

密级:



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

硕士学位论文

基于迁移学习的小样本光学遥感目标检测

作者姓名: 黄显淞

指导教师: 叶齐祥 教授 中国科学院大学

学位类别: 工程硕士

学科专业: 工业工程

研究所 : 中国科学院大学 工程科学学院

2017 年 05 月

**Optical Remote-Sensing Object Detection of Small
Samples Based On Transfer Learning**

By

Huang Xian Song

A Thesis Submitted to

University of Chinese Academy of Sciences

In partial fulfillment of the requirement

For the degree of

Master of Industrial Engineering

School of Engineering Science

University of Chinese Academy of Sciences

May, 2017

中国科学院大学直属院系 研究生学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的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，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对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或致谢。

作者签名：黄星松
日期：2017.5.23

中国科学院大学直属院系 学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。涉密的学位论文在解密后适用本声明。

作者签名：黄星松
日期：2017.5.23

导师签名：叶新峰
日期：2017.5.23

摘 要

光学遥感目标检测是利用计算机视觉算法检测遥感影像中特定的目标，是遥感领域中重要的应用问题之一。在军事战略侦查、智能交通、遥感图像解译和无人机应用领域中有着广阔的应用前景。近年来，大多数的遥感目标检测中采用的都是监督学习方法，通过大量标注样本信息，训练出能充分表示训练样本分布的模型，对测试样本具有良好的泛化能力。但就实际应用而言，许多特定的目标在遥感影像中只能获取到一个极小的样本数据集。由于小样本数据集只能描述出目标样本整体分布中的一小块区域，所以，监督学习在这个场景下学习到的模型并不能泛化到未出现的数据。本文就此问题提出了基于迁移学习的小样本目标检测框架。域适应的迁移学习作为近年来目标检测中不同数据域之间模型适应的新兴研究方向，其最大的特点是能在训练目标域模型时，从源域数据中获得必要的先验知识，使学习到的目标域模型具有更好的泛化能力，从而提高检测性能。本文首先在源域数据集中建立先验模型，将先验模型与目标域小样本数据集结合，通过域适应算法获得目标域模型。在本文的高清车辆数据集上，建立虚拟合成的数据集作为源域，最终训练得到的模型在检测性能上接近大样本中学习到的模型。

本文的主要贡献如下：

- 1) 提出了一种基于相似性的合成虚拟样本集方法。针对遥感中的舰船、飞机、车辆等人造目标，通过分析其中的样本分布特性，利用三维模型对目标建模，生成与目标域样本近似的虚拟数据集。
- 2) 提出了一种在 DPM 模型中混合模型划分的方法。利用特征可视化工具分析样本集在特征上的分布，将样本在特征空间中进行聚类，使用聚类后的样本训练混合模型，能够提高模型在检测上的泛化能力。
- 3) 提出了一种小样本数据集的检测框架。以遥感小样本数据为基础，结合虚拟数据集生成、源域建模以及模型域适应方法，获得具有良好检测性能的目标域模型。

关键词：小样本学习，域适应算法，遥感图像，目标检测，虚拟数据集

Abstract

Object detection in optical images is one of the fundamental problems in the field of remote sensing. It has potential capabilities in applicable prospects of military strategic investigation, intelligent transportation, remote sensing image interpretation and UAV applications. In recent years, supervised learning methods have been proposed in most of the remote sensing object recognition task. Supervised learning method is a classical algorithm in machine learning, which can learn models to represent the distribution of training samples by a large number of samples. However, in many remote sensing detection tasks, only a few training samples can be acquired. In general, small training set do not have the ability to describe the distribution of the overall sample domain. In this paper, we propose a transfer learning based detection framework which can be used in small training se. Domain Adaptation (DA) is getting more and more attention to deal with learning task between different data domain in transfer learning field. The most advantage of DA is the capabilities to train a better target model with prior knowledge from source domain rather than only use target sample, and make the model better in generation which can improve the detection performance. In this paper, we first establish the prior model from source domain, and then train a target model combined a small training set of target domain using the DA method. By applying the framework in SDL vehicle dataset with small training set, we successfully train a model which has high performance as better as the training set with abundant samples.

Our contributions are as follows:

- 1) A method of synthesizing virtual data sets has been proposed. Aiming at the artificial targets such as ships, aircrafts and vehicles in remote sensing images, we analyzes the feature distribution of samples, and uses the 3D model to simulate samples. And finally generates the virtual data set which is similar to the target sample sets.
- 2) A method of mixed model partition in DPM model training. By analysing feature distribution of the sample using visualization tools, we cluster samples

in the feature space, and use cluster to train a mixture model. This improves the generalization ability of the model.

- 3) A detection framework for small training set. Combined with virtual dataset, we build the model of small set by domain adaptation, which has similar performance as the full training set.

Key Words: learning with small sample, domain adaptation, remote sensing images, object detection, synthesize data

目录

摘 要.....	I
Abstract.....	III
图目录.....	VII
表目录.....	IX
第 1 章 绪论	1
1.1 课题背景和研究意义.....	1
1.2 国内外研究现状.....	3
1.2.1 候选区域提取.....	3
1.2.2 目标表示.....	4
1.2.3 分类算法.....	6
1.2.4 迁移学习.....	6
1.3 本文研究内容.....	7
1.4 本文的组织结构.....	8
第 2 章 相关技术研究	11
2.1 图像特征提取.....	11
2.1.1 HOG 特征.....	11
2.1.2 SIFT 特征.....	13
2.1.3 卷积神经网络.....	14
2.2 目标表示方法.....	15
2.2.1 基于样本特征表示.....	16
2.2.2 基于样本结构表示.....	16
2.3 隐变量 SVM.....	22
2.4 域适应的 SVM.....	23
2.5 本章小结.....	25
第 3 章 基于相似性判定的样本生成研究	27
3.1 问题描述及难点分析.....	27
3.2 目标数据集.....	28
3.3 基于统计分析的样本合成.....	29
3.4 合成样本集的相似性验证.....	31
3.4.1 基于特征的相似性验证.....	31
3.4.2 基于检测性能的相似性验证.....	32
3.5 本章小结.....	33

第 4 章 域适应的小样本目标检测研究	35
4.1 问题描述及难点分析	35
4.2 模型训练	37
4.2.1 目标聚类分析	37
4.2.2 混合模型训练	38
4.3 模型域适应	41
4.4 实验结果与分析	44
4.4.1 实验方法	44
4.4.2 混合模型数目与性能分析	45
4.4.3 域适应模型实验结果与分析	45
4.5 本章小结	46
第 5 章 结论与展望	49
参 考 文 献	51
致 谢	55
个人简历、在学期间发表的论文与研究成果	57

图目录

图 1-1 遥感检测应用	1
图 1-2 HOG、DPM 与 CNN 示意图	5
图 1-3 迁移学习基本流程	7
图 1-4 本文研究框架	8
图 2-1 HOG 特征提取示意图	12
图 2-2 神经元结构	14
图 2-3 具有隐藏层的神经网络	14
图 2-4 卷积神经网络工作流程图 ^[3]	15
图 2-5 行人的 DPM 模型表示 ^[30]	20
图 2-6 Adaptive SVM 分类平面	24
图 3-1 样本合成流程	28
图 3-2 数据集标注示意	29
图 3-3 样本合成示意	31
图 3-4 多目标特征可视化聚类分析	32
图 3-5 小样本与合成样本检测曲线对比	33
图 4-1 目标检测通用流程	35
图 4-2 小样本训练流程	36
图 4-3 车辆样本不同角度特征聚类	38
图 4-4 DPM 模型参数结构化域适应	42
图 4-5 域适应后的模型性能对比	46

表目录

表 3-1 中科院大学高清航拍数据集	28
表 3-2 角度分布直方图	30
表 3-3 长宽比分布直方图	30
表 3-4 像素和分布直方图	30
表 4-1 模型训练过程	40
表 4-2 模型域适应算法	44
表 4-3 不同混合模型下 mAP 与时间消耗	45

第1章 绪论

1.1 课题背景和研究意义

近年来，我国高分辨率遥感卫星、载人航天以及探月工程的成功，推动了国内遥感技术的进展，也使得遥感技术成为未来空间科技发展的重要技术之一。光学遥感影像是从离开地面的平台上通过光学成像技术获得的对地俯视图。随着无人机航拍和遥感覆盖范围、扫描频率、设备分辨率、存储媒介的不断发展，遥感影像的数据规模急剧增大。如何有效利用如此大规模的数据来发现有意义的军事民用目标是一个具有显著意义的问题。

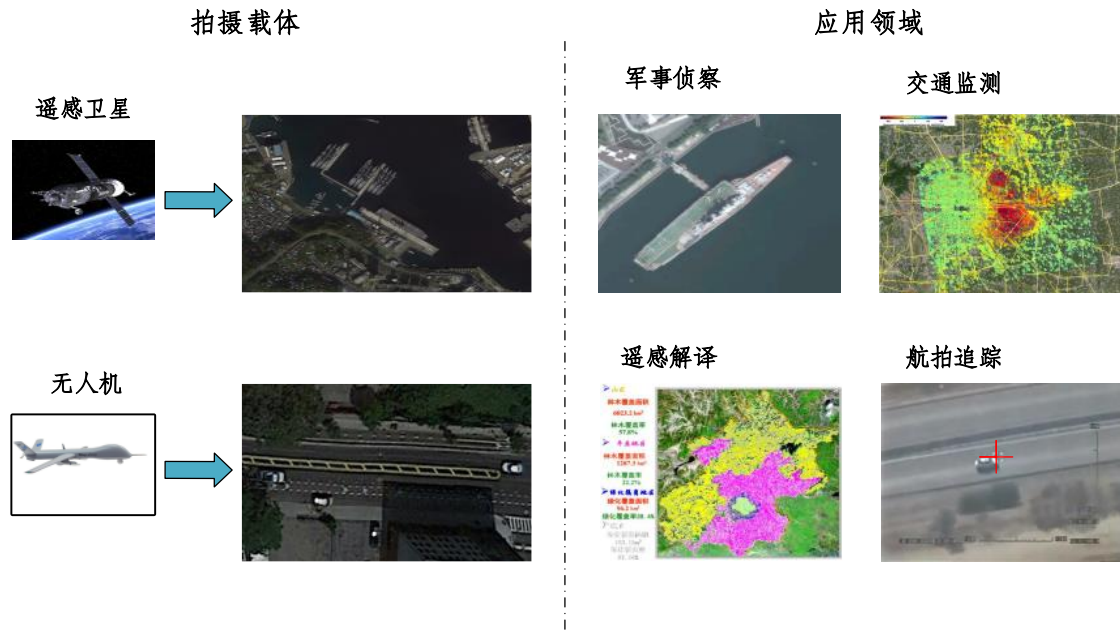


图 1-1 遥感检测应用

目标检测是通过机器学习方法与计算机视觉算法对图像、视频中的一类或多类物体进行精确定位或识别的过程。目标检测是计算机视觉中的高级任务，通常作为活动分析、事件识别、场景理解和目标追踪的研究基础^[1]。在遥感应用中，遥感目标检测处于关键的地位：（1）在军事战略侦查领域，快速精确地定位关键区域的飞机、舰船等目标能为整个作战信息系统提供核心数据；（2）在智能交通领域，通过航拍检测道路车辆等交通情况，能够进行合理的交通疏导及管制；（3）在遥感解译领域，通过目标检测可以对地面物体，如桥梁、作物、道路等，进行

分类；(4) 在无人机系统中，通过目标检测技术可以实时观测地面目标并进行跟踪。

过去的十年是目标检测领域发展的黄金时期，涌现出了诸多标志性的检测算法，并在实际项目中取得重大意义：人脸，车牌等检测方法已经走向实际应用；深度学习在超大规模的图像数据集 ImageNet 上的识别准确率也超越人类^{[2][3]}。在此基础上，许多研究机构和学者对现有检测算法作了大量适应性研究工作，使众多目标检测方法成功地应用到航拍遥感领域分析中，推进了航拍遥感事业的发展。但是，由于航拍遥感影像的各类目标有着与常规图像目标完全不同的姿态以及应用场景限制，因此仍存在很多挑战性的难题。

光学遥感影像目标检测中，作为建模基础的目标样本获取是一个显著的问题，也是所有学习算法的基础。

目标检测中的模型建立方法大多依赖于监督学习，建模过程中需要大量标注的训练样本以使得算法尽可能地逼近目标概念，获取具有良好泛化能力的检测模型。但在航拍与遥感应用中，对大规模数据进行人工标注通常是非常费时费力的。遥感影像包含地区的各时间段的数据，首先要估计目标可能在哪一个时间段和区域出现，再从杂乱的背景中寻找标注样本。以上的标注都带来了人工的消耗，在人力成本越来越贵的时间，想要得到大规模的标注数据是几乎不可能的。

为解决这个问题，有学者提出用弱监督学习的方法来进行半自动标注工作来代替人工标注^{[4][5][6][7]}。使用少量标注的样本，建立假设的目标模型，通过迭代的方式从未标注的图像集中挖掘样本，并将之扩充到已知数据中，作为下一次迭代的建模训练样本。通过这样的方法，能够在一定程度上解决人工标注的问题。但仍存在缺陷，即弱监督的方法是建立在未标注数据中有充足样本的情况下才能够正常工作的。当处于难以估计未标注数据中的目标样本数目情况下，弱监督的方法并不能确保解决问题。在不能利用未标注数据信息的前提下，我们需要面对的就是小样本目标检测问题。

因此，如果能够在小样本情况下提出一个光学遥感目标检测的高效率框架，将会进一步扩大遥感目标检测的应用范围。小样本数据集往往难以描述整个目标样本数据分布情况，如何从小样本中训练一个好的检测模型变得更加有挑战性。虽然国内外已有诸多学者对这个问题进行了长期的研究，提出不少解决方法并以此发表诸多论文，但小样本的目标检测算法仍然难以应用于实际问题中^[8]。

本文的目的就是在小样本条件下，通过建立一个辅助的合成虚拟数据集，利用模型域适应的方法，建立一个能用于实际样本的检测框架，解决遥感目标检测中对小样本建模的应用技术问题。本文的方法是模式识别和机器学习领域的结合，并提供解决快速检测的方案。

1.2 国内外研究现状

遥感目标检测中所采用的方法大多来自于经典目标检测算法，作为经典算法的扩展，也有与之不同的应用特性区别。光学遥感目标检测中主要面临以下几个具有挑战性的问题：a) 多样性的目标种类。随着对地观测技术的发展，遥感图像能够提供的纹理和细节信息越来越多，使得辨别更多的地物目标成为可能，从建筑、作物、道路、舰船、飞机以及车辆都成为识别的目标。b) 目标在尺度、角度的多变性。与通常直立视角不同，俯视视角下的目标具有任意的旋转角度，并且根据目标形状大小不同以及成像高度的差异，在尺度上也有极大变化。c) 复杂的背景因素。遥感成像极易受到天气、设备、光照、噪声等干扰，以及目标应用场景的背景差异大，使遥感影像中背景纷乱繁杂，给检测模型的训练和使用带来了极大的难度。d) 数据域差异。遥感影像因成像分辨率、设备和成像角度不同，多个来源的相同目标并不能视为相同数据域上的数据，使训练数据和应用场景数据在目标分布上有较大差异。针对这些问题，下面从经典目标检测以及遥感中的候选区域提取、目标表示、分类算法以及迁移学习等方面来探讨国内外的研究进展。

1.2.1 候选区域提取

候选区域提取是指在待检测的图像中产生一系列区域，作为有可能的目标定位区域。在检测中，会将这些候选区域投入目标分类器中，进而判断该区域是否为所需要检测的目标。

传统的候选区域提取方法是通过在多尺度图像特征金字塔上逐像素扫描窗口，并用分类器对每个窗口的特征向量进行分类，确定其是目标物体还是背景。在扫窗框架中，为了减少每个窗口计算特征所需要的时间，需要首先在图像金字塔上提取特征构建特征金字塔，然后在图像金字塔每层上逐像素扫描窗口并从其在特征金字塔上对应的位置提取特征。该方法遍历图像中的所有位置和尺度，能够精确的定位出目标真正所在的位置区域。但是会产生大量窗口，对于一幅标

