probe Set	466	(2~6) $\times 466 = 2520$
		Gallery	466	1×466

2.4.2 不同迁移问题下利用 TAME 迁移的人脸识别实验

为了验证本章提出的方法有效性, 本研究将在上述介绍的实验协议下, 对比 TAME 方法与国际上发表的迁移学习方法在这几个迁移问题上的性能。

由于在人脸识别领域, 大部分的方法都是基于判别性特征提取的(后续采用最近邻分类器(NN)进行识别), 且本章工作亦是基于特征提取的方式, 故本研究只对比基于特征提取的迁移学习方法。在本研究中将由目标域上数据的训练得到的模型称之为*目标域模型*; 将由源域上数据训练得到的模型称之为*源域模型*; 将由源域与目标域上的数据共同训练得到的模型称之为*迁移学习模型*。

目标域模型的 PCA 是指仅利用目标域上无监督样本作为训练数据, 基于 PCA 学习降维矩阵得到的模型。本研究中采用的 PCA 模型保留能量 99%。该方法代表了一种无监督的目标域模型。

源域模型的 FLD 是指在源域带监督信息的训练样本上, 经过 PCA 降维后(保留能量 99%), 再采用 LDA 方法提取有判别性的特征, 保留 $c_s - 1$ 维。该方法代表了一种有监督的源域模型。

迁移学习模型中, 本研究选用近年来发表在国际顶级会议上的基于特征提取的迁移学习方法——TDR[54]、SGF[31]、ITL[52]、TCA[34], 这些方法的介绍详见第 1.2 节。为了降低方法的计算复杂度且满足这些方法中要求源域与目标域维数相同的情况, 本研究对各个库上的训练数据进行了 PCA 降维, 各保持了 400

维数作为预处理（400 维的投影矩阵在所有实验的数据库上都至少保持了 98% 的能量，大部分达到 99% 的能量）。TDR 的方法中拉普拉斯矩阵采用热核函数计算，该函数中的方差采用两个域上所有样本对间欧氏距离的平均值。SGF 方法中采样了 10 个中间的子空间（包含起点与终点），SGF 后续采用原文[31]中推荐的 PLS 方法提取特征。这是因为应用 SGF 的方法后源域样本的特征扩增了 10 倍，若采用 LDA 方法提取特征，则需对 $4,000 \times 4,000$ 维的矩阵进行奇异值分解，需要非常大的计算量，且本研究在几个库上的实验亦发现 SGF 方法后接 PLS 方法提取的特征确实可得到更好的人脸识别性能。因此由于 SGF 方法的特殊性，本研究中关于 SGF 方法汇报的实验结果都是基于 PLS 的。TCA 的方法中的 RKHS 空间采用的线性核函数映射。这些方法中剩下的参数，采用遍历调优的方式确定。

由表格 2-4、表格 2-5 和表格 2-6 的结果可见，TAME 方法在各个实验中都获得了较源域模型 FLD、目标域模型 PCA 更好的性能。这表明 TAME 方法提取的判别性特征较仅基于源域或仅基于目标域提取的特征更适合目标域上的任务。

TAME 在跨种族、跨采集条件、跨姿态的问题下，都得到了比 SGF[31]、ITL[52]、TCA[34]更高的性能。对比的迁移学习方法中，仅有 TDR 的方法在某些迁移问题（比如 OFD→XM2VTS）下，得到最高可高出 TAME 方法 1% 左右的性能，但是在大部分的迁移学习问题中（表格 2-5、表格 2-6），TAME 方法大都较 TDR 方法有更好的性能，特别是在目标域为 FRGC 数据库的迁移问题下，由表格 2-5 可见 TAME 方法较 TDR 方法高出了约 8% 的性能。由表格 2-6 可见，在跨姿态的问题下，TAME 方法在大部分的实验上都得到了较 TDR 更高的性能（1%~4% 左右的提升），仅在由姿态 45° 迁移至 -45° 的方向上未能超过 TDR 的性能，但也取得了相当可观的性能。

与直接在目标域上采用无监督聚类（Kmeans）估计类别信息后再利用 FLD 提取的判别性特征相比，TAME 方法在大部分的实验上都表现出更为鲁棒的性能，特别是在表格 2-5 中目标域为现实条件下采集的 FRGC 数据库以及表格 2-6 中姿态跨度较大的迁移实验中，TAME 方法都能取得较好的性能。相反 Kmeans+FLD 的方法仅在目标域上样本空间位置关系与类别信息较为一致的情况下能得到不错的性能，但是在不一致的情况下不如 TAME 的性能，比如目标域为 MultiPIE 上 45°、-45° 的迁移实验。

通过在不同迁移问题上方法的对比，TAME 方法大都得到了最高的性能，说明了本方法的有效性与优越性。

表格 2-4 测试 TAME 方法在跨种族问题下的人脸识别性能

迁移问题		OFD→XM2VTS	XM2VTS→OFD
目标域模型	PCA	0.745	0.268
	Kmeans+FLD	0.857	0.981
源域模型	FLD	0.768	0.593
迁移学习模型	TDR[54]	0.858	0.971
	SGF[31]	0.783	0.653
	ITL[52]	0.715	0.110
	TCA[34]	0.740	0.517
	TAME	0.844	0.979

表格 2-5 测试 TAME 方法在跨采集条件问题下的人脸识别性能

迁移问题		XM2VTS→FRGC	FRGC→XM2VTS
目标域模型	PCA	0.012	0.745
	Kmeans+FLD	0.161	0.857
源域模型	FLD	0.083	0.767
迁移学习模型	TDR[54]	0.115	0.840
	SGF[31]	0.073	0.791
	ITL[52]	0.041	0.731
	TCA[34]	0.101	0.809
	TAME	0.190	0.840

表格 2-6 测试 TAME 方法在跨姿态问题下的人脸识别性能

源域		-45 °	45 °	0 °	0 °	-45 °	-30 °	30 °
目标域		45 °	-45 °	45 °	-45 °	30 °	30 °	-30 °
目标域模型	PCA	0.530	0.535	0.530	0.535	0.582	0.582	0.584
	Kmeans+FLD	0.662	0.658	0.662	0.658	0.751	0.751	0.740
源域模型	FLD	0.644	0.637	0.467	0.586	0.691	0.692	0.696
迁移学习模型	TDR[54]	0.658	0.665	0.678	0.645	0.708	0.712	0.723
	SGF[31]	0.631	0.608	0.631	0.609	0.664	0.635	0.676
	ITL[52]	0.664	0.64	0.671	0.62	0.658	0.666	0.698
	TCA[34]	0.620	0.588	0.562	0.521	0.655	0.645	0.669
	TAME	0.691	0.644	0.693	0.673	0.756	0.758	0.743

2.4.3 参数实验分析

为了测试 TAME 方法对于参数的敏感性，我们在以 FRGC 数据库为源域，XM2VTS 数据库为目标域的迁移问题上测试了不同参数组合下 TAME 方法的性能，具体的实验结果如图 2-6 所示。其中图 2-6 的横轴为参数 τ 、纵轴为人脸识别率，不同颜色的曲线为参数 λ 取不同值下的 TAME 方法随参数 τ 变化的曲线。在该实验中，TAME 方法在 $\lambda = 1e-3$ ， $\tau = 1e-1$ 时性能最好。在 τ 取最优值的情况下，TAME 方法在参数 λ 取 $1e-3$ 时最高，大于或小于 $1e-3$ 其性能慢慢开始下降。当 TAME 在参数 λ 取最优值的情况下，亦有类似的性能变化。总的来说，TAME 方法其性能随参数的变化较有规律。在图 2-6 上亦可注意到当 λ 与 τ 取很小时，模型的性能也不太低，这是因为在 λ 与 τ 取较小值时，源域的监督信息起到的作用较小，则这时候的 TAME 方法提取得到的特征非常接近目标域上进行无监督聚类估计分类再提取判别性特征的效果。一般来说，域上样本间的空间位置关系一定程度上反映了类别信息，故采用这种方法提取的特征亦具有一定的判别性，故其性能也不太低。类似地性能如图 2-7 所示。

在 MultiPIE 数据库上的一个跨姿态迁移问题上，TAME 方法随着参数变化的曲线如图 2-7 所示。在该实验上，TAME 方法受参数 τ 与 λ 的影响并不大，大约在 $\tau = 1e-3$ ， $\lambda = 1e-5$ 时最优。在源域为 0° 、目标域为 -45° 的跨姿态迁移问题中，由于源域与目标域上的人脸图像存在较大的表观差异，适当降低 TAME 目标函数中低秩表示时出现残差的惩罚，能得到较好的性能。这是因为，当源域与目标域存在较大差异时，较难找到一个特征子空间使得源域与目标域样本在该空间上相似的样本空间位置能非常紧致。在这种情况下，应适当降低对重构误差

的惩罚。

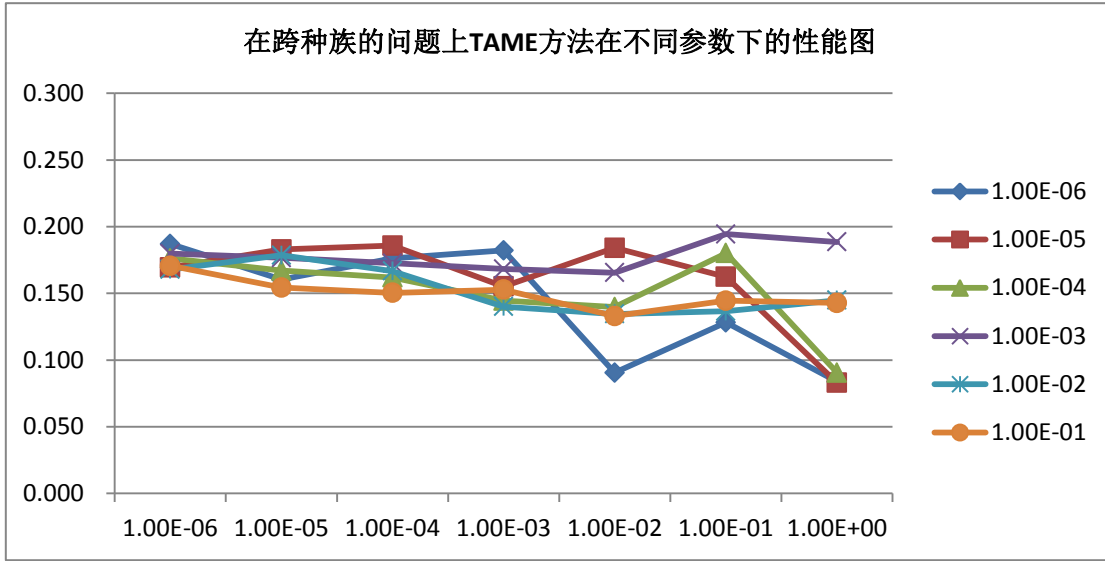


图 2-6 TAME 方法在以 XM2VTS 数据库为源域，FRGC 数据库为目标域的实验中，其性能随参数变化的曲线。其中横轴为参数 τ ，曲线的不同颜色表示取不同的 λ 值。

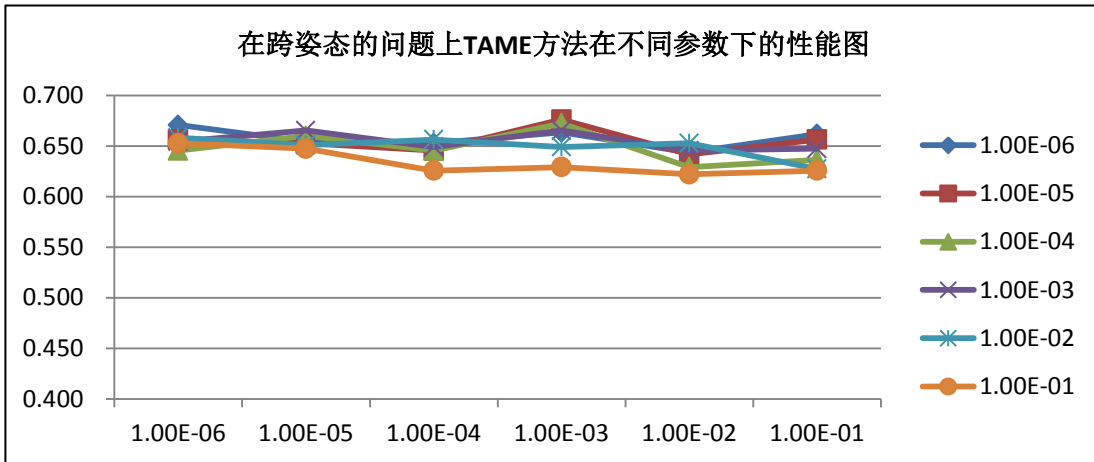


图 2-7 TAME 方法在 MultiPIE 数据库上源域为 0° ，目标域为 -45° 的实验中，其性能随参数变化的曲线。其中横轴为参数 τ ，曲线的不同颜色表示取不同的 λ 值。

2.5 总结与分析

本章提出的 TAME 方法利用源域上的监督信息来指导目标域上判别性相似

度矩阵估计，在此基础上提取判别性的特征。该方法基于源域与目标域共享同一分类准则（身份信息）这一先验，将其形式化成这一目标：估计一目标域相似度矩阵，使其能体现与源域上标注信息较为一致的类别信息。该类别信息的一致性采用域上样本空间位置关系判别性程度的一致性来衡量。关于域上样本空间位置关系的判别性程度，源域上采用 Fisher 准则度量；无监督目标域上采用自身低秩表示的系数矩阵的秩来衡量。在此基础上，TAME 提出了一个基于低秩矩阵估计的目标函数，采用拉格朗日乘子的方法进行求解。在本研究设计的跨库实验上，TAME 大都取得了最好的性能，体现了该方法的优越性。

TAME 方法中假设源域与目标域是通过同一个投影矩阵到达能同时使两个域具有判别性空间位置关系的特征子空间，这就要求在原始空间上源域与目标域的特征维数相同，这一定程度上限制了该方法在某些迁移问题上的应用。

