

专著 学术专著
信号与信息处理

ZHUANZHU

局部多核学习与图像分类

韩一娜 著

ZHUANZHU

西北工业大学出版社

学术专著·信号与信息处理

局部多核学习与图像分类

韩一娜 著



西北工业大学出版社

【内容简介】 本书系统地论述了局部多核学习的理论、优化算法及其在图像分类中的应用。全书分为9章,主要叙述了图像分类中的核方法,局部多核学习的正则风险框架与算法,局部多核学习的贝叶斯视角及其在构建生物视觉模型中的应用。书中融入了作者多年的研究成果,同时也吸纳了国际上在该领域的最新研究成果。本书对问题叙述详尽,理论分析深入浅出,是该领域一部内容新颖、系统全面、实用性强的学术著作。

本书可供从事局部多核学习理论及应用的研究人员阅读和参考。

图书在版编目(CIP)数据

局部多核学习与图像分类/韩一娜著. —西安:西北工业大学出版社,2014.2
ISBN 978-7-5612-3934-6

I. ①局… II. ①韩… III. ①信息处理系统—研究 IV. ①TP391

中国版本图书馆 CIP 数据核字(2014)第 030982 号

出版发行:西北工业大学出版社

通信地址:西安市友谊西路 127 号 邮编:710072

电话:(029)88493844 88491757

网址:www.nwpup.com

印刷者:陕西向阳印务有限公司

开本:727 mm×960 mm 1/16

印张:15.25

字数:273 千字

版次:2014 年 10 月第 1 版 2014 年 10 月第 1 次印刷

定价:45.00 元

前 言

图像分类是根据图像中出现的特定视觉内容,对其在语义层进行类别划分的过程,有着广泛的应用价值。美国麻省理工学院 Freeman 教授指出,图像分类中最为本质的问题就是设计性能优良的图像表征。但是,在现实世界中,图像表征往往呈现出同类间的多样性与不同类间的相关性,这使得图像分类问题的解决仍然极具挑战性。

局部多核学习的思想,即根据多种图像表征在当前分类任务下对每一幅图像的区分能力进行分类,并对其进行样本级的最优组合,可极大地减少同类间的多样性与不同类间的相关性对图像分类的负面影响。且近年来多核学习的理论与优化算法设计均取得了长足的进步,因此,采用局部多核学习这种特例来解决现实世界图像分类这一极具挑战性的问题成为可能。

尽管国内外已有很多优秀的书籍分别讨论了图像分类或核学习,但是至今还没有一本著作专门针对局部多核学习的理论与方法及其在图像分类中的应用进行全面而系统的阐述。本书的初衷是从不同的视角展示近年来局部多核学习领域的发展,即这一新兴技术在不同理论体系下与不同形式下的本质联系。同时,深入揭示局部多核学习与图像分类之间的关系。

本书不仅叙述了国际上现有的局部多核学习方法,而且同时还纳入了著者近年来在国际重要期刊上发表的多篇学术论文。书中融入了著者在西安交通大学攻读博士学位和在法国国立高等邮电学院(Telecom Paris Tech)接受联合培养博士期间的大部分研究成果,以及在西北工业大学开展教学与研究工作以来进一步取得的科研成果。

本书涉及的许多研究成果是在西安交通大学电子与信息工程学院长期从事科研工作积累的基础上完成的,在此对学校有关部门、有关老师和同学深表感谢。

本书在写作过程中,得到了西北工业大学航海学院马远良院士、杨益新院

长和杨坤德教授的支持与鼓励,在此表示深深的感谢。本书涉及的研究工作还得到了中科院声学所声场声信息重点实验室基金(SKLA201201)与国家自然科学基金(61301198)的资助。

限于著者的水平与经验,书中难免存在一些疏漏,恳请专家、同行和读者批评指正。

著 者

2013 年 11 月

目 录

第 1 章 绪论	1
1.1 研究背景与意义	1
1.2 基于多核学习的图像分类	4
1.3 研究现状	7
第 2 章 图像表征	17
2.1 场景要点	17
2.2 方向梯度直方图	21
2.3 尺度不变特征变换	25
2.4 直线特征	33
2.5 自相似性描述符	40
2.6 纹理元	43
2.7 几何概率映射	48
第 3 章 图像分类中的核方法	55
3.1 图像表征与核	55
3.2 风险最小化学习	62
3.3 核与线性分类器	67
3.4 支持向量机分类器	75
第 4 章 局部多核学习的正则风险框架	81
4.1 多核学习的正则风险框架	81
4.2 流行的多核学习形式	84
4.3 局部多核学习的正则风险框架简介	89
4.4 图像数据库	92