清的图像，构建一个四阶金字塔，每阶八层，共会产生百万级别的窗口。如此多的窗口，严重影响了目标分类的速度，不利于更复杂表示方式的应用。

因此，近几年来涌现而出了更多的优秀候选区域提取方法。Uijlings^[9]等人按照一定的准则将分割获得的小像素区域进行分层组合，提出了基于图像分割的 Selective Search 方法。程明明等人根据认知科学中人类对图像小区域信息聚焦的特点，提出基于区域梯度特征的 BING^[10]方法。Piotr 等人认为候选窗口一定要满足其内部有很多完整轮廓，穿越窗口的轮廓应该大部分位于窗口内部，并基于轮廓预判方式提出了 Edge Boxes^[11]方法。这些方法能极大的减少候选窗口的数目，得到了越来越多的关注。

1.2.2 目标表示

目标特征是从原始图像中经过一定处理后获得的向量式表示，一个具备良好判别性的特征往往能给检测带来更高的性能，因此，如何获取判别性强的特征是目标检测领域的一个重要研究方向。

在早期的目标检测中，特征的提取主要在对目标纹理进行，Jain 提出了 Gabor 特征^[12]，这是一种典型的纹理特征，参考人类视觉系统中对局部空间方向和频率特点，能很好的表示纹理信息，并用于分类。通过一个 Gabor 滤波器作用于图像，提取 x 和 y 方向上最强的频域响应组成特征，其特点是对图像的边缘敏感，并具有一定的尺度不变性。

LBP^[13]是 Ojala 提出的一个简单而有效的纹理描述。通过将中心像素与周围像素的比较，得到中心像素点的 LBP 值，用该值来描述此像素局部区域的纹理特性。在此基础上发展的一些圆形以及等价 LBP 算子，用直方图的方式表示后，具有良好的旋转不变性和灰度不变性的特点

SIFT (Scale-invariant feature transform) ^[14]也是一种局部特征，通过求图像中的特征点以及和尺度金字塔、方向相关的描述子得到特征。SIFT 特征具有对尺度和光照具有不变性的特点，在光照变化，视角变化中也具有一定的稳定性。（加应用示例以及缺陷分析）

Papageorigiou 提出了基于图像上多个形状大小形同的矩形区域像素灰度之差的 Haar-like 特征^[15]，并用于人脸特征表示上。Biola 和 Jones 在此基础上，使用了边缘特征，线性特征，中心和对角线特征来进一步表示人脸。Haar 特征能很好的反映图像的灰度变化趋势。通过改变三种特征模板的大小和位置，可

以在图像子窗口中穷举出大量的特征。

Dalal 在 2005 年提出了 HOG (Histogram Of Gradient) 特征^[16]，其特点是将图像局部出现的梯度信息统计为直方图，能够很好的描述目标的形状边缘。通过将图像分解为小块的区域 Cell，在单独统计区域 Cell 的梯度直方图基础上，将相邻的 cell 组成一个块，块内部的 cell 具有重叠区域，最终将各区域的直方图信息归一化后串联为 HOG 特征。HOG 特征及其改进版本 (C-HOG^[17]、R-HOG^[18]) 在行人检测上得到了很广泛的应用。

上面这些特征都属于手工设计的特征，能够在某一方面很好的描述目标。但随着目标的类别增多，以及背景的复杂变化，利用手工设计的特征去遍历匹配多种目标是难以实现的，基于学习的自动特征提取逐渐在目标检测领域兴起。

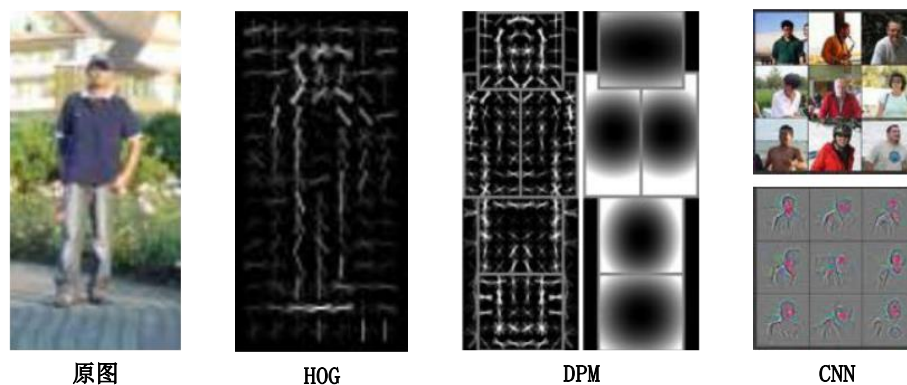


图 1-2 HOG、DPM 与 CNN 示意图

卷积神经网络 (CNN) 是近年来兴起的学习型特征提取方法^{[31][20]}。卷积神经网络是从神经网络发展而来的，克服了神经网络的反向传播算法造成的梯度扩散，通过卷积层与 pooling 层的多次连接，能从样本的底层特征中提取局部特征，并利用特征映射得到高层特征表示。深度网络提取的特征能对多种目标具有很好的判别能力，并在平移、比例缩放、倾斜或者其他形式的变形上具有高度不变性。然而，卷积神经网络的内部权值形式是将输入的样本强行记住，对于多角度的遥感目标并不能很好的处理。在这个缺陷的研究上，zhu^[21]在分析了卷积神经网络在处理航拍目标时的各层输出，提出了利用卷积神经网络的中间层的 FC5 作为输出特征，使其具有一定程度的旋转不变性。Zhou^[22]则是在神经网络的输入部分改变滤波器的模板，采用多角度滤波器显式编码方向信息，产生带有角度通道的特征图，得到角度响应网络，在多方向卷积过程中主动获取类内角度不变特征，在反向传播过程中利用误差更新滤波器，获得旋转不变性的网络，同时维持

分类任务中不同类别之间的差异。

1.2.3 分类算法

在目标检测中，分类器或者分类算法的任务是将输入的图像或特征向量划分到某一个类别。具有较强分类泛化能力的分类算法能够从训练集中有效的学习其内在特性，并在检测中具有更好的性能表现，因此分类器设计以及优化算法一直都是目标检测领域中关注的热点。

在分类器设计上，主要可以分为概率生成模型和判别模型两类^{[23][24]}。概率生成模型的基本想法是建立样本的联合概率密度模型 $P(X, Y)$ ，然后计算后验概率 $P(Y|X)$ ，根据后验概率来对目标进行分类。其代表性的算法主要有贝叶斯模型、隐马尔可夫模型、马尔科夫随机场和 Fisher 判别等。虽然生成模型一般具有比较好的性能，但在建模的过程中，需要额外学习一些与目标类别无关的隐信息，会使得计算量更大，所以后来的目标检测中更倾向于使用直观的判别模型。

判别模型是从样本或特征中建立一个判别函数，利用判别函数直接输出特征的目标分类。Adaboost^{[25][26][27]}是比较流行的一个判别学习算法，它迭代训练一系列的弱分类器，调整每个分类器的权重，然后将其合并成一个良好的分类器，但如果过分对分类器进行强化，可能导致数据的过拟合。SVM^[28]是使用最广泛的判别学习算法，通过最大化判别分界面和支持向量检测距离实现对目标的分类，不但可以解决线性可分的问题，还能通过核函数实现非线性分类。Felzenszwalb^[29]等人提的 DPM 是目标检测领域中的最优算法之一，通过训练出样本中的局部子滤波器并按照其相对于理想锚点的偏移量进行惩罚，该方法特别适合处理非刚性模型，能有效检测出多姿态变化的目标。在此基础上，Felzenszwalb^[30]进一步加入混合模型来处理多视角问题，并在 PASCAL 行人检测中取得了出色的性能。之后的深度学习也促使了 DPM 的进一步发展，Girshick^[31]将 DPM 框架中的特征提取、形变模型和混合模型均转换成 CNN 结构。这些模型的算法的发展，使视觉目标检测的性能进一步接近人类视觉。

1.2.4 迁移学习

在传统机器学习的框架下，分类器学习的任务就是在给定充分训练数据的基础上学习一个分类器，然后利用学习到的模型对测试样本进行检测。但对于遥感影像而言，由于分辨率、场景、视角的不同，用于训练和测试的样本往往不是

来自同样分布的。这样的训练和测试集分布差异也称为源域和目标域区别，会使分类模型的性能有较大下降。

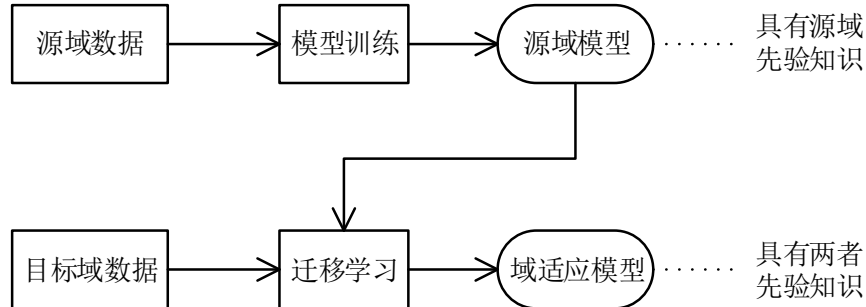


图 1-3 迁移学习基本流程

为了解决这个问题，研究人员主要提出了多种模型适应学习的算法。Kan^[8]从样本特征着手，利用非线性的转换方式，将源域的样本从特征上转换到目标域，通过将问题变为监督学习的方式来解决。更进一步的[33]将非线性的转换方式用流型结构来分析，提供新的样本选择方法。在[34]中，作者采用成份转换分析（transfer component analysis, TCA）的方式去学习特征，进而使得不同域样本在特征分布上的差异进一步减少。在 TCA 的基础上，从域自适应的角度^[34]去设计深度网络结构，从而使深度网络具有很好地表达不同域样本特征的能力。在最近的研究中，Tsai^[36]提出了异构的域自适应（Heterogeneous Domain Adaptation, HAD）方法。其主要思想是虽然源域与目标域的数据具有特征分布差异，但是可以通过特征投影来寻找共同的特征子空间，然后通过共同的特征投影子空间对未标号样本进行识别的方法来解决域自适应问题。在学习算法上，Vazquez 和 Xu^{[37][38]}通过在模型优化空间建立源域和目标域参数平面的约束，从源域模型参数着手去学习适用于目标域模型。Ghifar^[39]等人结合无监督学习，构建了无监督域自适应的深度学习算法。该算法采用多任务学习（multitask learning）^[40]的方式，一方面学习有标号的源域分类任务，另一方面学习无标记的目标域样本重建任务，通过这两个任务的联合学习，能够使得深度网络学习到具有判别能力且包含源域信息的特征。

1.3 本文研究内容

本课题研究的目標是在光学遥感场景下的小样本模型学习，即在给定较少数量的目标样本条件下去学习检测模型，使得性能与大数量的训练样本条件下得到的模型近似。本文采用了模型域适应的方法，首先通过一个与目标样本近似

的辅助数据集训练作为参考的检测器。之后在模型域适应算法中以辅助模型参数为基础，结合目标样本特征，在参数空间中调整分类超平面，重新训练出适用于目标领域的检测器。

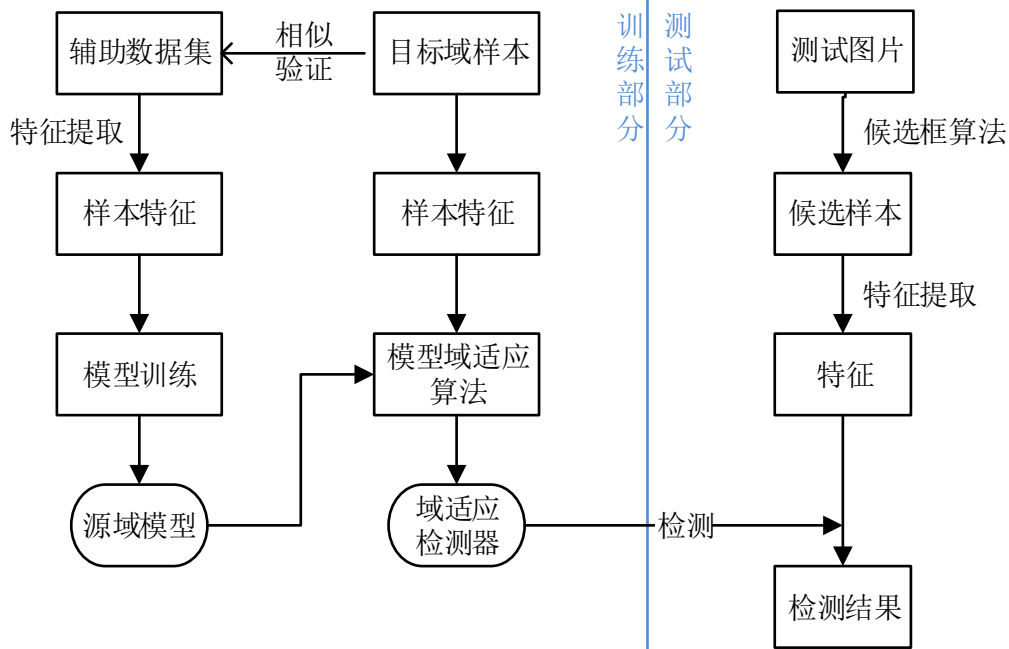


图 1-4 本文研究框架

- 1) 在生成辅助数据集的过程中，本文研究了基于相似性分析目标样本分布的方法。根据样本属性统计，采集了车辆 3D 模型，与可轻松获取的实际遥感背景结合，生成了大量俯视视角下的遥感合成影像。并在合成的样本上做模糊和加噪处理，使之与实际样本近似。
- 2) 在辅助模型训练中，运用目标样本特征聚类的方式，分析目标聚类数目对混合模型训练的影响。根据浅层特征可视化的验证，选取具有最良好性能的聚类方式，训练用于迁移学习的基础模型。
- 3) 针对传统的手工模型检测器，提出一个域适应模型训练算法，在源域样本中训练得到作为基准的先验模型。通过域适应算法在参数空间对分类面进行重训练，得到适用于目标域样本的新检测器。

1.4 本文的组织结构

第一章：绪论。介绍了小样本遥感目标检测的研究意义和研究价值。总览国内外近年来的研究现状和研究方法，同时比较了这些方法的优势与劣势。对本文

基于模型域适应算法的小样本学习目标检测框架进行了简要阐述说明。

第二章，相关工作与技术。介绍了本文使用到的当前目标检测领域中的相关工作与技术。主要包括候选区域选取方法、目标特征介绍、语义模型以及域适应算法。

第三章，基于相似性的样本生成研究。介绍了本文使用的遥感车辆数据集，并利用特征可视化的工具分析实际数据集的特点。根据这类特点采用合成样本的方式合成样本。以生成的数据集作为源域数据集，从特征聚类以及初步模型性能上验证合成样本的相似性。

第四章，基于迁移学习的小样本光学遥感目标检测研究。本章主要研究在 HOG 特征以及 DPM 模型下的域适应算法研究，并展示检测框架的细节。证明了以虚拟样本作为辅助数据集的情况下，域适应得到的模型能够获得与大量目标样本训练的模型相近的性能。

第五章，总结与展望。主要是对本文工作的总结，并探讨一些尚未解决的问题，以及下一步工作的展望。

第2章 相关技术研究

上一章主要介绍了光学遥感影像中目标检测的意义、国内外研究现状和本文的主要研究内容。本章作为上一章节的扩展与补充，将详细讲述研究中所使用到的相关技术和算法，作为后续内容的基础。主要从图像特征提取、目标表示方法、隐变量优化算法、域适应的 SVM 等方面进行介绍。

2.1 图像特征提取

图像特征，可以认为是在原始图像基础上通过一定计算后得到原始像素信息的线性或非线性向量表示。图像特征对于要分类的目标，应当具有良好的判别性和类内聚集性，并使特征向量维数尽可能低，良好的特征往往能给检测带来更高的性能。因此，如何获取判别性强的特征是目标检测领域的一个重要研究方向。近年来许多学者在特征提取算法上做了大量工作，提出许多优秀的特征提取方法，下面将简要介绍几种常用的提取方法。

2.1.1 HOG 特征

方向梯度直方图（Histogram of Oriented Gradient, HOG^[41]）是一种在计算机视觉和图像处理中用来进行物体检测的特征描述子，通过计算和统计图像局部区域的梯度方向直方图来构成特征。HOG 特征结合 SVM^[28]分类器的在目标检测领域是一种很广泛应用的方法，在行人检测中取得了很好的性能。