第3章 基于源域样本目标化的人脸识别

上一章提出的 TAME 方法无法处理源域与目标域位于不同子空间上的情况。此外，源域中可能包含部分对目标域上的人脸识别任务有害的信息。更为优秀的迁移学习方法应能同时应对这两个问题。带着这些考虑，本章提出借助公共子空间学习可靠迁移系数进而利用该系数将源域样本迁移至目标域的迁移方法，其中原始的源域与目标域可位于不同的特征子空间上。由于在迁移的过程中，并不影响源域上样本间的类别关系，故目标域上间接获得了一批可用于监督学习的带标注训练样本，进而可训练得到适合目标域上人脸识别问题的模型。

3.1 相关背景知识

3.1.1 局部保持映射

局部保持映射[65] (Locality Preserving Projection, 简称 LPP) 是一种通过线性投影矩阵来求解流形低维嵌入的方法。与 PCA 这种保持空间上样本间欧氏距离的方式不同, LPP 的方法认为样本落在一个非线性的流形上, 并且采用样本的局部近邻结构来刻画该流形。LPP 是对非线性的拉普拉斯映射[66] (Laplacian Eigenmap) 的一种线性逼近。与之类似的工作有局部投影嵌入[67] (Locally Linear Embedding, 简称 LLE) 等。

具体来说, 给定样本矩阵 $X = [x_1 \ x_2 \ \cdots \ x_n]$, 其中每列为一个样本的特征向量 $x_i \in R^p$, 根据 LPP 求解其 $d (d < p)$ 维的低维嵌入方法如下。

算法 3-1 局部保持投影 (LPP) 算法

1. 构建邻接矩阵图 (adjacency graph)

记 G 为该节点数为 n 的图, 若样本 x_i 与 x_j 满足近邻条件, 在节点 i 与 j 间连一条边。样本 x_i 与 x_j 是否为近邻, 可由以下两种常用的方法判断。

- 1) ε -近邻范围。若样本点 x_i 与 x_j 满足 $\|x_i - x_j\|_2 < \varepsilon$ 的条件, 则 x_i 与 x_j 互为近邻; 否则 x_i 与 x_j 互不为近邻。
- 2) k 近邻。若样本 x_i 是 x_j 的 k 个近邻样本之一, 则点 i 与 j 存在一条无向的

边；若样本 x_i 不属于 x_j 的 k 个近邻样本之一，则 i 与 j 之间无边。
<p>2. 估计样本间权重矩阵 W</p> <p>在确定样本间的近邻关系后，对有边连接的样本对赋予非零权重，无边连接的样本对赋予 0 权重，最后可得到对称且稀疏的权重矩阵。选用不同的函数可定义不同的权重，下文介绍两个常用的权重计算函数。</p> <p>1) 热核函数</p> <p>若节点 i 与 j 有边连接，其权重计算方式如下，</p> $W_{ij} = e^{-\frac{\ x_i - x_j\ _2^2}{t}} \quad (3-1)$ <p>其中 $t \in R$，经验上可将 t 设置为所有样本点对欧氏距离的平均值。其它无边连接的节点间权重为 0。</p> <p>2) 简单 0-1 权重</p> <p>若节点 i 与 j 有边连接，则权重为 1；若无边连接，则权重为 0</p>
3. 本征映射 (Eigenmap)
<p>LLP 的目标函数为，</p> $\arg \min_a a^T X L X^T a, \text{s.t. } a^T X D X^T a = 1 \quad (3-2)$ <p>该目标函数的求解等价于对下列目标函数的求解，</p> $X L X^T a = \lambda X D X^T a. \quad (3-3)$ <p>这是一个广义特征值求解问题。其中 D 为对角矩阵，$D_{ii} = \sum_{j=1}^n W_{ji}$；$L$ 即为所谓的拉普拉斯矩阵，可由 $L = D - W$ 计算得到。</p>

3.1.2 稀疏编码 (Sparse Coding)

稀疏编码涉及一个隐表示 (latent representation) h 和线性映射 D ，在稀疏编码中一般将线性映射 D 称之为词典。稀疏编码可分别从概率和非概率角度来解释其原理。

从非概率角度来说，稀疏编码可以看成是在词典 D 下对输入的信号 x 进行编码或者特征表示，其目标函数如公式(3-4)所示。该目标函数其实是一个 lasso 回归问题，可采用最小角度回归的方法 (Least Angle Regression, 简称 LAR) [68] 进行求解。

$$h^* = f(x) = \arg \min_h \|x - Dh\|_2^2 + \lambda \|h\|_1. \quad (3-4)$$

相应地,若已知信号的稀疏编码,则可通过以下目标函数来优化求解词典 D 。

$$J_{sc} = \sum_t \|x^{(t)} - Dh^{*(t)}\|_2^2 \quad (3-5)$$

其中 $x^{(t)}$ 为样本 t 的输入特征向量, $h^{*(t)}$ 为由公式(3-4)决定的稀疏系数。由于词典 D 的第 i 个列向量长度的缩放,可以对应在稀疏编码 $h^{*(t)}$ 第 i 个系数上的缩放;且稀疏编码的值大小直接影响到 L_1 惩罚的大小,故通常约束词典 D 中的每个列向量长度为 1。在文章[29]中,作者基于稀疏编码约束,通过迭代优化得到词典 D 。

从概率角度来说,稀疏编码可由隐表示 h 需遵循拉普拉斯分布的先验,如(3-6)所示,推理得到。给定稀疏编码时输入信号的似然函数如(3-7)所示,则基于最大后验准则 (Maximum a Posteriori, 简称 MAP),可推导得到如公式(3-4)所示的目标函数。

$$p(h) = \prod_i^{d_h} \lambda \exp(-\lambda |h_i|) \quad (3-6)$$

$$p(x|h) = N(x; Dh + \mu_x, \sigma_x^2 I) \quad (3-7)$$

稀疏编码的一个重要优点就是可解释性。这是因为即使在过完备词典 (包含许多冗余的码本) 的情况下,稀疏编码仍然可以找到最相关的那部分基,并将其它基上的系数置为零。

稀疏编码传统上作为一种特征提取的方法[29],但在 2009 年 Ma Yi 等人[69]将稀疏编码应用在了人脸识别的分类问题上,并取得了令人瞩目的性能。具体来说,这篇文章[69]中作者直接将库中带身份信息的样本 (人脸识别中通常作为 Gallery) 作为词典,然后对待分类的样本进行稀疏编码,从而找到与待分类样本最为相关的部分样本,再基于待分类样本与词典中每一对象的重构误差来进行分类。这么做的原理是,这篇文章[69]解释说明当词典中每一对象包含一定数量的样本时,即可描绘出该对象所有样本所构成的子空间,那么任意一个该对象的样本都可由这些样本线性重构得到,且待分类样本与其真实身份对象的样本所在空间的距离最近。此时稀疏编码的可解释性在这种情况下起到了很好的作用。一般认为稀疏系数的越大,所对应的样本与待分类的样本之间相似性越大;系数非零的样本与待分类样本极为相关。在稀疏编码所确定的相关样本基础上,这篇文章中提出采用重构误差作为待分类样本与词典中某一对象距离的度量,其定义如公式(3-10)所示。这种距离一定程度上定义了样本点到子空间的距离。文章[69]中认为待分类样本应是来自与之最近的对象子空间。该文中将这种利用稀疏编码进行分类的方法称之为基于稀疏表示的分类器 (Sparse Representation-based

Classification, 简称 SRC), 其具体的流程如算法 3-2 所示。

稀疏表示方法可视为最近邻分类器 (Nearest Neighbor, 简称 NN) 或最近子空间 (Nearest Subspace, 简称 NS) 方法的泛化推广, 兼具两者的优点。具体来说, 最近邻分类器是基于样本到样本距离的分类器, 该分类器认为待分类样本的标签应与其最近的标注样本相同。最近子空间分类器是基于样本到对象子空间距离的分类器, 该分类器认为待分类样本的标签应是与其最近的对象子空间相同。稀疏表示方法在这两种方法之间进行了折中, 采用所有对象的样本对待分类样本进行表示, 通过所有样本对待分类样本进行表示时的竞争, 首先找到与待分类样本最为相关的部分词典样本, 进而根据待分类样本到这些相关样本所在的对象子空间的距离进行分类。该分类过程中既考虑了样本对之间的相关性, 又考虑了样本与对象子空间的相关性, 因此是对最近邻分类器与最近子空间方法的良好平衡。

算法 3-2 基于稀疏编码的分类器

输入: Gallery 样本矩阵 $X = [X_1 \ X_2 \ \dots \ X_k] \in R^{d \times n}$, k 为 Gallery 上总的类别数, $X_i|_{i=1}^k$ 表示来自第 i 个对象的所有样本所构成的矩阵; 待分类样本 $y \in R^d$, 容忍误差 $\varepsilon > 0$

输出: 待分类样本 y 的标签 l_y

步骤:

1. 对矩阵 x 的每个列向量进行长度单位化。

2. 将矩阵 x 作为词典, 对待分类样本 y 进行稀疏编码。稀疏编码可通过求解下述的目标函数得到,

$$\alpha^* = \arg \min \|\alpha\|_1 \text{ s.t. } X\alpha = y, \quad (3-8)$$

或者求解,

$$\alpha^* = \arg \min \|\alpha\|_1 \text{ s.t. } \|X\alpha - y\|_2 < \varepsilon, \quad (3-9)$$

亦可。

3. 计算待分类样本的到各个对象子空间的距离。该距离采用各个对象的样本对待分类样本的重构误差来估计, 即

$$r_i(y) = \|y - X\alpha_i\|_2, i = 1, 2, \dots, k. \quad (3-10)$$

其中 α_i 表示来自第 i 个对象的所有样本所对应的稀疏系数。

4.对待分类样本进行标签估计。稀疏表示分类器认为待分类样本应来自与其具有最短距离的对象子空间，即，

$$l_y = \arg \min_i r_i(y)。 \quad (3-11)$$

3.2 基本思路与方法框架

3.2.1 基本思路

目标域上由于缺少带标注的样本，因此无法进行监督学习；但源域上却具备一批可用于稳定训练模型的带标注数据，若能将源域上的带标注样本迁移至目标域，则可解决目标域上缺少带标注样本的现状，进而可采用监督学习方法，提取可靠有用的判别性信息，从而训练得到能更好应对目标域上任务的模型。将源域与目标域视为两个流形，则域上的信息主要体现在样本的空间位置关系上，那么源域与目标域间的差异主要体现在样本空间位置分布的差异上。要想将源域样本迁移至目标域，其本质是将源域上样本的空间位置关系目标化。本章将该思路提炼为寻找源域样本在目标域上的空间位置这一先验。要估计该空间位置关系，直接的作法是在原始空间上进行估计。但由于在原始空间上，源域与目标域所在流形距离较远，差异较大，无法得到较为可靠的空间位置关系。本研究认为，源域与目标域在公共子空间上的空间位置关系，非常逼近源域样本在目标域原始空间上的空间位置关系。因此本章提出的方法为了将源域样本迁移至目标域，需估计源域与目标域的公共子空间，并在该子空间上学习源域相对于目标域的空间位置关系。

当前基于优化公共子空间的迁移学习方法，其一般流程如下所述：1) 定义度量域间距离的函数，在此基础上搜索最小化源域与目标域距离的公共子空间（第三方子空间）[34, 48]；2) 将带标注的源域投影至公共子空间，训练有监督模型；3) 处理目标域任务时，先将目标域亦投影至公共子空间，然后采用 2) 中训练得到的模型处理目标域上的任务。基于公共子空间的方法其基本原理是由于公共子空间上的源域与目标域无明显差异，因此源域在该子空间上的知识就能可靠地帮助目标域上任务的学习。

基于公共子空间的方法其最关键的部分是如何度量源域与目标域间的距离，一般有基于样本分布或特征子空间的方法。基于样本分布的方法较难应用在高维空间上。这是因为，在高维数空间上要估计样本的分布需要大量的样本；此外，即使在有大量样本的情况下，样本分布的估计仍然是一个很难的问题。因此基于样本分布估计域间距离的方法较难取得较好性能。基于特征子空间的方法，由于

对整个域只采用其特征子空间来粗略描述,即使缩小了两个域所在的特征子空间距离,亦不能保证这两个域上的样本能较好的交织分布。这是因为,同一个特征子空间上样本可以有許多不同的分布,特征子空间距离较小,并不代表源域与目标域之间就不存在差异,比如源域与目标域的特征子空间重合,但样本各自集中分布在子空间的两端。

一个好的公共子空间应该能使得源域与目标域上的样本在其上交织分布,样本的空间位置关系无明显的域间差异。这一性质既要求源域与目标域位于同一特征子空间,又要求两个域具有较为接近的样本空间位置关系。基于这样的目标,本章采用稀疏重构误差来度量两个域的距离。这是因为:1)稀疏重构误差从全局上度量了样本点到子空间的距离[69],且在稀疏编码时要求词典的少量样本即可表示好输入的信号,使得在约束点到子空间距离时要求更为严格。2)稀疏编码一定程度上约束了样本的空间位置关系,即稀疏编码中系数非零的样本与待表示样本在空间位置上较为接近。若在某一特征子空间上,源域与目标域样本能够互相比较好地稀疏表示,则表明源域与目标域具有非常好的相关性,且两个域上的样本交织分布。此外,稀疏编码具有协同表示的能力,当待表示样本中包含部分噪声信息时仍然得到稳定的编码。

此外,一个好的公共子空间应能尽量保存域上的信息,尽量少的丢失域上信息。将源域与目标域视为两个流形,则域上的信息主要体现在样本的空间位置关系上[65]。学术上通常采用拉普拉斯矩阵来描述这一信息,尽量保持域上的局部结构。

基于上述分析,可通过最小化稀疏重构误差与局部结构保持的两个约束来优化搜索公共子空间,该约束可形式化如公式(3-25)所示的目标函数。该目标函数既避免了复杂困难的样本分布估计,又能保证优化得到的特征子空间上两个域无明显差异,同时尽可能地减少信息丢失。此外,该目标函数中优化得到的稀疏编码蕴含了源域样本相对目标域样本的位置关系,故可利用该编码确定源域样本在目标域上的位置,从而实现将源域样本迁移至目标域的目标,本研究将该过程称之为目标化源域样本(Targetize Source-Domain Samples)。由于该过程中并未改变源域上样本间的类别关系,因此目标化的源域样本上仍然携带类别信息,可用于有监督模型的训练。由于该方法在目标化源域样本时是借由公共子空间学习得到稳定的迁移系数,因此将其称之为基于源域样本目标化的人脸识别方法,即 **Targetize Source domain images bridged by Common subspace for Face Recognition**,简称 TSC。