4.5	核矩阵构造及图像特征	94
4.6	实验设置与评价准则	96
第 5 章	基于样本级优化的方法	98
5.1	原-对偶等价形式	102
5.2	样本级的交替优化	103
5.3	样本级的目标函数求解	105
5.4	实验结果与分析	107
第 6 章	基于半定规划的方法	124
6.1	l_p 范数 LMKL 框架	124
6.2	单约束 QCQP 的 SDP 松弛	126
6.3	SDP 求解局部化 γ	129
6.4	实验结果与分析	132
第 7 章	基于门函数的方法	147
7.1	Gonen & Alpaydin 的方法	147
7.2	分组敏感的方法	151
7.3	基于概率置信核的方法	155
7.4	实验结果与分析	161
第 8 章	局部多核学习的贝叶斯视角	174
8.1	贝叶斯框架	174
8.2	高斯过程	179
8.3	多核学习的贝叶斯视角	185
8.4	局部多核学习的贝叶斯视角简介	194
第 9 章	局部多核学习在构建生物视觉模型中的应用	198
9.1	引言	198
9.2	场景分类	200

9.3 任务引导的生物启发式 Gist 模型(BT - Gist)	202
9.4 数据库与实验设置	206
9.5 实验结果与分析	208
9.6 与现有方法的比较	214
9.7 BT - Gist 模型讨论	219
参考文献	223

第 1 章 绪 论

1.1 研究背景与意义

图像分类是指根据图像中出现的特定视觉内容,对其在语义层进行类别的划分^[1]。图像分类是对图像理解或图像识别的延伸,它们通常需要解决相同的问题。图像分类或图像理解的中心问题即是识别图像中出现的目标物^[2]或图像所呈现的场景^[3]。

1.1.1 图像分类的意义

识别能力是人类视觉系统中最有价值的功能之一。例如,人眼瞬间便能够识别出目标对象的质地是大理石还是橘皮,表面是粗糙的还是光滑的,目标物是汽车还是柳树,所处环境是森林还是厨房。而且人眼既能够区分同一类别内的不同个体,如自己的卧室或他人的卧室,又能识别不同的类别,如人脸和青蛙^[1]。人类 6 岁时便能识别出 10 000 多类目标^[4]。视觉识别能力对于人类的生存至关重要,因此,在过去的数千万年间,人类进化出如此善于识别却又高度复杂的神经认知系统^[2]。

近年来,随着数字成像技术的快速发展和成本的降低,人们能够廉价地获取大量以电子形式存储的数字图像,且数字图像已成为一种广泛应用的媒体。同时,计算机技术与互联网的兴起极大地加快了包括数字图像在内的各种信息的传播速度,人们可访问和获取的图像数量呈现爆炸式增长。在这样一个“信息富集的时代”,如果能够为计算机赋予一定的识别能力,必将深刻地影响人类生活的方方面面。利用计算机进行图像分类,通常是指根据图像的某种视觉表征/特征(如像素、纹理、颜色等)以一定的准则将其划分到预定义的目标物或场景类别。这既是人工分类的延续与发展,也是自动提取视觉语义信息的重要方式,有着许多潜在的应用。

(1)视频搜索。许多电视台一年 365 天且每天 24 小时不间断地向因特网输送视频数据。针对如此海量的视觉信息,除非工作人员亲自观看这些视频,并对其内容进行繁杂的手工标识外,目前尚无有效的搜索方法。如果计算机