其基本提取流程如下：

- a) 将输入图像转化为灰度。
- b) 通过 gamma 校正法做图像颜色空间的标准化，调节图像对比度。
- c) 用水平和竖直两个方向的 $[+1, 0, -1]$ 滤波器计算得到每个像素的梯度，并求出大小和方向。
- d) 将图像按照诸如 8×8 像素大小划分为 cell, cell 之间没有相互覆盖的像素区域。
- e) 在每个 cell 中按照给定的方向数目 bin 统计梯度直方图，构成该 cell 的特征描述子，其中梯度的方向指离散化后所在区间，梯度大小为直方图的增加量。
- f) 将相邻的几个 cell 组成一个较大的 block, 如 2×2 或者 4×4 , cell 之

间会有重叠的区域。并将 block 内所有 cell 的特征描述子串联作为该 block 的描述子。

- g) 将图像中所有 block 描述子依次串联起来，归一化后形成图像的 HOG 特征向量。

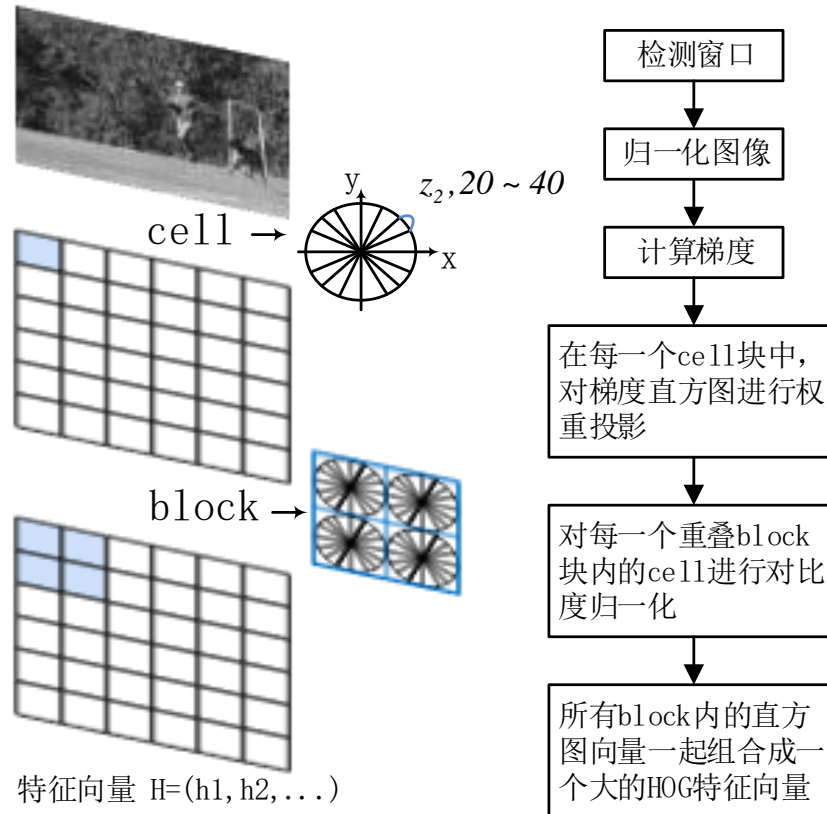


图 2-1 HOG 特征提取示意图

上面是 HOG 一般化的提取流程，但在实际使用中，还会根据应用场景做不同的调整，以提高特征的判别能力。

Dalal 和 Triggs^[41]在使用 HOG 特征做行人检测时，对每个 cell 提取的特征采用了四种不同的归一化方式，将相邻的 cell 特征描述子联系起来。假定 cell 的特征向量的 $C(i, j)$ ，可以将这四种归一化因子表示如下，

$$N_{\delta, \gamma}(i, j) = (\|C(i, j)\|^2 + \|C(i + \delta, j)\|^2 + \|C(i, j + \gamma)\|^2 + \|C(i + \delta, j + \gamma)\|^2)^{1/2} \quad \text{其中 } \delta, \gamma \in \{-1, +1\} \quad (2-1)$$

之后使用一个下限值来截断各个方向中的最小值，假设 $T_{\alpha}(v)$ 为将向量 v 中所有小于 α 的分量都置为 α 。最终得到的 HOG 特征为将上述四种归一化因子处理后的向量经截断后连接起来，即：

$$H(i, j) = \begin{pmatrix} T_{\alpha}(C(i, j)/N_{-1,-1}(i, j)) \\ T_{\alpha}(C(i, j)/N_{+1,-1}(i, j)) \\ T_{\alpha}(C(i, j)/N_{+1,+1}(i, j)) \\ T_{\alpha}(C(i, j)/N_{-1,+1}(i, j)) \end{pmatrix} \quad (2-2)$$

在 P.Fez^[30]的处理中，为了能够检测到尺度差异较大的目标，则在图像特征金字塔的末尾层增加了 2*2 以及 4*4 大小的 cell，并对这些特殊层的偏置值做了校正处理。

作为一种经典的提取方式，HOG 特征能够充分保留图像的梯度信息，对具有固定外形轮廓的目标具有极大优势。但因只提取图像梯度而舍弃了其他信息，HOG 特征的表达能力也因此受到一定限制。

2.1.2 SIFT 特征

尺度不变特征变换 (Scale-invariant feature transform, SIFT^[14]) 是基于图像中局部外观的兴趣点提取特征的算法。该方法于 1999 年由 David Lowe 首先提出，现在已被广泛应用于图像匹配、图像融合和三维重建等应用中。SIFT 具有尺度和旋转不变的特性，而且对光照、视角变化的容忍能力也较强。特征的提取主要由以下几个步骤组成：

- a) 输入图像构建尺度空间。用高斯卷积核实现尺度变换后建立高斯金字塔。由图片的大小决定建立几个塔，以及每个塔中有几层图像。0 塔的第 0 层为原始图像，之后每往上一层都是对下层的拉普拉斯变换。不同塔之间的对应层则是用下采样的方式得到。
- b) 在前面得到的高斯金字塔中，计算高斯差分金字塔。
- c) 寻找特征点。即每一个点都和其所有相邻点做比较，如果该点与同尺度的相邻 8 个点以及上下尺度相应的 9*2 个点相比，是最大或者最小值时，则认为点是图像在该尺度下的特征点。
- d) 选择特征点的主方向。利用关键点领域像素对其进行描述，根据邻域像素点的梯度方向特性计算出关键点的梯度方向和幅值，使算子具备旋转不变性。之后通过对关键点周围图像区域分块，计算块内梯度直方图，进行统计并寻找主方向和辅助方向。
- e) 关键点特征描述子生成。计算关键点周围 16*16 范围内每一个像素的梯度，在每个 4*4 的 16 象限中，计算一个梯度直方图。可以对每

一个特征形成 $4*4*8=128$ 维的特征描述子。

SIFT 特征在于尺度不变和旋转上有强大的优势，能够提供高速的提取算法，并可以和其他形式的特征向量联合使用。但对图像中具有光滑边缘的区域无法得到比较准确的特征点进行描述。

2.1.3 卷积神经网络

随着机器学习算法的发展以，在目标检测中，越来越多的人开始使用从样本中学习特征的方式。在 Hinton^[3]提出利用非监督逐层初始化来训练深度网络的方式后，Lecun 等人提出了卷积神经网络（Convolutional Neural Networks, CNN），克服了普通神经网络全连接的结构，成为当前目标检测领域最好的特征提取方式。下面将从神经网络开始，介绍 CNN。

首先介绍神经元，它是如下所示的一种结构：

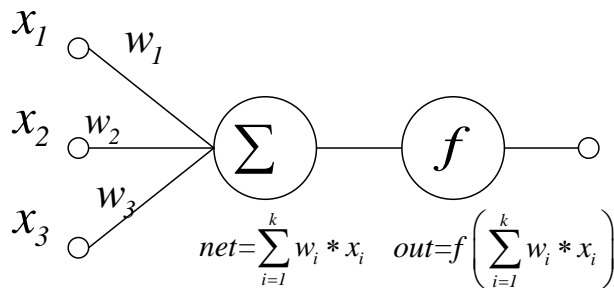


图 2-2 神经元结构

其数学公式上的表达就是 $out = f(\sum_{i=1}^k w_i * x_i)$ 。接受多个输入，并通过一个激活函数 f 给出这些输入的一个输出响应。其中激活函数 f 一般选用 Sigmoid 或者 ReLu 函数，以使得输入和输出具有相同的范围。

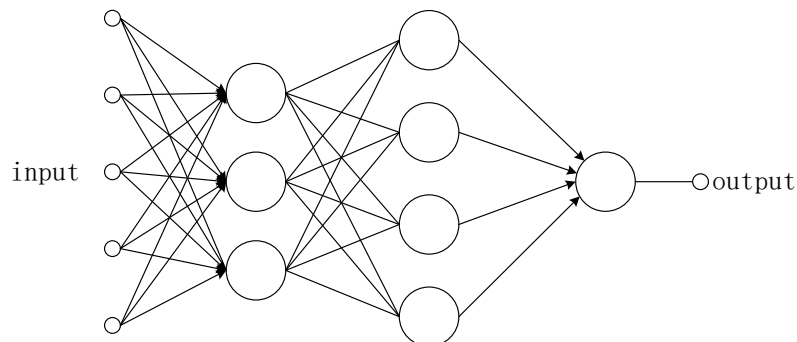


图 2-3 具有隐藏层的神经网络

神经网络是由多个神经元按照一定层次结构连接起来的网络，一般来说包括输入层、隐藏层和输出层。如上图所示，每个神经元的输出作为下一层的输入，

使输入数据经过网络中的每一层，逐渐向前传导，最终获得输出的结果。

普通神经网络的输入是向量的形式，在用于图像领域时，需要把输入的图像做一个特征提取处理，再将特征作为网络的输入参数。这样的过程使得图像的特征与网络使用分离，具有较大的冗余操作。而卷积神经网络则不同于此，利用卷积层可以直接将二维图像数据作为输入，使特征提取和目标分类结合在一起。

在网络连接上，直观来说层与层之间可以采用全连接，但网络会随着图像的增大以及网络层数增多呈现指数型增长，使参数学习所需要的资源大幅增加。为了解决这个问题，卷积神经网络引入了使用了局部感受野和网络参数之间权值共享的方法，极大减少了所需的参数个数。

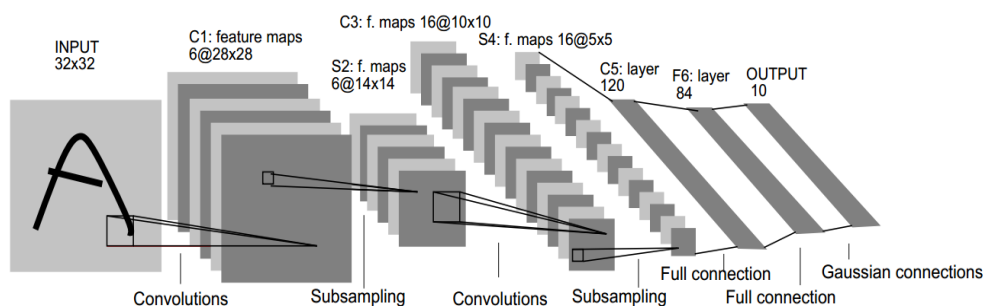


图 2-4 卷积神经网络工作流程图^[3]

如上图所示，卷积神经网络的基本结构包含卷积层、池化层和全连接层。卷积层的定义来自于局部感受野的方法，即输出层的每一个点都与输入的一块区域相关。此外，对图像来说，输出的每一个点都是由原图和滤波器模板卷积而来，这就是所谓的权值共享，每一层的神经元共享相同连接参数，使整个网络的参数大幅减少。之后通过池化层，计算图像一个区域上的某个特定特征的平均值，进行降采样，聚合上一层的特征，使网络具备一定的平移、缩放不变性。最后，卷积神经网络连接到全连接层，将前面形成的特征投影到目标分类空间。

2.2 目标表示方法

上文对图像特征提取做了介绍，但对于特定的目标，用整张图片的特征向量去描述是十分不明智的，常用的方法是利用这些特征提取的方法对目标进行表示。根据表示方法的不同，可以分为基于样本特征的表示和基于样本结构的表示这两种。

2.2.1 基于样本特征表示

基于样本特征表示的方法，即用带有目标的图像提取特征，用以表示样本。通常来说，在目标检测中，样本是指用边界框包围起来的目标区域。对于这样的样本，可以通过前面提到的图像特征提取算法，将目标区域作为一个图片进行计算，得到的特征向量作为样本向量。在具有分类能力的卷积神经网络中，则一般提取某一个卷积层或者池化层的输出作为样本特征。这样的表示方法具有较强的直观性，但其特征向量仅具备提取算法的性质，没有关于样本本身的特点。

2.2.2 基于样本结构表示

与用特征直接表示样本不同，基于结构的表示方法，则是以特征提取算法为基础，利用样本局部信息的特征，建立整体与局部的拓扑联系，可以更好的表达样本性质。关于这样的表示方法，下面先介绍语法模型，接着介绍基于语法的DPM模型。

2.2.2.1 语法模型

对于一个高度抽象的表述来说，目标表示的语法结构是指通过语法结构，用目标自身的形式以及生成规则来递归的描述目标的系统。这种框架可以认为是用模型化的语言来表示模型^[19]。

为了更好的说明图像中的语法模型，首先给出以下定义：

- a) 符号，这是语法模型中最基本的元素，有终止和非终止元素两种，所有的操作都是由符号指向符号的。
- b) 位置符号，指的是带有位置信息的符号。假设 Ω 是一个符号在图像中所有可能的位置，位置符号 $Y(\omega)$ 表示将符号 Y 放置于 ω 位置，其中 Y 为终止或非终止符号， $\omega \in \Omega$ ， $Y(\omega)$ 也称为实例化的符号。
- c) 生成规则，指明了符号之间的生成关系。

由以上定义，假设 N 是一个非终止的符号集合， T 是终止符号的集合。一个终止符号可以认为是自身组成一个目标模块，不能再细分成更小的组合。例如，对于一辆汽车来说，可以假设轮子、车身等是终止的元素，通过这些终止元素来表明汽车检测的语法。即终止符号有着直接的外形含义，或者有着不具备分解因素的模板形状。与终止符号相对的，非终止的符号表示则是由其他语义元素组成的，这些语义元素称之为非终止符号的部分，可以是终止或者非终止的符号。在

没有相互嵌套关系的情况下，从非终止符号的迭代关系，所有的非终止符号最终都能用一系列的终止符号表示。回到汽车的例子，将汽车作为一个非终止的元素，它就是由车身和车轮组成，而车身可以更进一步分解为前后挡风玻璃、后视镜等，如此迭代下去，最终一辆汽车的可以用多个不可再分割的符号组成。

为了了解组成关系，可用以下公式说明非终止符号可以通过生成规则扩展为一系列的其他符号：

$$X \xrightarrow{\beta} \{Y_1, \dots, Y_n\} \quad (2-3)$$

其中， $X \in N$ ， $Y_i \in N \cup T$ ，并且 $\beta \in R$ 是偏置量。上式表明了一个符号 X 是由 Y_1, \dots, Y_n 所组成的，所有的非终止符号都可以由这样的公式间接由终止符号表示出来。

对于有着多种不同子类的目标，即一个符号可以通过多种生成规则来描述，如下所示：

$$\begin{aligned} X &\xrightarrow{\beta_1} \{Y_1, \dots, Y_n\} \\ X &\xrightarrow{\beta_2} \{Y_1, \dots, Y_n\} \end{aligned} \quad (2-4)$$

举例来说，如果 X 表示一个通用的车辆分类，上面其中的一个生成式表示小型的汽车，而另一个生成式可以是表达大型卡车。偏移量 β_i 表示一个预先定义倾向值，用来从不同的扩展中区分出特定的一个。

如果一个语法在其生成规则的展开过程中生成了自身，则称之为循环的语法。这样的规则会使整个语法模型变得非常复杂，在对图像的语法模型中，限制只使用非循环的语法。

为了描述目标是如何通过语法模型表示出来的，首先需要在图像中定义符号的实例化。假设 Ω 是图像中符号可能存在的所有位置集合，即图中的坐标和尺度信息。则对于任意的 $\omega \in \Omega$ ，以及 $Y \in N \cup T$ ，实例化的符号 $Y(\omega)$ 表示符号出现在位置 ω 上。

图像中清晰的目标语法可以用一系列的生成规则与实例化的语义符号表示，如下所示：

$$X(\omega_0) \xrightarrow{\beta} \{Y_1(\omega_1), \dots, Y_n(\omega_n)\} \quad (2-5)$$

其中， $X \in N$ ， $Y_i \in N \cup T$ ， $\omega_i \in \Omega$ ， $\beta \in R$ 。这个式子表示了处于图像位置 ω_0 的目标 X ，可以由生成规则中的 Y_1, \dots, Y_n 在相应的 ω_i 位置表示。

