

基于加权子空间生成的迁移 学习方法研究

(申请清华大学工程硕士专业学位论文)

培养单位： 电子工程系

工程领域： 电子与通信工程

申请人： 陈 硕

指导教师： 廖庆敏 教授

二〇一七年五月

基于加权子空间生成的迁移学习方法研究

陈
硕

Research on Transfer Learning Algorithm based on Generating Weighted Subspaces

Thesis Submitted to

Tsinghua University

in partial fulfillment of the requirement

for the professional degree of

Master of Engineering

by

Chen Shuo

(Electronics and Communication Engineering)

Thesis Supervisor : Professor Liao Qingmin

May, 2017

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：（1）已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；（2）为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容。

本人保证遵守上述规定。

（保密的论文在解密后应遵守此规定）

作者签名： _____

导师签名： _____

日 期： _____

日 期： _____

摘要

传统机器学习中，往往假设训练集和测试集数据分布是相同的。然而在实际应用中，测试集的数据分布和训练集很有可能是不同的。域适应主要解决的是源域和目标域数据特征分布不同的问题。域适应分为两种情况：无监督域适应和半监督域适应。

在无监督域适应中，目标域样本没有标签信息。在这种应用情况，本文提出了基于加权子空间生成的无监督域适应算法。该算法主要包含两个部分：子空间生成和子空间对齐。在子空间生成这一步骤中，首先计算权重，目的是为了衡量每个源域样本的重要性。如果一个源域样本的分布越接近目标域，则它越有利于训练有域适应能力的分类器，相应的权重越高。因此权重计算是以源域数据特征到目标域数据特征的距离为依据的。获得源域样本权重后，实行加权PCA得到源域的加权子空间，目标域实行PCA得到子空间。接下来进行子空间对齐操作，目的是得到一个转换矩阵将源域子空间的基进行转换，使转换后的基尽可能地接近目标域子空间的基。借助已有的数学定理，可以得到该转换矩阵的闭式解，计算简便快捷，且有理论依据。在后来的工作中，本文还对权重计算进行了改进，使通过计算式得到的权重取值更合理。最后我们做了实验对加权子空间对齐算法进行了验证，实验任务是图片分类，实验结果证明了该算法在性能上的优越性。

在半监督域适应中，目标域样本有标签信息。在这种应用情况，本人主要是研究基于深度神经网络的方法，应用场景是行为识别。在行为识别领域，目前性能最好的模型是双流卷积网络。在现有的双流卷积网络中本算法加入了权重机制，目的是为了将帧中有益于分类的区域进行增强，从而使训练得到的网络具有域适应的分类能力。接下来进行了实验，实验任务是行为识别，本文针对现有的一个数据库建立了一个辅助数据库，实验结果显示本算法在分类结果上有提高。

关键词：域适应；迁移学习；表达学习；图像分类；行为识别

Abstract

In traditional machine learning, it is generally assumed that the training set and test set follow the same distribution. While in practical applications, the distribution of training and test set might be different. Domain adaptation mainly aims at solving this issue. Domain adaptation is usually differentiated into two scenarios, unsupervised domain adaptation and semi-supervised domain adaptation.

There is no labeled target instance in unsupervised domain adaptation. In the scenario, we proposed weighted subspace alignment algorithm. Our algorithm has two main steps, subspace generation and subspace alignment. In subspace generation, each source instance is given a weight to quantize the importance. If a source instance is close to target instance, it is beneficial to train a classifier capable of domain shift. And its weight should be high. We calculate the weight based on the feature distance between source instance and target instance. After gaining the source weights, we perform weighted PCA to get the weighted subspace of source domain and PCA to get the subspace of target domain. Then we perform subspace alignment. Our goal is to learn transformation matrix projecting source subspace to target subspace. At the same time, we need to minimize the error between transformed source subspace and target subspace. The transformation matrix has a closed-form solution and can be solved efficiently. In following work, we improved the weighted subspace alignment algorithm by refining the weight calculation. The experiment results about image classification show that our algorithm has a significant performance improvement.

In semi-supervised domain adaptation the target domain has labeled instances. In this scenario, we focus on deep learning methods about action recognition. The state-of-the-art methods in action recognition are based on two stream convolutional networks. We add weight layer into network aiming to augment image areas that are helpful to classification. Thus the network has the ability of domain adaptation. We perform an experiment on a standard data-set and corresponding auxiliary data-set. The experimental result demonstrates that our network has a significant performance improvement.

Key words: domain adaptation; transfer learning; representation learning; image classification; action recognition

目 录

第 1 章 绪论	1
1.1 研究背景	1
1.2 研究目的	3
1.3 研究意义	4
1.4 论文的组织结构	5
第 2 章 域适应算法研究现状	6
2.1 引言	6
2.2 定义	6
2.3 目前主要域适应算法	6
2.3.1 基于特征增强的方法	7
2.3.2 基于特征转换的方法	7
2.3.3 参数适应方法	9
2.3.4 基于字典学习的方法	10
2.4 本章小结	11
第 3 章 基于加权子空间生成的域适应算法	12
3.1 子空间对齐算法	12
3.2 加权子空间对齐算法	13
3.2.1 加权子空间生成	14
3.2.2 加权子空间对齐	16
3.2.3 数据准备	17
3.2.4 实验结果	19
3.3 对加权子空间对齐算法的改进	21
3.3.1 权重改进	21
3.3.2 数据准备	21
3.3.3 实验结果	23
3.4 本章小结	25

第 4 章 基于深度神经网络的域适应算法	26
4.1 基于深度神经网络的域适应算法.....	26
4.1.1 双流深度卷积网络	26
4.1.2 权重机制.....	27
4.2 实验.....	29
4.2.1 数据准备.....	30
4.2.2 实验结果.....	31
4.3 本章小结	33
第 5 章 总结与展望	37
5.1 总结.....	37
5.2 展望.....	38
参考文献	39
致 谢	43
声 明	44
个人简历、在学期间发表的学术论文与研究成果	45

第1章 绪论

1.1 研究背景

在传统的机器学习中，有一个很重要的假设是训练数据和测试数据的概率分布是相同的。然而，在很多现实生活的应用中，这个假设不一定成立。例如在模式识别和计算机视觉领域中，经常会遇到一种情况：用于训练模型的数据分布和真实数据分布是不同的。如图1.1所示，是四个不同数据库中显示屏的图片。另外，当数据的分布随着时间的推移发生变化时，分类器的效果也会下降。为了能够解决以上问题，迁移学习应运而生。



图 1.1 数据分布不同示例

近年来越来越多的人开始进行关于迁移学习的研究，在不同的阶段它也被赋予了不同的名字：学会学习（learning to learn）、终生学习（life-long learning）、知识迁移（knowledge transfer）、归纳迁移（inductive transfer）、多任务学习（multitask learning）、知识巩固（knowledge consolidation）、上下文敏感学习（context-sensitive learning）、基于知识的归纳偏置（knowledge-based inductive bias）、元学习（metalearning）以及增量学习（incremental learning）等^[1]。在这些学习方法中，多任务学习是最接近于迁移学习本质的。多任务学习的目标是能

够同时处理多个学习任务，即使这些学习任务之间是不相同的。如今，在很多顶级会议中都有迁移学习方法的身影，如数据挖掘领域中的ACM KDD、IEEE ICDM和PKDD，机器学习领域中的ICML、NIPS、ECML、AAAI和IJCAI等。

在机器学习和计算机视觉领域中有很多优秀的用来分类的监督学习算法，并且已经应用到实际生活中。但这些算法在取得良好效果的同时，也有其局限性：它们往往需要大量带标签样本来进行训练。而对样本的标定工作是非常耗费时间精力的，成本很大。因此，利用已有的无标签样本来训练分类器就成为一种可行的且节约成本的方式。以图片和视频为例，现在网上有很庞大的图片和视频数据库，如Flickr和YouTube，现实生活中也有很多的监控摄像头拍摄下来的视频画面。这些样本来自于不同的数据源，所以在利用这些数据来训练分类器时就要注意处理不同数据源带来的样本分布不同问题。举例来说，物体识别中在光线良好的环境下训练得到的分类器被应用在弱光环境下，将检测X光图像病态区域的模型应用在核磁共振图像（MRI）中以及将彩色人脸图像上训练得到的分类器应用在红外线人脸图像中等。

这些问题通常被称为covariate shift^[2]或者data set bias^[3,4]问题。在算法训练完成后，任何数据分布上的变化或域位移都会造成测试性能的下降。例如，在人脸识别这个问题上，为了在真实场景中也实现良好的识别性能，人脸识别算法必须学会适应每个应用场景中不同的数据分布情况。如图1.2所示，(a)表示各个角度下的人脸，(b)是有表情变化的人脸，(c)表示五官姿势变化的人脸，(d)是人脸素描图。在真实场景下的物体识别算法，例如人脸识别，必须学会适应图1.2中(a)-(d)所示不同域下的特征分布不同的数据^[5-7]。尽管迁移学习问题在不同领域以不同的名称被研究过，如covariate shift^[2]、class imbalance^[8]、sample selection bias^[9,10]等，但最近才开始在计算机视觉领域收到广泛关注。在机器学习领域也有一些与迁移学习相近的问题被广泛研究，例如多任务学习^[11]、自学习^[12]、半监督学习^[13]以及多角度分析^[14]。文章^[15]是关于域适应在机器学习及自然语言处理领域应用的综述。

在迁移学习中，域适应问题是一个很重要的子问题，其主要目的是使源域数据能够自动地适应目标域数据，这样利用源域数据的样本标签信息训练而得到的模型，在目标域的数据上也能取得好的效果^[16]。域适应有很多应用场景，在目标识别、目标检测、文档分类、情感分析、行为识别、语音识别等领域均有应用，如图1.3所示。域适应通常分为两种情况：第一类是半监督域适应，在这种情况下，目标域数据中的标签信息是可用的；第二类是无监督域适应，在无监督域适应中，目标域数据中的标签信息是不可用的。从以上表述中，可以看出，无监督域适应比半监督域适应更复杂。在第一种情况，部分目标域样本中的标签信息是



图 1.2 人脸识别数据库图片示例

可用的，有些研究工作是基于此做的^[17-19]。虽然目标域样本有标签信息可用，但往往也只有很少的带标签样本，数目比源域样本要少很多。这些少量的带标签目标域样本在训练应用于目标域的分类器上能起到很大作用。对于更复杂的第二种情况，目标域的标签信息是不可用的。最近，无监督域适应也是迁移学习中的研究热点，在各大顶会顶刊中都能看到研究无监督域适应的文章^[20-29]。目标域无标签信息而又希望得到应用于目标域的分类器，就需要好好研究源域样本特征和目标域样本特征之间的关系，或者好好研究分类器本身，在分类算法上做文章。由此引申出不同种类的无监督域适应算法。

1.2 研究目的

将域适应算法应用在分类问题中，在研究现有算法的基础上，对算法进行改进。检测分类算法的指标主要体现在分类的准确率上，即分类正确的样本占总样本的比例。改进后的算法要能够在准确率上有所提高。如图1.4所示，同一个黑框下的目标识别数据库拥有共同的类标签，但很明显这些图片的特征分布是不一样的。在域适应问题中，源域和目标域样本的特征分布是不同的，所以分类器需要有相应的适应能力。我们想训练得到能应用在目标域的分类器，如果直接在目标域上训练，就需要对目标域样本进行标定，需要大量人力物力和时间。域适应算法就能够很好的解决这个问题，域适应的目的是利用已有的带标签的源域样本训练分类器，得到的分类器在目标域上也能有良好的表现，如图1.5所示。这样就不需要目标域样本的标签信息，节约了大量成本。

在计算机视觉领域，域适应算法的应用场景很多。我主要针对两个问题进行研究：图片分类和行为识别。本文将在第3章和第4章分别阐述在这两个问题上进



图 1.3 域适应应用举例

行的工作。

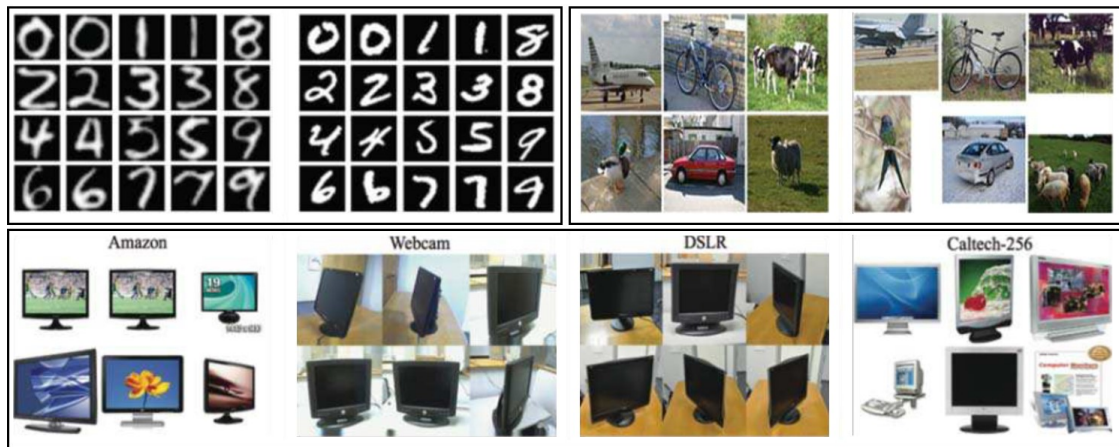


图 1.4 目标识别数据库示例

1.3 研究意义

传统的机器学习算法往往需要大量带标签的数据来驱动，而数据的标签往往是人工标定的，无疑会大量耗费人力物力。迁移学习能够解决带标签的数据量不足的问题，其利用现有的大量无标签数据，来辅助训练模型，从而达到迁移的效果。域适应作为迁移学习下面的一个分支，主要应用于目标识别领域中。目标识别主要是根据图片中的物品进行分类，在这个领域中已经有很多经典的算法。经