能够自动完成上述工作,则能帮助人们更为方便有效地搜索视频数据,如:找出自己最喜欢的节目或者演员。

(2)网络搜索。因特网图像搜索引擎是目前在网络上找寻图像数据的唯一途径。但其性能至今仍不十分令人满意,究其原因,主要还是因为现有搜索是基于图像名称或者图像周围的 HTML 信息来进行的。在进行图像搜索时,更为自然的方式应该是基于图像自身的视觉内容。考虑到直接描述一幅拟查询图像的视觉内容存在相当的难度,所以用户可首先选择少量几幅样本,使其在本质上类似于所需的图像;接着,通过这几幅样本来训练得到一个视觉模型,使其能够有效地用于搜索。因此,所需要的是一个能够从样本中学习图像视觉特征的模型。

(3)图像数据库搜索。越来越多的人拥有越来越多的家庭照片,而这些照片往往被松散地组织并存放于计算机中。例如,如果想要搜到所有关于“爸爸”或“妈妈”的照片,没有前期细致的手工标注是无法顺利进行的。此外,很多公司也希望能够在其所拥有的大量图像档案中进行有效的搜索。

(4)生物特征识别。随着计算机技术及网络技术的发展,信息安全彰显出前所未有的重要性。身份识别是保证安全的前提,而传统的方法存在诸多隐患。比较而言,生物特征具有很高的安全性。当前常见的生物特征识别技术主要包括人脸识别、虹膜识别、指纹识别、步态识别等。图像分类技术在这些领域中已有广泛的应用。

(5)在线约会。目前的在线约会网站仅能根据手动录入的文本信息(如发色、眼睛颜色等)来帮助搜索特定人物。更为实用的搜索形式应该利用个人向网站所提交的照片信息。例如,在搜索某个魅力人物时,可以搜索那些看起来像他或她的人。这需要建立你(或者与你具有相似品位的人)认为有魅力的人物模型。

(6)安全监控。在安全监控视频流中查询人物或交通工具。许多大城市都装有成千上万台闭路电视摄像机。对此,需配备大量的工作人员进行仔细的镜头检查。如果采用自动监控系统则不会出现目前因人力监控镜头的疲倦感或是精神不集中而造成的误判或漏检现象。

(7)机场行李检查。目前,主要利用 X 射线仪对飞机乘客的行李进行检测。每小时需检测大量行李,以人工的方式对其 X 射线图进行解释必将带来较高的误检率。设计能够自动识别诸如刀子、枪支和炸药等物品的视觉识别系统,必将进一步提高民航业的安全性。

(8)车辆安全。汽车制造商希望能够为他们的汽车安装具有一定视觉能

力的装置,即能够察觉其他车辆或者行人的靠近或者意识到已进入车道。具有这种视觉能力的车辆能够为司机报警或者自动避免事故,从而有助于减少目前高发的交通事故率。

1.1.2 图像分类的挑战

人类不但能够轻而易举地识别出一组由有限个特定视觉部件紧凑组织而成的几何布局,即目标物;还能毫不费力地识别出由若干个实体(如汽车、房子、建筑物、人脸、墙壁、树木、森林、岩石等)以某种不可预测的共生形式和组织形式所构成的空间布局,即场景^[3,5]。这种看似自然的能力极大地掩盖了隐藏于其中异常复杂的处理机制。事实上,不同的拍摄角度、混乱的背景、潜在遮挡、变化的光照条件以及同一类别目标所具有的视觉多样性等内、外因素极大地改变了图像数据的视觉呈现。一方面,同一类别的图像具有视觉属性上的语义不确定性;另一方面,不同类别的图像又具有底层视觉属性(如颜色、纹理和形状等)上的某些相关性。图1-1为8幅来自文献[6]数据库的场景图像。以类别“郊外”为例,一方面,同属于该类的图像样例具有不同的视觉呈现;另一方面,其他类的图像样例(如海岸、山脉、森林和高速公路)又呈现出与某些“郊外”图像相似的视觉属性。存在于同类图像间的视觉多样性与不同类图像间的视觉相关性正是利用计算机进行图像分类的困难所在。但是,人类却非常善于忽略那些变化无常的情况,以可靠地感知潜在的“语义一致性”属性,进而准确无误地识别它们。