但上式中生成规则的范围与 Ω 相关，通常情况下是极大的。为了避免采用枚

举的方法来表示生成规则，可以采用较小的带有模式量的生成规则隐式表达，将具有位置的函数模板生成规则转化为具体的生成规则。假设 $z \in Z$ ，定义了函数模式下所能产生的生成规则，则参数化后的生成规则表示如下：

$$\forall z \in Z: X(\omega_0(z)) \xrightarrow{\beta(z)} \{Y_1(\omega_1(z)), \dots, Y_n(\omega_n(z))\} \quad (2-6)$$

函数 $\omega_i(z): Z \rightarrow \Omega$ 将参数 z 映射到具体的实例化坐标位置，生成偏置函数中有 $\beta(z): Z \rightarrow R$ ，将参数 z 映射到生成偏置。

在视觉目标表示中，语义模型的生成迭代一般不会很多，更一般化的方式是将生成规则右边的位移表示为左边的位移偏置。定义 δ 为位置偏置，符号 \oplus 表示位移偏置操作，则目标的语法模型也可以表示如下：

$$X(\omega) \xrightarrow{\beta} \{Y_1(\omega \oplus \delta_1), \dots, Y_n(\omega \oplus \delta_n)\} \quad (2-7)$$

上述语法模型可以很好的表示目标与其子部件的位置关系。

有了表示目标的语法模型，下面考虑的是如何将模型运用到目标检测中去。

首先要考虑的是符号在图像中的响应得分。每一个终止符号 A 都有一个与之关联的滤波器模板 F_A ， H 是特征图， $\phi(H, \omega)$ 表示特征图中的向量，则符号 A 在图中 ω 位置的响应可以定义为：

$$score(A, \omega) = F_A \cdot \phi(H, \omega) \quad (2-8)$$

有了终止符号的得分之后，则可以由此定义非终止符号的响应得分。由前面推导，非循环情况下，所有的非终止符号 X 都可以最终形成一颗派生树 T ， T 的根是 X ，中间节点表示生成规则与其他非终止节点，叶子节点均是由终止符号表示。则非终止符号的响应即为派生树 T 的响应：

$$\begin{aligned} score(X, \omega) &= score(T, \omega) \\ &= \sum_{r \in int(T)} \beta_r + \sum_{i \in leaf(T)} score(A_i, \omega_i) \end{aligned} \quad (2-9)$$

其中 $int(T)$ 表示中间节点， $leaf(T)$ 表示叶子节点，即响应由派生树中的所有叶子节点的响应加上所有生成规则的偏置量。

考虑叶子节点与根节点是具有形变偏移的，还需要定义带有形变的响应得分规则， $X(\omega) \xrightarrow{\alpha \cdot \phi(\delta)} \{Y(\omega \oplus \delta)\}$ 。其中 $\phi(\delta)$ 是形变惩罚特征向量， α 是形变参数，如定义 $\phi(\delta) = \{dx, dy, dx^2, dy^2\}$ ，则 $\alpha \cdot \phi(\delta)$ 就代表了对于位移形变惩罚的二次方程。

则整体的检测流程就是在图像中所有位置应用派生树，并寻找当前派生树下叶子节点取得最大响应的位置，求和该位置的得分。

$$score(T, z) = \max_{z \in Z} (\sum_{r \in \text{int}(T)} \beta_r(z) + \sum_{i \in \text{leaf}(T)} score(A_i, \omega_i(z) \oplus \delta_1)) \quad (2-10)$$

之后通过阈值去除低于某个得分的所有位置，对于重叠派生树 T 所代表的检测框，则运用非极大值抑制，即可得到在该位置下最适合代表目标的语法模型派生树。

2.2.2.2 DPM 模型

在前面介绍了语法模型之后，这段将详细介绍语法模型在目标检测中的实例可变部件模型（Deformable Part-Base Models, DPM^[30]）模型。DPM 是一种具有多类混合及多尺度子部件的模型，能够用来表示具有一定程度形变的目标类别。最初是在 HOG 特征的基础上提出，之后 Girshick 通过语法模型将其扩展到 CNN 特征^[35]，在多种应用场景中都取得极好的检测效果。下文介绍基础的 DPM 模型，即基于 HOG 特征发展的线性参数模型。

首先需要定义线性滤波器 F 应用于特征图 G 的形式。特征图 G 即经过前面小节特征提取算法处理后得到的图像特征，以 HOG 特征为例，特征图 G 中的一小块矩形区域表示的是该区域对应的 block 所组成特征向量。接下来定义滤波器，滤波器 F 是一个区域的权值矩阵，可以拉伸成为与对应特征维数一致的向量。滤波器 F 放置于特征图 G 中 (x, y) 位置后，可以与特征图进行卷积操作，或称为乘积操作，即将两者从左上角位置开始，所有对应点的乘积之和，如下式所表达。

$$\sum_{x', y'} F[x', y'] \cdot G[x + x', y + y'] \quad (2-11)$$

上式也叫做滤波器 F 在特征图 G 中的响应得分，这个操作适用于任意尺度和位置上的特征图。

DPM 模型是一种星型的结构，与语法模型类似，由整体和部件按照一定层次组成。它有两个层级的滤波器，第一层是根滤波器模板，用来描述目标粗略的外形轮廓；第二层有多个子部件滤波器模板，处于比根模板更高分辨率的特征层，用来描述详细的局部信息。下图是行人在 HOG 特征金字塔中的 DPM 模型可视化示意图，其根滤波器定义了一个检测框，部件滤波器所涉及的所有像素都被包含在这个检测框之中。部件滤波器在金字塔中的位置比根滤波器更往下 λ 层， λ 是使得该层具有根滤波器所在层的两倍分辨率的任意数值。

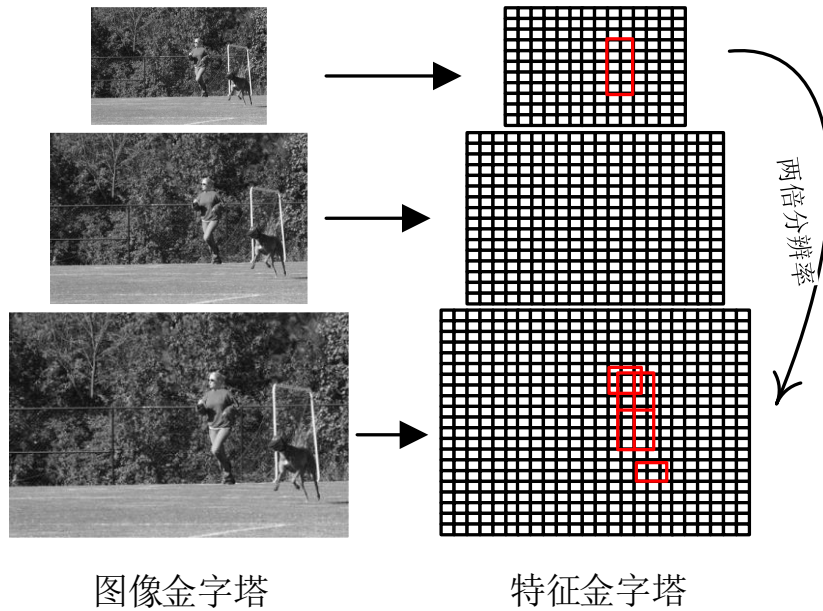


图 2-5 行人的 DPM 模型表示^[30]

使部件滤波器处于更高分辨率的特征层能有更好的检测性能，比根滤波器获得更加精细特征信息，也有着更好的区分能力。就如行人的模型，根滤波器就是整个人的外形，比较粗糙的情况下难以区分这是人还是其他物体。在具有了头部、躯干等子滤波器之后，就能够轻松的在更高分辨率特征层中辨别目标的细节是否具有这类子部件。

模型对目标的表示形式中，假设目标具有 n 个子部件，通常来说，可以使用一个 $n+2$ 的元组进行表示 $(F_0, P_1, \dots, P_n, b)$ ，其中 F_0 表示的是根滤波器， P_i 是表示第 i 个部件的模型， b 是表示偏置值的实数项。每个部件模型 P_i 是由一个三元组来进行表示的，即 (F_i, v_i, d_i) 。这里 F_i 是第 i 个子部件对应的滤波器， v_i 是一个两维的变量，指示部件滤波器 F_i 与根滤波器 F_0 相对应的锚点， d_i 是四维的向量，指明了当子部件位置偏移其理想位置后，用来计算位置惩罚的二次函数相关系数。

基于上面的描述，一个目标的假设应是能够指明根模板和所有子部件在特征金字塔中位置的变量。 $z = (P_0, \dots, P_n)$ ，其中 $P_i = (x_i, y_i, l_i)$ 明确指出第 i 个部件在特征金字塔中所处的层和位置， l_i 需要确保其所在的层是根模板两倍分辨率所在，即 $l_i = l_0 - \lambda, \forall i > 0$ 。

目标位置的假设 z 所得到的响应分数就是每一个滤波器在与根滤波器对应平移后的相关位置所得分数与偏移量的和：

$$score(P_0, \dots, P_n) = \sum_{i=0}^n F_i' \phi(H, P_i) - \sum_{i=1}^n d_i \cdot \varphi_d(dx_i, dy_i) + \mathbf{b} \quad (2-12)$$

其中 $(dx_i, dy_i) = (x_i, y_i) - (2(x_0, y_0) + v_i)$ 给出了第 i 个子部件与其理想位置的位移, $\varphi_d(dx, dy) = (dx, dy, dx^2, dy^2)$ 给出了位移形变的特征。值得注意的是, 当 $d_i = (0, 0, 1, 1)$ 时, 第 i 个子部件与相对于根模板下的理想锚点位置的形变惩罚就是偏移量的欧式距离。通常来说, 这个形变惩罚可以是以位移为变量的任意二次方程。

目标假设 z 的得分可以用向量点积的形式来表达, 用模型参数向量 β 和特征向量 $\psi(H, z)$ 表示即为 $\beta \cdot \psi(H, z)$, 其中

$$\begin{aligned} \beta &= (F_0, \dots, F_n, d_1, \dots, d_n, b) \\ \psi(H, z) &= (\phi(H, P_0), \dots, \phi(H, P_n), \\ &\quad -\varphi_d(dx_1, dy_1), \dots, -\varphi_d(dx_n, dy_n), 1) \end{aligned} \quad (2-13)$$

这阐释了模型与线性分类器之间的相互联系, 在后面涉及到模型参数学习的隐变量 SVM^[29]时, 可以使用与线性分类器相类似的学习优化算法。

在使用模型检测图像中的目标时, 对每一个可能的根模板位置, 都计算使总体响应得分最大的部件位移。

$$score(P_0) = \max_{P_1, \dots, P_n} score(P_0, \dots, P_n) \quad (2-14)$$

根滤波器得到的高响应分数同时局部的部件也能有高响应时, 说明根模板所处位置所定义的检测框是目标的概率较大。

对于上面的检测方式, 如果采用每一个可能的位置都计算子部件的得分, 由于图像中根模板以及子部件的可能位置集合很大, 直接计算的方式具有极高的复杂度, 可以采用了一个动态规划的过程进行计算。

这个动态规划过程是将每个部件的最大响应位置作为根模板位置的一个函数, 对于每一个位置, 只需要 $O(nk)$ 的复杂度即可得到最佳的位置响应。其中 n 是部件数目, k 是根模板在特征金字塔中可能位置的数目。

令 $R(x, y) = P_i \cdot \phi(H, (x, y, l))$ 做为一个存储着 i 子部件滤波器在特征金字塔中的第 1 层所有位置响应的得分矩阵, 检测算法从计算这些子部件的得分矩阵开始。当计算完所有的响应之后, 则子滤波器的响应得分就可以通过相应的矩阵直接计算带有位移的值:

$$D_{i,j}(x, y) = \max_{dx, dy} (R_{i,j}(x + dx, y + dy) - d_i \cdot \varphi_d(dx, dy)) \quad (2-15)$$

这样的转换, 使滤波器的分数在局部区域有高响应的传播, 能充分考虑到形

变惩罚的减分项。 D 的值是当根模板处于 (x, y, l) 位置时, 第 i 个子部件对其的最大得分贡献。这个转换矩阵可以用 **generalized distance transform algorithm**^[14] 在线性时间内计算出来。

则根模板在所有层的得分可以通过转换后的 D 矩阵和根模板在该层的得分表达, 即

$$score(x_0, y_0, l_0) = R_{0,l_0}(x_0, y_0) + \sum_{i=1}^n D_{i,l_0-\lambda}(2(x_0, y_0) + v_i) + b \quad (2-16)$$

注意到 λ 是需要从根滤波器所在特征层往下的数目, 以取得子模板里精确的两倍分辨率特征。为了更好的理解上式, 假设根模板处于一个比较合适的坐标, 因为子部件之间没有相互的干涉关系, 我们可以独立的找出每一个子滤波器所得到的最佳响应。转换矩阵 $D_{i,j}$ 以子滤波器的位置为变量, 给出了第 i 个部件对总体分数的贡献函数。故可以通过累积根滤波器与子滤波器在 $D_{i,j}$ 中提前计算好的所有对响应贡献来获得在该位置的总响应得分。

考虑 (2-16) 的形式, 可以直接计算第 i 个分布对于以其理想位置为变量, 对于偏移量最优化的函数, 如下所示:

$$P_{i,l}(x, y) = \arg \max_{dx, dy} (R_{i,l}(x + dx, y + dy) - d_i \cdot \varphi_d(dx, dy)) \quad (2-17)$$

根据上式, 在找到具有较高响应的根滤波器坐标 (x_0, y_0, l_0) 后, 可以直接从 $P_{i,l}(x, y)$ 查表得到相应的第 i 个子滤波器在理想位置的最优化位移得分。

2.3 隐变量 SVM

前面介绍了 DPM 模型与其检测方法, 以 β 表示模型的参数, $\psi(H, z)$ 表示在特征图 H 上以 z 为所有部件位置提取的特征, 则目标假设 z 的响应得分如下所示:

$$f_{\beta}(x) = \max_{z \in Z(x)} (\beta \cdot \psi(H, z)) \quad (2-18)$$

注意到上式是一种并不是线性的形式, 当想用训练样本学习模型参数 β 的时候, 因为隐变量 z 的存在, 所需要解决的问题就是隐变量 SVM^[29] 最优化问题。

参考没有隐变量的线性优化问题, 通常的做法是使用支持向量机 (Support Vector Machines, SVM) 来获得参数 β 的解。假设给定的所有训练样本为 $D = (\langle x_1, y_1 \rangle, \dots, \langle x_n, y_n \rangle)$, 其中 x_i 是样本特征, $y_i \in \{-1, +1\}$ 是样本标签, 则需要最小化的目标函数为

$$L_D(\beta) = \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n \max(0, 1 - y_i \cdot f_{\beta}(x_i)) \quad (2-19)$$

上式中, $\max(0, 1 - y_i \cdot f_{\beta}(x_i))$ 是损失函数, 常量 C 是该正则项的权重。这样典型的线性优化问题, 可以通过梯度下降法来进行求解。

回到带有隐变量的问题上, 如果对于任意的训练样本, z 是一个已知的常量, 则其方程可以视为线性, 通过 SVM 就能够获得参数的最优化解。

那么在优化之前, 可以增加一个过程, 将所有正样本的隐变量位置确定下来。令 Z_p 表示样本集 D 中每个正样本的位置隐变量, 定义如下辅助的训练目标函数, $L_D(\beta, Z_p) = L_{D(Z_p)}(\beta)$ 。其中 $D(Z_p)$ 是将 D 中所有正样本被 Z_p 所限定后的新样本集, 即对于正样本 x_i , 有对应的 $z_i \in Z_p$ 使 $Z(x_i) = \{z_i\}$ 。由此注意到有:

$$L_D(\beta) = \min_{Z_p} L_D(\beta, Z_p) \quad (2-20)$$

这个辅助的方程将隐变量 SVM 限制成为 SVM 的形式, 以最小化 $L_D(\beta, Z_p)$ 的方式改变了样本的假设。在实际使用中, 通常使用以下的坐标下降法来求解出 $L_D(\beta, Z_p)$ 的最小值:

- a) 正样本重标注: 通过在每一个正样本中选择能使响应得分最高的位置隐变量, 最优化 $L_D(\beta, Z_p)$ 中的位置变量, 即对任意 x_i

$$Z_i = \arg \max_{Z_p} \beta \cdot \psi(x_i, z) \quad (2-21)$$

- b) 优化 β : 通过解凸优化问题 $L_{D(Z_p)}(\beta)$, 求解得到的最优化解 β 作为 $L_D(\beta, Z_p)$ 的最优化解。