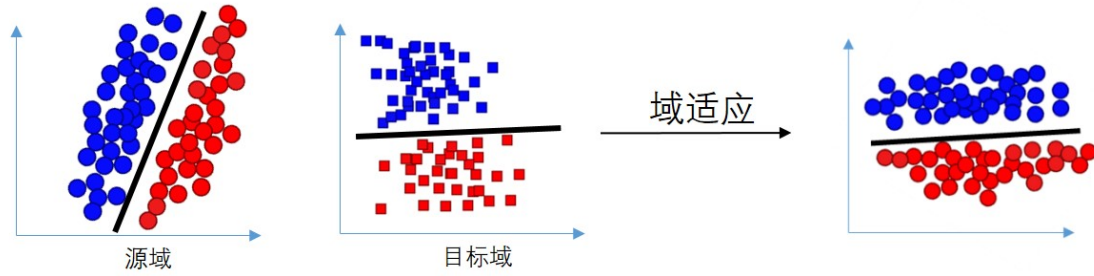


图 1.5 域适应算法示例

典的分类器算法都有一个假设，即训练样本和测试样本的分布是相同的。而域适应并没有这个假设，其训练样本和测试样本的特征分布不同，因此域适应的应用范围更广^[28]。除此之外，传统分类器需要大量带标签样本进行训练，以保证训练得到的分类器有良好的鲁棒性。而域适应可以大大减少带标签样本的数量需求，因为训练样本和测试样本的分布不同，域适应可以利用已有的数据库来训练分类器，纵然已有数据库的样本特征分布和测试样本不同，但域适应的分类器算法具有迁移能力，所以就不需要获取特征分布和测试样本相同的数据来进行训练，这大大减少了人工标定的成本。

1.4 论文的组织结构

本文的组织结构如下。

第一章主要介绍迁移学习的研究背景、目的和意义，并简要介绍本文的研究问题和方向。

第二章主要介绍了域适应算法的分类、定义，以及目前主流的几种域适应方法。

第三章详细介绍了在图片分类问题上进行的域适应算法研究，首先介绍了子空间对齐算法，然后介绍了本文提出的加权子空间对齐算法，并进行了实验，最后介绍了改进的加权子空间对齐算法，并给出了实验结果。

第四章详细介绍了在行为识别问题上进行的域适应算法研究，主要是基于深度神经网络。首先介绍了双流卷积网络模型，然后介绍了加入到模型中的权重机制。最后在改进后的网络模型上进行了实验。

第五章为总结与展望。首先总结了本论文域适应算法的工作，然后对域适应算法的未来进行了展望。

第2章 域适应算法研究现状

2.1 引言

域适应是迁移学习中的一个子问题，在机器学习和计算机视觉领域中引起了很大关注^[30]。众所周知，对于监督分类算法，其分类器性能主要取决于训练数据量的大小。然而，对数据进行标定是一项费时费力的工作。因此，如果仅利用有限的训练数据就能训练出高性能的分类器，就显得很有意义。在计算机视觉领域中，域适应方法主要解决的问题是：在有标签的源域下训练得到的分类器能迁移到目标域上。这里可以简单地将源域理解为训练集，将目标域理解为测试集。在第1章中，本文已经提到，源域和目标域的数据分布是不同的。根据应用场景，域适应算法课分为两类：一类是半监督域适应算法，另一类是无监督域适应算法。在半监督域适应算法中，目标域中有少部分样本是带标签的。而在无监督域适应算法中，目标域中的样本均是没有标签的。因为目标域中没有可利用的标签信息，因此无监督域适应算法相对于半监督域适应算法，其应用场景更难一些，也更接近真实情况。

2.2 定义

假设数据样本的形式为 (\mathbf{x}_i, y_i) ，其中 \mathbf{x}_i 表示第 i 个样本的特征，为 D 维向量，即 $\mathbf{x}_i \in \mathbb{R}^D$ 。而 y_i 表示第 i 个样本对应的标签，共有 K 类，即 $y_i \in \{1, \dots, K\}$ 。

在无监督域适应算法的应用场景中，源域中有大量带标签样本，而目标域没有带标签样本。源域样本用 (\mathbf{x}^s, y^s) 表示，数目为 N_s ，其分布为 $\mathcal{D}_s = P(\mathbf{x}^s, y^s)$ 。目标域样本用 (\mathbf{x}^t, y^t) 表示，数目为 N_t ，其分布为 $\mathcal{D}_t = P(\mathbf{x}^t, y^t)$ 。并且，源域和目标域的分布不同，即 $\mathcal{D}_s \neq \mathcal{D}_t$ 。

在半监督域适应算法的应用场景中，源域中的数据样本带标签，目标域中有少量样本带标签。源域样本用 (\mathbf{x}^s, y^s) 表示，数目为 N_s 。目标域少量带标签样本用 \mathbf{x}^{tl}, y^{tl} 表示，数目为 N_{tl} ，无标签样本用 \mathbf{x}^{tu}, y^{tu} 表示，数目为 N_{tu} 。在本论文中，所有向量默认均为列向量。

2.3 目前主要域适应算法

域适应是迁移学习中的一个基本问题，在自然语言处理、计算机视觉、统计

等领域中受到了广泛的关注。目前主要的领域适应算法可以被分为以下几类。

2.3.1 基于特征增强的方法

在基于特征增强的方法中，目前最简单直接的方法是将每个域中的原始特征进行拷贝^[31]。通过直接拷贝特征向量，原始域中的每个特征的维度都变为了原来的3倍。源域和目标域的增强特征定义如式2-1所示。

$$\Phi^s(\mathbf{x}_i^s) = \begin{bmatrix} \mathbf{x}_i^s \\ \mathbf{x}_i^s \\ \mathbf{0} \end{bmatrix}, \quad \Phi^s(\mathbf{x}_j^t) = \begin{bmatrix} \mathbf{x}_j^t \\ \mathbf{0} \\ \mathbf{x}_j^t \end{bmatrix} \quad (2-1)$$

其中， $\Phi^s(\mathbf{x}_i^s)$ 和 $\Phi^s(\mathbf{x}_j^t)$ 分别表示增强后的源域数据特征和目标域数据特征， $\mathbf{0}$ 表示零向量。增强特征的第一行向量代表了源域和目标域之间的共通性，第二行代表了源域，而第三行代表了目标域。通过使用增强特征，源域和目标域的特征被扩展到了一个增强的特征空间中，在这个增强的特征空间中训练而得到的分类器能够在两个域中都有较好的表现。

2.3.2 基于特征转换的方法

文章^[32]是最早应用领域适应来解决目标识别问题的方法之一，其主要思想是通过学习转移矩阵将源域特征进行转化，使之尽可能的接近目标域特征，来达到迁移的效果。假设有源域特征向量 \mathbf{x}^s 和目标域特征向量 \mathbf{x}^t ，文章^[32]将学习一个线性转换矩阵 \mathbf{W} 。 \mathbf{x}^s 和转换后的 \mathbf{x}^t 的内积相似性函数定义为：

$$sim_{\mathbf{W}} = (\mathbf{x}^s)^T \mathbf{W} \mathbf{x}^t \quad (2-2)$$

式2-2可以看作是 $\mathbf{W} \mathbf{x}^t$ 和 \mathbf{x}^s 的内积，我们的目标是为了通过学习得到线性转换矩阵。为了避免过拟合，对 \mathbf{W} 加入了一个正则项 $r(\mathbf{W})$ ，由此得到一个带约束项的优化问题：

$$\begin{aligned} & \min_{\mathbf{W}} r(\mathbf{W}) \\ & s.t. \quad c_i(\mathbf{x}^s)^T \mathbf{W} \mathbf{x}^t \geq 0, \quad 1 \leq i \leq c \end{aligned} \quad (2-3)$$

其中函数 $c_i(\cdot)$ 用于约束目标函数。式2-3可以被写成无约束项的形式：

$$\min_{\mathbf{W}} r(\mathbf{W}) + \lambda \sum_i c_i(\mathbf{x}^s \mathbf{T} \mathbf{W} \mathbf{x}^t) \quad (2-4)$$

在文章^[32]中的约束项为：

$$\mathbf{r}(\mathbf{W}) = \text{trace}(\mathbf{W}) - \log \det(\mathbf{W}) \quad (2-5)$$

接下来文章^[32]利用度量学习方法来解决此优化问题。不过该算法的不足之处是其只能应用于源域样本和目标域样本维数相同的情况下。基于该算法的不足，文章^[33]拓展了文章^[32]算法的应用情况，使其不再被限制在源域和目标域样本维数相同的情况。因此，文章^[33]的算法适用于更一般的域适应情况。实际上，文章^[32]提出的算法是文章^[33]的一种特殊情况，是其为对称半正定矩阵的情况。

最近，一种基于低秩近似的半监督域适应算法在文章^[34]中提出。该算法的目的是通过一个矩阵 $\mathbf{W} \in \mathbb{R}^{D \times D}$ 将源域数据映射到一个中间表达，然后这个中间表达由目标域样本的线性组合构成，其表达式如下：

$$\mathbf{W} \mathbf{x}^s = \mathbf{x}^t \mathbf{Z} \quad (2-6)$$

其中 $\mathbf{Z} \in \mathbb{R}^{N_t \times N_s}$ 是协方差矩阵。该低秩问题通过优化下式解决：

$$\begin{aligned} (\hat{\mathbf{W}}, \hat{\mathbf{Z}}, \hat{\mathbf{E}}) &= \min_{\mathbf{W}, \mathbf{Z}, \mathbf{E}} \text{rank}(\mathbf{Z}) + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s.t. } \mathbf{W} \mathbf{S} &= \mathbf{x}^t \mathbf{Z} + \mathbf{Z}, \mathbf{W} \mathbf{W}^T = \mathbf{I} \end{aligned} \quad (2-7)$$

这里 $\text{rank}(\cdot)$ 表示矩阵的秩， λ 是参数， $\mathbf{E} \in \mathbb{R}^{D \times D}$ 是误差项， $l_{2,1}$ 范数的定义如下：

$$\|\mathbf{E}\|_{2,1} = \sum_{i=1}^D \sqrt{\sum_{j=1}^{N_s} E_{ij}^2}$$

式2-7是一个常规的秩最小化问题，因此 \mathbf{Z} 的秩可用其核范数来代替^[34]，该优化问题可用增广拉格朗日乘子法来解决。

在获得 $(\hat{\mathbf{W}}, \hat{\mathbf{Z}}, \hat{\mathbf{E}})$ 的解后，被转换后的源域数据为：

$$\hat{\mathbf{W}}\mathbf{S} - \hat{\mathbf{E}} \quad (2-8)$$

转换后的源域数据和目标域数据混合作为训练数据来学习分类器，然后分类器再对目标域中无标签样本进行分类。

在基于特征转换的方法中，最近有一些基于子空间的方法取得了良好的效果^[20,22,23,27,28]。由于原始数据特征往往存在于高维空间，而对高维空间的数据直接进行分类比较困难，所以在原始数据的子空间上进行操作就比较合理有效。文章^[27]是最早提出在源域和目标域的子空间上进行域适应操作的文章。文章^[20]在文章^[27]的基础上进行了改进，引入了核机制，运用流型方法将源域一步步映射到目标域上。

2.3.3 参数适应方法

在现有的文献中，有些方法是通过更改SVM算法来解决领域适应问题^[35]。在文章^[35]中，作者提出了一种新的具有迁移能力的SVM算法，在源域数据上训练得到的分类器 $f_s(x)$ ，被迁移到无标签目标域数据上，从而得到具有迁移能力的新的分类器 $f_t(x)$ 。决策函数定义为：

$$f_t(x) = f_s(x) + \delta f(x) \quad (2-9)$$

在式2-9中， $\delta f(x)$ 是扰动函数。在文章^[35]中扰动函数为 $\delta f(x) = \theta^T \phi(x)$ ， ϕ 用来将 x 映射到高维空间。扰动函数 $\delta f(x)$ 是通过目标域中的带标签样本 $\{(\mathbf{x}^t, y^t)\}_{i=1}^{N_t}$ 学习得到的。为了学习扰动函数 $\delta f(x)$ 的参数 \mathbf{w} ，需要优化下式：

$$\begin{aligned} \min_{\theta} \quad & \frac{1}{2} \|\theta\|^2 + \alpha \sum_{i=1}^{N_t} \xi_i \\ \text{s.t.} \quad & \xi_i \geq 0, \\ & y_i^t f_s(\mathbf{x}_i^t) + y_i^t \theta^T \phi(\mathbf{x}_i^t) \geq 1 - \xi_i, \quad \forall (\mathbf{x}_i^t, y_i^t) \end{aligned} \quad (2-10)$$

这里 ξ_i 是松弛变量， α 是SVM的惩罚参数， α 值大时对误分类的惩罚增大， α 值小时对误分类的惩罚减小。式2-10中的第一项试图减小新的判决边界与旧的判决边界的偏差，第二项主要控制目标域训练样本误分类的惩罚程度。