3.2.2 方法框架

TSC 方法从原始空间上（这里采用的是灰度图像空间），将源域样本迁移至目标域，并且不改变源域样本的类别信息，从而使得目标域上获得了一批可用于有监督学习（LDA、PLS[70]、SVM[12]等）的带标注样本。采用 TSC 的方法进行迁移学习时，其应用框架如图 3-1 所示。首先，利用 TSC 方法学习将源域样本迁移至目标域的迁移系数；其次，利用 TSC 估计的迁移系数将源域样本目标化；进而在目标化的源域样本上进行有监督的模型学习。经过这系列步骤，即可训练得到一个更适合目标域上任务处理的模型。

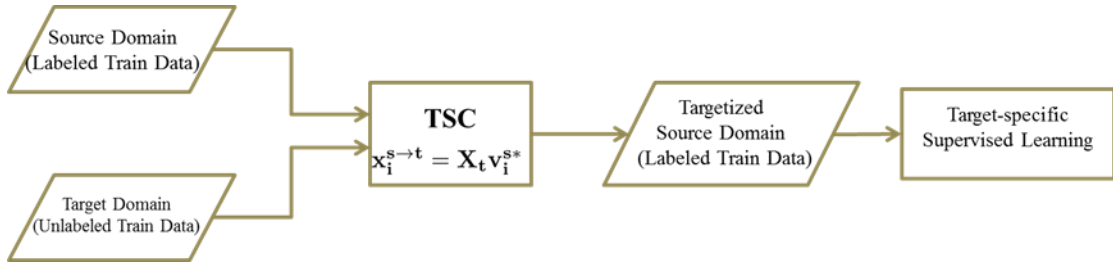


图 3-1 基于源域样本目标化的人脸识别方法（TSC）的方法框架

3.2.3 目标函数

源域记为 \mathcal{D}_s ，是由 n_s 个带类别标签的样本组成，即如下所示，

$$\mathcal{D}_s = \{(x_1, y_1) \ (x_2, y_2) \ \cdots \ (x_{n_s}, y_{n_s})\} \quad (3-12)$$

其中 n_s 为源域上的样本总数， $x_i|_{i=1}^{n_s}$ 表示源域上的第 i 个样本的特征向量，其对应的类别标签为 y_i 。目标域记为 \mathcal{D}_t ，是由 n_t 个无类别标签信息的目标域样本构成，即如下所示，

$$\mathcal{D}_t = \{u_1 \ u_2 \ \cdots \ u_{n_t}\}, \quad (3-13)$$

其中 n_t 为目标域上的样本总数， $u_j|_{j=1}^{n_t}$ 表示目标域上第 j 个样本的特征向量，无任何类别标签信息。

一般基于公共子空间的迁移学习方法都假设源域与目标域通过同一个投影矩阵即可到达公共子空间。但其实源域与目标域存在差异，不同的特征子空间到达同一个第三方特征子空间理应采用不同的投影矩阵。此外，假设源域与目标域共用同一个投影矩阵无法应对源域与目标域上特征维数不同的情况。因此，本研究认为源域与目标域需要域特殊（Domain-Specific）的投影矩阵，分别记为 W_s 、 W_t ，方可到达公共子空间。

基于第 3.2.1 节的分析，提炼出结合局部结构保持与最小化稀疏重构误差这两个约束来优化公共子空间的方法。下文分别对这两个约束进行形式化。

局部结构保持：记 X_s 、 U_t 分别为源域与目标域的样本特征矩阵，每个列向量为一个样本的特征向量。记 Ω_s 、 Ω_t 分别为源域与目标域上样本间的近邻权重矩阵，常用的有热核函数等计算方式。在此基础上，投影矩阵 W_s 、 W_t 应尽可能地保持对应域上的局部结构信息，本研究采用与 LPP 方法（详见第 3.1.1 节）相同的目标函数来约束这两个投影矩阵，即这两个投影矩阵可尽可能满足以下的目标函数，

$$\begin{aligned} W_s^* &= \arg \min_{W_s} W_s^T X_s L_s X_s^T W_s, \\ L_s &= I_s - D_s^{-\frac{1}{2}} \Omega_s D_s^{-\frac{1}{2}} \text{ s.t. } W_s^T W_s = I \\ W_t^* &= \arg \min_{W_t} W_t^T U_t L_t U_t^T W_t, \\ L_t &= I_t - D_t^{-\frac{1}{2}} \Omega_t D_t^{-\frac{1}{2}} \text{ s.t. } W_t^T W_t = I \end{aligned} \quad (3-14)$$

其中， D_s 、 D_t 为对角矩阵，其计算方式如下，

$$\begin{aligned} D_s(i, i) &= \sum_{j=n_s} \Omega_s(j, i) \\ D_t(i, i) &= \sum_{j=n_t} \Omega_t(j, i) \end{aligned} \quad (3-15)$$

最小化稀疏重构误差：稀疏重构指的是在给定词典的情况下，先对输入信号进行稀疏编码，再利用此编码系数在该词典上重构信号，该重构的信号即为所述的稀疏重构。相应地，该稀疏重构与原始输入信号间的差异称之为稀疏重构误差（Reconstruction Error）。根据上述的分析，本研究采用稀疏重构误差作为样本点到域的距离 $\text{dis}(x, \mathcal{D})$ 估计，即如下公式所示，

$$\text{dis}(x, \mathcal{D}) = \|x - A\alpha\|_2. \quad (3-16)$$

其中 α 为采用域 \mathcal{D} 上的所有样本作为词典（记为 A ）对样本 x 进行稀疏编码时的系数向量，即通过以下的目标函数得到，

$$\arg \min_{\alpha} \|x - A\alpha\|_2 \text{ s.t. } |\alpha|_0 < \tau. \quad (3-17)$$

其中 τ 为稀疏编码时的调节参数，用于控制系数向量 α 中非零的个数。那么域之间的距离，即可由该域上的样本到另外一个域的平均距离来估计，即

$$\text{Dis}(\mathcal{D}_s, \mathcal{D}_t) = \frac{1}{n_s} \sum_{i=1}^{n_s} \text{dis}(x_i, \mathcal{D}_t). \quad (3-18)$$

其中 $\text{Dis}(\mathcal{D}_s, \mathcal{D}_t)$ 为域 \mathcal{D}_s 到域 \mathcal{D}_t 的距离。一般的距离函数都是对称的，即源

域到目标域的距离应与目标域到源域的距离相等，故应有如下关系，

$$\text{Dis}(\mathcal{D}_s, \mathcal{D}_t) = \text{Dis}(\mathcal{D}_s, \mathcal{D}_t) = \text{Dis}(\mathcal{D}_t, \mathcal{D}_s) \quad (3-19)$$

其中 $\text{Dis}(\mathcal{D}_s, \mathcal{D}_t)$ 为域 \mathcal{D}_s 与域 \mathcal{D}_t 之间的距离。上述的性质一般通过取平均值来实现，即，

$$\text{Dis}(\mathcal{D}_s, \mathcal{D}_t) = \frac{1}{2} (\text{Dis}(\mathcal{D}_s, \mathcal{D}_t) + \text{Dis}(\mathcal{D}_t, \mathcal{D}_s)) \quad (3-20)$$

综上所述，采用目标域样本作为词典，对源域样本进行稀疏编码后，如公式(3-21)所示，再重构得到的信号与源域样本间的差异总合即为 $\text{Dis}(\mathcal{D}_s, \mathcal{D}_t)$ ，如公式(3-22)所示。

$$\begin{aligned} v_{s_i}^* &= \arg \min_{v_{s_i}} \|W_s^T x_{s_i} - W_t^T U_t v_{s_i}\|^2 \\ \text{s.t. } v_{s_i} &< \tau, i = 1, 2, \dots, n_s \end{aligned} \quad (3-21)$$

$$\begin{aligned} \text{Dis}(\mathcal{D}_s, \mathcal{D}_t) &= \frac{1}{n_s} \sum_{i=1}^{n_s} \|W_s^T x_{s_i} - W_t^T U_t v_{s_i}\|^2 \\ &= \frac{1}{n_s} \|W_s^T X_s - W_t^T U_t V_s\|_F^2 \\ V_s &= \begin{bmatrix} v_{s_1} & v_{s_2} & \cdots & v_{s_{n_s}} \end{bmatrix} \end{aligned} \quad (3-22)$$

相应地，采用源域样本作为词典，对目标域样本进行稀疏编码，如公式(3-23)所示，再重构回去得到的信号与目标域样本间的差异总和即为 $\text{Dis}(\mathcal{D}_t, \mathcal{D}_s)$ ，如(3-24)所示。

$$\begin{aligned} v_{t_j}^* &= \arg \min_{v_{t_j}} \|W_t^T u_{t_j} - W_s^T X_s v_{t_j}\|^2 \\ \text{s.t. } v_{t_j} &< \tau, j = 1, 2, \dots, n_t \end{aligned} \quad (3-23)$$

$$\begin{aligned} \text{Dis}(\mathcal{D}_t, \mathcal{D}_s) &= \frac{1}{n_t} \sum_{j=1}^{n_t} \|W_t^T u_{t_j} - W_s^T X_s v_{t_j}\|^2 \\ &= \frac{1}{n_t} \|W_t^T U_t - W_s^T X_s V_t\|_F^2 \\ V_t &= \begin{bmatrix} v_{t_1} & v_{t_2} & \cdots & v_{t_{n_t}} \end{bmatrix} \end{aligned} \quad (3-24)$$

最小化稀疏重构误差的约束即可形式化最小化为如上定义的 $\text{Dis}(\mathcal{D}_s, \mathcal{D}_t)$ 。

目标的投影矩阵 W_s 、 W_t 应同时能满足局部结构保持与最小化稀疏重构误差的约束，结合上述对这两个约束的形式化，可得到如公式(3-25)所示的整体目标

函数，前面两项即为域结构保持约束，后面两项为最小化重构误差的约束。

$$\begin{aligned}
 & \arg \min_{W_s, W_t, V_s, V_t} \frac{1}{n_s} W_s^T X_s L_s X_s^T W_s + \frac{1}{n_t} W_t^T U_t L_t U_t^T W_t \\
 & + \lambda \left(\frac{1}{n_s} \|W_s^T X_s - W_t^T U_t V_s\|_2^2 + \frac{1}{n_s} \|W_t^T U_t - W_s^T X_s V_t\|_2^2 \right) \quad (3-25) \\
 & \text{s.t. } W_s^T W_s = I, W_t^T W_t = I, \\
 & \|v_{s_i}\|_0 < \tau, i=1, 2, \dots, n_s, \|v_{t_j}\|_0 < \tau, j=1, 2, \dots, n_t
 \end{aligned}$$

该目标函数中含有 4 个未知量：两个投影矩阵 W_s 、 W_t ，两个迁移系数矩阵 V_s 、 V_t 。该目标函数可通过迭代优化的方式同时估计求解这 4 个未知量。其中投影矩阵 W_s 、 W_t 用于将源域与目标域分布投影至公共子空间，迁移系数矩阵 V_s 、 V_t 分别表示了公共子空间上源域样本与目标域样本之间的相对空间位置关系，如图 3-2 所示。

源域样本目标化：通过迁移系数矩阵 V_s 即可在原始空间上，将源域样本迁移至目标域，其具体的计算方式如公式(3-26)所示。其中 $X_{s \rightarrow t}$ 就是目标化后的源域样本，该目标化操作并不影响源域样本上的类别信息，因此在这批数据上可采用有监督的学习方法来训练得到更接近目标域模型。

$$X_{s \rightarrow t} = U_t V_s \quad (3-26)$$

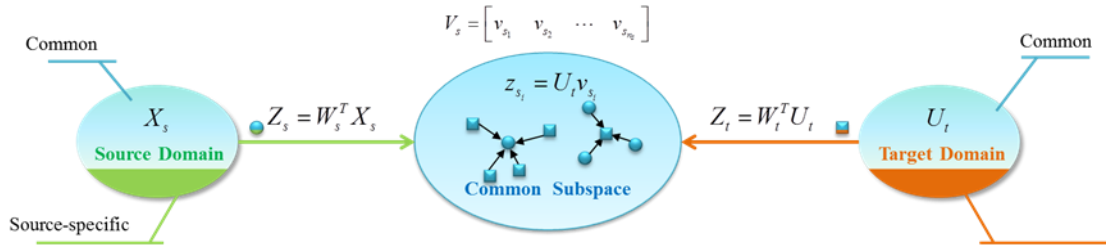


图 3-2 基于源域样本目标化（TSC）原理示意图

3.3 优化方法与算法流程

3.3.1 优化方法

本方法采用迭代优化的方式对目标函数(3-25)进行求解。记源域与目标域所在的原始特征子空间空间维数分别为 p_s 、 p_t ，公共子空间的维数为 d ，那么 W_s 、

W_t 分别为 $p_s \times d$ 、 $p_t \times d$ 的投影矩阵。

固定 W_s 、 W_t 求解 V_s 、 V_t ：由目标函数(3-25)可知，在给定 W_s 、 W_t 的情况下， V_s 、 V_t 的求解相互独立，故可分别独立优化。以求解 V_s 为例，在给定 W_s 、 W_t 时，保留目标函数(3-25)中关于 V_s 的项，整理后可得到如公式(3-27)所示的关于 V_s 的目标函数。又由于 V_s 的每一列求解都是独立的，故可分解成对每列的信号进行独立稀疏编码的优化问题，如公式(3-33)所示。

$$\begin{aligned} V_s^* &= \arg \min_{V_s} \|W_s^T X_s - W_t^T U_t V_s\|_F^2 \\ \text{s.t. } \|v_{s_i}\|_0 &< \tau, i=1, 2, \dots, n_s \end{aligned} \quad (3-27)$$

固定 V_s 、 V_t 求解 W_s 、 W_t ：保留目标函数(3-25)中关于 W_s 和 W_t 的部分，可化简成如下所示的目标函数，

$$\begin{aligned} \arg \min_{W_s, W_t} W_s^T L_{ss} W_s + W_t^T L_{tt} W_t + W_s^T L_{st} W_t + W_t^T L_{ts} W_s \\ \text{s.t. } W_s^T W_s = I, W_t^T W_t = I \end{aligned} \quad (3-28)$$