这两步通常能够提升或者维持 $L_D(\beta, Z_p)$ 的值, 因 a) 中搜索了正样本中隐变量指数大小的取值空间, b) 中则遍历了所有可能的模型, 隐含地考虑了负样本指数级别的取值空间。在收敛之后, 就可以得到一个具有局部最优的解。

模型参数的良好初始化是很必要的, 否则在 1) 中很难以选择出正样本合适的隐变量位置, 并在后续导致一个比较差的模型。

DPM 模型参数求解问题, 除了上述隐变量 SVM 的过程外, 往往还会使用一个难反例挖掘的方法减少内存消耗。

2.4 域适应的 SVM

当训练数据集和测试数据集的分布不一样时, 从训练数据中学习到的模型往往不能直接用于测试集, 这两个域之间的差异会使得检测性能有比较大的下降。为解决这种域适应的问题, 学者们提出了许多解决方案, 这里介绍一种比较通用的从 SVM 分类器中发展出来域适应 SVM 算法 (Adaptive SVM) [42]。

首先做一些基本的定义。在域适应中，一般将训练数据集称为源域，测试数据集成为目标域。源域中训练得到的分类器参数为 W_S ，目标域中的分类器参数为 W_T 。域适应 SVM 的目标在于利用 W_S 来辅助训练 W_T ，从而使 W_T 具备仅使用目标域数据训练时更高的性能。

域适应 SVM 的做法是在 W_S 和 W_T 之间添加一个扰动函数，从而使两者联系起来。对源域来说，样本 x 的响应为 W_S 与 x 特征向量的点积 $f^s(x) = W_S \cdot \varphi(x)$ 。对目标域，通过扰动方程后，可以如下表示：

$$f(x) = f^s(x) + \Delta f(x) = f^s(x) + \Delta W \phi(x) \quad (2-22)$$

其中， $\Delta f(x) = \Delta W \phi(x)$ ，即 $W_T = W_S + \Delta W$ 表示出两个域之间的关系。下图采用特征空间中分类面参数示意图表示出这个公式的含义。

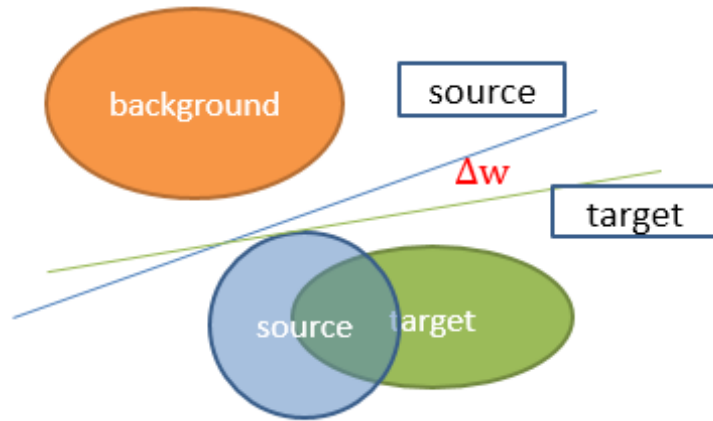


图 2-6 Adaptive SVM 分类平面

为了求解扰动函数中的参数 ΔW ，进而求得目标域的参数，上式可使用给定的样本与以下决策函数表达。

$$\begin{aligned} \min_w & \frac{1}{2} \|\Delta \mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s. t. } & \xi_i \geq 0 \\ & y_i f^s(x_i) + y_i \mathbf{w}^T \phi(x_i) \geq 1 - \xi_i, \forall (x_i, y_i) \in D_i^t \end{aligned} \quad (2-23)$$

其中惩罚函数 C 代表源域分类器对目标域样本的重要性， C 越大则源域分类

器对于目标域而言越不重要。

通过求解上式的拉格朗日对偶式，得到以下对偶方程的解。

$$L_D = \sum_{i=1}^N (1 - y_i f^s(\mathbf{x}_i)) \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (2-24)$$

其中 $\alpha_i (i = 1, \dots, N)$ 是拉格朗日系数。从上式中可以看出，当 $y_i f^s(\mathbf{x}_i) < 0$ 即源域分类器对目标域样本分错时候， $f(\mathbf{x})$ 的分类面会远离 $f^s(\mathbf{x})$ 分类面。 α_i 也会取更大的值，从而使分错的样本对新分类面的影响更大。

2.5 本章小结

这一部分主要是详细介绍了本文所涉及的相关知识，从图像特征提取、目标表示方法、隐变量优化算法、域适应的 SVM 等方面。

特征提取算法介绍了 HOG 和 SIFT 两个手工特征，以及基于学习的深度特征。HOG 和 SIFT 是比较传统的特征，其优点在于能快速计算并具有很好的表达能力，给人一种直观的感觉，但提取算法本身的形式限制了它们的性能上限。深度特征在消耗较多计算资源情况下，能够提取出比手工设计更好的特征。

目标表示方法上，介绍了直接基于特征提取的表示，随后介绍具有层次表达能力的语法模型，之后在语法模型的基础上详细介绍可变部件的 DPM 模型。DPM 模型在处理遥感目标多视角旋转上有着比较好的优势。

隐变量优化算法中，主要推导了 DPM 这一类具有位置隐变量的目标方程优化算法，并介绍了其优化算法 latent SVM 的计算过程。

模型域适应算法是不同数据域之间模型参数进行迁移的算法，能联合两个数据域的分布特性，在面对比较多的应用领域时是及其有效的方式。本章以 SVM 为例，介绍了一个通用的域适应 SVM 算法。

第3章 基于相似性判定的样本生成研究

在前面的两章中，主要明确了本文的背景是在于小样本目标检测，并对国内外研究现状和一些相关技术做了详细介绍。本文的主要工作是基于迁移学习相关技术上来进行的，在第一章末尾，提出了整体的研究框架，其中的训练部分主要包括了辅助数据集建立和模型域适应两个步骤。

本章主要介绍了基于相似性验证的样本生成研究，从样本生成的问题分析开始，讨论样本生成中的技术，并在特征和实验层面对生成的样本相似性做了初步讨论。

3.1 问题描述及难点分析

在所有的迁移学习或者模型域适应中，都涉及了源域和目标域两个概念。源域是指作为参考数据集训练先验模型的数据领域，目标域是指将基础模型适应过去的的数据领域。模型迁移过程可以是将单个或多个源域作为先验模型对目标域做适应学习，从而使目标域模型具有与源域数据相关的泛化能力。

源域与目标域的样本数据集在特征空间中必须具有相关性，否则难以通过迁移学习算法来提高目标域模型的检测性能。从直观上说，这种关联的必要性是显而易见的。例如，从源域中训练了一个具有很好性能的飞机检测模型，这样的先验模型，不管使用怎样的迁移算法，都难以将先验知识适应到检测车辆的目标域中，只能具备一个随机初始化的效果。

然而，当前的迁移学习领域中，存在的一个巨大问题是没有算法可以直接判定两个数据域是否相似。存在多个可选源域数据集时，理想的情况是应该选择“最好”的源域用于训练先验模型，使域适应之后的目标域模型具有最佳的检测性能。但是，如何通过直接计算来判定什么是“最好”的源域，至今还没有一个确定的解决方法。实际应用中，通常只能对可能的数据源都进行迁移学习，通过目标模型的检测性能间接选取合适的源域。

回到遥感检测中，因目标数据难以获取，通常意义上从多个源域中都学习先验模型来选择最佳源域的方法并不适用。受到行人检测提出的采用虚拟合成样本的方式生成源域样本^{[37][38]}的启发，以及第一章研究框架中所述的遥感检测中大部分目标是人工物体，利用人工物体建模技术，能够较轻易获得目标的近似三

维模型，本文因此采用了虚拟合成样本的方式进行数据增强。在合成样本中，需要面对的问题即为怎样生成与目标域样本近似的源域样本数据，并通过哪些方式能够验证这些相似性。

为解决这个问题，本文提出了一种样本生成与验证方法，能够给相似性做一定的验证。流程如下图所示：

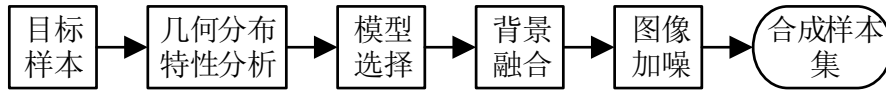


图 3-1 样本合成流程

如上图所示，首先选择对目标样本进行数据分析，获取关于分辨率，视角，长宽比，目标类型等粗略数据，并根据数据选择合适的三维模型，进行样本合成。之后利用特征可视化技术以及模型的初步检测结果作为判定，证实具有的相似性质。

3.2 目标数据集

为验证本文迁移学习框架的有效性，这里采用中科院大学高清航拍数据集作为目标域数据。

中科院大学高清航拍数据集是在 Google Earth 采集遥感图片后，用标注工具制作的遥感数据集，分别包括飞机、车辆和舰船三种目标，具体的目标明细如下表所示：

表 3-1 中科院大学高清航拍数据集

数据集	正例图像/幅	正样本/个	反例图像/幅
民用飞机数据集	1000	7482	492
民用车辆数据集	510	7114	408
民用船舶数据集	102	1229	-
军用舰船数据集	565	6749	513
总计	2177	22574	1413

数据集中的图像来源主要是：民用机场中停泊的各种型号飞机，成像高度为 600m 到 2km，分辨率为 0.5~1.5m；城市、郊区道路场景中的常见民用车辆，成像高度 400m，分辨率为 0.2~0.5m；各国知名军事港口中的包括航空母舰、驱逐舰、护卫舰、潜艇等目标，成像高度为 1km 到 1.5km，分辨率约为 5~10m；民用港口中停泊的货船只、渔船，成像高度 1km，分辨率为 5~10m。

所有样本的标注信息是包含目标样本的矩形包络框四个顶点坐标，以及目标中轴线与水平轴的夹角。标注示例如下所示：

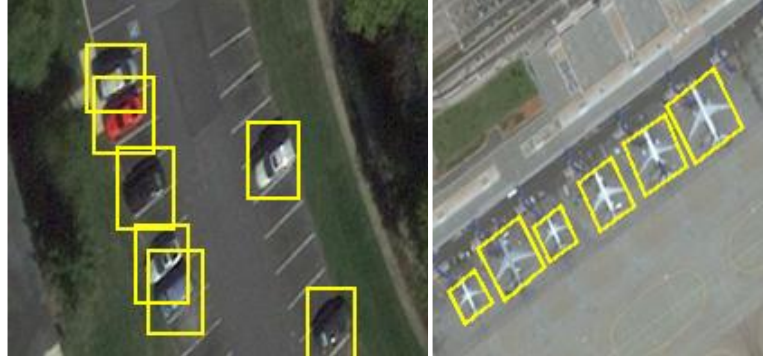


图 3-2 数据集标注示意

在本文的后续内容里，主要实验集中在车辆样本数据中：飞机具有机翼和机尾等明显特征，基于外形的检测难度较低；舰船由于其长宽比较大，且大多出于较为纯净的海域背景，用于检测中不具有太大泛化性；车辆因其外形接近于矩形，无明显突出特征，且与周围建筑具有较大相似性，用于验证检测框架能有更好的代表性。

在车辆数据集中，随机选取了 2700 样本作为验证集，从中学习到的模型作为检测性能基准线。选取 300 样本为小样本集，作为检测框架中的目标数据集。

3.3 基于统计分析的样本合成

上一小节中介绍了作为研究对象的车辆数据集，本节内容主要是根据目标数据集进行样本生成，包括目标数据分析以及样本生成两个步骤。

为了生成与目标数据相似的样本，需要对目标数据集的一些属性进行统计分析，这里选取了样本角度、长宽比以及样本像素和三个值作为分析对象。

在遥感视角下，目标的角度变化可以是任意方向的，对于车辆样本来说，由于对称性，其主方向只有 180° 的变化范围。为了统计分析的便利，这里将样本按照角度间隔 30° 进行聚类，考虑到在虚拟样本生成时，目标的角度会有一定随机性，这个间隔仍能够较好的表示出目标样本的分布情况。样本长宽比的含义在于实际车辆样本中的车型不同，分析长宽比的分布对于车辆三维模型的选取具有重要意义。样本像素和指的是样本在图像中的面积大小，即矩形区域所占的像素之和。通常来说，样本像素和与遥感成像高度与分辨率有关，但在检测中，这两者最终都体现在像素级别，因而只需要对样本像素和分析即可得到成像高

度与分辨率的重现。

由上面的描述，我们对目标车辆数据集相关性质做了直方图统计，结果如下表所示：

表 3-2 角度分布直方图

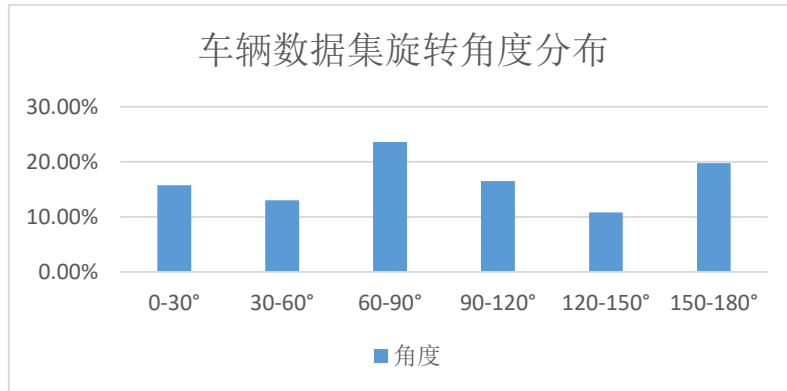


表 3-3 长宽比分布直方图

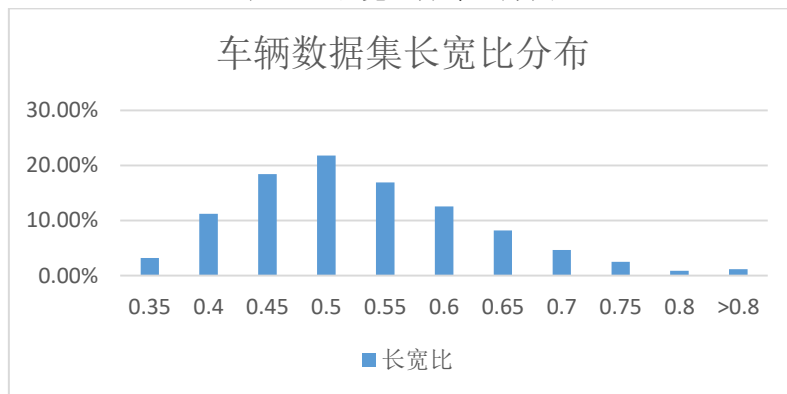
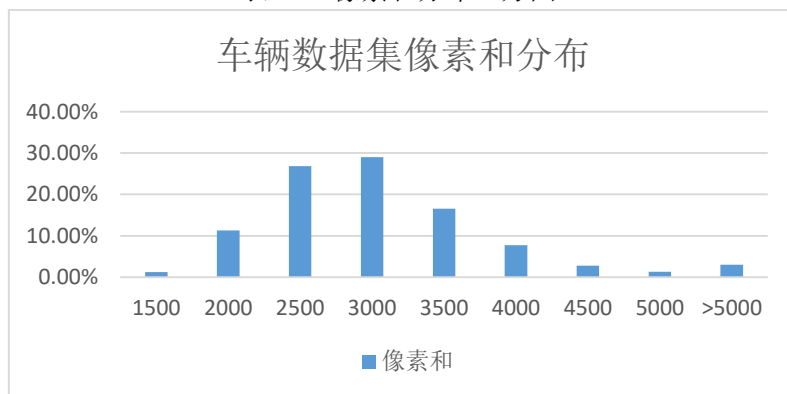


表 3-4 像素和分布直方图



由上表的统计数据可以做如下简要分析：角度的分布虽然具有一定波动，但从实际背景中考虑，可以认为是随机均匀分布的；车辆的长宽比近似是以 0.5 为

中心的正态分布，故应选择长宽比 0.4~0.7 的车辆模型，按照样本长宽比的百分数进行生成；对于像素和，指定了样本在合成图像中的面积大小，分布分析类似于长宽比，并与长宽比一起决定车辆的大小，例如长宽比 0.5、3000 像素和情况下，样本的平均大小近似为 40*80。