文章^[36]对这项工作进行了改进并应用到目标检测中，文章^[37]将其应用到了目标识别中。文章^[38]提出了域迁移SVM算法，通过最小化源域和目标域的最大平

均偏差 (Maximum Mean Discrepancy), 来匹配源域和目标域不同的特征分布。其它基于SVM的域适应算法可在文章^[39-43]中找到。

2.3.4 基于字典学习的方法

最近几年, 稀疏表达在图像领域引起了广泛的关注。这主要是因为原始图片作为一个高维信号, 拥有很多冗余信息, 从而可以通过学习字典进行表达。同样, 在域适应领域, 也有很多基于字典学习的算法。文章^[44]提出了一个学习框架: 在源域样本上学习字典, 通过字典构建一个具有域不变性的稀疏表达。假设有源域样本 $\mathbf{x}^s \in \mathbb{R}^{D_s \times N_s}$ 和带标签目标域样本 $\mathbf{x}^{tl} \in \mathbb{R}^{D_t \times N_{tl}}$, 该算法希望学习一个 K 维原子字典 $\mathbf{D} \in \mathbb{R}^{l \times K}$, 然后将 $\mathbf{W}_1 \in \mathbb{R}^{l \times D_s}$ 和 $\mathbf{W}_2 \in \mathbb{R}^{l \times D_t}$ 映射到一个共同的低维子空间上。在映射过程中, 希望重建误差能够尽可能小, 所以最小化下式:

$$\begin{aligned} \mathbf{C}_1(\mathbf{D}, \mathbf{W}_1, \mathbf{W}_2, \mathbf{X}_1, \mathbf{X}_2) &= \|\mathbf{W}_1 \mathbf{x}^s - \mathbf{D} \mathbf{X}_1\|_F^2 + \|\mathbf{W}_2 \mathbf{x}^{tl} - \mathbf{D} \mathbf{X}_2\|_F^2 \\ \text{s.t. } \mathbf{X}_1 &\in \mathbb{R}^{K \times N_s}, \mathbf{X}_2 \in \mathbb{R}^{K \times N_{tl}} \end{aligned} \quad (2-11)$$

这里假设映射矩阵 \mathbf{W}_1 和 \mathbf{W}_2 的列向量已被标准正交化, 使算法利用核方法变成了可能。

为了确保将源域数据映射到子空间时不丢失太多的信息, 提出了一个类似PCA的正则项, 定义如下:

$$\mathbf{C}_2(\mathbf{W}_1, \mathbf{W}_2) = \|\mathbf{x}^s - \mathbf{W}_1^T \mathbf{W}_1 \mathbf{x}^s\|_F^2 + \|\mathbf{x}^{tl} - \mathbf{W}_2^T \mathbf{W}_2 \mathbf{x}^{tl}\|_F^2 \quad (2-12)$$

可以看出 \mathbf{C}_1 和 \mathbf{C}_2 在忽略常量 \mathbf{Y} 后, 可以写成:

$$\mathbf{C}_1(\mathbf{D}, \hat{\mathbf{W}}, \hat{\mathbf{X}}) = \|\hat{\mathbf{W}} \hat{\mathbf{Y}} - \mathbf{D} \hat{\mathbf{X}}\|_F^2 \quad (2-13)$$

$$\mathbf{C}_2(\hat{\mathbf{W}}) = -\text{trace}((\hat{\mathbf{W}} \hat{\mathbf{Y}})(\hat{\mathbf{W}} \hat{\mathbf{Y}})^T) \quad (2-14)$$

这里

$$\hat{\mathbf{W}} = [\mathbf{W}_1, \mathbf{W}_2], \hat{\mathbf{Y}} = \begin{pmatrix} \mathbf{x}^s & 0 \\ 0 & \mathbf{x}^{tl} \end{pmatrix}, \text{ and } \hat{\mathbf{X}} = [\mathbf{X}_1, \mathbf{X}_2]$$

因此，总的优化问题如下：

$$\begin{aligned} \{\mathbf{D}^*, \hat{\mathbf{W}}^*, \hat{\mathbf{W}}^*\} &= \arg \min_{\mathbf{D}, \hat{\mathbf{W}}, \hat{\mathbf{W}}} \mathbf{C}_1(\mathbf{D}, \hat{\mathbf{W}}, \hat{\mathbf{W}}) + \lambda \mathbf{C}_2(\hat{\mathbf{W}}) \\ &s.t. \mathbf{W}_i \mathbf{W}_i^T = \mathbf{I}, i = 1, 2 \text{ and } \|\hat{\mathbf{x}}_j\|_0 \leq T_0, \forall j \end{aligned} \quad (2-15)$$

这里 λ 是一个常数。该优化问题在文章^[45]中由一个两步的操作解决。

2.4 本章小结

本章主要介绍了域适应算法的两种应用场景以及目前主要的域适应算法。在计算机视觉中域适应算法大致可分为以下几类：基于特征增强的方法、基于特征转换的方法、参数适应方法和基于字典学习的方法，在本章中对这几种方法都进行了说明。

第3章 基于加权子空间生成的域适应算法

本章主要介绍基于加权子空间生成的域适应算法。首先介绍基于子空间对齐的域适应算法，该算法在文章^[23]中提出。接着介绍在子空间对齐算法基础上，由我提出的加权子空间对齐算法，实验结果显示该算法与当前主流算法相比性能有提升，相关成果已发表在文章^[46]中，本人是第一作者。之后对加权子空间对齐算法又进行了相应改进，改进后的算法的性能也有相应提升。

3.1 子空间对齐算法

如2.2小节所定义，源域有 N_s 个数据样本 \mathbf{x}^s ，目标域有 N_t 个数据样本 \mathbf{x}^t ，且 $\mathbf{x}^s, \mathbf{x}^t \in \mathbb{R}^D$ 。即使源域和目标域数据均位于 D 维空间，它们的边际分布也不同。如果直接在原始特征空间上进行迁移训练，得到的分类器普适性比较差，容易过拟合。因此，要对源域和目标域数据更鲁棒的特征表达上操作^[23]。首先，将 D 维源域样本和目标域样本进行归一化操作（均值为0，标准差为1）。然后，利用PCA降维，按由大到小顺序，选出前 d ($d < D$) 个特征值对应的特征向量。这前 d 个特征值对应的特征向量作为源域和目标域数据子空间的基，分别用 \mathbf{S} 和 \mathbf{T} 表示，这里 $\mathbf{S}, \mathbf{T} \in \mathbb{R}^{D \times d}$ 。因为 \mathbf{S}^\top 和 \mathbf{T}^\top 是标准正交的，所以 $\mathbf{S}^\top \mathbf{S} = \mathbf{I}_d$ 、 $\mathbf{T}^\top \mathbf{T} = \mathbf{I}_d$ ， \mathbf{I}_d 表示 $d \times d$ 的单位矩阵。

在子空间对齐算法中，首先将源域样本和目标域样本分别映射到其对应的子空间中，这一步通过如下操作实行： $\mathbf{x}^s \mathbf{S}$ ， $\mathbf{x}^t \mathbf{T}$ 。然后，在子空间上学习线性映射将源域子空间对齐到目标域子空间上。为了实现这一任务，使用子空间对齐方法。子空间对齐方法通过转换矩阵 \mathbf{M} 将源域子空间的基 \mathbf{S} 对齐到目标域子空间的基 \mathbf{T} 。转换矩阵 \mathbf{M} 通过最小化下面式子的Bregman散度得到：

$$F(\mathbf{M}) = \|\mathbf{S}\mathbf{M} - \mathbf{T}\|_F^2, \quad (3-1)$$

$$\mathbf{M}^* = \operatorname{argmin}_{\mathbf{M}}(F(\mathbf{M})) \quad (3-2)$$

这里 $\|\cdot\|_F$ 表示Frobenius范数。因为 \mathbf{S} 和 \mathbf{T} 是前 d 个特征值对应的特征向量组成的，所以 \mathbf{S} 和 \mathbf{T} 本身已被正则化，故在式3-1中不需要再加正则项。因此在式3-2中

对 \mathbf{M} 可以获得一个简单的闭式解。由Frobenius范数在正交变换下的不变性，可得：

$$F(\mathbf{M}) = \|\mathbf{S}^\top \mathbf{S} \mathbf{M} - \mathbf{S}^\top \mathbf{T}\|_F^2 = \|\mathbf{M} - \mathbf{S}^\top \mathbf{T}\|_F^2 \quad (3-3)$$

从式3-3可以得到 \mathbf{M} 的解为 $\mathbf{M}^* = \mathbf{S}^\top \mathbf{T}$ 。由此可以得到源域子空间对齐到目标域子空间后的基为 $\mathbf{S}_a = \mathbf{S} \mathbf{S}^\top \mathbf{T}$ 。值得注意的是，如果源域和目标域相同，那么 $\mathbf{S} = \mathbf{T}$ 并且 \mathbf{M}^* 是单位矩阵。

为了对源域样本 \mathbf{x}^s 和目标域样本 \mathbf{x}^t 进行比较，需要一个相似性度量函数 $\text{sim}(\mathbf{x}^s, \mathbf{x}^t)$ 。我们先将 \mathbf{x}^s 和 \mathbf{x}^t 映射到其对应的子空间上，得到 $\mathbf{x}^{s,d}$ 和 $\mathbf{x}^{t,d}$ ，接着利用转换矩阵 \mathbf{M}^* 在子空间上将其对齐。因此，相似性度量函数 $\text{sim}(\mathbf{x}^s, \mathbf{x}^t)$ 的定义如下所示：

$$\text{sim}(\mathbf{x}^s, \mathbf{x}^t) = (\mathbf{x}^s \mathbf{S} \mathbf{M}^*) (\mathbf{x}^t \mathbf{T})^\top = \mathbf{x}^s \mathbf{S} \mathbf{M}^* \mathbf{T}^\top \mathbf{x}^{t \top} = \mathbf{x}^s \mathbf{A} \mathbf{x}^{t \top} \quad (3-4)$$

这里 $\mathbf{A} = \mathbf{S} \mathbf{S}^\top \mathbf{T} \mathbf{T}^\top$ 。式3-4看起来像一个泛化的内积形式，但 \mathbf{A} 并不是半正定的， \mathbf{A} 的主要作用是表达原始空间中特征向量各部分的重要程度。

通过式3-4得到了在对齐后的子空间上源域样本 \mathbf{x}^s 和目标域样本 \mathbf{x}^t 的相似性度量，接下来可以利用最近邻算法进行分类。综上，子空间对齐算法的流程如算法1所示。

算法 1: 子空间对齐算法

输入: 源域样本 \mathbf{x}^s , 目标域样本 \mathbf{T} , 源域样本标签 y^s , 子空间维数 d

输出: 预测目标域样本标签 y^t

$\mathbf{S} \leftarrow \text{PCA}(\mathbf{x}^s, d)$;

$\mathbf{T} \leftarrow \text{PCA}(\mathbf{x}^t, d)$;

$\mathbf{S}_a \leftarrow \mathbf{S} \mathbf{S}^\top \mathbf{T}$;

$\mathbf{x}^{s,d} = \mathbf{x}^s \mathbf{S}_a$;

$\mathbf{x}^{t,d} = \mathbf{x}^t \mathbf{T}$;

$y^t \leftarrow \text{classifier}(\mathbf{S}_a, \mathbf{x}^{s,d}, \mathbf{x}^{t,d})$;

3.2 加权子空间对齐算法

上文提到的子空间对齐算法中，其认为每个源域样本都同等重要。实际上并

非所有的源域样本都同样重要。源域样本的特征越接近目标域样本，其对于训练具有迁移能力分类器的作用越大。针对以上不足，本文对子空间对齐算法进行了改进，设计了权重表达式，对每一个源域样本赋予权重，并创造性地将权重运用于子空间的生成中。改进后算法的主要思想如图3.1所示：(a)原始源域，(b)原始目标域，(c)由源域样本生成的加权子空间，(d)对齐后的加权子空间。改进后的算法被称为加权子空间对齐算法，主要由两步组成：第一步是加权子空间生成，第二步是加权子空间对齐。接下来将对这两个步骤进行详细说明。

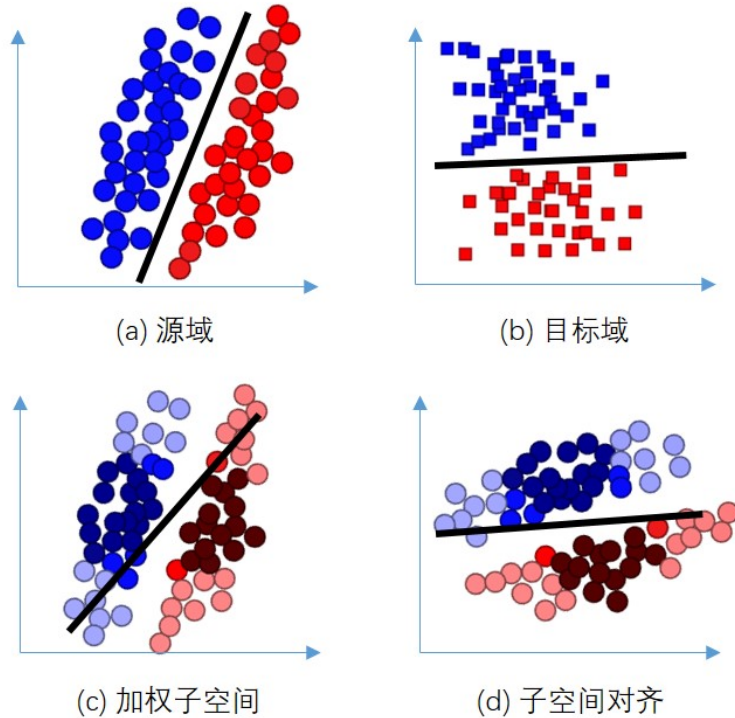


图 3.1 加权子空间对齐算法示意图

3.2.1 加权子空间生成

图3.2是加权子空间对齐算法的框图，虚线框部分的主要步骤是生成加权子空间，也是本算法的创新部分。这一步骤的目的主要是使生成的源域子空间能够接近目标域子空间，从而有益于接下来的子空间对齐操作。生成加权子空间主要包含两个操作：计算权重和生成子空间。计算权重的目的是把那些较重要的源域样本筛选出来，赋予更高的权重，接下来生成子空间。这样做的目的是使构造出的源域子空间更接近目标域子空间，迁移效果更好。