其中 L_{ss} 、 L_{tt} 、 L_{st} 、 L_{ts} 的计算方式如下，

$$\begin{aligned} L_{ss} &= \frac{1}{n_s} X_s X_s^T + \frac{\lambda}{n_t} I + \frac{\lambda}{n_t} V_t V_t^T, \\ L_{tt} &= \frac{1}{n_t} U_t U_t^T + \frac{\lambda}{n_t} I + \frac{\lambda}{n_s} V_s V_s^T, \\ L_{st} &= -\left(\frac{\lambda}{n_s} V_s^T + \frac{\lambda}{n_t} V_t \right), \\ L_{ts} &= L_{st}^T. \end{aligned} \quad (3-29)$$

将 W_s 与 W_t 写成一个矩阵 $W = \begin{bmatrix} W_s^T & W_t^T \end{bmatrix}^T$ ，整理后目标函数(3-28)可重写关于 W 的目标函数，即

$$\begin{aligned} \arg \min_W W^T \Sigma W \\ \text{s.t. } W^T W = I \end{aligned} \quad (3-30)$$

其中 Σ 的计算方式如下所示，

$$\Sigma = \begin{bmatrix} L_{ss} & L_{st} \\ L_{ts} & L_{tt} \end{bmatrix}. \quad (3-31)$$

目标函数(3-30)与第 3.1.1 节介绍的局部保持映射中的优化目标函数(3-2)、(3-3)相似，其最优解为等式(3-32)所要求的特征向量及特征值，取其前 d 个最小的特征值对应的特征向量构成的矩阵即为 W 的最优解。

$$\Sigma w = \lambda w \quad (3-32)$$

在得到 W 的最优解后，取前 p_s 行构成 W_s 的最优解，取余下的行构成 W_t 的最优解。

3.3.2 算法流程

根据第 3.3.1 节介绍的优化方法，TSC 的实现流程如算法 3-3 所示。

算法 3-3 目标化源域样本 (TSC) 的算法

输入： 源域样本矩阵 X_s ，目标域样本矩阵 U_t ，收敛精度 ε ，最大循环次数 T
输出： 目标化的源域样本矩阵 $X_{s \rightarrow t}$
1. 初始化：随机初始化 W_s 、 W_t ，约束其每列长度为 1；循环次数 $t = 0$
2. 循环 <div style="margin-left: 20px;"> a) 优化稀疏编码矩阵 V_s、V_t <div style="margin-left: 20px;"> i. 遍历 $i = 1, 2, \dots, n_s$ $v_{s_i}^* = \arg \min_{v_{s_i}} \ W_s^T x_{s_i} - W_t^T U_t v_{s_i}\ ^2, \text{s.t. } \ v_{s_i}\ < \tau \quad (3-33)$ 结束 </div> <div style="margin-left: 20px;"> ii. 遍历 $j = 1, 2, \dots, n_t$ $v_{t_j}^* = \arg \min_{v_{t_j}} \ W_t^T u_{t_j} - W_s^T X_s v_{t_j}\ ^2, \text{s.t. } \ v_{t_j}\ < \tau \quad (3-34)$ 结束 </div> </div> <div style="margin-left: 20px;"> b) 根据公式(3-29)(3-30)(3-31)(3-32)更新 W_s、W_t。 c) 循环次数更新 $t = t + 1$。 d) 检查收敛条件。 若满足 $\ W_s - W_s\ _F^2 < \varepsilon$ 且 $\ W_t - W_t\ _F^2 < \varepsilon$ 则收敛，跳出循环；或循环次数超过最大循环次数，即 $t > T$ 则跳出循环。除以上情况外，仍未收敛，返回步骤 a)。 结束 </div>
3. 目标化源域样本 $X_{s \rightarrow t} = U_t V_s$ 。

3.4 实验及性能比较

3.4.1 不同迁移问题下利用 TSC 迁移的人脸识别实验

为了测试采用 TSC 方法迁移的源域样本作为训练数据的有效性，我们在不同的迁移问题上测试其性能，关于实验的迁移问题设置详见第 2.4.1 节。本研究将在这些迁移实验上与发表在顶级会议上的迁移学习方法 TDR[54]、SGF[31]、ITL[52]、TCA[34]以及第 2 章提出的基于目标域相似度矩阵估计的判别性特征提取方法（TAME）方法进行对比。这些方法的参数设置如第 2.4.2 节中介绍的相同。因此为了能与比对的迁移方法具有可比性，本研究中将所有域上的样本先进行保留 400 维的 PCA 降维预处理，然后再应用算法 3-3 目标化源域样本。但其实 TSC 方法可适用于源域与目标域维数不同的迁移问题下。同样地，TSC 后续的监督学习方法亦采用 LDA 提取特征，然后再应用最近邻分类器进行人脸识别的方法训练模型。

由表格 3-1、表格 3-2 和表格 3-3 可见，虽然采用相同的学习方法（FLD+NN），在原始的源域样本上训练得到的模型，其性能远远不如由 TSC 目标化后的源域样本所训练得到的模型性能。该结果表明经 TSC 方法迁移的源域样本确实更为接近目标域，从而训练得到的模型在目标数据库上可得到更好的人脸识别性能。

对比的他人迁移学习方法中，仅有 TDR 的性能与本章提出的 TSC 方法有一定的可比性。具体来说，TSC 的方法虽然在跨种族的迁移问题下（性能见表格 3-1）并没有超过 TDR 特别多的性能，但是在跨采集条件的问题中，当目标域为较为复杂的 FRGC 数据库时，TSC 较 TDR 高出 6% 的性能。特别是在跨姿态的问题中，TSC 在所有的跨姿态迁移问题上都较 TDR 有更好的性能，姿态差异越大的迁移问题上，TSC 方法提升的更为明显。这是因为，TSC 是借助公共子空间的方式来学习稳定可靠的迁移系数，同时考虑了源域与目标域可能距离较远的情况；TDR 的方法一来是在原始空间上融合源域与目标域的信息，并未考虑到域间差异可能导致这部分信息中包含不可靠的部分；另一方面 TDR 目标函数优化得到的特征投影矩阵，是在源域与目标域间进行了折中的结果。当源域与目标域差异较小的时，TDR 的方法亦能得到不错的性能，但是当源域与目标域有明显差异时，TDR 的方法性能下降较快。

与第 2 章提出的 TAME 方法相比，TSC 方法又有了进一步的提升。在跨姿态的迁移学习实验中，TSC 的方法较 TAME 有大约 4% 左右的提升。这是因为 TSC 方法中采用两个投影矩阵来将源域与目标域投影至同一个公共子空间，；而

TAME 方法只采用一个投影矩阵将源域与目标域投影至公共子空间,当源域与目标域之间差异较大时,这种投影方式可能使得源域和目标域在公共子空间上仍然存在一定差异,在此基础上利用源域的监督信息来指导目标域的相似度矩阵估计,可能存在不可靠的部分。

表格 3-1 测试 TSC 在跨种族问题下的性能及与其他方法的对比

迁移问题		OFD→XM2VTS	XM2VTS→OFD
目标域模型	PCA	0.745	0.268
源域模型	FLD	0.768	0.593
迁移学习模型	TDR[54]	0.858	0.971
	SGF[31]	0.783	0.653
	ITL[52]	0.715	0.110
	TCA[34]	0.740	0.517
	TAME	0.844	0.979
	TSC	0.859	0.983

表格 3-2 测试 TSC 在跨采集条件问题下的性能及与其他方法的对比

迁移问题		XM2VTS→FRGC	FRGC→XM2VTS
目标域模型	PCA	0.012	0.745
源域模型	FLD	0.083	0.767
迁移学习模型	TDR[54]	0.115	0.840
	SGF[31]	0.073	0.791
	ITL[52]	0.041	0.731
	TCA[34]	0.101	0.809
	TAME	0.190	0.840
	TSC	0.245	0.843

表格 3-3 测试 TSC 在跨姿态问题下的性能及与其他方法的对比

源域		-45°	45°	0°	0°	-45°	-30°	30°
目标域		45°	-45°	45°	-45°	30°	30°	-30°
目标域模型	PCA	0.53	0.535	0.53	0.535	0.582	0.582	0.584
源域模型	FLD	0.644	0.637	0.467	0.586	0.691	0.692	0.696
迁移模型	TDR	0.658	0.665	0.678	0.645	0.708	0.712	0.723
	SGF	0.631	0.608	0.631	0.609	0.664	0.635	0.676

	ITL	0.664	0.64	0.671	0.62	0.658	0.666	0.698
	TCA	0.62	0.588	0.562	0.521	0.655	0.645	0.669
	TAME	0.691	0.644	0.693	0.673	0.756	0.758	0.743
	TSC	0.738	0.731	0.738	0.720	0.780	0.798	0.779

3.4.2 实验分析

● 参数敏感性

TSC 的目标化函数(3-25)中, 最为重要的参数是 λ 。这是因为参数 λ 是在域自身信息量保持与缩减域差异进行调节的参数。 λ 越大则对域间差异的容忍性越低, 优化得到的公共子空间上域特有的信息可能越少, 丢失的信息可能也越多; λ 越小则对域间差异的容忍性越低, 虽然保持的域上信息越多, 但可能域间差异仍然较大, 可能会导致学习得到的迁移系数不可靠。因此 λ 的选取可能会对 TSC 的性能有一定影响。因此本研究在跨姿态的模拟实验室上测试了不同 λ 值下, TSC 方法的性能, 其结果如图 3-3 所示。

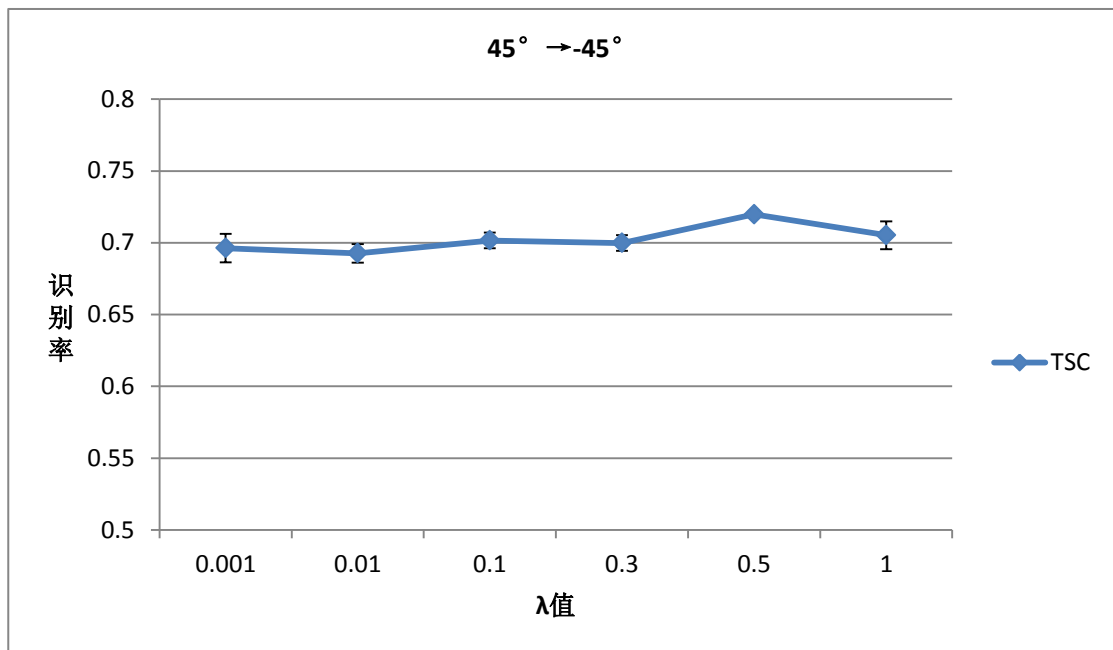


图 3-3 TSC 方法在源域为 45° 、目标域为 -45° 下, 随 λ 值变化的曲线, 以及随机初始化 W_s 、 W_t 矩阵时的标准差图。

其中图 3-3 中的曲线为 TSC 方法在源域为 45° 、目标域为 -45° 的跨姿态迁移实验上随参数 λ 变化的曲线。由该图可见，TSC 方法在参数 λ 取不同值时，其性能变化并不剧烈，可见 TSC 方法对参数 λ 具有鲁棒性。此外该图上代表性能的每一点上，“T”型线段表示投影矩阵 W_s 、 W_t 随机初始化 10 次的情况下 TSC 方法的方差，由图可见该方差俱较小。这表明投影矩阵 W_s 、 W_t 的随机初始化并不会很大大地影响 TSC 方法的性能。

● 迁移系数及其应用方式的有效性

TSC 方法中认为直接在原始空间上学习的迁移系数不够可靠，因此需先优化得到一个公共子空间，在此基础上方可学到稳定的迁移系数，即为目标化函数中的 V_s 。为了验证 TSC 方法中如此学习迁移系数的有效性，我们在几个跨姿态的迁移学习实验上分别对比了以下几个方法是性能：1) SC 方法。该方法直接在原始空间上利用稀疏编码学习源域至目标域的迁移系数；2) LCS 方法。该方法指的是在基于 TSC 目标函数优化得到的公共子空间上采用源域样本训练模型的方法。该方法可视为将 TSC 方法退化成基于公共子空间的迁移方法。3) 本章提出的 TSC 方法，即借助公共子空间的优化学习得到稳定的迁移系数，进而将其应用在原始空间上实现将源域样本迁移至目标域的目标，然后借助目标化后的源域样本训练模型的方法。

由表格 3-4 可见，在所有的跨姿态迁移实验中，SC 方法的性能远不如 LCS 的性能，证明了原始空间上直接学习迁移系数的不可靠性；另一方面也说明了 TSC 目标函数估计的公共子空间确实提取到了源域上可靠的有利信息。此外，该表格中，TSC 方法也较 LCS 得到更好的性能，这是因为 LCS 的方法仅利用到了目标域在公共子空间上的信息，而 TSC 是在原始空间上将源域样本迁移至目标域，迁移后的样本蕴含目标域整个特征子空间上的信息，可用于目标域上人脸识别的信息更多，因此训练得到的模型性能更好。

表格 3-4 TSC 迁移方法的有效性。分别对比了直接在原始空间上进行稀疏重构迁移的性能（SC），公共子空间模型（LCS）的性能，以及 TSC 方法的性能。

源域	0°	0°	-45°	45°	-30°	30°
目标域	-45°	45°	45°	-45°	30°	-30°
SC	0.481	0.564	0.436	0.488	0.515	0.523
LCS	0.617	0.707	0.629	0.600	0.653	0.671
TSC	0.725	0.731	0.738	0.738	0.798	0.779