根据上面的分析，我们选取了具有合适长宽比的车辆模型与遥感背景图像进行样本合成，流程示意如下图所示：



图 3-3 样本合成示意

图中，背景融合主要涉及设定车辆模型的主方向随机角度，以及按照长宽比与像素和分布对成像中目标的大小进行限制。最终通过将合成图像加噪后得到具有完备标注信息的 2700 个车辆样本，作为后续域适应的先验模型训练集。

3.4 合成样本集的相似性验证

经过上面的目标域数据分析与虚拟合成样本，我们获得了在样本外观与浅层属性分布上具有直觉上相似的源域车辆数据集。本节将从特征与检测性能两个更深层的属性上验证数据集的相似性。

3.4.1 基于特征的相似性验证

数据集中的样本，最终都是计算特征之后才用于分类的，在特征空间中，相似的样本会有聚集性，而不相似的样本则会相离较远。基于这样的思想，本文通过提取样本的特征，从特征空间的聚类性质中进行相似性的定性判别。

在特征提取中，我们采用了 ImageNet 预训练的 VGG 网络。VGG 网络是一种直线型的具有较高层数的深度神经网络，与前文介绍的 AlexNet^[3]相比，不仅具有更深的层，而且对多类目标有着更精确的检测结果。通过高维空间可视化工

具 t-sne 对提取的车辆样本深度特征进行可视化后，能够对特征的聚类性具有清晰的概念。

在对于特征的提取比较中，本文主要尝试了目标样本、虚拟样本、非目标样本以及背景的特征。因神经网络具有一定的尺度不变性，仅需要考虑样本角度的聚类因素，但这里的目的是为了验证上述几种目标的相似性，为此仅需做一个粗略的划分即可。将背景作为类别 0，非目标样本作为类别 1，目标样本作为类别 2，合成样本作为类别 3。其中，非目标样本选择了同为遥感数据集目标的飞机样本。所提取的深度特征经过可视化后的结果如下：

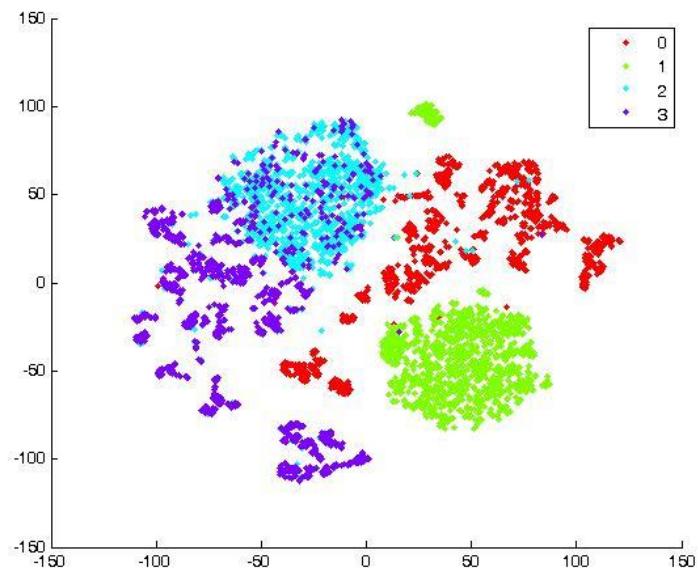


图 3-4 多目标特征可视化聚类分析

由上图可以看出，背景的与飞机的特征在可视化后与两种车辆的特征在聚类上有明显差异。同时合成的车辆样本特征在目标车辆特征的聚类中有一定的分布，具有与目标域数据类似的性质，合成样本还包括了小样本中不具备的先验知识，集图中左下角位置的紫色部分数据点。即合成的样本与目标样本具有相似性，并能提供目标样本未能包括的先验知识。

3.4.2 基于检测性能的相似性验证

在难以直接计算源域与目标域数据集相似的情况下，实际中另一种可行的方式是选择能够在目标域中获得较好性能的源域模型作为先验。这种方法从直

觉上来说是很明显的，对目标域有较好检测性能的先验模型，通常会来自于具有相近似分布的，并因此能提供一个良好的适应起点。

为了实现以上相似性假设，本文用生成的车辆样本学习了具有 6 个混合模型的 DPM 分类器，如果生成的样本与目标车辆数据集具有相似性，则这个分类器应当具有一定的检测性能。将 DPM 分类器用于测试集中的部分检测结果示意图如下所示：

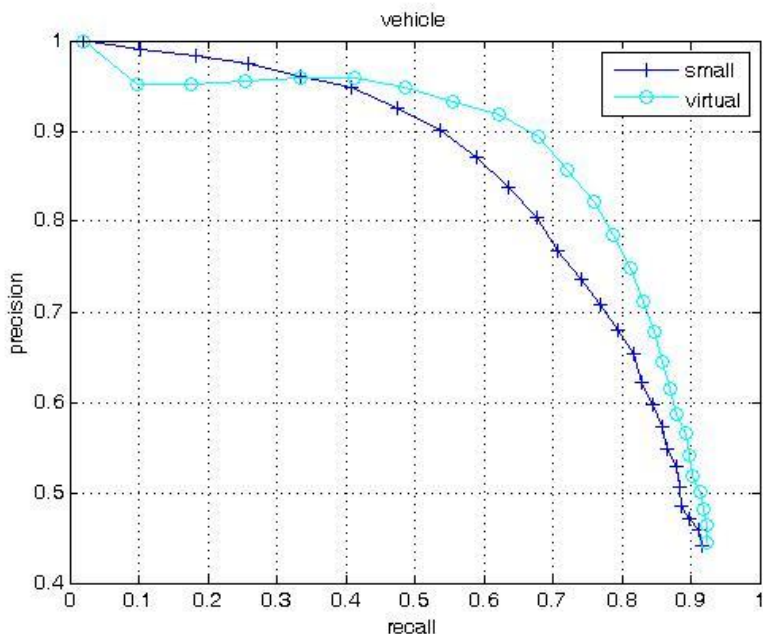


图 3-5 小样本与合成样本检测曲线对比

与前面特征可视化中的情况类似，合成样本训练的分类器能够在目标车辆数据集中对目标进行检测，并具有比小样本好一些的性能，但检测的性能受限。从直观的检测结果上验证了合成样本集作为源域数据时可以具有与目标域相似的性质。

3.5 本章小结

在迁移学习中，源域样本应当与目标域具有相似性，并能提供目标域不具备的先验知识。但源域与目标域之间相似性并没有一个确定性的计算方式，本章围绕这个问题，探讨了源域数据集的建立与相似性验证方法。首先拟定中科院高清航拍数据集中的车辆作为目标数据集，分析车辆在角度、长宽比、像素和等几何统计特性，在此基础上选择车辆的模型进行样本生成，得到具有相似几何特性的

合成样本集。之后，对源域与目标域的内在相似性做了验证：用 VGG 网络提取多种不同目标的深度特征，通过 t-sne 数据可视化工具对这些目标在深度特征空间中的聚类进行分析，从特征角度验证相似性；在源域数据中学习具有外观特性的 DPM 模型，通过分析模型在目标域测试集上的检测性能从实验角度验证样本集之间的相似性。

第4章 域适应的小样本目标检测研究

前面介绍了遥感目标检测的背景和所涉及到的相关技术，并对迁移学习中源域数据集的建立方式进行了研究。本章主要内容是在这个基础上提出小样本光学遥感目标检测的具体研究方案。在前面第一章的末尾，提出了本文的研究内容，主要包括以下几个方面：源域数据集的建立、源域先验模型训练、先验模型的域适应训练。本章将会逐渐对这些研究内容展开详细说明，并用实验来验证本方案的可行性。

本章结构安排如下：第一节是对背景的详细分析，指出小样本目标检测研究中所面临的问题；第二节介绍以合成的车辆数据集作为源域时先验模型的训练方法；第三节给出与结构相关的 DPM 模型域适应方法；第四节给出相应的实验验证和结果分析；最后一节则是对于本章内容的总结。

4.1 问题描述及难点分析

在目标检测中，其目的在于能够快速找出图像中感兴趣的目标。为实现检测的目的，首先要根据给定的目标样本与背景图片来训练一个检测模型。之后通过候选区域提取算法在待检测图像中提取目标可能存在的候选区域，并在候选区域中依次应用分类器进行检测，判定候选区域是否为所需要检测的目标。整体流程如下所示：

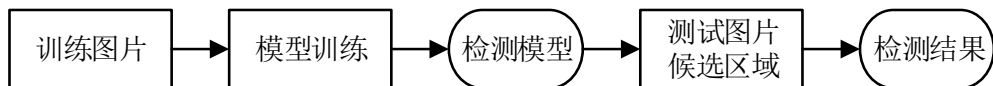


图 4-1 目标检测通用流程

本文根据上面的目标检测通用流程来分析小样本光学遥感目标检测中存在的问题与难点。

在通常的机器学习问题中，检测模型参数的训练是采用监督学习来完成的。通过大量标记的正负样本，对选择的模型进行训练优化，得到在该样本集下的最优化模型参数。

但在遥感目标检测中，存在着小样本的问题。遥感图像中难以直接获取到较大数目的样本，原因在遥感图像中目标是很稀疏的，没有目标的图占了绝大多数。而且目标种类多，标注难度大，导致标注样本较少，这是遥感目标检测中的直接问题。

由上分析，可以定义出小样本光学目标检测的应用场景，即我们所能获得的样本数目是有限的，这个有限是指难以从其他未标注的图像中得到样本数据。这与弱监督目标发现所不同，弱监督目标发现是假设在未标注的图像中会存在有较多的样本数据，通过弱监督的方式来将这些样本找出来标注。

大量的样本，从另一个侧面说，几乎能够较为清晰地描述出整个目标样本数据分布特点。从这些样本中学到的模型，对整个样本集上的目标都具有较高的检测泛化能力。而小样本数据集，则只能表达出样本空间中的部分区域，用于训练时难以使一些高效的检测模型在优化中有良好的参数收敛。如在深度神经网络中，几百个样本训练几乎不能使参数得到优化；即使选择的模型能够通过小样本集对参数进行训练，并得到一个合适的收敛参数结果，但模型的泛化能力并不是无限的，只能描述出训练样本分布域附近的样本，难以检测出在数据空间上相差较远的新样本。

综上所述，模型参数收敛以及检测模型的泛化能力限制就是小样本目标检测中所需要面对的核心问题。

为了解决这个问题，结合前面对目标检测领域的分析，本文采用了如下的方式进行研究

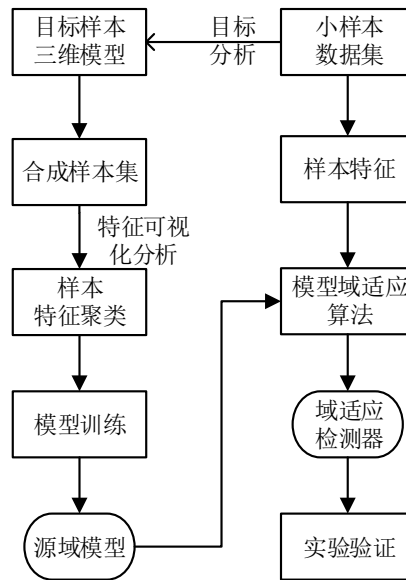


图 4-2 小样本训练流程

如图，在辅助数据集的建立过程中，本文采用了第三章介绍合成的方式建立合成车辆数据集，并对源域数据集作了相似性验证。之后采用在源域数据集

下训练具有先验知识的 DPM 模型，最后将这个模型经过域适应算法转换到目标数据域。

4.2 模型训练

有了前面对小样本目标检测中存在的问题与检测框架的分析，以及根据目标域样本建立的源域数据集。本节内容是考虑如何从源域数据集中训练一个具备较好性能的先验模型。

4.2.1 目标聚类分析

HOG 特征能够清晰的描述图像中的边缘信息，对于有外形轮廓的目标，提取的 HOG 特征也具备良好的表达能力。DPM 模型在直立视角下的行人检测中，采用了按照样本的长宽比将行人分类为正面和侧面两种类型，进行混合模型的初始化。但在遥感场景下的目标检测中，目标的长宽比与角度相关，在此并不能采用相同的形式进行混合模型初始化。而且，通过第三章对车辆数据集的角度直方图分析，目标的在俯视视角中的旋转是任意的，可以看作近似符合于随机均匀分布。

由于 HOG 特征在角度上的泛化能力有限，对于使用了 HOG 特征的 DPM 模型，并不能将所有样本视为单一类别。从行人检测按照长宽比对混合模型的初始化方式中得到启发，本文以类似的方式先将遥感目标按照角度进行分类。

在按照角度分类时，可以是将所有样本作为一类，或者每间隔 10° 作为类别区分，为了寻求较好的混合模型分类方式，需要考虑以什么样的角度对样本进行分类以最终得到较好的模型。

为了验证以什么样的角度来划分车辆样本会使得混合模型的类内聚类能力更好，提取样本的根模板的 HOG 特征做可视化，从特征中对角度的聚类进行初步分析，提出可行的划分方式。

由于遥感的俯视视角中，车辆自身具有对称性，样本主方向的旋转角度只有 180° 。首先将提取的特征按照样本角度从 0 到 180° 的 $1, 2, 3, 4, 6$ 以及 9 等分后用 t-sne 做可视化。得到的聚类信息如下：

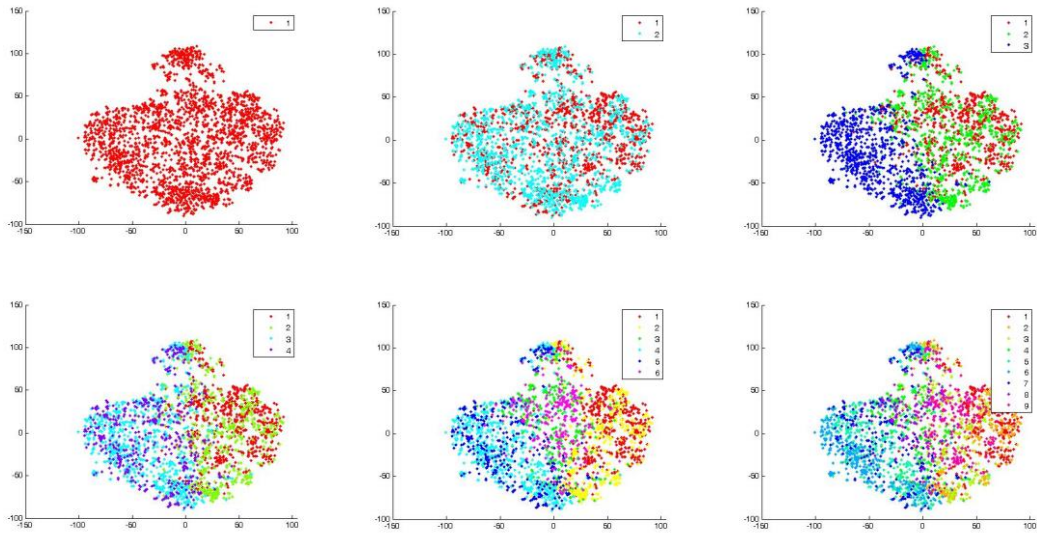


图 4-3 车辆样本不同角度特征聚类

从上面的特征可视化可以看出，对每一种角度划分方式，类内的特征点都有一定的聚集性，不同类的特征也有较为明显的区分，但因只提取了根模板的 HOG 做初步分析，数据之间的混杂程度较高。随着等分角度的增加，划分的数据在聚类上看起来就越准确。

从图中可以得出，具有三个以上等分角度的样本间的区分性会更好一些，在后面的实验中，我们会继续探讨混合模型数目与对应模型检测性能的具体情况。

4.2.2 混合模型训练

得到车辆不同聚类的样本之后，采用前面 2.2-2.3 所推荐的方式来进行模型训练。将坐标下降方式与数据挖掘结合，找出具有当前参数下最优的隐变量值，之后通过随机梯度下降方式求解本次迭代后的模型参数。

值得注意的是，在使用 2.3 的坐标下降法求解模型参数时，算法比较容易陷入局部最优化的结果，并对模型初始化比较敏感，这是其他所有具有隐变量算法的缺陷。因此必须考虑良好的模型初始化，降低取得局部最优解的概率。