我们对每一个源域样本赋予一个权重，设源域样本的权重向量 $\mathbf{w} = [w_1 \dots w_{N_s}]^T$ ，这里 $\mathbf{w} \in \mathbb{R}^{N_s \times 1}$ 。源域样本 \mathbf{x}^s 的分布离目标域越近，说明它越重

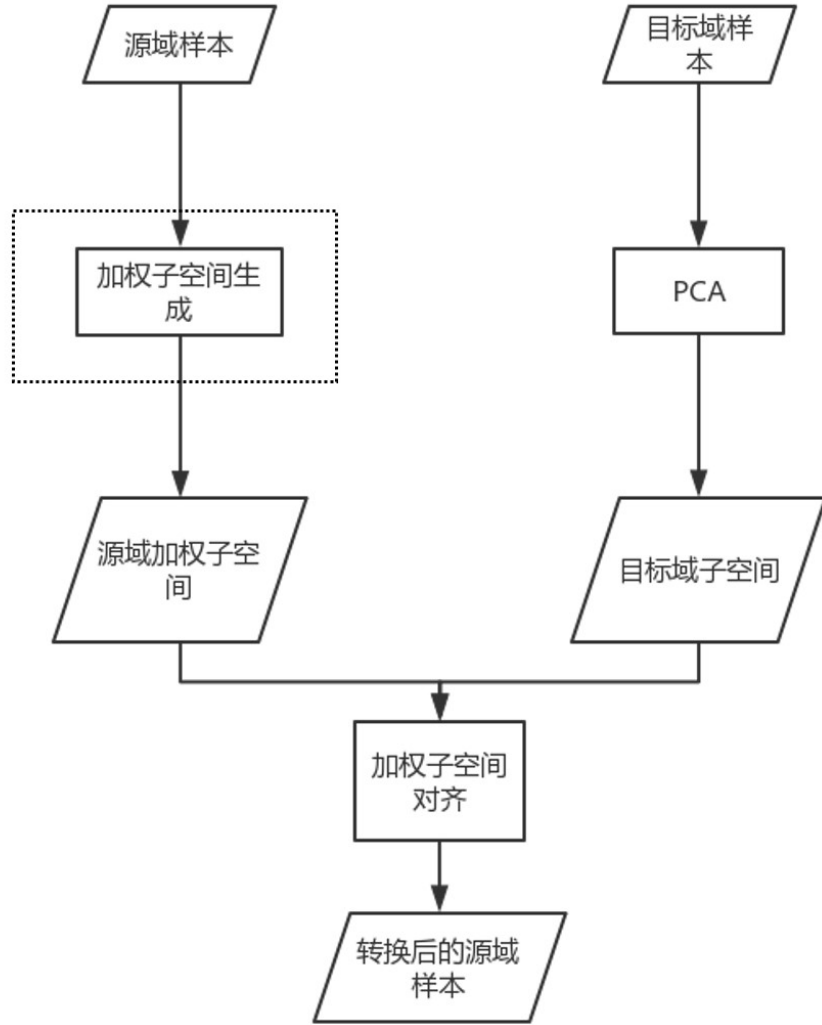


图 3.2 加权子空间对齐算法框图

要，相应的权重越高，即：

$$\mathbf{w} \propto \frac{1}{\|\mathcal{D}_S - \mathcal{D}_T\|^2} \quad (3-5)$$

依据式3-5，我们用源域样本与目标域样本间的距离作为衡量权重的准则。对于源域样本 \mathbf{x}_i^s ，首先计算它与每个目标域样本间的距离：

$$dist(\mathbf{x}_i^s, \mathbf{x}_j^t) = \|\mathbf{x}_j^t - \mathbf{x}_i^s\|_2^2 \quad (3-6)$$

式3-6定义了源域样本与目标域样本间的距离。首先计算源域样本 \mathbf{x}_i^s 与每一个

目标域样本的距离，然后计算最小距离，作为权重度量基准。

$$\min_dist_i = \min_{j=1, \dots, N_s} dist(\mathbf{x}_i^s, \mathbf{x}_j^t), \quad (3-7)$$

由式3-7得到最小距离 \min_dist 后，计算源域样本 \mathbf{x}_i^s 权重：

$$w_i = w_0 + \sum_{j=1}^{N_t} \delta(\min_dist_i - dist(\mathbf{x}_i^s, \mathbf{x}_j^t)) \quad (3-8)$$

这里 w_0 是初始权重（实验时将其设为1），而 $\delta(x) = \begin{cases} 1, & x = 0 \\ 0, & x < 0 \end{cases}$ 。根据 \min_dist_i 的定义可知，函数 $\delta(x)$ 的定义域是 $(-\infty, 0]$ 。从式3-8可知，如果源域样本 \mathbf{x}_i^s 与目标域样本 \mathbf{x}_j^t 的距离等于其与目标域样本距离的最小值 \min_dist_i ，则源域样本 \mathbf{x}_i^s 的权重增大。由式3-8可得到源域样本的权重向量 \mathbf{w} 。

接下来，我们通过加权PCA生成加权子空间。对于加权PCA，首先利用下式计算加权协方差矩阵：

$$\mathbf{C} = \frac{1}{N_s} \sum_{i=1}^{N_s} (\mathbf{x}_i^s - \mu)^\top w_i (\mathbf{x}_i^s - \mu) \quad (3-9)$$

这里 μ 是加权均值， μ 的计算方式如下：

$$\mu = \frac{1}{\sum_{i=1}^{N_s} w_i} \sum_{i=1}^{N_s} w_i \mathbf{x}_i^s. \quad (3-10)$$

然后我们对加权协方差矩阵 \mathbf{C} 进行特征分解，并选取前 d 个最大的特征值对应的特征向量。这些特征向量是源域子空间的基向量，组成矩阵 $\mathbf{S}_w \in \mathbb{R}^{D \times d}$ 。在得到源域的加权子空间后，在目标域样本上执行PCA，得到目标域子空间的基 $\mathbf{T} \in \mathbb{R}^{D \times d}$ 。至此，加权子空间生成这一步完成，接下来进行加权子空间对齐操作。

3.2.2 加权子空间对齐

通过加权子空间生成，我们得到了源域和目标域的子空间。如第3.1节所述，子空间对齐操作目的是找到一个线性转换矩阵 \mathbf{M} ，对源域子空间的基进行转换，

使转换后的子空间的基能尽可能地接近目标域子空间的基，从而起到域适应的功能。同样，在加权子空间对齐中，要寻找下式的最优解：

$$F(\mathbf{M}_w) = \|\mathbf{S}_w^\top \mathbf{S} \mathbf{M} - \mathbf{S}_w^\top \mathbf{T}\|_F^2 = \|\mathbf{M} - \mathbf{S}_w^\top \mathbf{T}\|_F^2 \quad (3-11)$$

$$\mathbf{M}_w^* = \operatorname{argmin}_{\mathbf{M}_w} (F(\mathbf{M}_w)) \quad (3-12)$$

由第3.1可知，转换矩阵 \mathbf{M}_w 有闭式解 $\mathbf{M}_w^* = \mathbf{S}^\top \mathbf{T}$ 。接下来我们得到矩阵：

$$\mathbf{U} = \mathbf{S} \mathbf{M}_w^* = \mathbf{S}_w \mathbf{S}^\top \mathbf{T}. \quad (3-13)$$

矩阵 $\mathbf{U} \in \mathbb{R}^{D \times d}$ 被用来表达源域样本特征。最后，用最近邻算法在转换后的源域样本特征上学习分类器。整个算法的伪代码如算法2所示。

3.2.3 数据准备

为了检验加权子空间算法的性能，我们将其与其它几个目前最优秀的算法进行比较，实验任务是图片分类。实验数据库是Office数据库^[32]和Caltech10数据集^[20]，其中Office数据集有3个子数据集：Amazon、DSLR以及Webcam。

- Amazon数据集中的图片是从网络上下载下来的，抓取网店（www.amazon.com）上的商品图片。这是一种目前很流行的获取数据的方式，因为很容易就能获得用于训练分类模型的大量数据。这些商品图片大多是在工作室光照环境下拍摄的，光照条件良好，图片背景干扰很少，图片分辨率质量中等。

- DSLR数据集里的图片是由一部数码单反相机（Digital SLR camera）在真实场景里自然光照条件下拍摄的。这些图片具有高分辨率（4288×2848）和低噪声的特点。

- Webcam数据集里的图片是用一个简易的网络摄像头拍摄的。图片分辨率较低（640×480）并且有噪声、颜色、白平衡等问题。

- Caltech10数据集是从Caltech-256数据库^[47]中抽取部分类的图片得到的。

每个数据集被看作一个域，从这四个数据集中取了共有的10类：BACKPACK、TOURING-BIKE、CALCULATOR、HEADPHONES、COMPUTER-KEYBOARD、LAPTOP-101、COMPUTER-MONITOR、COMPUTER-MOUSE、COFFEE-MUG、VIDEO-PROJECTOR。每个域中的每一类大约有8到151个图片

算法 2: 加权子空间对齐算法

输入: 源域样本 \mathbf{x}^s , 目标域样本 \mathbf{x}^t , 源域样本标签 y^s , 子空间维数 d

输出: 预测目标域样本标签 y^t

```

for  $i = 1$  to  $N_s$  do
  for  $j = 1$  to  $N_t$  do
     $dist(\mathbf{x}_i^s, \mathbf{x}_j^t) = \|\mathbf{x}_i^s - \mathbf{x}_j^t\|_2$ 
  end for
end for
for  $j = 1$  to  $N_t$  do
   $min\_dist_i = \min_{i=1, \dots, N_s} dist(\mathbf{x}_i^s, \mathbf{x}_j^t)$ 
end for
 $\mu = \sum_{i=1}^{N_s} w_i \mathbf{x}_i^s / \sum_{i=1}^{N_s} w_i$ 
for  $i = 1$  to  $N_s$  do
   $w_i = w_0 + \sum_{j=1}^{N_t} \delta(min\_dist_i - dist(\mathbf{x}_i^s, \mathbf{x}_j^t))$ 
end for
 $\mathbf{C} = \frac{1}{N_s} \sum_{i=1}^{N_s} (\mathbf{x}_i^s - \mu)^T w_i (\mathbf{x}_i^s - \mu)$ 
 $\mathbf{S}_w \leftarrow \text{weighted PCA}(\mathbf{x}^s, w)$ 
 $\mathbf{T} \leftarrow \text{PCA}(\mathbf{x}^t, d)$ 
 $\mathbf{M}_w^* = \mathbf{S}_w^T \mathbf{T}$ 
 $\mathbf{U} = \mathbf{S}_w \mathbf{M}_w^* = \mathbf{S}_w \mathbf{S}_w^T \mathbf{T}$ 
 $y^t \leftarrow \text{classifier}(\mathbf{x}^s, y^s, \mathbf{x}^t, \mathbf{U}, \mathbf{T})$ 

```

样本, 总共有2533张图片。因为总共有四个数据集: \mathcal{A} (Amazon)、 \mathcal{C} (Caltech10)、 \mathcal{D} (DSLRL)和 \mathcal{W} (Webcam), 对这四个数据集两两配对进行迁移, 共有12种迁移情况, 例如 $\mathcal{A} \rightarrow \mathcal{C}$ 。图3.3是四个数据集的图片示例, 从中可以看出不同域间的数据差异。

在得到图片后, 使用如下方法提取特征。首先所有图片均被调整为相同尺寸并转换成灰度图, 然后提取SURF特征^[48]。在之前的文献中, SURF特征已被证明在处理图片的噪声、位移、几何形变和亮度等问题上均有良好的鲁棒性, 对各种情况下的图片提取出的特征都比较稳定。在提取SURF特征时, 将团块阈值设为1000, 其余参数为默认值不变。本实验用一个64维的SURF特征描述子来描述兴趣点周围的图片块, 在得到图片的描述子之后, 将其向量化得到最终的特征向量。然后在Amazon数据集上随机选取一部分用k-means聚类算法构建一个800维的字典。最后依据该字典, 每张图片的特征向量都被转换为一个800维的直方图向