3.5 总结与分析

本章提出的 TSC 方法从解决目标域缺少带标注样本的问题出发, 提出将带标注的源域样本目标化的迁移学习方法。该方法中的目标函数可同时估计公共子空间与迁移系数, 估计得到的迁移系数可用于将源域样本迁移至目标域, 即源域样本的目标化。由于迁移的过程中, 并不影响源域上样本的类别信息, 因此目标化后的源域样本仍然携带有类别信息。通过该方法目标域上即可获得一批可用于训练模型的带标注样本, 从而可训练得到更适合目标域上任务的模型。在 XM2VTS 到 FRGC 等跨库的迁移实验中, TSC 方法较与之比对的方法有明显的性能提升, 展示了该方法的有效性与优越性。

TSC 方法中的迁移系数是基于公共子空间上估计得到的, 但却可直接应用在原始空间上实现对源域样本的目标化, 其潜在假设是公共子空间上源域相对于目标域样本的空间位置关系良好地逼近在原始空间上目标化的源域样本相对于目标域样本的关系。由于迁移学习的人脸识别问题中, 源域与目标域上都是关于人脸的图像, 两个域本身蕴含许多可共享的信息, 因此该假设一般情况下都是满足的。

第4章 融合公共与特有子空间上判别性特征的迁移学习方法

当前学术上主流的迁移学习方法大都是基于公共子空间的迁移学习方法,旨在通过寻找到一个源域与目标域无差异的特征子空间,在此基础上提取源域中对目标域上任务有可靠帮助的信息。但这些方法只考虑了公共子空间中所蕴含的对目标域有帮助的信息,而忽略了目标域除公共子空间外所剩下的目标域特有子空间(Target Domain Specific Subspace)中亦可能蕴含对目标域上分类有用的信息。若能全面地利用到蕴含在目标域所在的特征子空间中所有的信息,训练得到的模型势必能更好地应对目标域上的分类问题。因此本章提出了一种融合公共与特有子空间上判别性特征的迁移学习方法。该方法亦是基于公共子空间搜索的迁移学习方法,而不同的是该方法在优化得到公共子空间后,全面提取目标域原始空间中所有可能的有用信息。

4.1 相关背景知识

4.1.1 主成份分析

主成份分析(Principal Component Analysis, 简称 PCA[71])是一种无监督线性降维的方法,其目标为寻找一组线性无关的投影方向,同时尽可能地保留域上信息。PCA 中认为协方差矩阵是描述域上样本所涵盖信息的良好统计量,故而将其核心思想形式化为寻找最大化保持协方差矩阵的投影方向,其解为该协方差矩阵的特征向量。这些投影向量被称为主成份(Principal Components, 简称 PCs),每一投影方向上保留原始特征空间信息的多少可采用其所对应的特征值大小来衡量。特征值越大的特征向量越好地保持了协方差矩阵的大小,即越多地保留了原始特征子空间上的信息。在 PCA 中选取的投影方向所对应的特征值总和占协方差矩阵所有特征值总和的百分比称之为能量。基于 PCA 方法估计投影方向的方法如算法 4-1 所示。



图 4-1 主成份分析的前 8 维本征脸

实际应用中，通常将 PCA 作为降维、去噪等预处理，可大大减轻后续学习方法的计算复杂度。在 1991 年，由 Petland 等人在文章[72]中首次将其应用在人脸识别领域，使得人脸识别技术从基于简单的几何结构信息发展到基于人脸纹理比对的路线上。在这篇文章中 Petland 等人将选取的主成份可视化成图像，如图 4-1 所示，可视化后的主成份具有模糊的人脸轮廓，故将其称之为本征脸（Eigenface）。

算法 4-1 主成份分析（PCA）算法

输入：样本特征向量 $X_i \in R^p$ ， $i=1,2,\dots,n$ ，目标维数 d 或者保持的能量 e 。
输出：降维投影矩阵 W
<p>1. 计算所有样本的协方差矩阵 $\Sigma = \sum_{i=1}^n (X_i - u)(X_i - u)^T$，$u = \frac{1}{n} \sum_{i=1}^n X_i$。</p>
<p>2. 对协方差矩阵 Σ 进行奇异值分解[73]（SVD）分解，即 $\Sigma = UDV^T$，其中 D 为按特征值从大到小排序好的对角矩阵，即有，</p> $D = \text{diag}(\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_n)。$ <p style="text-align: right;">(4-1)</p> <p>其中 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$。相应地，$U$ 中的列向量即为按对应的特征值从大到小排序后的特征向量。</p>
<p>3. 根据目标维数 d 确定投影矩阵：选取 U 的前 d 个列向量，构成投影矩阵 W。</p> <p>根据保持的能量 e 确定投影矩阵：首先根据以下公式确定目标维数 d，</p>

$$d = \arg \min_d \left(\frac{\sum_{i=1}^d \lambda_i}{\sum_{i=1}^n \lambda_i} \geq e \right). \quad (4-2)$$

再得到投影目标维数 d 后，根据目标维数 d 确定投影矩阵。

4.2 基本思路与方法框架

4.2.1 基本思路

与利用充足的目标域带标注样本训练得到的模型相比，基于公共子空间的迁移方法只利用到了在公共子空间上带标注源域的监督信息，缺少对目标域特有子空间（Target Domain Specific Subspace）上所蕴含的信息进行提取。但是目标域特有子空间上的信息亦可能对目标域上的任务学习有帮助，因此更为合理的思路（如图 4-2 所示）是融合公共子空间与目标域特有子空间上的信息来训练模型，从而提升模型的性能。其中由于目标域与源域在公共子空间上无明显差异，且源域样本携带有标注信息，因此目标域在公共子空间上蕴含的有用信息可基于带标注信息的源域进行提取。另一方面，目标域特有子空间上的信息，可由目标域样本在该空间上的样本空间位置关系进行描述。这一部分信息可采用无监督聚类来估计目标域样本在该空间上体现出的类别关系，进而提取对目标域上分类任务有用的信息。该方法同时利用了公共子空间与目标域特有子空间上的信息来提取判别性特征，是一种新的迁移思路。本章将该方法称之为融合公共与特有子空间上判别性特征的迁移方法，即 **F**use **C**ommon and **S**pecific **D**iscriminative Feature for Transfer Learning，简称 FCSD。

具体来说，公共子空间上的模型训练可与一般基于公共子空间的迁移学习方法类似：先基于降低源域与目标域间差异的准则优化公共子空间，然后将源域样本投影至该公共子空间，进而利用源域上的监督信息提取对目标域分类任务有用的判别性信息。该空间上判别性特征提取最关键部分的是如何优化源域与目标域共享的公共子空间。第 3 章提出的 TSC 方法采用结合域上局部结构保持与最小化重构残差这两个约束来优化公共子空间，其中将局部结构保持的约束形式化为关于拉普拉斯矩阵（Laplacian Matrix）[65]的目标函数。但是由于拉普拉斯矩阵的计算需涉及较多的调优（权重计算函数调优、参数调优等），无形中增加了该目标函数的复杂度，因此本章采用刻画整个子空间上变化的协方差矩阵来替换拉

普拉斯矩阵作为域上信息保存的统计量。PCA 方法的广泛应用，体现了协方差矩阵作为域上信息保持统计量的可靠性。因此本章提出的关于公共子空间优化的两个约束为最大化协方差矩阵与最小化稀疏重构误差这两项，其形式化的目标函数如(4-7)所示。

目标域特有子空间上的判别性特征我们采用无监督聚类（Kmeans）的方法先提取蕴含在样本空间位置关系中的类别信息，在此基础上采用监督学习方法（比如 LDA）提取对分类有用的判别性特征。该空间上所蕴含的信息为一般基于公共子空间的方法所忽视，而本章提出的方法则首次显示地考虑了如何利用该空间上的信息。

该方法希望能够充分利用源域中有用的信息，同时也利用目标域特有子空间上的信息，以期能提取到更多对目标域上分类任务有用的判别性信息。

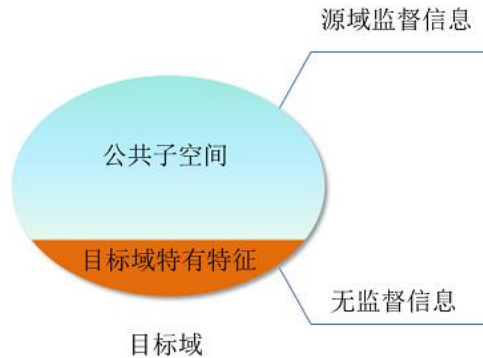


图 4-2 FCSD 方法的原理示意图

4.2.2 方法框架

本章节提出的 FCSD 方法旨在全面地利用目标域所有特征子空间上所蕴含的有用信息，从而学习得到能更好解决目标域上分类问题的模型。具体来说，首先需估计得到源域与目标域的公共子空间（估计方法见下节），然后分别在公共子空间与特有子空间上训练模型，进而融合这两个子空间上的模型得到最终模型。由于在人脸识别领域，大部分的模型都是基于判别性特征提取然后采用最近邻分类器进行人脸识别的方法，因此本章提出的 FCSD 方法中训练的模型亦采用这种思路，这里选用 FLD 作为提取判别性特征的方法，方法的应用框架如图 4-3 所示。

如图 4-3 所示，FCSD 方法在公共子空间上利用源域带标注的样本作为训练数据，然后采用有监督的学习方法（比如 FLD）学习判别性的特征子空间；在

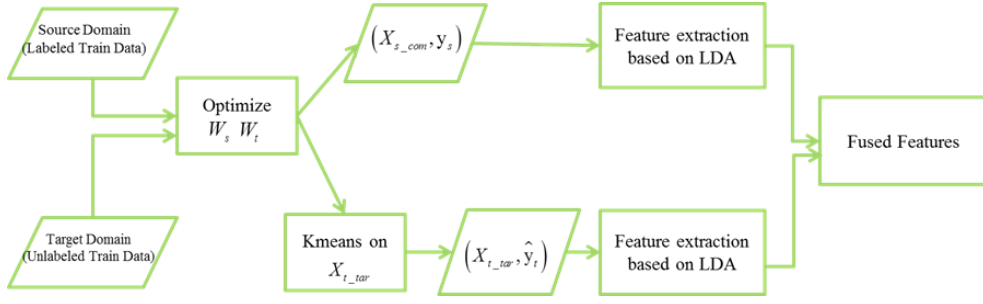


图 4-3 FCSD 方法框架图

特有子空间上，利用目标域上无标注的样本作为训练数据，然后采用无监督聚类方法（比如 **Kmeans**）提取无标注样本空间位置关系中所蕴含的类别信息，进而利用有监督的学习方法（比如 **FLD**）学习判别特征子空间，最后融合这两部分的判别性特征作为最终提取到的判别性特征。**FCSD** 方法具体的应用算法如算法 4-2 所示。

算法 4-2 公共子空间与特有子空间判别性特征融合（FCSD）的算法

步骤 1：根据目标函数(4-7)优化得到将源域与目标域投影至公共子空间的的投影矩阵 W_s 及 W_t

步骤 2：在公共子空间与特有子空间上分别提取判别性特征

1. 在公共子空间上提取判别性特征投影。

利用 W_s 将源域样本投影至公共子空间，得到源域样本在公共子空间上的样本矩阵，即

$$X_{s_com} = W_s^T \times X_s \quad (4-3)$$

由于源域样本投影至公共子空间并不影响其类别信息，故而公共子空间上存在一批带标注信息的样本 (X_{s_com}, y_s) ，其中 y_s 为源域所有样本标签所构成的向量。将 (X_{s_com}, y_s) 作为训练数据，采用 **FLD** 提取判别性的特征，即得到在公共子空间上的判别性特征投影矩阵 W_{com_LDA} 。

2. 在特有子空间上提取判别性特征投影。

将目标域样本中所包含的关于公共子空间上的特征去掉，即可获取仅包含特有子空间上特征的目标域样本，即有，

$$X_{t_tar} = X_t - (W_t^T)^{-1} W_t^T X_t \quad (4-4)$$

其中 X_{t_tar} 为特有子空间上目标域样本矩阵。将 X_{t_tar} 作为训练数据，先

利用 Kmeans 估计其类别信息，得到 (X_{t_tar}, y_t) ；进而采用 FLD 提取该空间上的判别性特征。记该空间上的判别性特征投影矩阵为 W_{LDA_tar} 。

步骤 4: 融合公共与特有子空间上的判别性特征投影矩阵，其融合方式如下所示，

$$W_{LDA_ALL} = \begin{bmatrix} W_t W_{LDA_com} & W_{LDA_tar} \end{bmatrix} \quad (4-5)$$

W_{LDA_ALL} 即为利用到目标域所在的全部特征子空间上的信息而估计得到的判别性特征投影矩阵。在对目标域上的进行分类时，可先利用 W_{LDA_ALL} 将目标域样本投影至具有判别性的特征空间，进而采用最近邻分类器进行分类。

4.2.3 目标函数

本节介绍 FCSD 方法中提出的优化公共子空间的目标函数。记源域、目标域上分别进行零均值操作后的样本矩阵为 X_s 、 U_t 。基于上述的分析，本章采用最大化协方差矩阵与最小化稀疏重构误差的两个约束来优化公共子空间，下文分别对这两个约束进行形式化。与第 3 章提出的 TSC 目标函数类似，本章亦采用两个投影矩阵 W_s 、 W_t 分别将源域与目标域投影至公共子空间。

最大协方差矩阵：由于 X_s 、 U_t 分别为源域与目标域均值为 0 的样本矩阵，那么源域与目标域上的协方差矩阵可分别由 $X_s X_s^T$ 、 $U_t U_t^T$ 表示。投影矩阵 W_s 、 W_t 应尽量地保持源域与目标域上的协方差矩阵，即可形式化成如下所示的目标函数，

$$\begin{aligned} W_s^* &= \arg \max_{W_s} W_s^T X_s X_s^T W_s \\ W_t^* &= \arg \max_{W_t} W_t^T U_t U_t^T W_t \end{aligned} \quad (4-6)$$