为了得到较好的初始化参数模型，采用了以下三个阶段来进行混合模型参数初始化工作。

A) 初始化根滤波器。对于具有 m 个聚类的目标混合模型，先将每一聚类中的样本按照边界框长宽比率的大小排列。为了确定 m 个聚类的根滤波器维度

F_i ，选择其平均长宽比，并使该线框限制在不超过最大样本像素大小的 80%。这样做能够确保对于聚类中的任意一个样本，都能够在特征金字塔中找到能够用 F_i 覆盖其大小位置。在训练 F_i 时，因为没有隐变量的信息，使用的是 SVM 训练方式即可，如[10]中所示。对于跟模板的反例样本，则从反例图片中随机抽取与 F_i 大小相同的区域，提取特征作为根滤波器的负样本。

B) 合并聚类。将所有的不带子部件隐变量信息的根模板合并到一个混合模型中。在合并训练中，每个样本的混合标签和根滤波器的位置是当前考虑的隐变量。可以将这个过程的梯度下降算法看做是判别性的分类过程，先判定每个正样本的混合标签，再估计分类的平均即根滤波器的值。

C) 初始化子部件滤波器。对每个子部件的初始化，采用了一个简单的启发式算法。用一个较小的矩形区域形状作为初始的子滤波器，用贪心算法在跟模板覆盖区域中寻找具备高能量的局部区域（以区域内部权重的 L1 范数作为能量的定义表示）。为了程序的简化，假设所有的子模板或者出现在中线位置，或者以中线位置为轴存在对称的另一个子部件。一旦确定子部件后，就将其覆盖区域的能量值清零，开始寻找次高能量的区域，直到所有的子部件都确定为止。

得到了初始化模型参数之后，将采用下面的过程来训练整个模型。在训练开始之前，首先定义一些符号。设想训练样本是由包含正样本的边界框 P 和由反例图片集 N 给定的。 P 是 (I, B) ，其中 I 是图像， B 是图像 I 中标志目标的边界框。令 M 作为具有合适结构的混合模型，回忆上文，模型的参数是有一个向量 β 定义的。为了得到 β 的最优解，用包含正样本和负样本的训练集 D 定义了一个隐变量的 svm 学习问题。每个样本 x 都有相关联的图像 I 和特征金字塔 $H(x)$ 。隐变量 $z \in Z(x)$ 定义了特征金字塔 H 中 M 的一个实例。

定义 $\psi(H, z) = \varphi(H(x), z)$ ，则 $\beta \cdot \psi(H, z)$ 就是实例假设 z 在 M 中的响应得分。对正样本来说，其得分应该尽可能的高，而负样本则应尽可能的低。注意到，根据 50% 的覆盖范围， P 中能产生很多位置满足这个要求，这种情况下，将根模板在 P 中的位置也作为一个隐变量对待，能够在一定程度上弥补位置误差。模型训练的整体过程如下：

表 4-1 模型训练过程

模型训练算法:

数据:

正样本集: $P = \{(I_1, B_1), \dots, (I_n, B_n)\}$

负样本集: $N = \{J_1, \dots, J_m\}$

初始化模型: β

输出: 新模型: β

```

1   $F_n := \emptyset$ 
2  for relabel:=1 to num-relabel do
3       $F_P := \emptyset$ 
4      for i:=1 to n do
5          Add detect – best( $\beta, I_i, B_i$ ) to  $F_P$ 
6      end
7      for datamine:=1 to num-datamine do
8          for j:=1 to m do
9              if  $|F_P| \geq \text{memory} - \text{limit}$  then break
10             Add detect – all( $\beta, J_i, -(1 + \delta)$ ) to  $F_P$ 
11         end
12          $\beta := \text{gradient} - \text{descent}(F_P \cup F_N)$ 
13         remove  $(i, v)$  with  $\beta \cdot v < -(1 + \delta)$  from  $F_N$ 
14     end
15 End

```

在这个过程中，最外层的迭代实现了 $L_D(\beta, Z_P)$ 上的坐标下降。3-6 是正样本重标注步骤，将当前迭代参数下最好的实例位置添加到正样本特征向量缓存 F 中。7-14 则是 β 的最优化过程，由于由反例图片 N 隐含定义的反例特征向量数目是巨大的，在优化中使用了多次难反例挖掘的过程。每一次迭代中，收集 F_N 中的难反例，用梯度下降方式训练一个新的模型，之后将 F_N 中的简单样本移除。函数detect – best(β, I_i, B_i)寻找 B 中重叠的具有最好响应的目标假设。detect – all($\beta, J_i, -(1 + \delta)$)计算每一个根模板位置下最高的响应假设，选择所有阈值大于 t 的假设位置。gradient – descent($F_P \cup F_N$)是在正负样本特征向量缓

存下学习模型参数，特别的，在子部件位置对称的情况下，参数的学习能够有较高的效率。

在上面的模型训练方法下，采用合成的虚拟样本集训练了辅助的参考模型参数。

4.3 模型域适应

在得到辅助模型参数之后，存在的问题就是如何将辅助模型与实际的小样本数据结合，以获得能够在小样本中使用的分类模型。

在经典的模型参数迁移算法中 Adaptive-SVM 中，在预训练的辅助分类器模型下，通过一个有效的 DA 算法从先验模型的扰动方程中找到新的分类界面。即给定辅助模型参数 W_S ，最终的分器由下式表示：

$$\begin{aligned} f(\mathbf{x}) &= \max_h (f^s(\mathbf{x}, \mathbf{h}) + \Delta f(\mathbf{x}, \mathbf{h})) \\ &= \max_h (f^s(\mathbf{x}, \mathbf{h}) + \Delta W \phi(\mathbf{x}, \mathbf{h})) \end{aligned} \quad (4-1)$$

其中 Δf 称为扰动函数， $\Delta W = W_S - W_T$ 是两个域之间参数差异， W_T 则是最终得到的适应模型参数。其基本的思想就是学习与原始辅助模型近似的新决策分界面。新的决策函数可以通过解下面最优化问题得到：

$$\min_{\Delta W} R(\Delta W) + CL(\Delta W, D_l) \quad (4-2)$$

其中 R 是正则化项， L 表示了目标小样本上的损失项， C 是惩罚系数。

更一般的，上式可以表达为这个最优化方程的形式：

$$\begin{aligned} \min_{\mathbf{w}, \xi} & \frac{1}{2} \|\Delta \mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t. } & \forall i, y, h \quad \xi_i \geq 0; \forall (x_i, y_i) \in D_l \\ & y_i f^s(x_i) + y_i \mathbf{w}^T \phi(x_i) \geq 1 - \xi_i, \forall (x_i, y_i) \in D_l^t \end{aligned} \quad (4-3)$$

这里 y_i 和 h_i 是实际的类别标签和目标假设位置。 y 和 h 则表示任意输出结果中的标签和目标假设位置。正则项表明 A-SVM 将辅助模型中训练得到的参数通过规范 W_T 和 W_S 中的距离来适应新的模型，即通过调整模型参数使决策分界面朝着目标样本领域调整。

上述模型迁移方法是将所有辅助模型参数视为整体，在整体上来进行调整的，但这会限制了辅助数据集与目标样本在特征空间上要具有极其近似的分布

表示。这对于具有混合的子部件模型来说是很难做到的。首先，其并没有考虑到模型内部的本质结构知识联系。其次，当辅助数据集和目标样本在特征空间上具有较大不同的时候，即模型决策分界面差距较大，训练的效率会变得很低。

考虑到使用联合的特征图，用 $\varphi(x, h)$ 来进行结构化的学习，训练得到的超平面参数自然包含了其中的结构编码。并且相应的 DPM 模型参数能够根据混合模型以及子部件将参数向量划分为多个不同的模块。这促使了能够用一种具有结构化的措施来将前期的模型做一些域适应的调整。

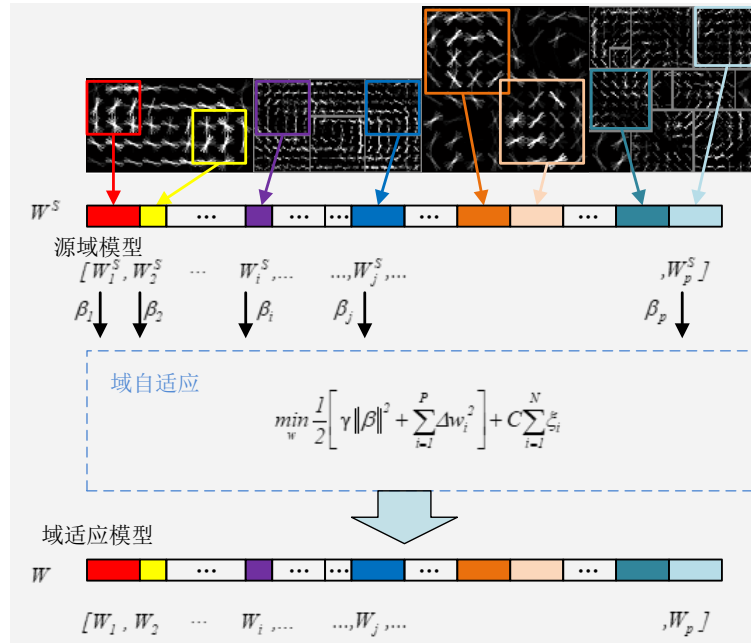


图 4-4 DPM 模型参数结构化域适应

上图表示了具有结构化的域适应方法在车辆的 DPM 模型上的划分示意。首先，利用辅助数据集训练得到辅助模型，设其参数为 W_S 。则 W_S 包含了按不同角度划分的车辆混合模型，每个混合模型都由相应的子部件组成。以将该预训练的模型参数迁移到一个不同的领域中，首先需要将参数向量分解为 P 个部分，P 是划分的混合模型数目。注意到对每个组成部分 W_{Sp} 来说，都会包含有性状相关的滤波器和变形参数。分解的模型参数通过不同的权重参数 $\beta_p \in [1, P]$ 应用到目标样本中。

决策函数可以写为如下形式：

$$f(x, h) = f^s(x, h) + \Delta f(x, h) = f^s(x, h) + \Delta w \phi(x, h)$$

$$\mathbf{w}_s = [w_{s1}, w_{s2}, \dots, w_{sp}] \quad (4-4)$$

$$\Delta \mathbf{w} = [\Delta w_1, \Delta w_2, \dots, \Delta w_p], \beta = [\beta_1, \beta_2, \dots, \beta_p]$$

$$\Delta w_i = w_i - \beta_i w_{si}$$

为了在训练中学习这些 β_p 参数，需要在目标函数中增加关于 β 的正则化项，并用参数 γ 来控制该正则项的相关惩罚系数。

定义 $\Delta \mathbf{w} = [\Delta w_1, \dots, \Delta w_p]$ ， $\beta = [\beta_1, \beta_2, \dots, \beta_p]$ ，其中 $\Delta w_p = w_p - \beta_p \cdot w_{sp}$ ， $p \in [1, P]$ 。上面 A-SVM 的正则化项可以修改为

$$R(\mathbf{w}, \beta, \mathbf{w}_s) = \frac{1}{2} (\gamma \|\beta\|^2 + \sum_{p=1}^P \|\Delta w_p\|^2) \quad (4-5)$$

对应的最优化方程则为

$$\begin{aligned} \min_{\mathbf{w}} \frac{1}{2} [\|\beta\|^2 + \sum_{i=1}^P \|\Delta w_i\|^2] + C \sum_{i=1}^N \xi_i \\ \text{s. t. } \forall (x_i, y_i, h), \xi_i \geq 0 \\ \mathbf{w}^T \phi(x_i, h_i) - \mathbf{w}^T \phi(x_i, h) \geq L(y_i, y', h') - \xi_i \end{aligned} \quad (4-6)$$

上式有两个参数需要优化， β 和 \mathbf{w} 。上面方程组的解难以直接通过现有的工具得到其最优化解。通过重新排列特征和参数的表达方式，可以将其转换为二次规划问题，之后用标准的 SVM 方法进行求解。采用了串联起来的增广向量 $\bar{\mathbf{w}} = [\Delta \mathbf{w}, \sqrt{\gamma} \beta]$ ，则有

$$\phi(x_i, h) = \left[\phi(x_i, h), \frac{1}{\sqrt{\gamma}} \theta^s(x_i) \right] \quad (4-7)$$

其中 $\theta^s(x_i) = [w_{s1} \cdot \phi(x_i, h), \dots, w_{sp} \cdot \phi(x_i, h)]$ 并且 $\phi(x_i, h)$ 代表了由候选 x 和假设 h 给出的部件 p 的特征。则(4.7)的最优化问题可以重写为下面形式：

$$\begin{aligned} \min_{\mathbf{w}} \frac{1}{2} \sum_{i=1}^P \Delta w_i' + C \sum_{i=1}^N \xi_i \\ \text{s. t. } \forall (x_i, y_i, h) \in D_i^t; \xi_i \geq 0 \\ \mathbf{w}^T \phi(x_i, h_i) + \mathbf{w}^T \phi(x_i, h) \geq L(y_i, y', h') - \xi_i, \forall (x_i, y_i) \in D_i^t \end{aligned} \quad (4-8)$$

注意到正则化项是凸函数，以及损失项也是凸函数，因此整个目标函数也是凸的。能够通过常规的方式进行求解。

部件级别的适应分析，与前面提到的 A-SVM 相比，当前算法是在部件上进

行正则化的迁移。将预训练的分类器分解为不同的子部件分类器，其最终的得分则是先验子部件分类器与扰动函数的权重之和。因此，这样的形式考虑了模型在子部件结构上的先验知识。

由上面的推导，最终本文采用了结构化分解与迁移的方式，将用辅助数据集训练的模型参数迁适应到目标小样本集中。

表 4-2 模型域适应算法

具有结构化的域适应训练算法：

输入： w_s , target 训练样本

$$w = w_s$$

Loop:

$$\text{寻找 } h_i^* = \underset{h}{\operatorname{argmax}}(w^T \phi_p(x_i, h))$$

根据找出的 h_i^* 更新 w 并解迁移学习的优化问题

判断是否达到结束阈值

4.4 实验结果与分析

4.4.1 实验方法

本文在 3.2 介绍的车辆数据集上做实验，以 DPM 作为先验模型，对迁移学习的一些效果进行了验证，并与基准模型作了对比。在数据集中随机选取了 2700 车辆样本作为参考数据，由此训练一个基准模型，并以其中的 300 个样本作为小样本目标数据集，作为域适应中的目标域数据。

检测返回结果是图像中的矩形框，实验中将与测试目标矩形框重叠率超过 50% 的检测结果作为正确，其余均为误检。

在对测试集做整体评估时，采用准确率(Precision)和召回率(Recall)来综合评判检测性能。准确率是指正确判别的样本数目在所有检测结果中所占的百分比，召回率是指结果中正确检测出的目标样本与测试集中所有目标样本的比值，计算公式如下：

$$\text{Precision} = \frac{\text{正确检测数目}}{\text{正确检测数目} + \text{错误数目}} \quad (4-9)$$

$$\text{Recall} = \frac{\text{正确检测数目}}{\text{测试集中所有目标数目}} \quad (4-10)$$

4.4.2 混合模型数目与性能分析

基于外观的特征用于检测时，在形状上的泛化能力是有限的，对同样目标的模型训练需要根据目标的外观聚类进行混合模型的训练。与行人检测中根据目标长宽比进行同类目标中的聚类情况类似，遥感视角下应按旋转的角度对目标聚类。根据 3.3 中对车辆样本在角度上的聚类分析，这里选择将所有样本按照角度做了多个划分，用于测试集中，从检测效果中寻找最佳的角度聚类，以期获得足够高性能的先验模型。随着混合模型的增加，检测所需要的时间会更多，因此还对比了检测所消耗的时间，将两者综合考虑后得到先验模型的训练参数。

表 4-3 不同混合模型下 mAP 与时间消耗

混合模型	1	2	3	4	6	9
mAP	0.737	0.9177	0.922	0.918	0.921	0.917
时间	2.15	3.61	5.09	6.88	10.79	13.9

由上表可以看出，在只使用一个混合模型时，平均准确率（mean Average Precision, mAP）较低，随着混合模型的数目增多，mAP 呈现上升状态，但在超过 3 时，基本不再有太大变化。同时，随着混合模型的数目增加，检测消耗的时间呈现增多的趋势。综合考虑下，可以选择 3 个或者更多的混合数目进行先验模型的训练。

4.4.3 域适应模型实验结果与分析

通过前面的工作建立了源域车辆数据集，采用上面的方法学习到了具有 6 个混合类别的车辆先验模型。在此基础上，根据参数结构化域适应方法，用目标数据对先验模型重训练，得到目标域的模型。下图展示了域适应后的检测结果对比：