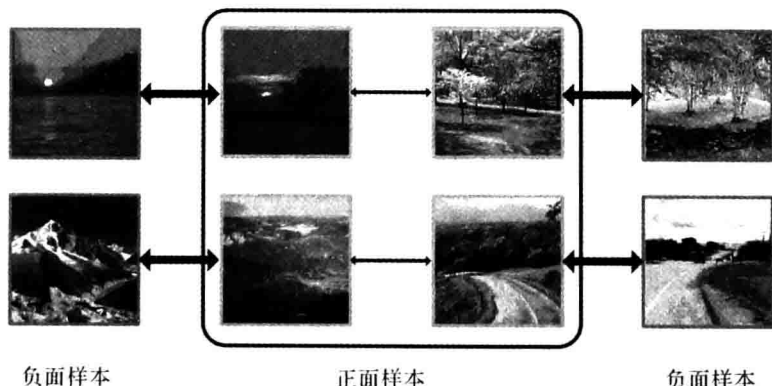


图1-1 关于场景“郊外”的类别内的多样性与类别间的相关性说明
(双向箭头代表了图像间的视觉相关性; 较粗的箭头表示更强的视觉相关性)

1.1.3 分类中的图像表征

执行图像分类任务时,常常要预定义一组类别,并给出相应的图像样本以组成训练集合。接着将每一幅图像实例表示为某种视觉特征,即图像表征。分类系统希望将手中类别未知的图像表征与大量类别已知的训练图像表征进行学习与比对,进而有效地推理出图像在目标级或场景级的类别。近年来,图像分类技术取得了巨大的发展,主要表现在各种新式的图像表征^[7]。那么,如何衡量一个图像表征的优劣呢? Freeman 指出^[8],在理想情况下,一个好的图像表征应以相同的数值来描述同一对象在不同条件下的呈现,例如,对同一片森林以不同的角度所成的像;类似地,在描述同一类别的不同实例时,例如,两幅不同的森林场景,也应该尽量保持数值上的相似。而对于不同类别的图像,相应的理想表征应具有足够大的数值差异以将它们区分开来。因此,针对图像分类中存在的“同类间的视觉多样性”和“不同类间的视觉相关性”^[9]这一挑战性问题,理想的图像表征应能有效地兼顾“类别内的不变性”和“类别间的区分性”,即捕获图像多样呈现下的“潜在语义一致性”,从而利用这些表征进行可靠的语义识别。

1.2 基于多核学习的图像分类

1.2.1 图像表征

图像分类中最为本质的问题即是设计性能优良的图像表征。现有工作为我们提供了许多成功的“手工制作/编码”的图像表征^[10]。回顾这些表征,其最本质的差异在于它们在保持类别内的不变性与刻画类别间的区分性上所取得的折中^[11]。例如,图像中的某一小块像素在使用标准欧氏距离进行比较时,几乎没有不变性,但是却具有极高的区分性;相应地,另一个极端情况则是以某个常量为表征,其具有绝对的不变性,但是却没有任何区分能力。绝大部分的图像表征具有某种确定的最优折中,置于由上述两种极端情况所界定的范围之间的某处。

但是,不变性与区分性的最优折中取决于当前特定的分类任务,并随训练数据和先验知识的变化而变化。因此,没有任何一种图像特征能够保证以最优的性能执行所有的图像分类任务^[11]。以对数字 6 与 9 的分类为例,在旋转不变性表征下,两者将被混淆。此时,完全的旋转不变性特征并不是一个好的

选择。如果分类任务为区分4和9这两类数字,其中这两类数字可能被旋转到任意方向,那么采用具有完全旋转不变性的特征则是可取的。但是,事实上4和9也是极易混淆的两类数字。因此,如果训练集足够丰富,即包含了大量不同方向的样本,那么要对两者进行卓有成效的分类还需回到具有较强的区分性而较弱的不变性的视觉表征上来。因为,在此情况下,丰富多样的数据本身会提供旋转不变性,所以即使采用随旋转而变的特征进行最近邻匹配也可取得令人满意的分类性能。如果照此进行,虽然针对某特定的分类任务通过手动调节可获得最优的视觉表征,但是一旦训练数据发生变化,该表征将不再是最优。

有时也可通过直观的视觉观察获得关于某个特定分类任务大致所需的不变性与区分性,但是要准确地指出其精确的折中值仍然存在相当的困难。回到对数字4和9的分类。如果训练数据仅由对这两类数字旋转所得的各种样本组成,那么一个泛泛的结论是旋转不变性特征适于这一分类任务。但是,如果对某些旋转后的样本再进行少许的尺度伸缩,使其恰好开始在这两类数字间产生混淆。此时,似乎应该考虑具有相似性或仿射不变性的视觉特征。但是,事实上这可能导致更差的分类性能,因为相似性与仿射不变性的区分能力都要比纯粹的旋转不变性特征低得多。