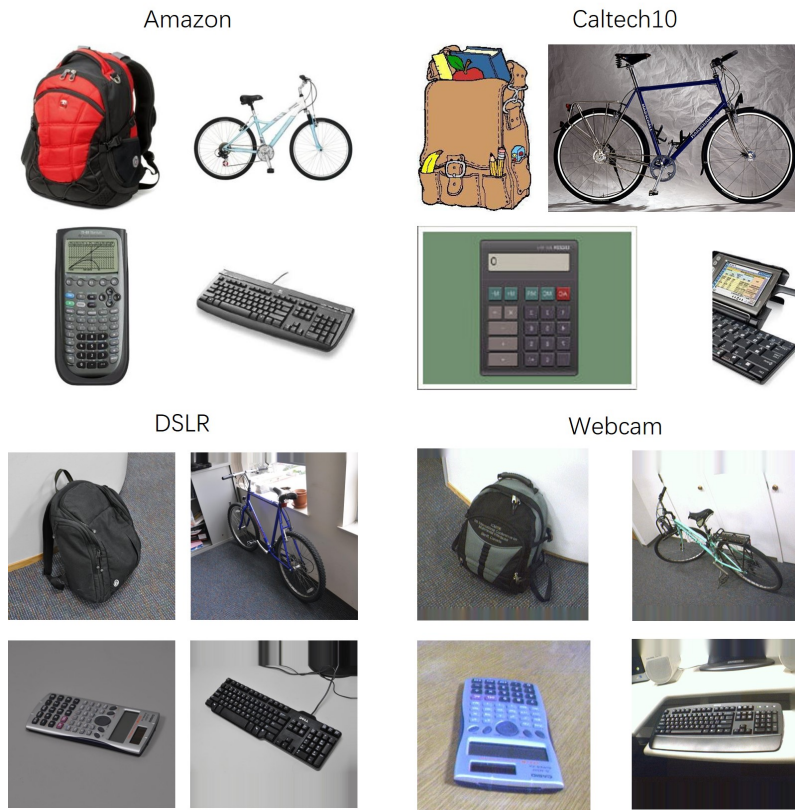


图 3.3 Amazon、Caltech10、DSLR、Webcam 数据集图片示例

量。因此，每个样本的特征维度是800。

3.2.4 实验结果

第3.2.2节中，已经说到用最近邻算法分类。因为最近邻算法没有参数，不需要调参，能够消除参数选取对实验结果的影响，适合用于进行比较实验。我们将子空间加权算法与三个目前最优秀的算法进行比较，三个算法分别是：Geodesic Flow Kernel (GFK)^[20]、Landmarks (LM)^[49]和Subspace Alignment (SA)^[23]。正如文章^[20]所建议的那样，将最近邻分类器作为基准，因为不需要验证集来调参。

表3.1是加权子空间对齐算法与其它域适应算法的实验数值结果，即在目标域样本上的分类准确率，粗体表示在该迁移情况下分类的最好结果。在表3.1中，为了简便，加权子空间对齐算法用WSA (Weighted Subspace Alignment)表示。同时为了更直观的观测实验结果，我根据表3.1制作了柱状图，如图3.4所示。值得注意的是，在没有进行域适应操作是，最近邻分类器的平均准确率只有31.73%。在所有12种迁移情况中，加权子空间对齐算法有9种情况准确率是最高的。而在剩下的3中迁移情况中，加权子空间对齐算法的准确率也能排在第二。在12中迁移情况中，加权子空间对齐算法的平均准确率是46.88%，相较于其它算法，有了2%的提高，说明算法效果良好、性能优异。在GFK算法中，子空间维数 d 需要足够小才

表 3.1 域适应算法在目标域样本上的准确率 (%)

算法	NN	GFK ^[20]	LM ^[49]	SA ^[23]	WSA
$\mathcal{A} \rightarrow \mathcal{C}$	26.00	40.25	40.16	39.80	39.98
$\mathcal{A} \rightarrow \mathcal{D}$	25.48	36.31	40.76	36.94	38.85
$\mathcal{A} \rightarrow \mathcal{W}$	29.83	38.98	38.98	37.63	39.32
$\mathcal{C} \rightarrow \mathcal{A}$	23.70	41.02	41.75	42.07	42.17
$\mathcal{C} \rightarrow \mathcal{D}$	25.48	38.85	39.49	45.86	46.49
$\mathcal{C} \rightarrow \mathcal{W}$	25.76	40.68	37.97	32.20	42.03
$\mathcal{D} \rightarrow \mathcal{A}$	28.50	32.05	30.97	34.24	35.38
$\mathcal{D} \rightarrow \mathcal{C}$	26.27	30.28	31.34	32.50	34.37
$\mathcal{D} \rightarrow \mathcal{W}$	63.39	75.59	84.75	88.47	89.49
$\mathcal{W} \rightarrow \mathcal{A}$	22.96	29.75	31.00	34.34	33.19
$\mathcal{W} \rightarrow \mathcal{C}$	19.86	30.72	29.21	28.76	30.89
$\mathcal{W} \rightarrow \mathcal{D}$	59.24	80.89	83.44	88.54	90.44
平均值	31.37	42.95	44.15	45.11	46.88

能保证源域样本在子空间上平稳地过渡到目标域子空间。然而，低维子空间不能很好地表达原始特征空间的数据，从而导致分类器性能下降。而在加权子空间对齐算法中，子空间维数 d 有理论上限^[23]，能够充分表达原始数据。LM算法对源域样本进行了筛选，只利用了部分源域样本数据，没有完全利用源域样本信息。而加权子空间对齐算法用上了所有的源域样本，每一个源域样本依据其重要性被赋予一个权重，因此，加权子空间对齐算法的分类器性能相比于LM算法有所提高。SA算法认为所有源域样本都同样重要，没有应用权重思想，而加权子空间对齐算法对每个源域样本进行了权重衡量，所以实验准确率较SA有所提升。

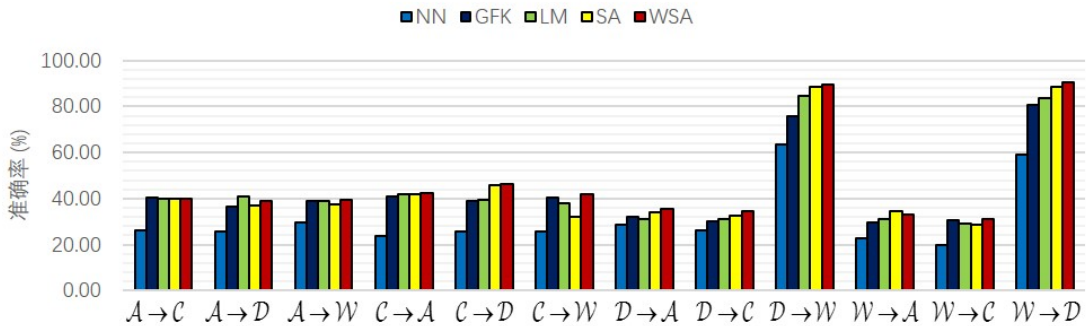


图 3.4 实验结果柱状图

3.3 对加权子空间对齐算法的改进

在提出加权子空间对齐算法后，我没有停止研究的步伐。从式3-8中函数 $\delta(x)$ 的定义可以看出源域样本 \mathbf{x}_i^s 的权重 w_i 为整数。这就产生了局限性，我们希望权重取值范围不仅是整数，还能是连续的，从而能更好的体现源域样本的重要性。

3.3.1 权重改进

在第3.2.1节中，源域样本 \mathbf{x}_i^s 的权重定义如下：

$$w_i = w_0 + \sum_{j=1}^{N_i} \delta(\min_dist_i - \text{dist}(\mathbf{x}_i^s, \mathbf{x}_j^t))$$

其中 $\delta(x) = \begin{cases} 1, & x = 0 \\ 0, & x < 0 \end{cases}$ 。为了使其具有连续性，进行了指数化操作，改进后的权重定义式如下：

$$w_i = w_0 + \sum_{j=1}^{N_i} \exp\left(-\frac{\text{dist}(\mathbf{x}_i^s, \mathbf{x}_j^t) - \min_dist}{\sigma}\right) \quad (3-14)$$

其中 σ 是一个常数。通过指数化操作后，权重的取值具有连续性，能更好的表达源域样本的重要性。赋予权重之后进行子空间生成和对齐操作，改进的加权子空间对齐算法的步骤如算法3所示。

3.3.2 数据准备

为了检验改进的加权子空间对齐算法的性能，该算法与其它域适应算法做了对比实验^[20,21,23,49]。除了在第3.2.3节用到的Office和Caltech-256数据库外，本实验还另外增加了四个数据库：USPS、MNIST、MSRC、VOC2007。

- USPS数据库包含了7291张训练图片和2007张测试图片，图片大小为 16×16 。

- MNIST数据库包含60000张训练图片和10000张测试图片，图片大小为 28×28 。

USPS数据库和MNIST数据库都是关于手写体数字的，共有10类。从图3.5可以看出，USPS和MNIST数据库图片的特征分布不同。按照文章^[21]的实验设置，本实验随机选取了USPS数据库中的1800张以及MNIST数据库中的2000张图片，

算法 3: 改进的加权子空间对齐算法

输入: 源域样本 \mathbf{x}^s , 目标域样本 \mathbf{x}^t , 源域样本标签 y^s , 子空间维数 d

输出: 预测目标域样本标签 y^t

```

for  $i = 1$  to  $N_s$  do
  for  $j = 1$  to  $N_t$  do
     $dist(\mathbf{x}_i^s, \mathbf{x}_j^t) = \|\mathbf{x}_i^s - \mathbf{x}_j^t\|_2$ 
  end for
end for
for  $j = 1$  to  $N_t$  do
   $min\_dist_i = \min_{i=1, \dots, N_s} dist(\mathbf{x}_i^s, \mathbf{x}_j^t)$ 
end for
 $\mu = \sum_{i=1}^{N_s} w_i \mathbf{x}_i^s / \sum_{i=1}^{N_s} w_i$ 
for  $i = 1$  to  $N_s$  do
   $w_i = w_0 + \sum_{j=1}^{N_t} \exp\left(-\frac{dist(\mathbf{x}_i^s, \mathbf{x}_j^t) - min\_dist}{\sigma}\right)$ 
end for
 $\mathbf{C} = \frac{1}{N_s} \sum_{i=1}^{N_s} (\mathbf{x}_i^s - \mu)^T w_i (\mathbf{x}_i^s - \mu)$ 
 $\mathbf{S}_w \leftarrow \text{weighted PCA}(\mathbf{x}^s, w)$ 
 $\mathbf{T} \leftarrow \text{PCA}(\mathbf{x}^t, d)$ 
 $\mathbf{M}_w^* = \mathbf{S}_w^T \mathbf{T}$ 
 $\mathbf{U} = \mathbf{S}_w \mathbf{M}_w^* = \mathbf{S}_w \mathbf{S}_w^T \mathbf{T}$ 
 $y^t \leftarrow \text{classifier}(\mathbf{x}^s, y^s, \mathbf{x}^t, \mathbf{U}, \mathbf{T})$ 

```

然后统一调整到了 16×16 的大小。调整之后，源域样本和目标域样本的特征空间的维度相同。我们先把USPS数据库作为源域，MNIST数据库作为目标域，然后再倒过来MNIST数据库作为源域，USPS数据库作为目标域。

- MSRC数据库包含4323张图片，共有18类。
- VOC2007数据库中训练集和验证集加在一起共有5011张图片，共有20类。

从图3.5中可以看出MSRC和VOC2007数据库图片的特征分布不同。这两个数据库的类中有6个是相同的：aeroplane、bicycle、bird、car、cow、sheep。遵循文章^[21]的实验设置，我们在MSRC数据库中选取了1269张图片，在VOC2007数据库中选取了1530张图片。图片被统一调整为256个像素大小，然后运用VLFeat开源包提取128维的SIFT特征^[50]，在此过程中，运用k-means聚类算法得到240维的字典。我们先用MSRC数据库作为源域，VOC2007数据库作为目标域，然后再倒过来VOC2007数据库作为源域，MSRC数据库作为目标域。

除了第3.2.3节提到的12种迁移情况，又多了4种迁移情况，所以该实验总共有16种迁移情况。



图 3.5 USPS、MNIST、MSRC、VOC2007、Office、Caltech-256 数据库图片示例

3.3.3 实验结果

为了简便起见，改进的加权子空间对齐算法称为IWSA (Improved Weighted Subspace Alignment)。本实验将IWSA算法与其它四个目前性能最好的域适应算法进行比较，任务是图片分类，分类器是最近邻算法NN (1-Nearest Neighbor

Classifier)。

- 1-Nearest Neighbor Classifier (NN)
- Geodesic Flow Kernel (GFK)^[20] + NN
- Landmarks (LM)^[49] + NN
- Subspace Alignment (SA)^[23] + NN
- Transfer Joint Matching (TJM)^[21] + NN

表3.2提供了IWSA算法与其它目前性能最好的域适应算法的实验结果。粗体表示该迁移情况下的最好结果。值得注意的是LM算法在前4种情况不适用，故没有实验结果。为了更好的观测实验结果，将表3.2可视化做成了柱状图，如图3.6所示。从实验结果可以看出，在没有域适应操作的情况下，NN算法在前4种迁移情况下的平均准确率只有45.92%，后12种迁移情况的平均准确率只有31.37%。在所有16中迁移情况中，IWSA在11种迁移情况下的分类准确率是最高的，而在剩下的5种，IWSA也仅比最好的结果稍微低了一点。在前4种迁移情况中，IWSA算法的平均准确率是50.57%，在后8种迁移情况中，IWSA算法的平均准确率是48.82%，相比于NN、GFK、SA、LM和TJM算法有了提高，证明了我们提出的IWSA算法在性能上的优越性。在第3.2.4节中，实验数据显示，WSA算法在后8种迁移情况的平均准确率是46.88%，改进后的IWSA比WSA在平均准确率上有了2%的提高，说明了算法改进后性能上有了明显的提高。在之前的第3.2.4节中，我们讨论了GFK、SA、LM算法实验结果较差的原因。接下来讨论一下TJM算法在实验结果上不如IWSA算法的原因。TJM算法在目标函数上引用了 $l_{2,1}$ 范数来对源域样本进行权重衡量。TJM根据源域样本和目标域样本的相关性来赋予权重。然而，TJM没有利用权重来生成子空间。IWSA相比于TJM在性能上有提高的主要原因是其利用源域样本的权重构造了加权子空间，取得了良好的域适应效果。