最小化稀疏重构误差：该约束的形式化详见第 3.2.3 小节介绍的“稀疏重构残差最小化”部分。

结合上述两个约束的形式化，可得到如下的目标函数，

$$\begin{aligned} &W_s^*, W_t^*, V_s^*, V_t^* = \\ &\arg \min_{W_s, W_t, V_s, V_t} \frac{Tr \left(\frac{1}{n_s} W_s^T X_s X_s^T W_s + \frac{1}{n_t} W_t^T U_t U_t^T W_t \right)}{\frac{1}{n_s} \|W_s^T X_s - W_t^T U_t V_s\|_F^2 + \frac{1}{n_t} \|W_t^T U_t - W_s^T X_s V_t\|_F^2} \quad (4-7) \\ &\text{s.t. } \|v_{s_i}\|_0 < \tau, \|v_{t_j}\|_0 < \tau, i=1, 2, \dots, n_s, j=1, 2, \dots, n_t \end{aligned}$$

该目标函数中通过最小化公共子空间上域间重构误差最小化这两个约束，共同优化得到公共子空间的投影矩阵 W_s 、 W_t ，这两个投影矩阵可分别将源域与目

标域投影至公共子空间,进而根据算法 4-2 分别在公共子空间与特有序子空间上学习具有判别性的特征投影。

4.3 优化方法与算法流程

4.3.1 优化方法

本方法采用迭代优化的方式对目标函数(4-7)进行求解。记源域与目标域所在的原始特征子空间空间维数分别为 p_s 、 p_t , 公共子空间的维数为 d , 那么 W_s 、 W_t 分别为 $p_s \times d$ 、 $p_t \times d$ 的投影矩阵。

给定 W_s 、 W_t 求解 V_s 、 V_t : 由目标函数(4-7)可知, 在给定 W_s 、 W_t 的情况下, V_s 、 V_t 的求解相互独立, 且与 W_s 、 W_t 无关, 故可分别独立优化。以求解 V_s 为例, 在给定 W_s 、 W_t 时, 由于 V_s 的每一列求解是独立, 故可分解成对每个输入信号的独立稀疏编码, 如公式(4-8)所示。

最小化稀疏重构误差的约束旨在降低源域与目标域间的差异, 使得两个域上所有样本的空间位置关系无明显的域差异。基于这样的先验, 本节提出的优化算法在求解稀疏系数时增加非零系数集中在某些样本上的惩罚, 以期能更快地促进两个域上样本的混合, 使优化过程能够更快收敛。以求解稀疏编码 v_{s_i} 为例, 增添惩罚后的目标函数如公式所示(4-8)。其中 h_s 为基的权重, 初始时 h_s 为一个元素值全为 1 的 n_s 维的向量, 每一维表示其对应基的权重, 每次稀疏编码后, 对 h_s 进行更新, 如公式(4-9)所示, $|v_{s_i}|$ 为对 v_{s_i} 的每一项元素取绝对值。

$$v_{s_i}^* = \arg \min_{V_{s_i}} \|W_s^T x_{s_i} - W_t^T U_t v_{s_i}\|_F^2 + \lambda \|1 - h_s^T v_{s_i}\|_F^2 \quad (4-8)$$

$$\text{s.t. } \|v_{s_i}\|_0 < \tau, i = 1, 2, \dots, n_s$$

$$h_s = \max(h_s - \frac{0.5}{\max |v_{s_i}|} |v_{s_i}|, 0) \quad (4-9)$$

根据公式(4-10)(4-11) 定义变量 \tilde{z}_{s_i} 和 Z_t , 将其带入目标函数(4-8), 则可将其整理如(4-12)所示的目标函数, 这是一个 lasso 回归问题, 可采用最小角度回归的方式进行求解。

$$\tilde{z}_{s_i} = \begin{bmatrix} W_s^T x_{s_i} \\ \sqrt{\lambda} \end{bmatrix} \quad (4-10)$$

$$Z_t = \begin{bmatrix} W_t^T X_t \\ \sqrt{\lambda} h_s^T \end{bmatrix} \quad (4-11)$$

$$\begin{aligned} v_{s_i}^* &= \arg \min \| \tilde{z}_{s_i} - Z_t v_{s_i} \|^2 \\ \text{s.t. } |v_{s_i}|_0 &< \tau, i=1,2,\dots,n_s \end{aligned} \quad (4-12)$$

给定 V_s 、 V_t 求解 W_s 、 W_t ：固定 V_s 、 V_t 时，保留目标函数(4-7)中关于 W_s 、 W_t 的项，并记 $W=[W_s^T \ W_t^T]^T$ ，将其带入目标函数(4-7)，则可将其整理成如下所示的目标函数，

$$W^* = \arg \max_W \frac{\text{Tr}(W \Sigma_b W^T)}{\text{Tr}(W \Sigma_w W^T)} \quad (4-13)$$

其中，

$$\Sigma_b = \begin{bmatrix} \frac{1}{n_s} X_s X_s^T & 0 \\ 0 & \frac{1}{n_t} U_t U_t^T \end{bmatrix} \quad (4-14)$$

$$\Sigma_w = \begin{bmatrix} X_s (\frac{1}{n_s} I + \frac{1}{n_t} V_t V_t^T) X_s^T & -X_s (\frac{1}{n_s} V_s^T + \frac{1}{n_t} V_t) X_s^T \\ -U_t (\frac{1}{n_t} V_t^T + \frac{1}{n_s} V_s) U_t^T & U_t (\frac{1}{n_t} I + \frac{1}{n_s} V_s V_s^T) U_t^T \end{bmatrix} \quad (4-15)$$

目标函数(4-13)是一个最大化 Rayleigh 商的问题，可通过求解广义特征值得到，优化得到的 W 矩阵中，取前 p_s 行构成 W_s ，余下的 p_t 行构成 W_t 。

4.3.2 算法流程

根据第 4.3.1 节介绍的优化方法，FCSD 中提出的公共子空间优化步骤可整理如算法 3-3 所示。

算法 4-3 FCSD 中公共子空间的优化算法

输入：源域样本 X_s ，目标域样本 U_t ，非零系数个数 τ ，惩罚调节参数 λ ，最大循环次数 T ，。
输出：投影矩阵 W_s 、 W_t
1. 随机初始化 W_s 、 W_t ，约束其每个列向量长度为 1；初始化循环次数 $t=0$ 。
2. 循环 <div style="margin-left: 20px;"> a) 优化稀疏系数矩阵 V_s、V_t <div style="margin-left: 20px;"> i. 遍历 $i=1,2,\dots,n_s$ </div> </div>

$$v_{s_i}^* = \arg \min_{v_{s_i}} \left\| W_s^T x_{s_i} - W_t^T U_t v_{s_i} \right\|_F^2 + \lambda \left\| 1 - h_s^T v_{s_i} \right\|^2, \text{s.t. } \|v_{s_i}\|_0 < \tau$$

$$h_s = h_s - \frac{0.5}{\max(|v_{s_i}^*|)} |v_{s_i}^*|$$

结束。

ii. 遍历 $j=1, 2, \dots, n_t$

$$v_{t_j}^* = \arg \min_{v_{t_j}} \left\| W_t^T u_{t_j} - W_s^T X_s v_{t_j} \right\|_F^2 + \lambda \left\| 1 - h_t^T v_{t_j} \right\|^2, \text{s.t. } \|v_{t_j}\|_0 < \tau$$

$$h_t = h_t - \frac{0.5}{\max(|v_{t_j}^*|)} |v_{t_j}^*|$$

结束。

b) 根据公式(4-13)(4-14)(4-15)优化 W_s 、 W_t 。

c) 循环次数更新 $t=t+1$ 。

d) 检查收敛条件。

若满足 $\|W_s - W_s\|_F^2 < \varepsilon$ 且 $\|W_t - W_t\|_F^2 < \varepsilon$ 或则收敛，跳出循环；或循环次数

超过最大循环次数，即 $t > T$ 则跳出循环。除以上情况外，返回步骤 a)。

结束

4.4 实验及性能比较

4.4.1 不同迁移问题下利用 FCSD 迁移的人脸识别实验

为了测试 FCSD 方法融合公共与特有子空间模型的有效性，我们在不同的迁移问题上测试其性能，关于实验的迁移问题设置详见第 2.4.1 节。本研究将在这些迁移实验上与发表在顶级会议上的迁移学习方法 TDR[54]、SGF[31]、ITL[52]、TCA[34]以及第 3 章提出的基于源域样本目标化的人脸识别（TSC）方法进行对比。这些方法的参数设置如第 2.4.2 节、第 3.4 节中介绍的相同。

由表格 4-1、表格 4-2 和表格 4-3 可知，在所有的迁移实验中，FCSD 方法提取得到的判别特征较直接利用源域样本作为训练数据基于 FLD 提取得到的判别性特征性能有明显的提升。

此外，本实验中还对比了在原始空间上直接进行无监督聚类进而提取判别性特征的方法（Kmeans+FLD）得到的性能。由表格 4-1、表格 4-2 和表格 4-3 上

的实验结果可知，在整个目标域子空间上采用无监督聚类存在一定的不稳定性，故而在不同的迁移问题中，该方法的性能有高有低。相比之下，FCSD 方法由于充分利用了目标域子空间中包含与源域共享的许多信息，在共享的子空间上利用带标注的源域样本学习可靠有用的判别性特征，进而得到的模型在大部分的迁移实验上都得到了较为理想的性能。

在大部分的实验上，FCSD 方法都得到了与 TSC 方法最为接近的性能。这是因为这两个方法都同时利用到了目标域整个特征子空间上可能包含的有用信息。当目标域特有子空间上的样本空间位置关系越能体现其类别样本信息时，FCSD 方法提取得到的信息则更为有效，比如表格 4-1、表格 4-2 中目标域为 XM2VTS 的迁移实验；若目标域样本在公共子空间与特有子空间上样本的空间位置关系越为接近，那么 TSC 方法在公共子空间上估计得到的迁移系数矩阵则能更为有效地将源域样本迁移至目标域，从而提取得到可靠的信息，这种情况下 TSC 方法可得到更好的结果。

FCSD 方法的优化过程中初始化投影矩阵 W_s 、 W_t 时采用随机初始化的方式，可能对 FCSD 方法的性能造成影响。为了估计随机初始化对 FCSD 方法的影响，本研究在源域为 XM2VTS 数据库、目标域为 OFD 数据库的迁移问题上，测试了随机初始化 W_s 、 W_t 10 次的情况下 FCSD 模型的性能，在此基础上计算得到性能的方差为 $3.0589e-5$ 。由此可见在投影矩阵随机初始化对于方法的求解影响并不大。

表格 4-1 测试 FCSD 在跨种族问题下的性能及与其他方法的对比

迁移问题		OFD→XM2VTS	XM2VTS→OFD
目标域模型	PCA	0.745	0.268
	Kmeans+FLD	0.857	0.981
源域模型	FLD	0.768	0.593
迁移学习模型	TDR	0.858	0.971
	SGF	0.783	0.653
	ITL	0.715	0.110
	TCA	0.740	0.517
	TSC	0.859	0.983
	FCSD	0.869	0.974 (3.0589e-5)

表格 4-2 测试 FCSD 在跨采集条件问题下的性能及与其他方法的对比

迁移问题		XM2VTS→FRGC	FRGC→XM2VTS
目标域模型	PCA	0.012	0.745
	Kmeans+FLD	0.161	0.857
源域模型	FLD	0.083	0.767
迁移学习模型	TDR	0.115	0.840
	SGF	0.073	0.791
	ITL	0.041	0.731
	TCA	0.101	0.809
	TSC	0.245	0.843
	FCSD	0.231	0.859

表格 4-3 测试 FCSD 在跨姿态问题下的性能及与其他方法的对比

源域		-45°	45°	0°	0°	-45°	-30°	30°
目标域		45°	-45°	45°	-45°	30°	30°	-30°
目标域模型	PCA	0.533	0.535	0.533	0.535	0.534	0.588	0.584
	Kmeans+FLD	0.662	0.658	0.662	0.658	0.751	0.751	0.74
源域模型	FLD	0.644	0.637	0.467	0.586	0.691	0.692	0.696
迁移模型	TDR	0.658	0.665	0.678	0.645	0.708	0.712	0.723
	SGF	0.631	0.608	0.631	0.609	0.664	0.635	0.676
	ITL	0.664	0.64	0.671	0.62	0.658	0.666	0.698
	TCA	0.62	0.588	0.562	0.521	0.655	0.645	0.669
	TSC	0.738	0.731	0.738	0.720	0.780	0.798	0.779
	FCSD	0.723	0.709	0.72	0.705	0.799	0.797	0.772

4.4.2 实验分析

公共子空间能量的敏感性: 本研究中将采用 FCSD 目标函数估计得到的公共子空间维数占目标域所在原始空间维数的比例称之为公共子空间的能量 e ，即有 $e = d/p_t$ 。理想情况下，应尽量多地提取到源域与目标域共享的公共特征，从而利用到更多源域上对目标域分类有用的可靠信息。FCSD 方法中可通过为公共子空间选择不同的维数来约束公共子空间上的能量。维数越高则学习的公共子空间占源域原始空间的比例越高，但若学习的公共子空间过高，引入源域中对目标域上的人脸识别有害的信息，既不利于源域与目标域差异的消除，亦不利于源域中 useful 信息的迁移，反而会得到较差的性能。为了验证这一推论，本研究在源域为 XM2VTS，目标域为 FRGC 以及 MultiPIE 上跨姿态的几个迁移实验上测试了当公共子空间选择不同的能量时，优化得到的公共子空间上所包含的信息对目标域分类问题的有用情况，如图 4-4、图 4-6 所示。其中图 4-4、图 4-5 为分别为参数 λ 、 τ 取不同值下，公共子空间上保留不同能量时对其保留有用信息的影响。公共子空间上有用的信息采用该子空间上训练得到的模型（FLD+NN）在目标域上人脸识别问题上的性能来衡量。由该图可见，公共子空间的能量并不是越高越好，当达到 100% 时性能反而猛得下降，这是因为在原始空间上，源域与目标域之间存在显著的差异，故而在此基础上学习得到的公共子空间可能存在许多不利、不可靠的信息。但是若公共子空间保留的能量太低，则提取到源域中对目标域上分类任务有用的信息越少，亦不利于目标域上任务的处理。由图 4-4、图 4-5 可见，在该迁移实验上，保留 75% 的能量可得到最好的性能，该结论亦可由图 4-6