若要描述图像中一组由有限的特定视觉部件拼凑组织而成的几何布局,即目标物,和由若干个实体(如汽车、房子、建筑物、人脸、墙壁、树木、森林、岩石等)以某种不可预测的共生形式和组织形式所构成的空间布局,即场景^[3,5]的区分性与不变性则更为困难。一种理想化的解决方案是为每一个特征都设置一个连续可调的参数,来控制相应特征的不变性与区分性程度。根据当前的分类任务,首先应确定大致所需的不变性与区分性的折中范围,适当调节相应参数以生成一组有限的基础表征。然后从该集合中选出与分类当前任务所需的最佳折中程度相对应的基础特征。对于图像分类而言,在由最佳特征所定义的特征空间中,图像间具有较小的类内距离和较大的类间距离。遗憾的是,绝大部分的视觉特征不具有如此连续可调的参数。

尽管如此,从具有不同折中级别的图像特征出发,总可以构成一组“离散采样”图像中所存在的不变性与区分性的基础特征。对这些基础特征进行适当的组合即可获得最优描述符。考虑到视觉分类任务的复杂性,采用先验准则为各个基础特征手动地指定组合权值,即使可行也存在巨大的实践障碍。要能处理不同的分类任务,应该采用一种更为灵活的机制。目前,学习机制已被广泛地应用于更高层的语义分类任务当中,在此是否能将它引入到图像表

征中来呢?即通过训练样本上的学习得到最适于当前分类任务的表征组合。

1.2.2 多核学习

近年来,鉴于核学习机制所具有的良好泛化性和可保证的计算可行性,被广泛地应用到数据量巨大、具有重要价值的领域,如计算机视觉^[9,11,12]。

核是一个函数 k , 对于任意的 $\mathbf{x}_i, \mathbf{x}_j \in X$, 满足

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle \quad (1-1)$$

其中, X 为图像表征 $\mathbf{x}_i$ 所取自的非空集合。

核的思想推广了输入空间的标准内积,从而能够利用线性学习器来学习某种非线性关系。首先根据需要,选择某种非线性映射 φ , 将原始的特征集合映射到某非线性集合:

$$\begin{aligned} \varphi: X &\rightarrow H \\ \mathbf{x} &\rightarrow \varphi(\mathbf{x}) \end{aligned} \quad (1-2)$$

接着,在映射后的非线性空间应用线性学习器。线性学习器的一个重要特性是可以表示成对偶形式,即训练点的线性组合。因此,决策函数可以用测试点和训练点的内积,即核来表示。学习虽然是在内积空间进行的,但是核的使用使我们不再需要了解潜在映射 φ 的具体形式,其以一种规范的方式从简单的线性机制中获取复杂的学习机制。

基于核的学习为我们提供了一种简单而巧妙的模块化方式,将学习最终判决器的算法和理论与应用领域的特性分离开来。一方面,对于学习最终判决器的问题,核定义了恰当的正则化项;另一方面,对于具体的应用领域,通过核的设计定义了特征间的相对位置,即指定了特征空间的几何特性。在核学习发展的十余年当中,前者已经取得了实质性进步;而对于提高性能至关重要的后者,却往往需要用户进行手动调节或提前确定。以本研究所特别关注的图像分类为例,给定 N 幅图像 $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, 每一幅图像 $\mathbf{x}_i$ 可由多种特征 $\mathbf{x}_i = \{\mathbf{x}_i^1, \mathbf{x}_i^2, \dots, \mathbf{x}_i^M\}$ 来表征。那么将各个特征分别与相应的距离度量 d_m , $m=1, 2, \dots, M$ 相结合,总可以为每一种特征构建一个数据 X 上的对称正定矩阵来描述数据间的相似性:

$$\mathbf{K}_m(\mathbf{x}_i^m, \mathbf{x}_j^m) = \exp(-\gamma d(\mathbf{x}_i^m, \mathbf{x}_j^m)) \quad m=1, 2, \dots, M \quad (1-3)$$