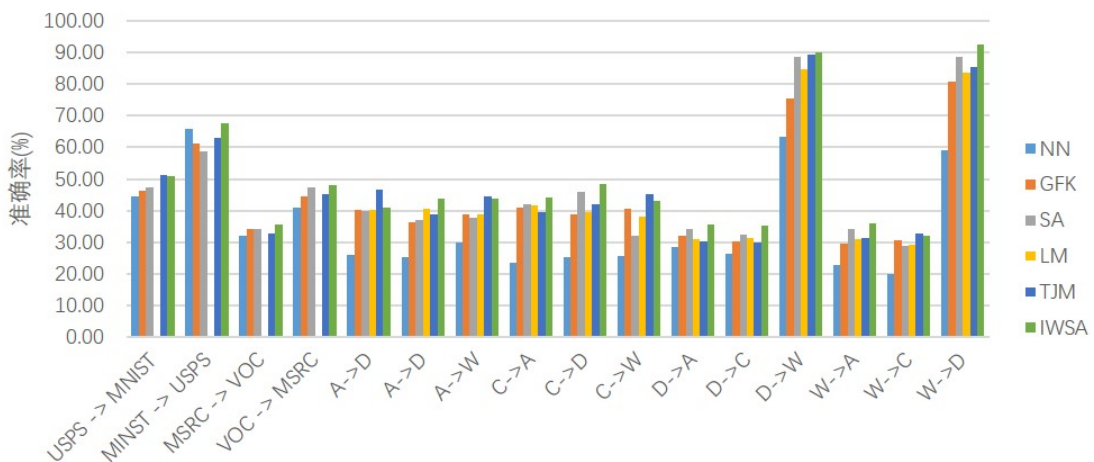


图 3.6 实验结果柱状图

表 3.2 域适应算法在目标域样本上的准确率 (%)

算法	NN	GFK ^[20]	SA ^[23]	LM ^[49]	TJM ^[21]	IWSA
USPS → MNIST	44.70	46.45	47.50	-	51.25	50.95
MNIST → USPS	65.94	61.22	58.83	-	63.00	67.61
MSRC → VOC	31.96	34.18	34.38	-	32.75	35.56
VOC → MSRC	41.06	44.47	47.36	-	45.15	48.15
平均值	45.92	46.58	47.02	-	48.04	50.57
A → C	26.00	40.25	39.80	40.16	46.76	40.87
A → D	25.48	36.31	36.94	40.76	38.98	43.95
A → W	29.83	38.98	37.63	38.98	44.59	43.73
C → A	23.70	41.02	42.07	41.75	39.45	44.05
C → D	25.48	38.85	45.86	39.49	42.03	48.41
C → W	25.76	40.68	32.20	37.97	45.22	43.05
D → A	28.50	32.05	34.24	30.97	30.19	35.80
D → C	26.27	30.28	32.50	31.34	29.96	35.26
D → W	63.39	75.59	88.47	84.75	89.17	90.17
W → A	22.96	29.75	34.34	31.00	31.43	36.01
W → C	19.86	30.72	28.76	29.21	32.78	32.15
W → D	59.24	80.89	88.54	83.44	85.42	92.36
平均值	31.37	42.95	45.11	44.15	46.33	48.82

3.4 本章小结

在本章中，本文首先介绍了文章^[23]提出的子空间对齐算法，其目的是在源域和目标域的子空间上学习线性映射，使映射后的源域数据特征尽可能地接近目标域。然后，介绍了我们提出的新的无监督域适应算法，称之为加权子空间对齐算法，该算法的主要步骤是将加权源域样本生成的子空间对齐。通过加权子空间对齐操作，源域样本中特征分布更接近目标域的被赋予更高的权重，权重是根据源域样本特征到目标域样本特征的距离计算的。接着将源域样本的加权子空间和目标域的子空间对齐，实现域适应的功能。本文提出的加权子空间算法在数个图片分类的数据库上均有良好的表现，实验结果也证明了该算法相较于其它无监督域适应算法在性能上的优越性。

最后我们对加权子空间对齐算法进行了相应改进，主要体现在权重设计上。本人对权重表达式进行了改进，使之更能体现源域样本的重要性。在获得源域样本的权重后，利用加权PCA生成加权子空间并进行子空间对齐操作。改进后的加权子空间对齐算法在性能上有了相应的提高，在实验上也取得了良好的表现。

第4章 基于深度神经网络的域适应算法

自从2012年神经网络在ImageNet竞赛上大放光彩后^[51]，计算机视觉领域掀起了一股深度学习的浪潮。在计算机视觉的很多问题上，深度学习算法都取得了比非深度学习算法更优异的成绩。而在域适应领域，也有很多学者开始使用深度学习的相关算法和技术，来解决源域和目标域数据特征分布不同的问题，并取得了良好效果^[52-55]。在本章中，本文将介绍一下神经网络在域适应中的研究，实验任务是行为识别。

4.1 基于深度神经网络的域适应算法

4.1.1 双流深度卷积网络

行为识别在计算机视觉领域一直受到广泛关注，相比于静止的图片，视频的时间组成部分又提供了额外的信息，所以有些工作是基于运动信息进行研究。双流深度卷积网络是对深度卷积网络的拓展，我们希望能将其应用到行为识别领域中。这项工作曾经在文章^[56]中被探索过，把一连串的视频帧图片直接送到网络中去，但实验结果却不好，甚至比一些用手工设计的特征的算法还要差^[57,58]。因此在文章^[59]中，作者提出了一个不同的基于双流（空间流和时间流）的架构模型。空间流网络用于训练静止的视频帧图片，而时间流网络用于训练视频的光流图，两个网络都是基于卷积网络。将空间流和时间流分开训练能使我们得以利用在大图图片库ImageNet上训练好的网络模型，这些模型因为是在大量标定图片上训练出的，其提取特征的能力都比较强。

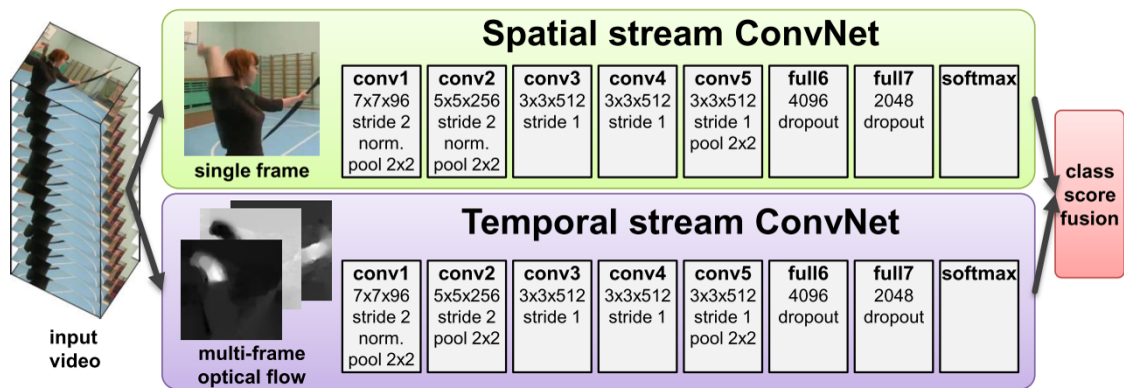


图 4.1 针对行为识别的双流架构模型

对于视频数据，其主要包含两个方面的信息，一个是空间上的，一个是时间

上的。对于空间部分，其主要体现在视频里一个个静止的帧图片，这些图片包含了视频中的场景和物体信息。对于时间部分，主要体现在帧间的运动，包含了观察者（相机）的运动以及物体本身的运动。针对这两个部分的信息，将行为识别的模型架构设计为双流（空间流和时间流）。如图4.1所示，空间流和时间流各用一个深度卷积网络进行训练，之后再两个网络得到的分类分数进行融合。

空间流卷积网络主要在单独的帧图片上进行操作，在静止图片上进行行为识别。图片上的静止信息是很有用的，因为有些行为是和特定物体强烈相关的。实际上，仅仅使用空间流卷积网络就能实现良好的分类效果。基于此，在后面的实验环节中只在空间流上进行了实验，来快速验证想法，毕竟训练深度神经网络很耗费时间。因为空间流卷积网络本质上就是对图片进行分类，所以可以利用已有的图片分类的卷积网络模型^[4]，这些模型在大型图片库ImageNet上有良好的分类效果，说明其提取图片特征的能力很强。而在这些模型的基础上进行再训练无疑可以提高网络本身的分类能力，从而提高行为识别的准确率。

时间流卷积网络主要是在光流图上进行操作，假设有连续的两帧 t 和 $t+1$ ，光流图表达的是帧间的运动情况，由一系列的位移向量组成。对于帧 t 上的点 (u, v) ，该点到帧 $t+1$ 的位移向量用 $\mathbf{d}_t(u, v)$ 表示。位移包含两个方向，水平方向和垂直方向，分别用 d_t^x 和 d_t^y 表示。如图4.2所示，(a)和(b)相邻连续两帧，青色方框标注的区域内有一个正在运动的手臂；(c)表示的是青色方框区域的光流；(d)表示的是水平方向的光流位移 d^x ；(e)表示的是垂直方向的光流位移 d^y 。

在训练深度卷积网络时， d_t^x 和 d_t^y 可以看作channel。为了表示一系列帧间的运动情况，我们将 L 张连续帧的光流位移 $d_t^{x,y}$ 叠在一起，总共有 $2L$ 个channel送到深度卷积神经网络中进行训练。

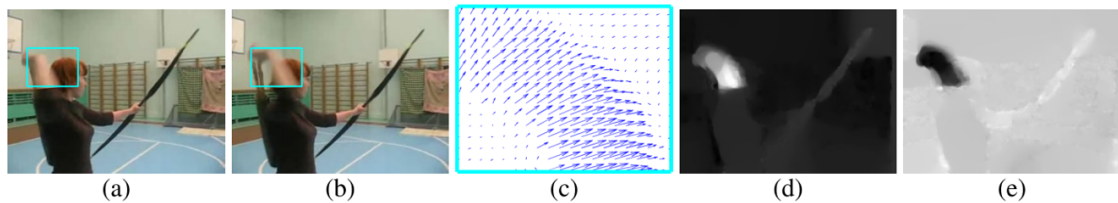


图 4.2 光流图

4.1.2 权重机制

在文章^[59]中，双流深度卷积网络使用的是CNN-M-2048网络模型^[60]。CNN-M-2048网络模型主要架构如图4.3所示，其中卷积层Convolutional Layer有5层，池化层Pooling Layer有1层，全连接层Fully-Connected Layer有2层，最后一层是Softmax层。

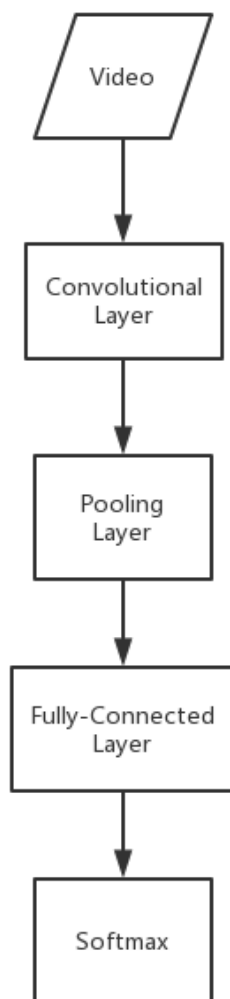


图 4.3 CNN-M-2048网络模型

在CNN-M-2048网络模型基础上，我加入了权重机制。这个权重机制被命名为**Attention Model**，其主要目的是对每张帧图片学习一个权重矩阵，以将图片中利于分类的区域通过权重矩阵高亮出来。**Attention Model**是在卷积层和池化层之间进行的，接下来详细介绍一下**Attention Model**的具体步骤。

假设第 k 张图片空间位置 (i, j) 上最后一层卷积层上出来的特征为 $\mathbf{x}_{ijk} \in \mathbb{R}^c$ ，这里 $i = 1, 2, \dots, d$ 和 $j = 1, 2, \dots, d$ 表示特征图的坐标， d 是特征图的高或宽（一般情况下特征图的高和宽是相同的）。

Attention Model会给每个图片的特征图一个权重分数 $s_{ijk} \in \mathbb{R}$ ，用于衡量权重，

定义如下：

$$s_{ijk} = f(\mathbf{w}^T \mathbf{x}_{ijk} + b) \quad (4-1)$$

这里 $\mathbf{w} \in \mathbb{R}^c$, $b \in \mathbb{R}^1$ 表示的是 Attention Model 学到的权重和偏置项。这两个参数作为整个网络参数的一部分，将在模型的训练过程中通过神经网络的优化学习得到。 $f(\cdot)$ 是 softplus 函数，定义如下：

$$f(x) = \ln(1 + \exp(x)) \quad (4-2)$$

因为加入 Attention Model 的目的是衡量图片中局部区域的重要性，对有利于分类的区域起到增强的作用，因此需要对 s_{ijk} 进行归一化操作：

$$a_{ijk} = \frac{s_{ijk} + \epsilon}{\sum_i \sum_j (s_{ijk} + \epsilon)} \quad (4-3)$$

其中 a_{ijk} 表示归一化后的权重分数， ϵ 是一个常量参数，对调节权重分数的整体分布具有重要作用，可以使 a_{ijk} 的分布更加合理。如果 s_{ijk} 小且没有常量参数 ϵ ，那么 a_{ijk} 就会较大，而这不是我们希望看到的，因为 s_{ijk} 小说明该区域对训练网络的重要性低。常量参数 ϵ 的设置就可以有效地解决这个问题，如果 ϵ 的设置合理， a_{ijk} 值小时会有 $a_{ijk} \approx 1/d^2$ 。