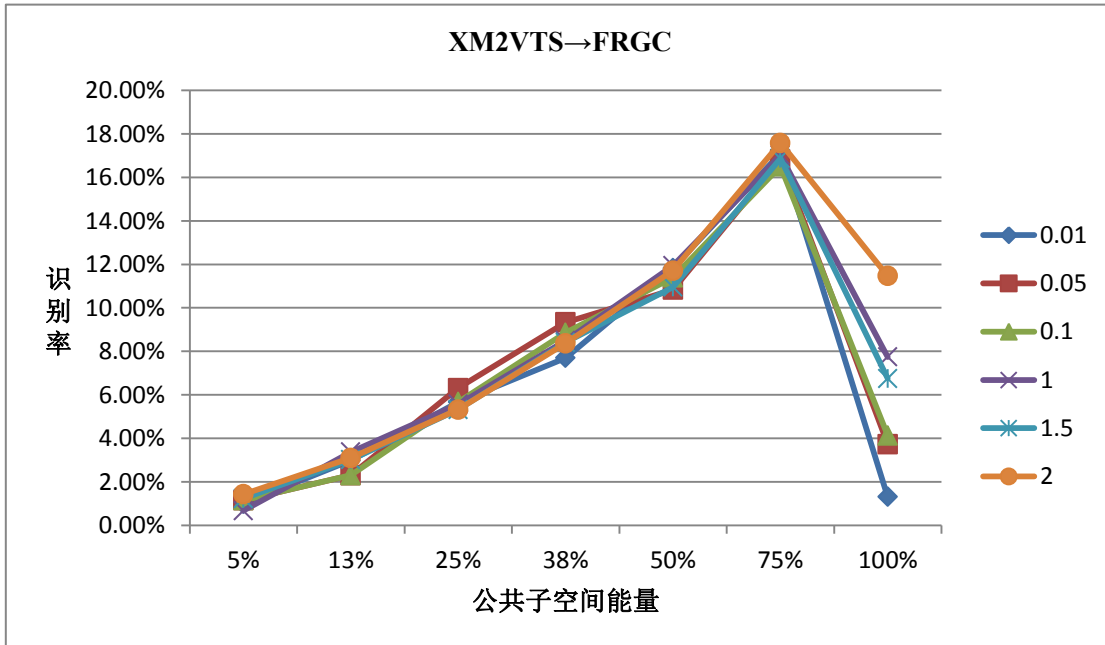


图 4-4 公共子空间模型在不同的 λ 下对公共子空间能量的敏感性。纵轴为人脸识别率，横轴为公共子空间保留的能量，曲线的不同颜色表示取不同的参数 λ 值

上的实验可见。此外，由图 4-4、图 4-5 可见，由 XM2VTS 迁移至 FRGC 的实验中，由 FCSD 目标函数估计得到的公共子空间上利用源域带标注样本训练得到的模型最好的性能为 17.58%，该性能不如 FCSD 方法在该迁移实验上得到的性能，这表明 FCSD 融合公共与特有子空间上信息的方法确实提取到更多的对目标域分类任务有用的信息，从而提升了模型的性能。

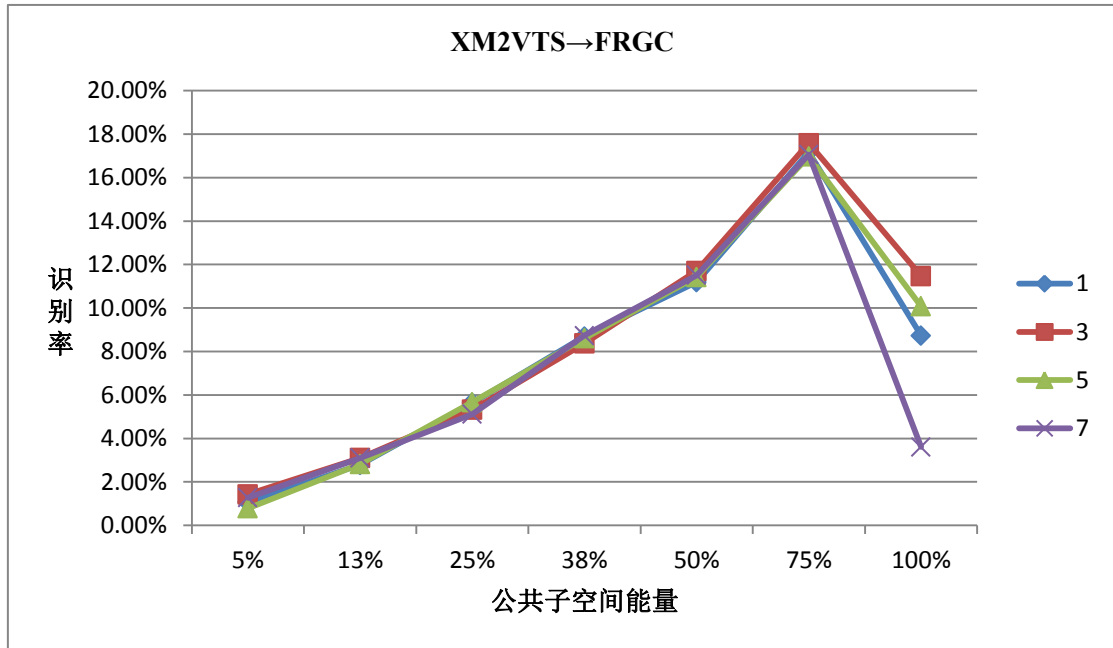


图 4-5 公共子空间特征模型性能在不同的 τ 下对公共子空间能量的敏感性。纵轴为人脸识别率，横轴为公共子空间保留的能量。曲线的不同颜色表示取不同的参数 τ 值。

参数 λ 与 τ 的敏感性：由图 4-7 可见在 MultiPIE 的两个姿态迁移问题上，FCSD 模型性能随着最近邻个数 τ 由小到大 (1-3-5) 慢慢有所增加，但若太高 ($\tau=7$) 则会有显著的下降。这是因为，若采用过多的样本来对另外一个域中的样本进行表示时，稀疏编码寻找最为相关部分样本的能力则开始退化，故而在非零系数过高的情况下，模型的性能则开始下降。其中在由 -45° 迁移至 30° 的实验中，近邻个数在取 5 时达到最佳；而在由 45° 迁移至 30° 的实验中，近邻个数在取 3 时达到最佳。这是因为当源域与目标域较远的情况下，采用一个域作为词典对另外一个域进行稀疏表述的时所需的样本应较多些，方可得到可靠稳定的编码。由这两个实验可见，当源域与目标域间的差异较大的情况下，非零系数的个数应适当多些，适当放松稀疏编码时的约束；反之若差异较小，则非零稀疏的个数应适当控制，以防降低稀疏编码寻找相关样本的能力。由图 4-7 上两个实验的结果亦

可见，参数 λ 一般取 1.5 时较好。参数 λ 是用于域间互相稀疏重构时对基进行重复利用时的惩罚，稍微取大些可避免寻找得到的公共子空间上域间样本仅与另一域的部分样本混合的情况。但是参数 λ 取太大（ $\lambda=2$ ）则对重复利用到的基惩罚较重，可能导致学习得到的域间样本关系不够可靠。

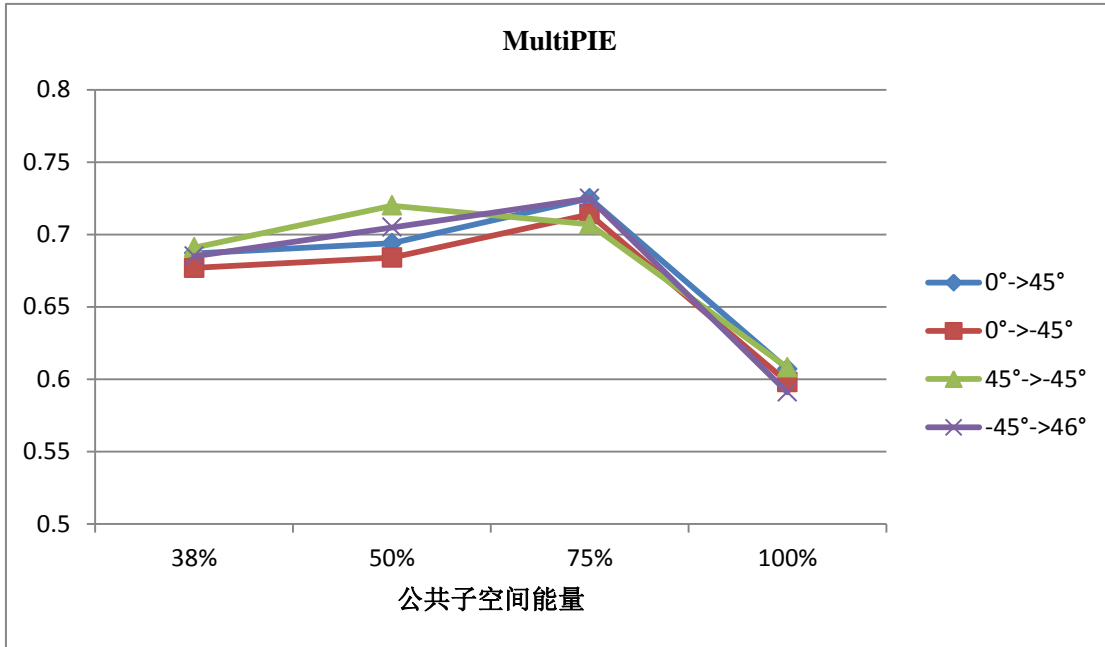


图 4-6 MultiPIE 不同迁移问题下，公共子空间模型对于公共子空间保留的能量敏感性。

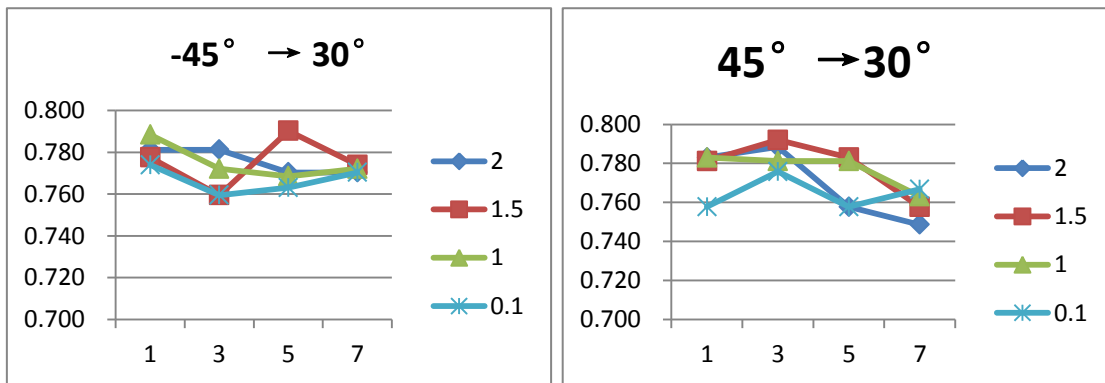


图 4-7 MultiPIE 两个不同的迁移问题下，FCSD 模型随参数 λ 和 τ 变化的性能曲线。

4.5 总结与分析

当前基于公共子空间的迁移学习方法只利用到源域在公共子空间上的信息，而忽略了目标域特有子空间上亦可能蕴含对目标域上分类问题有帮助的信息，基

于这一思路,本章提出了一种如何全面地利用目标域所在的特征子空间上所有可能有帮助信息的融合方法,其中公共子空间上利用带标注源域所携带的监督信息,而目标域特有子空间上利用无监督聚类提取无标注目标域样本空间位置关系中所蕴含的类别信息,进而提取对分类有用的判别性特征。该融合方式不仅仅可基于 **FCSD** 目标函数估计的公共子空间,亦可应用在其他基于公共子空间的方法上。在 **XM2VTS** 至 **FRGC** 等跨库的迁移实验上,本章提出的 **FCSD** 方法具都得到了较对比的他人迁移学习方法有更好的性能。

第5章 结束语

5.1 本文总结

本篇论文从解决实际应用中训练数据与目标数据之间存在差异而导致人脸识别模型无法得到理想性能的问题出发,与其学习一个非常强大可以处理各种问题的人脸识别模型,本篇文章模仿人类在面对类似问题时采用的迁移学习思维方式来应对该问题。迁移学习借助无监督目标域上所描述的目标域信息来指导对源域中所包含的监督信息(旧知识)的提取,从而更好地处理目标域上的人脸识别问题(新任务)。基于迁移学习的思路,本篇论文分别从两个不同的角度提出 3 种方法来提升模型的人脸识别性能。

1. 从利用源域样本所携带的监督信息来指导目标域上判别性特征提取的角度:
 - 1) 提出基于目标域相似度矩阵估计的判别性特征提取的方法。该方法在源域上采用 Fisher 准则度量其判别性,用于指导能反映与其具有一致判别性的目标域相似度矩阵的学习,在此基础上利用有监督的学习方法提取具有判别性的特征。基于对判别性相似度矩阵应具有低秩性质的约束,该方法提出了一个结合公共子空间优化与低秩相似度矩阵估计的目标函数。针对该方法本研究提出了相应的优化求解算法,并在相关的迁移问题上测试其有效性。
2. 从充分迁移源域中有效信息的角度:
 - 2) 提出基于源域样本目标化的人脸识别方法。该方法的目标是将源域上的监督信息通过迁移源域样本而转移至目标域,使得目标域上获得一批可用于训练模型的带标注样本。该方法提出了结合局部结构保持和最小化稀疏重构误差的目标函数来求解该迁移系数,其中拉普拉斯矩阵用于保持域上的信息,稀疏重构误差用于度量域间的差异,其编码用于将源域样本迁移至目标域。通过该方法实现对源域样本的目标化后,可采用不同的监督学习算法来提取判别性特征或分类器。本研究中采用人脸识别领域常用的 FLD+NN 的方法训练模型,用于评测目标化的源域样本作为训练样本的有效性。在不同的迁移实验上都证明了该方法比与之对比的迁移学习方法具有更好的性能。
 - 3) 提出融合公共与特有子空间上判别性特征的迁移学习方法。该方法提出了一种如何充分提取目标域整个原始特征空间上有效信息的判别性特征