根据 Mercer 定理可知, $\mathbf{K}_m$ 为核矩阵的充分必要条件是满足半正定性,所以上述构建的 $\mathbf{K}_m$ 为核矩阵。相比较于为任意分类任务自动构造出完美的核,一种更为可行的选择为,通过训练数据从一组候选核中自动选出满足当前分类任务需要的那一个或多个核的组合。这便是通常意义下的核学习问题。

在绝大多数情况下,首先在训练阶段针对多个已知特征,准备一组与之相应的核以组成备选核集合,即 $\{K_m\}_{m=1}^M$,这些备选核被称为基核。接着通过比较它们在一组确认集上的泛化性能来进行核的选择,如交叉确认的方法。但是,是否可以通过限定问题的一般性,来获得一个数学上既正确合理而又更为灵活的解决方案呢?近年来的研究指出,与单个核相比,采用多个核可以增强决策函数的可解释性并提高性能^[13]。对此,一种便捷而直接的方式即是考虑多个基核的组合,即

$$K(x_i, x_j) = f_\gamma(\{K_m(x_i^m, x_j^m)\}_{m=1}^M \mid \gamma) \quad (1-4)$$

式中, f_γ 为组合函数,其可以是线性的也可以是非线性的; γ 为核组合参数。

由式(1-4)可知,学得核 K 在某一具体的组合方式下(如:相加或相乘),唯一指定了与之相对应的特征空间。虽然无法确知该表征与已知表征的具体关系形式,但是通过核 K 却能够完全指定该特征空间的全部几何信息。因此,核学习隐含地学习了与之相对应的图像表征。

1.3 研究现状

1.3.1 图像表征研究现状

对图像最简单而直接的表征是采用图像本身的灰度与颜色分量。受心理学研究的启发,即人类视觉系统对于图像分辨率的降低有着非凡的包容能力^[14],Torralba^[14]等人直接将图像表征为与其相应的 32×32 低分辨率的小图,以此研究图像本身在目标分类与场景分类中的价值。该表征极大地降低了图像维数,从而使基于图像本身的计算变得可行,同时,该表征还具有对图像的对准问题不敏感的特性。

基于图像底层像素统计特性的直方图,也是广为使用的图像表征方法。常用的直方图特征有,联合图像像素在各个颜色通道统计分布的颜色直方图^[15]、统计图像边缘直线的直线角度直方图与直线长度直方图^[16]。

Oliva和Torralba^[6]提出一种描述图像全局布局的要点特征,即GIST,专门用于分类场景图像。其思想是,在无需任何分割的情况下,设计一种低维的图像表征。对此,文献[6]中提出了一组感知维数,分别为自然度、开阔度、粗糙度、延伸度和险峻度,以表征场景的主导空间结构。Oliva和Torralba指出通过对频谱施以粗糙的空间约束,便能够可靠地估计这些维数。GIST特征不仅可以有效地对具有标准空间布局的室外场景^[6]分类,而且还在大规模

的图像搜索中^[17,18]以高效的计算和较低的存储需求成为一种引人注目的方法。

由 David Lowe^[19]所设计的 SIFT(Scale Invariant Feature Transform)特征以善于发现图像中的重复内容而著称。SIFT 特征通过采用方向直方图,避免了传统灰度直方图对光照条件敏感的不足;同时又允许在关于图像细节的精确布局上存在一定的小波动。这些特性使得 SIFT 能够很好地匹配从不同视角所拍摄的相同图像区域,甚至还能够确定类别相同的目标物。SIFT 特征既可以在一组特别确定的图像位置上(如 Hessian-affine 和 MSER)进行稀疏抽取,又可在由空间网格所确定的位置上进行密集抽取。密集 SIFT 特征在区分 15 类场景数据集时表现不俗,而稀疏 SIFT 则在“Video Google”中发挥了可观的效用。

与 SIFT 的思想类似,Dalal 和 Triggs^[20]提出在一个密集的空间网格上计算由其划分的小图像细胞(cell)上的梯度方向直方图。并对其根据更大空间块(block)上的所有局部直方图的累积能量进行归一化,以去除光照与阴影的影响。经过归一化的描述符块被称为 HOG(Histogram of Oriented Gradient)特征。在目标检测和人物检测的应用中,该特征彰显了卓越的性能。