在得到归一化后的权重分数后，我们可以得到加权后的特征表达，通过如下操作：

$$\hat{\mathbf{x}}_{ijk} = a_{ijk} \odot \mathbf{x}_{ijk} \quad (4-4)$$

这里 \odot 表示对应元素相乘， $\hat{\mathbf{x}}_{ijk}$ 是加权后的特征表达，即加权特征图。

式4-1到4-4是 Attention Model 的主要步骤，执行完 Attention Model 后，加权特征图接入池化层和全连接层，最后进入 Softmax 层。综上，加入 Attention Model 的网络模型构造如图4.4所示。

4.2 实验

为了检验加入 Attention Model 的网络性能，我们进行了实验。实验任务是行为识别：对视频里人的行为进行分类，如图4.5所示。

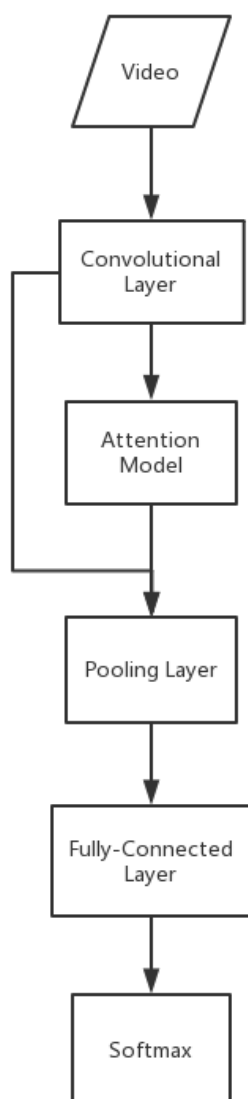


图 4.4 加入Attention Model的CNN-M-2048网络模型

4.2.1 数据准备

实验使用了UCF101数据库，这是中佛罗里达大学研究人员公布的一个行为识别的数据库。近年来行为识别算法大都在该数据库上进行实验检验，是公认的权威数据库。该库总共收集有101类行为，如图4.6所示，这101类可以划分为以下五种类型：

- Human-Object Interaction（人物交互）
- Body-Motion Only（人体交互）
- Human-Human Interaction（人人交互）
- Playing Musical Instruments（演奏乐器）

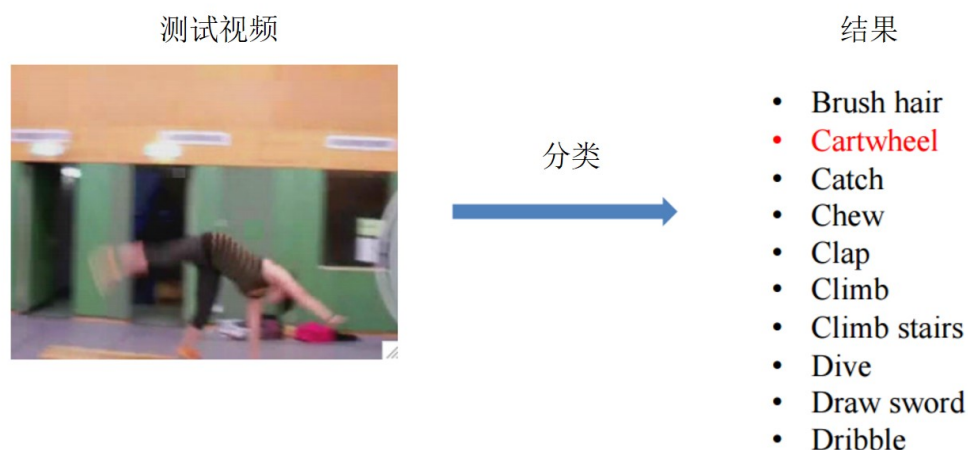


图 4.5 行为识别

• Sports（运动）

图4.7是UCF101数据库视频时长的直方图，从中可以看出该数据库各类视频的分布。UCF101数据库共有3中划分训练集和测试集的方式：split1、split2和split3，本实验是在split1上进行的。在split1中，有9537个训练视频和3783个测试视频。

除了UCF101数据库外，本人还在网上依据这101类行为下载了一些视频，建立了一个辅助数据库，称之为Web数据库。我主要是从YouTube和Flickr这两个网站上根据关键词搜索结果下载视频的，每个类别从YouTube中收集40个视频或从Flickr中收集100个小视频。从网上下载下来的视频通常较长，我对长视频进行了裁剪，最后获得42187个视频片段。图4.8是UCF101和Web数据库一些视频截图，从中可以看出两个数据库的视频分布还是有很大区别的。在这种情况下，前面提到的Attention Model就能发挥其作用，将帧中有益于分类的区域进行增强，训练得到识别能力更强的网络。

4.2.2 实验结果

实验是在Caffe上跑的，图4.9是构建的网络框架图。Caffe是一个开源的深度学习软件包^[61]，在学术界和工业界使用的人很多。为了加快训练速度，训练网络是在GPU上跑的。我们是在CNN_M_2048模型上再训练的，具体训练参数如下所示：

```
test_iter: 151
test_interval: 100
base_lr: 0.0001
```



图 4.6 UCF101数据库

```

lr_policy: "step"
stepsize: 2000
gamma: 0.1
display: 20
max_iter: 10000
momentum: 0.9
weight_decay: 0.0005
snapshot: 1000
iter_size: 1
solver_mode: GPU
    
```

表 4.1 行为识别实验结果

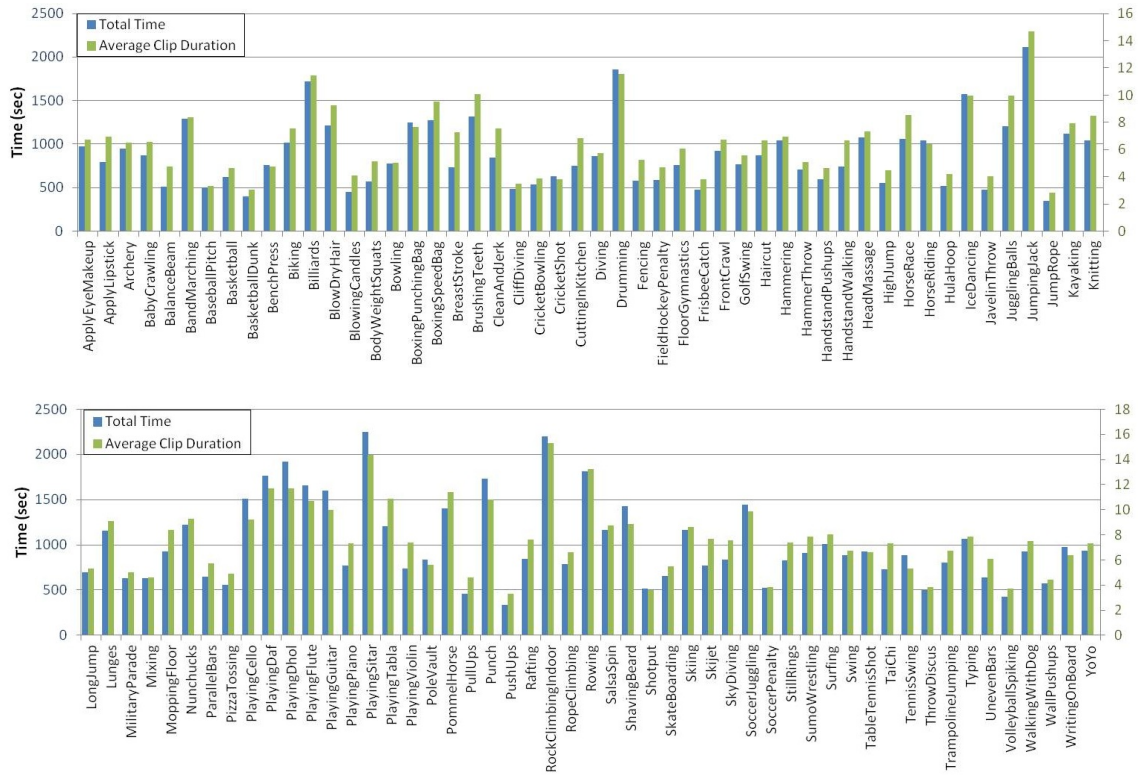
数据库	网络模型	准确率
UCF101	CNN_M.2048	67.12%
UCF101	CNN_M.2048	71.05%
UCF101 + Web	CNN_M.2048 + Attention Model	71.93%

表4.1是实验结果。在只用UCF101数据库时，训练后的CNN_M.2048网络的准确率是67.12%。在加了Web数据库后，准确率达到71.05%。而对网络进行了改进，加入Attention Model后，准确率升到了71.93%，相比于改进前升了将近1%。测试视频总共有3783个，说明识别正确的视频个数增加了30多个。从实验结果可以看出权重机制对于分类效果有提升，由此证明了Attention Model网络的性能优越性。

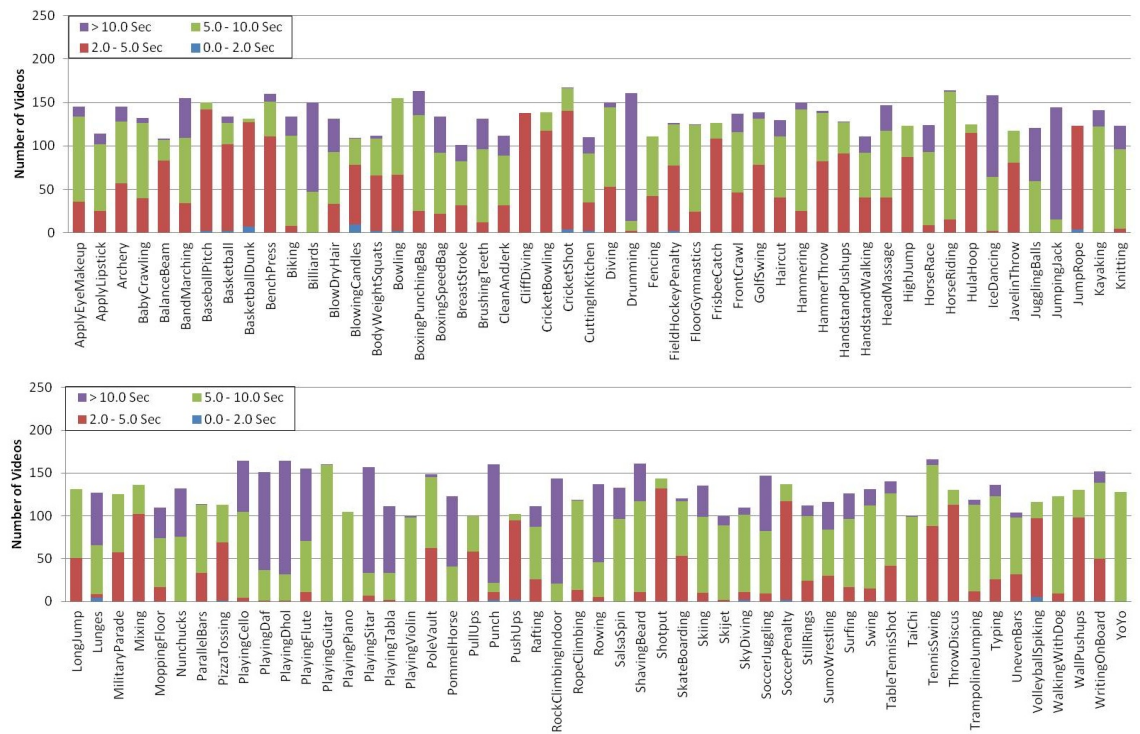
4.3 本章小结

在本章中，本文主要介绍了基于深度神经网络的域适应算法。首先介绍了双流深度卷积网络，这是目前行为识别领域中深度学习的主流框架。接着介绍了我提出的权重机制，即Attention Model网络，其主要目的是通过赋予权重将图片中有益于分类的区域进行增强，从而提高网络分类性能。最后我们进行了实验，实验任务是行为识别，使用的是UCF101数据库和其对应的辅助数据库，结果证明了我们提出的Attention Model网络在分类效果上有提高。

第 4 章 基于深度神经网络的域适应算法



(a) UCF101数据库视频平均时长和总时长



(b) UCF101数据库视频时长分布

图 4.7 UCF101数据库视频时长统计



图 4.8 UCF101和Web数据库示例

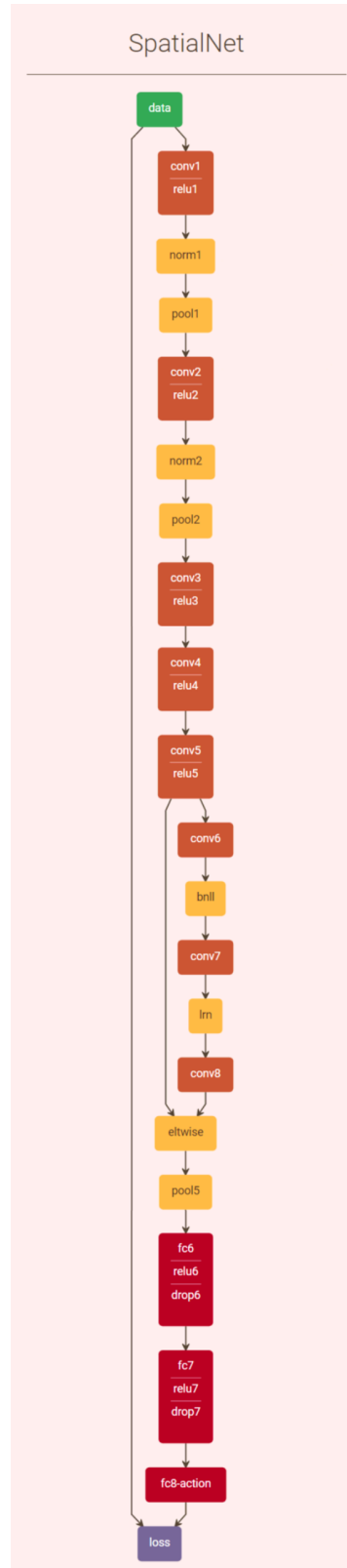


图 4.9 Attention Model Caffe实现框架

第5章 总结与展望

5.1 总结

传统机器学习中，训练集和测试集数据分布是相同的。然而在实际应用中，由于数据本身的变化以及待测试集没有标签训练样本等原因，测试集的数据分布和训练集很有可能是不同的。域适应主要解决的是源域和目标域数据特征分布不同的问题，在很多领域都有应用。域适应分为两种情况：无监督域适应和半监督域适应。本人对这两种应用情况都进行了相应研究。