提取方法。该方法将目标域原始特征空间分成与源域共享的公共子空间和目标域特有的特征子空间两部分，在各个子空间上有针对性地进行判别性特征提取，进而融合这两部分特征得到最后的判别性特征。此外，该方法对目标化源域样本方法中的目标函数进行改进，采用协方差矩阵替代拉普拉斯矩阵用于域上信息的保持，减少了所需估计的参数，从而降低目标函数的复杂度。本研究在设计的迁移实验上与其他迁移学习方法进行了比对，实验结果可见该方法提出的融合方法较只利用到公共子空间的迁移学习方法有更好的性能。

5.2 展望

在开展本论文提出的方法过程中，对迁移学习的相关工作进行了深入的思考研究。除了本论文介绍的工作外，总结将来可从以下几个思路继续拓展人脸识别中的迁移学习方法。

1. 本论文提出的方法只考虑了存在单个源域的情况下如何进行迁移学习，但实际应用中，可能存在多个源域的情况。将来的研究方向可从将本篇提出的方法进行拓展方面展开，使方法可应对多个源域的问题。
2. 将源域与目标域视为两个流形，将迁移学习问题形式化成对齐两个流形的问题。与当前基于公共子空间的迁移学习相比，该思路旨在寻找公共流形从而进行迁移学习。
3. 更为充分分析在迁移学习中针对人脸识别问题有用的先验，提出有针对性的人脸识别迁移学习方法。

参考文献

- [1] P. Viola and M. Jones, "Robust real-time face detection," [C], in International Conference on Computer Vision, 2001.
- [2] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," [J], International Journal of Computer Vision, vol. 1, pp. 321-331, 1988.
- [3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," [J], Computer Vision and Image Understanding, vol. 61, pp. 38-59, 1995.
- [4] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 681-685, 2001.
- [5] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 2037-2041, 2006.
- [6] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," [J], Journal of the Optical Society of America, A: Optics and Image Science, vol. 2, pp. 1160-1169, 1985.
- [7] H. Zhou, Y. Yuan, and C. Shi, "Object tracking using SIFT features and mean shift," [J], Computer Vision and Image Understanding, vol. 113, pp. 345-352, 2009.
- [8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," [C], in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp. 886-893.
- [9] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," [J], Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 61, pp. 611-622, 1999.
- [10] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," [J], Journal of Computational and Graphical Statistics, vol. 15, pp. 265-286, 2006.
- [11] W. Zhao, R. Chellappa, and P. J. Phillips. (1999). Subspace linear discriminant analysis for face recognition.
- [12] S. R. Gunn, "Support vector machines for classification and regression," [J], ISIS Technical Report, vol. 14, 1998.
- [13] A. Martinez and R. Benavente, "The AR face database," [J], Rapport technique, vol. 24, 1998.
- [14] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, "The CAS-PEAL large-scale Chinese face database and baseline evaluations," [J], IEEE Transactions on Systems,

- Man and Cybernetics, Part A: Systems and Humans, vol. 38, pp. 149-161, 2008.
- [15] P. J. Phillips, H. Moon, P. Rauss, and S. A. Rizvi, "The FERET evaluation methodology for face-recognition algorithms," [C], in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997, pp. 137-143.
- [16] H. Moon and P. J. Phillips, "The FERET verification testing protocol for face recognition algorithms," [C], in IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 48-53.
- [17] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," [C], in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp. 947-954.
- [18] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," [C], in Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, 2008.
- [19] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," [C], in IEEE International Conference on Automatic Face & Gesture Recognition, 2008, pp. 1-8.
- [20] OFD 数据库. Available: <http://www.aiar.xjtu.edu.cn/dfslsjk5.htm>
- [21] A. S. Georgiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 643-660, 2001.
- [22] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," [C], in International Conference on Audio and Video-based Biometric Person Authentication, 1999, pp. 965-966.
- [23] S. J. Pan and Q. Yang, "A survey on transfer learning," [J], IEEE Transactions on Knowledge and Data Engineering, vol. 22, pp. 1345-1359, 2010.
- [24] X. Chen, W. Pan, J. T. Kwok, and J. G. Carbonell, "Accelerated gradient method for multi-task sparse learning problem," [C], in IEEE International Conference on Data Mining, 2009, pp. 746-751.
- [25] H. Daumé III, "Bayesian multitask learning with latent hierarchies," [C], in Proceedings of the Conference on Uncertainty in Artificial Intelligence, 2009, pp. 135-142.
- [26] S. Han, X. Liao, and L. Carin, "Cross-Domain Multitask Learning with Latent Probit Models," [J], arXiv preprint arXiv:1206.6419, 2012.
- [27] N. Quadrianto and C. H. Lampert, "Learning multi-view neighborhood preserving projections," [C], in Proceedings of the the International Conference on Machine Learning 2011, pp. 425-432.

- [28] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Ng, "Multimodal deep learning," [C], in Proceedings of the International Conference on Machine Learning, 2011, pp. 689-696.
- [29] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: transfer learning from unlabeled data," [C], in Proceedings of the International Conference on Machine Learning, 2007, pp. 759-766.
- [30] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," [C], in Proceedings of the International Conference on Machine Learning, 2011, pp. 513--520.
- [31] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," [C], in IEEE International Conference on Computer Vision (ICCV), 2011, pp. 999-1006.
- [32] D. Uribe, "Domain Adaptation in Sentiment Classification," [C], in International Conference on Machine Learning and Applications, 2010, pp. 857-860.
- [33] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A DASVM classification technique and a circular validation strategy," [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, pp. 770-787, 2010.
- [34] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," [J], IEEE Transactions on Neural Networks, pp. 1-12, 2009.
- [35] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," [C], in Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, 2006, pp. 120-128.
- [36] M. Zeng and J. Ren, "Domain transfer dimensionality reduction via discriminant kernel learning," [J], Advances in Knowledge Discovery and Data Mining, pp. 280-291, 2012.
- [37] L. Duan, I. W. Tsang, D. Xu, and S. J. Maybank, "Domain transfer svm for video concept detection," [C], in IEEE Conference on Computer Vision and Pattern Recognition 2009, pp. 1375-1381.
- [38] D. Lopez-Paz, J. M. Hernández-Lobato, and B. Schölkopf, "Semi-Supervised Domain Adaptation with Non-Parametric Copulas," [C], in Advances in Neural Information Processing Systems, 2012, pp. 674-682.
- [39] M. Sugiyama, M. Krauledat, and K.-R. Müller, "Covariate shift adaptation by importance weighted cross validation," [J], The Journal of Machine Learning Research, vol. 8, pp. 985-1005, 2007.
- [40] M. Sugiyama, S. Nakajima, H. Kashima, P. Von Buenau, and M. Kawanabe, "Direct importance estimation with model selection and its application to covariate shift adaptation," [J],

- Advances in Neural Information Processing Systems, vol. 20, pp. 1433-1440, 2008.
- [41] H. Shimodaira, "Improving predictive inference under covariate shift by weighting the log-likelihood function," [J], Journal of Statistical Planning and Inference, vol. 90, pp. 227-244, 2000.
- [42] J. Huang, A. J. Smola, A. Gretton, K. M. Borgwardt, and B. Schölkopf, "Correcting Sample Selection Bias by Unlabeled Data," [C], in Advances in Neural Information Processing Systems, 2006.
- [43] M. Dudík, R. E. Schapire, and S. J. Phillips, "Correcting sample selection bias in maximum entropy density estimation," [J], Advances in Neural Information Processing Systems, vol. 17, pp. 323-330, 2005.
- [44] B. Zadrozny, "Learning and evaluating classifiers under sample selection bias," [C], in Proceedings of International Conference on Machine Learning, 2004, p. 114.
- [45] J. J. Heckman, "Sample selection bias as a specification error," [J], Econometrica: Journal of the Econometric Society, pp. 153-161, 1979.
- [46] J. Yang, R. Yan, and A. G. Hauptmann, "Cross-domain video concept detection using adaptive svms," [C], in Proceedings of the International Conference on Multimedia, 2007, pp. 188-197.
- [47] A. Smola, A. Gretton, L. Song, and B. Schölkopf, "A Hilbert space embedding for distributions," [C], in Algorithmic Learning Theory, 2007, pp. 13-31.
- [48] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction," [C], in Proceedings of the National Conference on Artificial Intelligence, 2008, pp. 677-682.
- [49] P. A. Absil, R. Mahony, and R. Sepulchre, "Riemannian geometry of Grassmann manifolds with a view on algorithmic computation," [J], Acta Applicandae Mathematicae, vol. 80, pp. 199-220, 2004.
- [50] P. Turaga, A. Veeraraghavan, and R. Chellappa, "Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision," [C], in IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-8.
- [51] K. A. Gallivan, A. Srivastava, X. Liu, and P. Van Dooren, "Efficient algorithms for inferences on Grassmann manifolds," [C], in 2003 IEEE Workshop on Statistical Signal Processing, 2003, pp. 315-318.
- [52] Y. Shi and F. Sha, "Information-Theoretical Learning of Discriminative Clusters for Unsupervised Domain Adaptation," [C], in International Conference on Machine Learning, 2012.
- [53] 姜丹, 信息论与编码: 中国科学技术大学出版社, 2001.
- [54] Z. Wang, Y. Song, and C. Zhang, "Transferred dimensionality reduction," [J], Machine

- Learning and Knowledge Discovery in Databases, pp. 550-565, 2008.
- [55] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," [J], Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 28, pp. 100-108, 1979.
- [56] J. Jiang and C. Zhai, "Instance weighting for domain adaptation in NLP," [C], in Annual Meeting-Association For Computational Linguistics, 2007, p. 264.
- [57] Y. Chen, G. Wang, and S. Dong, "Learning with progressive transductive support vector machine," [J], Pattern Recognition Letters, vol. 24, pp. 1845-1855, 2003.
- [58] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," [J], Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 19, pp. 711-720, 1997.
- [59] E. J. Candes and Y. Plan, "Matrix completion with noise," [J], Proceedings of the IEEE, vol. 98, pp. 925-936, 2010.
- [60] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," [C], in Proceedings of the International Conference on Machine Learning, 2010.
- [61] R. T. Rockafellar, "Augmented Lagrange multiplier functions and duality in nonconvex programming," [J], SIAM Journal on Control, vol. 12, pp. 268-285, 1974.
- [62] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," [J], arXiv preprint arXiv:1009.5055, 2010.
- [63] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," [J], SIAM Journal on Optimization, vol. 20, pp. 1956-1982, 2010.
- [64] Z. Wen and W. Yin, "A feasible method for optimization with orthogonality constraints," [J], Mathematical Programming, pp. 1-38, 2010.
- [65] P. Niyogi and X. He, "Locality preserving projections," [C], in Advances in Neural Information Processing Systems, 2004, p. 153.
- [66] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," [J], Advances in Neural Information Processing Systems, vol. 14, pp. 585-591, 2001.
- [67] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," [J], Science, vol. 290, pp. 2323-2326, 2000.
- [68] B. Efron, T. Hastie, I. Jhonstone, and R. Tibshirani, "Least Angle Regression," [J], The Annals of Statistics vol. 32, pp. 407-499, 2004.
- [69] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," [J], IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 31, pp. 210-227, 2009.
- [70] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," [J], Analytica

Chimica Acta, vol. 185, pp. 1-17, 1986.

[71] I. T. Jolliffe, Principal component analysis: Springer Verlag, 2002.

[72] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," [C], in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991, pp. 586-591.

[73] 程云鹏, 张凯院, and 徐仲, 矩阵论: 西北工业大学出版社, 2006.

致 谢

时光不待，再回首已经到了研究生阶段的最后一段时光，值此论文快完成之际，谨向这三年来给予我无私帮助、关心和支持的老师、同学们、亲友和家人致以诚挚的谢意。

首先要诚挚地感谢我的导师，山世光恩师。山老师渊博的学识、严谨的科研态度以及学无止尽的观念都是我学习的榜样。非常感谢山老师对我科研方向的把握与指导，提供了宽松自由的学术氛围，培养我发现、思考与解决问题的能力。这些都将会为我将来的发展打下夯实的基础。

感谢实验室的陈熙霖老师，陈老师在学术科研上高瞻远瞩的鞭策与反思常常使我跳出当前着眼的小问题，从更开阔的角度来思辨问题。

感谢实验室的常虹老师、苗军老师，两位老师对于学术的执着与造诣以及在读书会议上思考问题的角度都使我深叹两位老师的治学严谨、慧眼独道。

感谢实验室的卿来云老师、梁路宏老师、柴秀娟老师和王瑞平老师，你们的言谈身教都让我获益良多。

感谢实验室的崔振师兄。谢谢崔振师兄对初入实验室凡事还一知半解的我给予了细致的指导，帮助我一窥如何进行科研。师兄对于国际先进技术方法的及时关注与深入理解都深深地影响着我。每一次与师兄的讨论都使我受益匪浅。师兄深厚的知识与见解，都鞭策着我进一步的学习。

感谢实验室的阚美娜师姐。谢谢美娜师姐对我的指导与经验分享，使我收获颇丰。师姐在方向、方法的可行性上给予了我严谨科学的分析，使我既少走了很多弯路，又学习到了如何分析问题、排除问题的方法。

感谢实验室的已毕业的陈静师姐、黄煜师姐、李哲师姐、王崇秀师姐、狄晓斐师姐、都大龙师兄、项翔师兄、李浩师兄、马兴师兄、洪晓鹏师兄对我的帮助。感谢吉娜烨师姐、王丹师姐、马志国师兄、王瑞平师兄、赵小伟师兄、李绍兴师兄、黄智武师兄、郑伟师兄、在科研上的指导与帮助。

感谢实验室的王汉杰、唐毅力、孟令勋、淮静、宋松林、李瑞平、李岩，这三年的同窗为伴、科研为伴，各位带给了我无限的收获。感谢室友董智对我的照顾与帮助。

最后特别感谢我的爸爸、妈妈，感谢你们对我的养育之恩，给我一片广阔的天空，又永远守护身旁，祝你们永远健康快乐！

吴俊婷

作者简历

基本情况

姓名：吴俊婷 性别：女 出生日期：1989-04-22 籍贯：福建莆田

教育状况

2010.09—至今 中国科学院计算技术研究所 攻读硕士学位

2006.09—2010.07 浙江大学 获学士学位

攻读硕士期间的科研项目

2012.01-至今 参与国家自然科学基金项目“跨姿态人脸识别研究”。

2010.11-2012.12 参与计算所与 Omron 合作项目“人脸识别技术与开发”。