人类对视野之内的纹理特征能够以并行的方式进行迅速处理,而纹理元又是影响人类感知纹理的基本要素。因此,Renninger 和 Malik^[21]将纹理元形式化为一组现实世界图像对滤波器组的响应输出,并研究了它对于场景图像的分类区分能力。

局部二值模式(LBP)^[7]是一种强大的纹理描述特征,该特征基于局部二值模式的出现直方图。在文献[7]中,Ojala 等人提出一种能够反映纹理最为本质属性的“统一”二值模式,该模式提供了大量的主要局部特征,对应于诸如边缘等的微结构。通过估计这些微结构,将纹理的结构化分析与统计分析相结合。该特征具有理论与计算上的简单性与高效性。

作为对图像潜在外观特性(如像素、灰度、边界、滤波器响应的梯度等)的补充性描述,Shechtman 和 Irani^[22]提出了一种具有多尺度特性的 SSIM(Self-Similarity)特征。该特征刻画了图像中某局部区域上特定灰度模式间的相对几何布局,Shechtman 和 Irani 称之为自相似的局部内在布局;而生成这些自相似的模式也许并不具有共同的图像外观特性。

为了得到一种对视觉呈现具有不变性而对空间布局较为敏感的特征,Hoiem 等人^[10]提出了一种用于标定图像区域几何类别的方法。其思想是,分

别学习这些几何类别的视觉模型,来粗略描述每一幅图像区域的三维场景布局。该分割框架为我们提供了一种鲁棒的空间支持,其中可以使用各种视觉线索(如颜色、纹理和视角等)来构建每一个几何类别标签的置信。

1.3.2 多核学习研究现状

Lanckriet 等人^[23]为建立一个实际可行的核学习模型迈出了第一步。Lanckriet 等人在文献[23]中指出,给定一组备选核的集合,当这组核的线性组合正定,且其迹被归一化时,同时学习支持向量机(Support Vector Machine, SVM)与这组核的线性组合具有计算上的可行性:当对权值不施加任何约束时,可采用半定规划(Semi-Definite Programming, SDP)进行优化;当将权值约束为非负时,可采用二次约束二次规划(Quadratically Constrained Quadratic Program, QCQP)进行优化。该框架被称为多核学习(Multiple Kernel Learning, MKL)。在后续几年中,涌现了大量有关多核学习的研究工作。虽然它们在具体的核组合形式(如线性的、非线性的)、目标函数(如核配准^[24]、最大化边界分类误差^[25-27]、Fisher 区分分析^[12])以及优化算法(如半定规划^[23]、半无限线性规划^[26]、梯度下降^[27])等方面存在很大差异,但是根据解决的问题主要可分为两类:一类针对线性核组合与对核权值的稀疏正则相结合,关注设计高效的优化算法;另一类则探索能够为实际应用带来性能增益的新形式,包括核权值的非稀疏正则化和非线性的核组合方式。

1. 线性核组合

(1)核权值的稀疏正则化。虽然 Lanckriet 等人在其颇具深远影响的著作^[23]中,提出了基于 SDP 和 QCQP 策略的求解算法,但是这些算法无法有效地推广到对大规模问题的求解。对此, Bach 等人^[25]对 Lanckriet 等人^[23]所建立的原问题进行改进,为不同特征空间施以加权的 l_1 范数,而在每一个特征空间内部施以 l_2 范数,一般称这种方式为分块 l_1 范数正则化。所得的新的形式化可视为一个二阶锥规划(Second Order Cone Programming, SOCP)问题。相较于文献[23]中的 QCQP 形式,基于 SOCP 形式, Bach 等人^[25]通过为原问题中的目标函数增加 Moreau-Yosida 正则化项,并推导出相应的对偶形式,设计出了一种类似于序列最小化优化(Sequential Minimization Optimization, SMO)的算法。这种方法的提出首先突破了对中等规模数据的处理能力。但是, SOCP 仍然具有较高的计算耗费,因此,在处理更多的核与训练数据时存在局限性。

保持上述分块 l_1 范数的正则化形式, Sonnenburg 等人^[26]重新形式化了