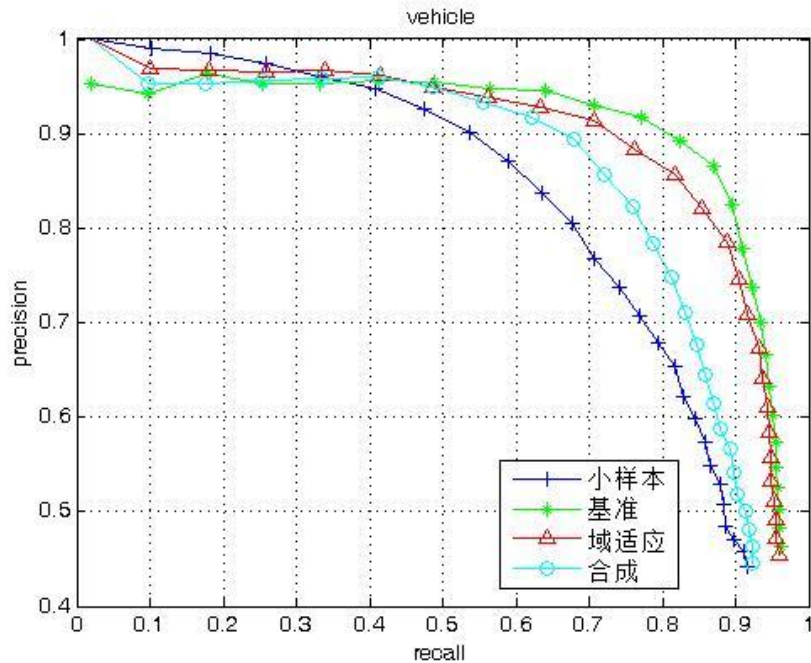


图 4-5 域适应后的模型性能对比

其中，小样本模型指的是车辆数据集中的随机 300 个小样本训练得到的模型，基准模型采用了 2740 样本，合成模型是用 2700 虚拟样本训练的源域先验模型，域适应模型则为结合源域模型与 300 个小样本后训练得到的模型。

在上一章对小样本和合成样本的特征可视化分析中，合成样本与小样本在特征空间中是有一定交集的，并且合成样本包括了许多小样本不具备的特征信息。以合成样本的先验模型作为基础对小样本进行域适应的再学习，使小样本中所具有的目标数据域细节信息可以参与到模型训练中，并在一定程度上拟合了合成样本的分布情况，最终得到的模型在性能上比两者都要好，接近作为基准的大样本所训练的模型性能。域适应的模型并没有达到基准的性能，原因在于合成样本的分布并没有能够覆盖整个目标域的数据，作为先验模型提供的信息不足以使域适应模型具有泛化到整个目标域的能力。

4.5 本章小结

本章主要介绍了基于域适应算法的小样本目标检测的研究框架。根据目标域的样本分析，合成源域数据集作为迁移学习的先验模型训练集。之后将学习到的先验模型用于域适应算法中，由小样本数据集重训练得到适应性的模型，

并在给定的测试数据集上做了实验验证。

首先从通用目标检测流程出发，分析了遥感目标检测中的小样本存在的问题。在现有的分类器下，必须有大量能够描述整个样本特征分布的数据集时才能具有很好的性能，而小样本训练得到的模型不具备良好的泛化能力。为了解决这个核心问题，提出从具备先验知识的源域数据中进行参数域适应的研究方法框架。其次，在源域数据集学习先验模型时，通过对目标样本角度参数分割的探讨，寻找检测性能与检测时间之间的平衡点。之后，以源域数据集中建立先验检测模型为起点，并根据模型子部件的特征，采用具有参数结构化的域适应算法在小样本目标数据中进行重训练，获得目标域适应的检测模型。

最后一节给出了前面部分相关的实验，从实验结果可以看出，本文提出的检测框架能显著提高小样本条件下的检测性能，甚至接近作为基准的完备样本训练的检测模型。

第5章 结论与展望

遥感目标检测作为目标检测的子课题，是计算机视觉和机器学习领域中重要的研究内容。近年来，随着遥感技术的快速发展，提出了许多具有挑战性的问题。许多机构和学者对现有的检测算法做了大量适应性研究工作，将众多目标检测方法成功地应用到遥感领域分析中，推进了遥感事业的发展。然而，由于实际遥感影像的各类目标样本不易获取，小样本问题影响着传统方法的推广使用。本文提出使用模型域适应算法检测框架，探索了辅助数据集的获取，建模和模型域适应方法。来解决光学遥感小样本的模型训练问题，并获得与大样本集相近的检测性能。

本文首先简述了遥感目标检测的研究背景和应用意义，在此基础上简述小样本遥感目标检测的重要性，并总览国内外近年来的研究现状和研究方法，同时比较了这些方法的优势与劣势。在第二章中，详细介绍了图像特征提取、目标表示方法、隐变量模型最优化以及模型域适应的方法。第三章介绍了本文的主要工作，本文提出了一个小样本目标检测框架的流程。从分析实际遥感数据集开始，利用车辆的三维模型生成与目标数据集近似的辅助虚拟数据集，并利用高维数据可视化工具在特征空间验证了生成方法的可行性。在此基础上，利用特征可视化分析样本聚类的最佳角度，再使用基于 HOG 特征的 DPM 模型对辅助数据集建模，得到域适应中的源域模型。在向小样本目标域进行迁移时，针对 DPM 模型参数的拓扑结构划分，采用具有子部件结构化的参数划分，使模型在域适应中更加合理。最后，在车辆数据集上进行了相应的实验，结果表明，本文所使用的检测框架得到的效果远比直接使用小样本训练模型好，接近于大样本下的训练模型。

本文在高清遥感小样本的目标检测中取得了一定的阶段性成果，主要贡献总结如下。一是利用人造目标的形状特性，通过重建其三维模型，结合相应背景图片，虚拟生成特定场景下的大量数据。之后采用模型域适应的方法，将虚拟数据集训练的模型迁移到目标域。通过验证这种虚拟数据到目标领域框架的有效性，为后续遥感目标检测提供了解决思路，能够用于无监督目标发现的模型初始化以及样本自动标注。二是在模型训练算法中，将 HOG 特征金字塔的上下两层进行调整，使尺度变化较大的目标也能用于训练。模型域适应算法

中，利用 DPM 模型的拓扑结构，将结构化的域适应方法用于该模型，实现模型的重训练。但因为时间关系，依然存在不足。在特征方面，只尝试了 HOG 等浅层特征，并没有充分利用深度特征的分类性能优势，以及具有方向校正的深度网络。

未来遥感场景下的小样本目标检测算法进一步研究的方向可能会包括如下方面：域适应的算法中，会将适应部分前移，把当前在模型上的适应过程直接提前到样本或者样本特征中，直接生成目标域样本；在虚拟生成数据的基础上，建立自动化的标注与分类检测框架；提取更具表达能力的鲁棒性深度特征，以及深度神经网络在遥感场景的特化精简，在遥感成像设备中实时训练与检测。

参 考 文 献

- [1] Cheng G, Han J. A Survey on Object Detection in Optical Remote Sensing Images[J]. *Isprs Journal of Photogrammetry & Remote Sensing*, 2016, 117:11-28.
- [2] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. *Computer Vision and Pattern Recognition*. IEEE, 2015:1-9.
- [3] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]. *International Conference on Neural Information Processing Systems*. Curran Associates Inc. 2012:1097-1105.
- [4] Cheriyyadat A M. Unsupervised Feature Learning for Aerial Scene Classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(1): 439-451.
- [5] Yao J, Zhang Z. Object detection in aerial imagery based on enhanced semi-supervised learning[C]. *International Conference on Computer Vision*, 2005: 1012-1017..
- [6] Zhong L, Wu J, Tang X, et al. The Semi-supervised Classification Method of LiDAR Data Integrating with Aerial Images[J]. *Information Science & Engineering International Symposium*, 2010, 15(4):386-390.
- [7] Han J, Zhang D, Cheng G, et al. Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-Level Feature Learning[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2015, 53(6):3325-3337.
- [8] Kan M, Shan S, Chen X. Bi-Shifting Auto-Encoder for Unsupervised Domain Adaptation[C]. *IEEE International Conference on Computer Vision*. IEEE, 2016:3846-3854.
- [9] Van d S K E A, Uijlings J R R, Gevers T, et al. Segmentation as selective search for object recognition[C]. *International Conference on Computer Vision*. IEEE Computer Society, 2011:1879-1886.
- [10] Cheng M M, Zhang Z, Lin W Y, et al. BING: Binarized Normed Gradients for Objectness Estimation at 300fps[C]. *Computer Vision and Pattern Recognition*. IEEE, 2014:3286-3293.
- [11] Zitnick C L, Dollár P. Edge Boxes: Locating Object Proposals from Edges[C]. *European Conference on Computer Vision*. Springer, Cham, 2014:391-405.
- [12] Jain A K, Ratha N K, Lakshmanan S. Object detection using gabor filters[J]. *Pattern Recognition*, 1997, 30(2):295-309.
- [13] Wang X, Han T X, Yan S. An HOG-LBP human detector with partial occlusion handling[J]. 2009, 30(2):32-39.
- [14] Lowe D G. Object recognition from local scale-invariant features[C]. *International Conference on Computer Vision*. IEEE Computer Society, 1999:1150.
- [15] Viola P, Jones M, Snow D C, et al. Detecting Pedestrians Using Patterns of Motion and Appearance[J]. *International Journal of Computer Vision*, 2005, 63(2): 153-161..
- [16] Hyun J, Baek J, Kim J, et al. Proposing a fast circular HOG descriptor for detecting rotated

- objects[C]. International Joint Conference on Neural Networks. IEEE, 2015:1-6.
- [17] Skibbe H, Reiser M. Circular Fourier-HOG features for rotation invariant object detection in biomedical images[C]. International Symposium on Biomedical Imaging. IEEE, 2012:450-453.
- [18] He S, Lau R W H. Oriented Object Proposals[C]. International Conference on Computer Vision. IEEE, 2015:280-288.
- [19] Felzenszwalb P. Object detection grammars[C]. International Conference on Computer Vision Workshops. IEEE, 2011:691-691.
- [20] Donahue J, Jia Y, Vinyals O, et al. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition[J]. Computer Science, 2013, 50(1):815-830.
- [21] Zhu H, Chen X, Dai W, et al. Orientation robust object detection in aerial images using deep convolutional neural network[C]. International Conference on Image Processing. IEEE, 2015:3735-3739.
- [22] Zhou Y, Ye Q, Qiu Q, et al. Oriented Response Networks[C]. Computer Vision and Pattern Recognition. IEEE, 2017.
- [23] Tuzel O, Porikli F, Meer P. Pedestrian detection via classification on Riemannian manifolds.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2008, 30(10):1713-27.
- [24] 邓乃扬, 田英杰, 数据挖掘中的新方法-支持向量机[M]. 北京:科学出版社, 2004
- [25] Viola P, Jones M. Robust real-time face detection[C]. international conference on computer vision, 2001, 2(2): 747-747.
- [26] Davies W G. Sensation and Perception[M]. Sensation and perception. Brooks/Cole Pub. 1996:407.
- [27] Schapire R E. A Brief Introduction to Boosting[C]. Sixteenth International Joint Conference on Artificial Intelligence. Morgan Kaufmann Publishers Inc. 1999:1401-1406.
- [28] Hearst M A. Support Vector Machines[J]. IEEE Intelligent Systems & Their Applications, 1998, 13(4):18-28.
- [29] Felzenszwalb P, Mcallester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]. Computer Vision and Pattern Recognition. IEEE. 2008:1-8..
- [30] Felzenszwalb P F, Girshick R B, Mcallester D, et al. Object Detection with Discriminatively Trained Part-Based Models[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(9):1627-1645.
- [31] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]. Computer Vision and Pattern Recognition, 2014: 580-587.
- [32] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
- [33] Baktashmotlagh M, Harandi M T, Lovell B C, et al. Domain Adaptation on the Statistical Manifold[C]. Computer Vision and Pattern Recognition. IEEE. 2014: 2481-2488
- [34] Pan S J, Tsang I W, Kwok J T, et al. Domain Adaptation via Transfer Component Analysis[J]. IEEE Transactions on Neural Networks, 2011, 22(2): 199-210.

- [35] Girshick R, Iandola F, Darrell T, et al. Deformable part models are convolutional neural networks[C]. Computer Vision and Pattern Recognition, 2015: 437-446.
- [36] Tsai Y H, Yeh Y, Wang Y F, et al. Learning Cross-Domain Landmarks for Heterogeneous Domain Adaptation[C]. Computer Vision and Pattern Recognition, 2016: 5081-5090.
- [37] Vazquez D, Ramos S, Lopez A M. Domain Adaptation of Deformable Part-Based Models[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36(12):2367-2380.
- [38] Ponsa D, Geronimo D, Marin J, et al. Virtual and Real World Adaptation for Pedestrian Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 36(4):797-809.
- [39] Ghifary M, Kleijn W B, Zhang M, et al. Deep Reconstruction-Classification Networks for Unsupervised Domain Adaptation[C]. European Conference on Computer Vision. Springer, Cham, 2016:597-613.
- [40] Caruana R. Multitask Learning[J]. Machine Learning, 1997, 28(1):41-75.
- [41] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. Computer Vision and Pattern Recognition, 2005: 886-893.
- [42] Yang J, Yan R, Hauptmann A G. Cross-domain video concept detection using adaptive svms[C]. ACM International Conference on Multimedia. ACM, 2007:188-197.
- [43] Der Maaten L V, Hinton G E. Visualizing Data using t-SNE[J]. Journal of Machine Learning Research, 2008: 2579-2605.

致 谢

转眼间，在中国科学院大学攻读硕士学位的三年学习生活即将结束。与刚入学相比，我收获了不仅仅是专业上的知识，还有良师益友，还有自信与勇气。在毕业论文的尾声，在此借这个机会衷心地感谢我的老师、实验室的兄弟姐妹以及我的家人。

本课题的研究工作及学位论文是在焦建彬教授、叶齐祥教授、韩振军副教授、秦飞副教授的悉心指导下完成的。首先，我要感谢焦老师给予珍贵的学习机会，感谢在我硕士期间每一次的指导和鼓励；感谢叶老师在我研究的选题、框架设计、论文撰写等方面提出指导和建议，至今我仍记得您曾说，对待研究和工作要有一种无所畏惧的勇气，必须盯着目标去做，否则很多的杂事会一直干扰科研的进度，感谢您在我对自己毫无信心的时候给予了鼓励；感谢韩老师在理论知识和论文写作的过程中给予的指导和帮助，以及您在生活中对每个学生无微不至的照顾；感谢秦老师用严谨的思考和认真的态度以及开阔的眼界影响着我。衷心地感谢各位老师三年来对待学生的真心实意，用实际行动诠释春风化雨，使得实验室的每一位学生都在愉快而严谨的氛围中学习和成长。

在学三年半，我同样要感谢朝夕相处的实验室的兄弟姐妹们。首先，感谢陈孝罡师兄、邹佳凌师兄、魏鹏旭师姐、崔妍婷在我刚进入实验室的时候，通过不断的小组讨论指导我学习计算机视觉和机器学习领域的基础知识；感谢朱海港师兄、戴蔚群师姐在我课题研究一筹莫展的时候，与我一同探讨。感谢晓丹师姐、高山师兄、柯炜师兄以及其他师兄师姐，给予我学习上的指点，生活中的无私帮助。感谢我同届好友刘嫣然、赵恒、施梦楠，一起克服困难，也一起度过美好的时光。感谢实验室的每一位成员，因为有你们，我的三年生活才能如此多彩斑斓。

特别要感谢我的家人，在我失去方向处于迷茫的几年中，给予我极大的自由去寻找生活的意义。感谢过去的自己，即使在暗淡中也能怀着期望去不断尝试，使现在的我能够有拥抱新生活的勇气和信念。

感谢参与我开题和中期评审的每一位老师，你们丰富的经验和无私地帮助我把握论文方向和研究进度。

黄显淞
2017年5月

个人简历、在学期间发表的论文与研究成果

姓名：黄显淞 性别：男 出生日期：1988年06月05日

教育经历：

- 2008年9月至2012年7月 中国科学技术大学 精密机械与精密仪器 学士
- 2014年9月至2017年7月 中国科学院大学 工业工程 硕士

工作经历：

- 2012年7月至2013年8月 深圳市普联技术有限公司 SMT 工艺工程师

研究成果：

- 《一种基于域自适应算法的小样本遥感目标检测方法》专利（已申请）