在无监督域适应中，目标域样本没有标签信息。在这种应用情况，本文提出了基于加权子空间生成的无监督域适应算法。该算法主要包含两个部分：子空间生成和子空间对齐。在子空间生成这一步骤中，首先计算权重，目的是为了衡量每个源域样本的重要性。如果一个源域样本的分布越接近目标域，则它越有利于训练有域适应能力的分类器，相应的权重越高。因此权重计算是以源域数据特征到目标域数据特征的距离为依据的。获得源域样本权重后，我们实行加权PCA得到源域的加权子空间，目标域实行PCA得到子空间。接下来进行子空间对齐操作，目的是得到一个转换矩阵将源域子空间的基进行转换，使转换后的基尽可能地接近目标域子空间的基。借助已有的数学定理，我们可以得到该转换矩阵的闭式解，计算简便快捷，且有理论依据。在后来的工作中，本人还对权重计算进行了改进，使通过计算式得到的权重取值更合理。最后做了实验对加权子空间对齐算法进行了验证，实验任务是图片分类，实验结果证明了该算法在性能上的优越性。

在半监督域适应中，目标域样本有标签信息。在这种应用情况，我主要是研究基于深度神经网络的方法，应用场景是行为识别。在行为识别领域，目前性能最好的模型是双流卷积网络。本人在现有的双流卷积网络中加入了权重机制（Attention Model），目的是为了将帧中有益于分类的区域进行增强，从而使训练得到的网络具有域适应的分类能力。接下来进行了实验，实验任务是行为识别，我针对现有的一个数据库建立了一个辅助数据库，实验结果显示Attention Model网络在分类结果上有提高。

本文针对不同的应用场景提出了两个算法，分别进行了实验，结果均证明了算法在性能上有提高。

5.2 展望

最近几年，深度学习在计算机视觉领域的研究进行地如火如荼，在域适应领域也有越来越多的人利用深度学习的方法进行研究。现有的工作已经证明了深度神经网络在提取特征上的有效性，如果能提出好的特征，那么接下来的域适应操作会容易很多。虽然我在半监督域适应中进行了一些研究，但基于深度学习的方法还有很多地方可以研究，也是未来的研究方向。

参考文献

- [1] Pan S, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10):1345–1359.
- [2] Shimodaira H. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of statistical planning and inference*, 2000, 90(2):227–244.
- [3] Torralba A, Efros A. Unbiased look at dataset bias. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. 1521–1528.
- [4] Khosla A, Zhou T, Malisiewicz T, et al. Undoing the damage of dataset bias. *Computer Vision–ECCV 2012*, 2012. 158–171.
- [5] Gross R, Matthews I, Cohn J, et al. Multi-pie. *Image and Vision Computing*, 2010, 28(5):807–813.
- [6] Wang X, Tang X. Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(11):1955–1967.
- [7] Huang G B, Ramesh M, Berg T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst, 2007.
- [8] Japkowicz N, Stephen S. The class imbalance problem: A systematic study. *Intelligent data analysis*, 2002, 6(5):429–449.
- [9] Heckman J J. Sample selection bias as a specification error (with an application to the estimation of labor supply functions), 1977.
- [10] Zadrozny B. Learning and evaluating classifiers under sample selection bias. *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004. 114.
- [11] Caruana R. Multitask learning. *Learning to learn*. Springer, 1998: 95–133.
- [12] Raina R, Battle A, Lee H, et al. Self-taught learning: transfer learning from unlabeled data. *Proceedings of the 24th international conference on Machine learning*. ACM, 2007. 759–766.
- [13] Chapelle O, Scholkopf B, Zien A. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, 2009, 20(3):542–542.
- [14] Sharma A, Kumar A, Daume H, et al. Generalized multiview analysis: A discriminative latent space. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012. 2160–2167.
- [15] Jiang J. Domain adaptation in natural language processing. University of Illinois at Urbana-Champaign, 2008.
- [16] Patel V, Gopalan R, Li R, et al. Visual domain adaptation: A survey of recent advances. *IEEE, Signal Processing Magazine*, 2015, 32(3):53–69.
- [17] Pardoe D, Stone P. Boosting for regression transfer. *Proceedings of the 27th international conference on Machine learning (ICML-10)*, 2010. 863–870.

-
- [18] Donahue J, Hoffman J, Rodner E, et al. Semi-supervised domain adaptation with instance constraints. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013. 668–675.
- [19] Yao T, Pan Y, Ngo C W, et al. Semi-supervised domain adaptation with subspace learning for visual recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015. 2142–2150.
- [20] Gong B, Shi Y, Sha F, et al. Geodesic flow kernel for unsupervised domain adaptation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012. 2066–2073.
- [21] Long M, Wang J, Ding G, et al. Transfer joint matching for unsupervised domain adaptation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 1410–1417.
- [22] Aljundi R, Emonet R, Muselet D, et al. Landmarks-based kernelized subspace alignment for unsupervised domain adaptation. *Proc. CVPR*, 2015. 56–63.
- [23] Fernando B, Habrard A, Sebban M, et al. Unsupervised visual domain adaptation using subspace alignment. *IEEE Conference on Computer Vision (ICCV)*. IEEE, 2013. 2960–2967.
- [24] Ni J, Qiu Q, Chellappa R. Subspace interpolation via dictionary learning for unsupervised domain adaptation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 692–699.
- [25] Fernando B, Tommasi T, Tuytelaars T. Joint cross-domain classification and subspace learning for unsupervised adaptation. *Pattern Recognition Letters*, 2015, 65:60–66.
- [26] Long M, Wang J, Ding G, et al. Transfer feature learning with joint distribution adaptation. *Proceedings of the IEEE International Conference on Computer Vision*, 2013. 2200–2207.
- [27] Gopalan R, Li R, Chellappa R. Domain adaptation for object recognition: An unsupervised approach. *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2011. 999–1006.
- [28] Gong B, Grauman K, Sha F. Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation. *Proc. ICML*, 2013. 222–230.
- [29] Fernando B, Habrard A, Sebban M, et al. Subspace alignment for domain adaptation. *arXiv preprint arXiv:1409.5241*, 2014..
- [30] Jiang J. A literature survey on domain adaptation of statistical classifiers. URL: <http://sifaka.cs.uiuc.edu/jiang4/domainadaptation/survey>, 2008, 3.
- [31] Daumé III H. Frustratingly easy domain adaptation. *ACL 2007*, 2007. 256–263.
- [32] Saenko K, Kulis B, Fritz M, et al. Adapting visual category models to new domains. *European Conference on Computer Vision*. Springer, 2010: 213–226.
- [33] Kulis B, Saenko K, Darrell T. What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011. 1785–1792.
- [34] Jhuo I H, Liu D, Lee D, et al. Robust visual domain adaptation with low-rank reconstruction. *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on. IEEE, 2012. 2168–2175.
- [35] Yang J, Yan R, Hauptmann A G. Cross-domain video concept detection using adaptive svms. *Proceedings of the 15th ACM international conference on Multimedia*. ACM, 2007. 188–197.

- [36] Aytar Y, Zisserman A. Tabula rasa: Model transfer for object category detection. *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011. 2252–2259.
- [37] Jiang W, Zavesky E, Chang S F, et al. Cross-domain learning methods for high-level visual concept classification. *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008. 161–164.
- [38] Duan L, Tsang I W, Xu D, et al. Domain transfer svm for video concept detection. *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009. 1375–1381.
- [39] Bergamo A, Torresani L. Exploiting weakly-labeled web images to improve object classification: a domain adaptation approach. *Advances in Neural Information Processing Systems*, 2010. 181–189.
- [40] Bruzzone L, Marconcini M. Domain adaptation problems: A dasvm classification technique and a circular validation strategy. *IEEE transactions on pattern analysis and machine intelligence*, 2010, 32(5):770–787.
- [41] Duan L, Xu D, Tsang I W H. Domain adaptation from multiple sources: A domain-dependent regularization approach. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, 23(3):504–518.
- [42] Ma A J, Yuen P C, Li J. Domain transfer support vector ranking for person re-identification without target camera label information. *Proceedings of the IEEE International Conference on Computer Vision*, 2013. 3567–3574.
- [43] Duan L, Xu D, Chang S F. Exploiting web images for event recognition in consumer videos: A multiple source domain adaptation approach. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012. 1338–1345.
- [44] Qiu Q, Patel V M, Turaga P, et al. Domain adaptive dictionary learning. *European Conference on Computer Vision*. Springer, 2012. 631–645.
- [45] Shekhar S, Patel V M, Nguyen H V, et al. Generalized domain-adaptive dictionaries. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013. 361–368.
- [46] Chen S, Zhou F, Liao Q. Visual domain adaptation using weighted subspace alignment. *Visual Communications and Image Processing (VCIP)*, 2016. IEEE, 2016. 1–4.
- [47] Griffin G, Holub A, Perona P. Caltech-256 object category dataset. 2007..
- [48] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features. *European Conference on Computer Vision*. Springer, 2006: 404–417.
- [49] Gong B, Grauman K, Sha F. Learning kernels for unsupervised domain adaptation with applications to visual object recognition. *International Journal of Computer Vision*, 2014, 109(1-2):3–27.
- [50] Vedaldi A, Fulkerson B. VLFeat - an open and portable library of computer vision algorithms. *ACM International Conference on Multimedia*, 2010.
- [51] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 2012. 1097–1105.
- [52] Long M, Zhu H, Wang J, et al. Unsupervised domain adaptation with residual transfer networks. *Advances in Neural Information Processing Systems*, 2016. 136–144.

-
- [53] Hoffman J, Wang D, Yu F, et al. Fcns in the wild: Pixel-level adversarial and constraint-based adaptation. arXiv preprint arXiv:1612.02649, 2016..
 - [54] Bousmalis K, Silberman N, Dohan D, et al. Unsupervised pixel-level domain adaptation with generative adversarial networks. arXiv preprint arXiv:1612.05424, 2016..
 - [55] Ge W, Yu Y. Borrowing treasures from the wealthy: Deep transfer learning through selective joint fine-tuning. arXiv preprint arXiv:1702.08690, 2017..
 - [56] Karpathy A, Toderici G, Shetty S, et al. Large-scale video classification with convolutional neural networks. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2014. 1725–1732.
 - [57] Peng X, Wang L, Wang X, et al. Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice. Computer Vision and Image Understanding, 2016, 150:109–125.
 - [58] Wang H, Schmid C. Action recognition with improved trajectories. Proceedings of the IEEE International Conference on Computer Vision, 2013. 3551–3558.
 - [59] Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos. Advances in neural information processing systems, 2014. 568–576.
 - [60] Chatfield K, Simonyan K, Vedaldi A, et al. Return of the devil in the details: Delving deep into convolutional nets. British Machine Vision Conference, 2014.
 - [61] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding. Proceedings of the 22Nd ACM International Conference on Multimedia, New York, NY, USA: ACM, 2014. 675–678.

致 谢

首先感谢我的导师廖庆敏教授。在我的硕士研究生期间，廖老师给了我充分的学术自由，让我可以根据自己的兴趣和自身情况选择研究方向。每隔一段时间，廖老师就会跟我沟通，了解我的科研进度。廖老师性格沉稳，待人随和，是我学习的榜样。

感谢周飞博士对我的悉心指导和鼓励，我经常与周老师进行学术上的交流讨论，每次讨论都能从周老师那里得到科研上的灵感和启发。周老师作为一名奋斗在科研一线的年轻学者，其学术态度严谨，思维活跃，能从多种角度看待一个问题。

感谢杨文明老师在组会报告上对我科研工作上的指导和建议，感谢卢宗庆老师在我实习期间对我的帮助和照顾。特别感谢实验室的师兄师姐、同学及师弟师妹们，能和他们一起学习生活是我的荣幸。

最后，但最重要的是，在深圳求学的三年间，感谢在深工作的父母平日在物质和精神上对我的支持及鼓励。我会一如既往地奋斗，他们的期望是我不断向上的动力源泉。

声 明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名：_____ 日 期：_____

个人简历、在学期间发表的学术论文与研究成果

个人简历

1992年11月出生于安徽省淮北市。

2010年9月考入南京邮电大学物联网学院网络工程（物联网与传感网技术）专业，2014年6月本科毕业并获得工学学士学位。

2014年9月考取研究生进入清华大学电子工程系攻读工程硕士学位至今。

发表的学术论文

- [1] S. Chen, F. Zhou and Q. Liao, "Visual domain adaptation using weighted subspace alignment," 2016 Visual Communications and Image Processing (VCIP), Chengdu, 2016, pp. 1-4. (EI收录, 检索号: 20170503316920)